STIMULATING ANTIBIOTIC DEVELOPMENT BY TARGETING VIRULENCE AND FACILITATING NATURAL PRODUCT DISCOVERY

BY

TUCKER MAXSON

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemistry
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

Associate Professor Douglas A. Mitchell, Chair
Professor Paul J. Hergenrother
Professor Martin D. Burke
Professor Steven R. Blanke

Abstract

Antibiotics are a cornerstone of modern medicine and have drastically reduced the burden of infectious diseases. Unfortunately, resistance to all clinically used antibiotics has become a major challenge that is exacerbated by numerous difficulties surrounding the development of new drugs. However, inventive strategies to overcome resistance as well as discover novel antibiotics are increasingly being explored. Whereas traditional antibiotics were generally designed to directly kill as many species of bacteria as possible, several new approaches have focused on narrower spectrum agents that have significant potential benefits. Antibiotics active against only one or a small group of pathogens would spare the microbiome, which may decrease the risk of secondary infections and slow the spread of resistance. One such narrow-spectrum strategy is to target the virulence factors employed by pathogens during an infection. In chapter 2, I demonstrate that the FDA approved HIV protease inhibitor nelfinavir can be repurposed as an inhibitor of the biosynthesis of the *Streptococcus pyogenes* cytolytic toxin streptolysin S. Nelfinavir was utilized to explore the proteolytic processing step in streptolysin S biosynthesis and was also shown to inhibit toxin production in other pathogens known to harbor similar biosynthetic clusters. Another approach to the problem of finding new antibiotics can be found in facilitating natural product discovery. Many antibiotics are derived from natural products but continuing to find new compounds has become increasingly difficult, especially due to rediscovery of known natural products. To help circumvent this problem, I developed a probe for identifying natural products containing aldehydes and ketones from microbial extracts based on the chemical reactivity of those carbonyl functional groups (chapter 3). This method is agnostic to the activity of the product and allows for the rapid identification of low abundance compounds that may be missed through activity-based screening. I demonstrate the utility of this probe by screening a collection of bacterial extracts, leading to the discovery of an analog of the protease inhibitor antipain.

To mom and dad for always being enthusiastic about chemistry - the apple doesn't fall far from the tree.

ACKNOWLEDGEMENTS

First, I'd like to thank my advisor, Professor Doug Mitchell, for taking me into his lab and for his advice and guidance over the years. I joined your group at the start of your career when you were just transitioning from post doc to faculty, and I think you've done a great job keeping us all pushing forward over the years. I'd like to thank the rest of my committee as well, Professor Marty Burke, Professor Paul Hergenrother, and Professor Steven Blanke. Your feedback on my work may have been painful at times but definitely helped me grow as a scientist. This work would not have been possible without some awesome collaboration, and I'd like to thank Evelyn Molloy, Caitlin Deane, Shaun Lee, Andrew Markley, Victor Nizet, and Mary Hensler for your contributions to the nelfinavir story. It has been a pleasure working with all of you. I am also extremely grateful to Jonathan Tietz for all your help getting NMR set up and advice on how to interpret the data.

Beyond the people I have directly collaborated with, having great colleagues both in and out of the Mitchell lab has been immeasurably important for me. You all have been fantastic sounding boards for ideas and sources of help when I have no idea what I am doing. Perhaps more importantly, you have all helped keep me sane by being great friends over the years. I would especially like to thank Kyle Dunbar, Joel and Kate Melby, and Spencer Peck – I could never have memorized every single card in Bang! without you!

Finally, I'd like to thank my mom and dad. Although n = 1 for this experiment, I am sure having two PhD chemists for parents as well as numerous family friends in the sciences helped get me interested in chemistry and encouraged me to keep going with science. While I maybe haven't always been thrilled with what I've been working on, I am happy with my decisions overall. Dad, I hope you would have been proud.

Table of Contents

Chapter 1: Introduction

This chapter is taken in part from Maxson and Mitchell[207] and is reproduced with permission from Tetrahedron.

1.1 The impetus for antibiotic research

The advent of antibiotics in the early twentieth century catalyzed a medical revolution, drastically reducing mortality due to bacterial infections. Along with numerous other advances in healthcare, such as vaccines and improved sanitation, antibiotics have contributed to an extension in the average life expectancy in the USA from 59.7 years in 1930 to 78.7 years in 2010.[1] Antibiotics are a critical component of a number of modern medical procedures, including many surgeries and transplants, as the rate of severe complications and death from infection would otherwise be unacceptably high. Decades of research into antibiotic development have produced highly effective and safe antibiotics, giving clinicians a wide range of tools to prevent and fight bacterial infections. However, resistance has inevitably followed the release of each new drug,[2-7] and the rapid propagation of resistant pathogens has become a serious issue, resulting in at least 23,000 deaths in 2013 in the United States alone.[8] Of particular concern are the ESKAPE pathogens (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* species) which have high levels of resistance and frequently "escape" eradication by antibiotics.[9, 10] While drug resistance or insensitivity can arise in other therapeutic areas, only in infectious diseases is that resistance known to be directly transmissible from person to person. Conservative use of antibiotics to slow resistance can be beneficial to society by extending the lifetime of current drugs[11] but doctors still use them liberally to treat individual patients if alternative options are not available. This has created a unique situation where everyone is at risk of having fewer treatment options than in previous decades unless the development of new antibiotics keeps pace with the development of resistance.

Unfortunately, bringing new antibiotics into the clinic has proven to be challenging for a number of reasons. The most fruitful strategy to discover novel antibiotics has been natural product screening, but it is increasingly rare to find compounds with new scaffolds, with rediscovery of known antibiotics

posing a significant challenge.[12, 13] Synthetic small molecule libraries have not fared better as effective antibiotics tend to have different properties than drugs in other therapeutic areas. Antibiotics often contain complex structures with multiple stereocenters and can be much larger than the usually small, flat compounds that dominate synthetic libraries.[14] Thus, compound collections designed around Lipinski's "rule of five" are poor sources for lead compounds.[15] This has been especially evident in screening campaigns against essential targets identified through genomics in which hits from *in vitro* screens frequently failed to display activity in whole cell assays.[16] A frequently cited factor for the decline in antibiotic development is the potential for low financial returns.[17-19] Bringing new antibiotics to market is seen as a poor investment as they are often reserved as drugs of last resort with a short duration of treatment, yet are subject to a pricing scheme dictated by a market saturated with the generics of older antibiotics that are often still effective.[20, 21] Additionally, FDA regulations have made clinical trials for antibiotics difficult and expensive to perform.[22] Non-inferiority trials are usually required due to the ethical constraints of withholding antibiotics from patients with serious infections, and the very low non-inferiority margin requirements result in the need for large patient populations.[20, 22] This is exacerbated by the disqualification of patients who have received any prestudy antibiotic, although the FDA has started to ease the regulatory restrictions for antibiotic clinical trials over the last several years. These factors had two major consequences. First, pharmaceutical companies interested in antibiotic development focused more on broad-spectrum agents to widen the potential market. Second, there has been an overall drop in approvals for new antibiotics over the last 30 years (Figure 1.1), which is concerning since resistance to available drugs is rapidly climbing leading to some experts calling this a "perfect storm".[23-26] However, this drop in approvals may not be as dire as it initially appears. Only looking at the trend in the number of approvals conceals the fact that many of the antibiotics released in the 1980s and 1990s were perhaps rushed through development too quickly and have since been discontinued while newer drugs have thus far fared much better (Figure 1.1). Many of the second and third generation cephalosporins approved in the 1980s had overlapping clinical utility with better selling family members and were withdrawn due to poor sales, while several fluoroquinolones approved in the 1990's were withdrawn for safety reasons.[27]

Despite the declining trend in antibiotic approvals, 2014 witnessed an uptick with four new molecular entities receiving FDA approval (1. dalbavancin, 2. oritavancin; both vancomycin derivatives, 3. ceftolozane; a 5th generation cephalosporin, and 4. tedizolid; an oxazolidinone related to linezolid).

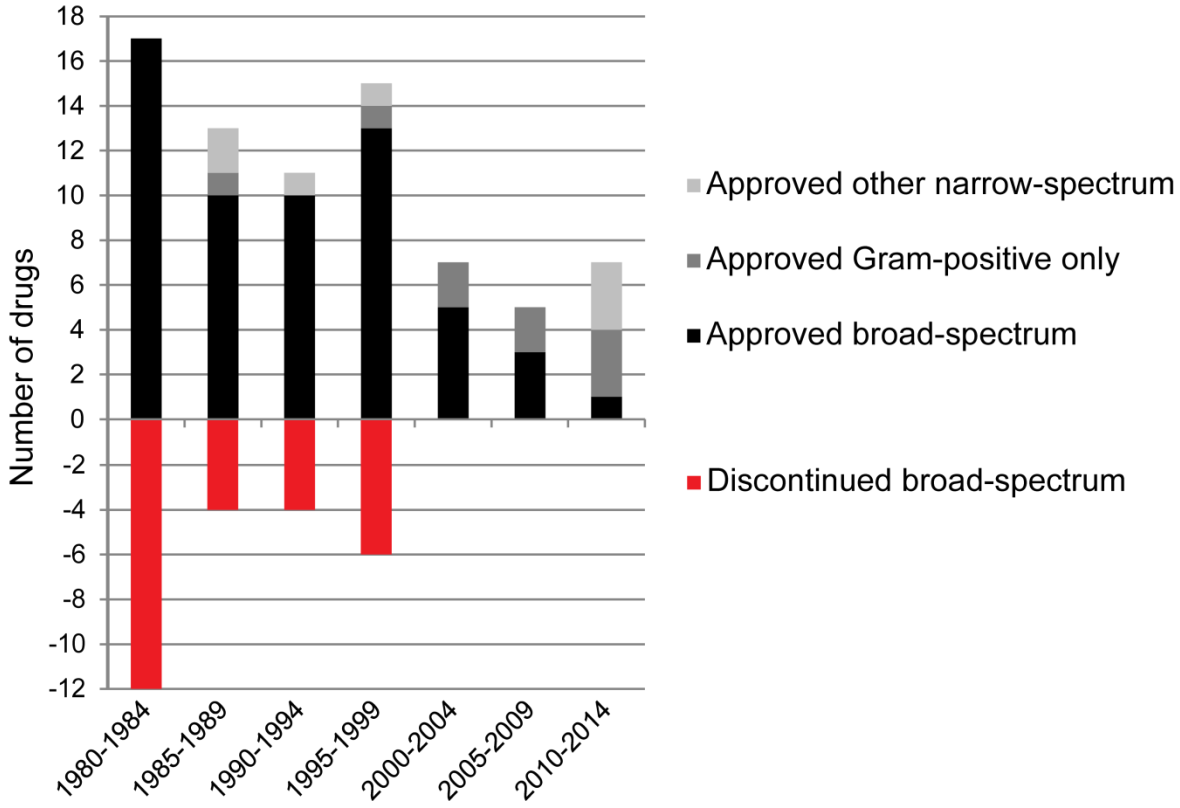


Figure 1.1. US FDA approval and discontinuation of antibiotics from 1980 to 2014. Approvals are shown in black. Discontinued antibiotics are shown in red as negative values.[27] Approvals are split by clinically useful activity into broad-spectrum, narrow-spectrum that target Gram-positive species only, and all other narrow-spectrum drugs (includes antimycobacterials). No narrow-spectrum antibiotics were discontinued during this time period.

The uptick of new approvals may have been spurred in part by a heightened academic and media interest in antibiotic development and conservation as a response to the dire consequences of antibiotic ineffectiveness. More recently, government organizations have been promoting antibiotic discovery as well. The United States congress passed the Generating Antibiotic Incentives Now (GAIN) Act in 2012 to financially incentivize antibiotic development[28, 29] followed by Executive Order 13676 (Combating Antibiotic-Resistant Bacteria), which directs government agencies to promote the development of new drugs and diagnostics, identify means of slowing resistance, and strengthen surveillance efforts for

resistant bacteria.[30, 31] A number of specific suggestions have been put forward to stimulate antibiotic development, including providing further financial incentives like research and development tax credits and grants, easing regulatory constraints, and promoting industry-academic collaborations.[23, 32, 33] New ideas to identify novel antibiotics are also appearing in the literature. Given previous successes with natural products serving as antibiotic lead structures,[34] strategies promoting natural product discovery with new cultivation techniques and molecular methods to reconstitute or activate specific gene clusters are especially promising.[15, 35-42] Beyond representing a promising source of new antibiotics, natural products have also served as drugs or inspirations for drugs for many other diseases[34] and have been used extensively as biological probes.[43, 44] Another potentially successful approach for antibiotic discovery is reexamining old drug leads that were previously abandoned during development, often because they were not considered broad-spectrum enough at the time.[45] Repurposing drugs designed for other indications can also produce new antibiotics or anti-virulence agents, with the added benefit of known safety and pharmacokinetic profiles.[46, 47]

Expanding the focus of antibiotic development to narrower spectrum compounds opens the door to the development of previously discarded leads, and makes it easier to find new ones by removing the requirement that a compound is growth inhibitory towards evolutionarily diverse bacterial pathogens.[15] Additional benefits associated with narrower spectrum drugs range from the possibility of slower resistance generation[48-50] to the demonstrated decreased risk of antibiotic-associated colitis (i.e. *Clostridium difficile* infections).[51-53] However, efficacious use of narrow-spectrum antibiotics requires rapid and sensitive diagnostics to identify the bacterial cause of infection. The current gold standard diagnostic is still traditional culture-based clinical microbiology, which is slow and often insensitive, such as with blood culture for sepsis patients.[54-57] Furthermore, determination of an antibiotic resistance profile adds even more time to this process. Ironically, the ready-availability of numerous, effective broad-spectrum drugs that saved so many lives also contributed to a stagnation in diagnostic advances for decades.[58] Only in the past 15 years have improved diagnostic techniques begun to gain momentum. The introduction of new molecular techniques, especially those that are polymerase chain reaction (PCR)- and

mass spectrometry (MS)-based, have primarily driven this innovation and, along with several as yet clinically unevaluated technologies, have already resulted in tests that shave hours or even days off traditional diagnostic methods.[59] The benefits and disadvantages of currently used molecular diagnostics, as well as several that are still in development, are discussed in Appendix A in this thesis.

Looking to the future, every avenue of antibiotic discovery and development will need to be explored in conjunction with efforts to conserve our current arsenal of therapeutics. Effective broad-spectrum antibiotics will likely always be needed for emergency empirical treatment and for polymicrobial infections. Introducing a wider variety narrow-spectrum drugs that do not disturb the microbiome to the clinic may provide an alternative, safer option. Ideally, narrow-spectrum antibiotics and rapid diagnostic tests will co-evolve. The effectiveness of this strategy has already been demonstrated in cancer therapy with companion diagnostics[60] and we argue that a similar effort is needed for bacterial infections. Regardless of the specifics of the approach, however, efforts to address antibiotic resistance are worthwhile and necessary endeavors that will hopefully maintain or enhance our ability to fight bacterial infections.

1.2 Narrow-spectrum antibiotics

1.2.1 Definition and therapeutic benefits

As a preface to a discussion on the benefits and drawbacks of antibiotics with different spectra of activity, it is important to define what exactly is meant by the terms "narrow-spectrum" and "broad-spectrum". The terms were introduced in the 1950's as comparators to describe the obvious differences between the original penicillins, such as penicillin G, and the broader spectrum tetracyclines and chloramphenicol.[61] When used to compare two antibiotics with different spectra of activity, broad- and narrow-spectrum are relatively easy to define. However, the terms have evolved over time into sweeping categories into which compounds with very different activities are lumped.[61] This has resulted in discrepancies in the literature wherein an antibiotic described as broad-spectrum in one paper may be called narrow-spectrum in another, often with no explanation as to why the label was chosen in either case. The phenotypic Gram stain is also frequently used as part of descriptions of narrow-spectrum

antibiotics; for instance, many agents act selectively on Gram-positive bacteria because the drug cannot penetrate the outer membrane of a Gram-negative organism. However, this can be misleading in the case of organisms that lack an outer membrane but are still phenotypically Gram-negative (e.g. mycobacteria).[62] Additionally, many antibiotics that target only Gram-positive or Gram-negative species are still broadly active within that category. We will reserve the term narrow-spectrum for antibiotics that have a reasonable likelihood of only affecting one or a small handful of species when administered to a patient, either through specifically targeting a single species or by targeting virulence factors that would be generally absent from non-pathogenic bacteria. However, we include a discussion of antibiotics that are selective for a single Gram stain phenotype as these are often considered to be narrow-spectrum[63, 64] and are certainly narrower spectrum than many other available agents.

Historically, the development of broader spectrum drugs by synthetically modifying existing antibiotics was the goal of pharmaceutical companies.[4] This can be readily seen with beta-lactam antibiotics, which were designed to achieve broader activity with the second and third generation compounds.[65] This approach is logical from a number of standpoints, most notably offering pharmaceutical companies the highest potential return on investment, as the drug could theoretically be used for multiple types of infections. Furthermore, broad-spectrum drugs allow doctors to treat infections empirically with a higher likelihood of success.[66] This is critical in life-threatening situations and broad-spectrum antibiotics will likely always have a place in medicine for this purpose.[66] The simplicity of the "one drug treats all" approach is appealing, but retrospectively we can no longer ignore antibiotic resistance and other associated problems like *C. difficile* infections. While a number of factors contribute to the rate at which resistance appears and spreads, including the frequency of use and misuse, the type of resistance mechanism required, and the fitness of resistant organisms,[67, 68] the use of broad-spectrum antibiotics appears to be correlated with increased emergence of resistance.[48-50, 69] Accumulation of resistance to antibiotics other than the one employed in treatment has been observed as well,[49] resulting from horizontal gene transfer (HGT) of DNA cassettes containing multiple resistance genes.[70-72] Administration of several broad-spectrum antibiotics, especially those that accumulate in the intestines,

can have a devastating impact on the microbiome that persists for months or even years after cessation of treatment.[73-76] As alluded to in section 1, disturbances of the gut microbiome can lead to a number of issues, most notably *C. difficile* infections.[77, 78] *C. difficile* thrives in the human colon when other bacteria are not present to suppress its colonization and growth.[79, 80] Secondary infections like those caused by *C. difficile* are directly linked to antibiotic usage, especially broad-spectrum cephalosporins, fluoroquinolones, and clindamycin.[77] Thus, *C. difficile* is difficult to eradicate with antibiotics owing to a catch-22-like scenario. As such, healthcare-associated strains of *C. difficile* have quickly become a global problem.[23] Additionally, antibiotics may be contributing to other issues linked to changes in the gut microbiome including chronic inflammatory diseases, diabetes, and asthma.[81-85] It is becoming increasingly obvious that the health of the microbiome must be taken into consideration during the development of future antibiotics.

Due to the drawbacks of broad-spectrum antibiotic usage and the general difficulties associated with finding new drugs capable of inhibiting the growth of a range of evolutionarily diverse bacterial pathogens, there have been calls in the literature to give renewed consideration to narrow-spectrum compounds.[58, 86] A number of possible benefits to using narrower spectrum drugs have been postulated, including those mentioned above, although some of the benefits are either theoretical, anecdotal, or are based on limited clinical data. In contrast to broad-spectrum treatment, the use of narrow-spectrum drugs may slow the spread of resistance through a lessened impact on the human microbiome, leading to reduced HGT of pre-existing resistance mechanisms.[50, 76, 87] Further studies are needed to confirm the trend, but large scale correlative evidence of this can be seen in Europe, where northern European countries tend to prescribe narrow-spectrum antibiotics more frequently than their southern neighbors on a proportional basis.[87-89] However, it is difficult to determine the direct impact of narrow- versus broad-spectrum antibiotic usage on a large scale since resistance rates also correlate heavily with frequency of usage. Countries that take a narrower spectrum approach also tend to be more conservative towards antibiotic usage overall and an increase in use with possible new narrow-spectrum antibiotics could still potentially cause a rapid rise in resistance. Sparing the gut microbiome has also been shown to correlate

with a lower rate of childhood obesity as compared with the use of broad-spectrum drugs.[90] Although direct evidence for the impact of broad- versus narrow-spectrum antibiotics on other conditions associated with the human microbiome is scarce, it is not difficult to imagine that significant links exist. As for the treatment efficacy, numerous studies have demonstrated that narrow-spectrum drugs can be just as effective as broad-spectrum ones in certain circumstances, especially in prophylaxis.[91-94]

Narrow-spectrum drugs have been developed and marketed since the advent of antibiotics, starting with salvarsan for syphilis. However, there has been a recent uptick in FDA approvals of antibiotics with narrower spectrums of activity, especially drugs targeting Gram-positive bacteria (Figure 1.1). Pathogen-specific antibiotics that target only one or a small set of species are also receiving increased interest, with fidaxomicin being recently approved as a selective agent for *C. difficile* that permits the gut microbiome to recover.[76, 95] Other strategies for fighting bacterial infections, such as targeting virulence[3, 96, 97] or treatment with antibodies or phage,[58, 98] are alternatives to growth-suppressive, small molecule antibiotics that spare the microbiome and possibly slow resistance. However, these approaches must be demonstrated to cure patients as effectively as traditional antibiotics or they will not gain FDA approval, let alone find clinical utility. Even if treatment with a broad-spectrum antibiotic leads to the spread of resistance in the longer-term, the immediate need of the patient will likely outweigh what may be best for the community at large.[99, 100]

1.2.2 Antibiotics selective for one Gram stain group

Until recently, narrow-spectrum generally meant that an antibiotic was only considered to be active against either Gram-positive or Gram-negative bacteria, with the exception of drugs for tuberculosis (TB) and salvarsan for syphilis. However, as noted in section 2.1, many of these drugs are still rather broad-spectrum within one Gram stain grouping. Many of the early antibiotics are Gram selective, but this shifted to a trend of even broader spectrum compounds in the 1980s and 1990s as medicinal chemists continuously tinkered with the properties of existing drugs.[4] This resulted in the release of ever more broad-spectrum fluoroquinolones and second and third generation cephalosporins during that time period.[34] More recently, there has been a shift towards the approval of more antibiotics

active only against Gram-positive bacteria (Figure 1.1), spurred in part by the prevalence and threat posed by multiple drug-resistant (MDR) *S. aureus* and *Streptococcus pneumonia*. These two pathogens alone were responsible for over 75% of deaths from antibiotic-resistant infections in the United States in 2013.[8] Discovering compounds with Gram-positive-only activity is theoretically easier than for Gram-negative activity, since the outer membrane present in most Gram-negative bacteria presents a formidable barrier.[101, 102] Focusing on a subset of bacterial species helps circumvent problems posed by the diversity among species that must be considered when attempting to develop broad-spectrum antibiotics,[103] though targeting an entire Gram stain grouping may not be specific enough to reap the potential benefits of narrow-spectrum agents.

Two notable examples of the renewed interest in narrowing the spectrum of antibiotics by targeting Gram-positive pathogens are the oxazolidinone and lipopeptide classes (exemplified by the first-in-class agents linezolid and daptomycin, respectively) (Figure 1.2).[103] Intriguingly, both compounds had been identified as potential leads, but discarded because of toxicity concerns.[104-106] However, in both cases, different pharmaceutical companies picked up the development of the compounds, overcame the toxicity issues, and brought them to market.[104, 106] Although both drugs are still broad-spectrum enough to be clinically useful in a number of infections, the fact that both have achieved blockbuster status is a testament to the potential for financial success with emerging narrow-spectrum antibiotics. A number of other Gram-positive selective antibiotics have been successfully released since, including glycopeptide and oxazolidinone family members.

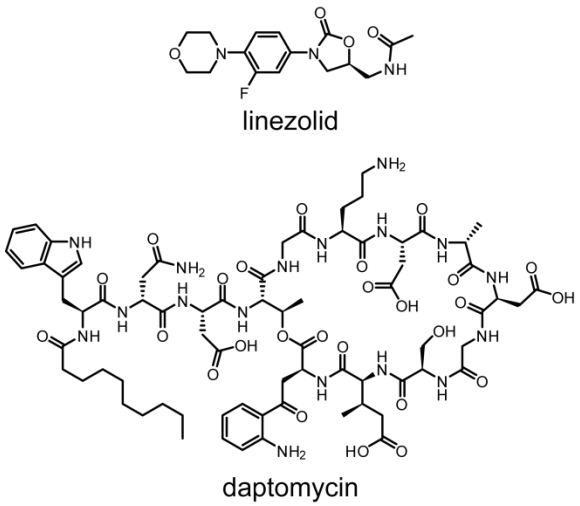


Figure 1.2. Structures of linezolid and daptomycin.

Despite clinical and financial success, drugs targeting a single Gram stain grouping share the limitations of broader spectrum antibiotics. For example, the clinical appearance of resistance to linezolid and daptomycin was not significantly slower than for other antibiotic classes.[5, 6] This may partly be a consequence of the success of the drugs, with high usage leading both to blockbuster status and the rapid development and dissemination of resistance. In theory, narrower spectrum antibiotics should have a reduced impact on the human microbiome, although this does not always appear to be the case for Gram selective drugs. Oral vancomycin treatment, although rare and generally reserved for *C. difficile*-associated diarrhea/colitis,[76] induces dramatic changes in the gut microbiome similar to broader spectrum antibiotics.[74, 107, 108] Indeed, orally administered vancomycin kills *Bacteroides* species that are not susceptible *in vitro*[109] which has been attributed to vancomycin concentrating to unusually high levels in the gut due to poor oral absorption;[110, 111] however, this may also stem from the knock-on effect of wiping out a subset of the microbiome that produces factors that other members rely on for stable colonization.[76] The latter possibility is potentially an issue for all narrow-spectrum antibiotics that affect even a single off-target species.[107, 112] Moving forward, the impact of Gram selective versus broader spectrum drugs on the human microbiome, as well as on resistance, will need to be examined in detail on a compound-by-compound basis to determine if they do provide an advantage.

1.2.3 Pathogen-specific antibiotics

In contrast to Gram selective antibiotics, pathogen-specific narrow-spectrum antibiotics would theoretically be used against a specific pathogenic species/genus rather than a general disease category like sepsis or urinary tract infections (UTI). This would constitute a type of personalized medicine with clear parallels to newer cancer drugs that are marketed with companion diagnostics. In cancer therapy, the diagnostic serves the purpose of determining whether the genotype of the patient's cancer cells is a match for the drug, while with an infectious disease the specific invading pathogen would be identified.[60, 113] The main advantage of such a personalized strategy is that treatment would be expected to minimize collateral damage to the microbiome[95, 107, 112] and perhaps even delay resistance acquisition by HGT. Pathogen-specific drugs have been used for decades in the treatment TB, with a number of FDA-approved drugs that are only active against *Mycobacterium tuberculosis* and other related mycobacteria (isoniazid, ethambutol, etc.).[114] The focus on TB stems in part from the staggering number of people infected (an estimated 2 billion worldwide) and the difficulty in treating the disease (6 months with multiple drugs).[114] While mycobacteria-specific antibiotics are effective at treating TB, a cocktail of several drugs that usually includes broad-spectrum antibiotics as well is necessary to prevent the rapid generation of resistance.[114] Thus, TB presents a well-studied counter-point to the idea that pathogen-specific antibiotics may help to slow the spread of resistance. Nonetheless, the other potential advantages to pathogen-specific treatment discussed in section 2.1 may still be beneficial, especially given the extended duration of treatment.[114] Since TB does not kill quickly in most cases, there is time to accurately diagnose the disease through culture-based methods and begin a treatment regimen that includes TB-specific antibiotics. It should be noted though, that the existence of TB-specific drugs certainly had more to do with clinical efficacy rather than with a conscientious attempt to spare the microbiome. Mycobacteria contain a number of unique targets related to the cell wall that facilitated the development of specific antibiotics.[115] Generalizing beyond TB selective drugs, there is no reason to think that pathogen-specific antibiotics couldn't be developed for other pathogens as well, although it is likely that differentiating between two Gram-negative species, for example, would be more difficult than differentiating between

mycobacteria and other pathogens. Very narrow-spectrum antibiotics are anticipated to be useful primarily for mono-microbial infections, and only then if the infecting pathogen can be rapidly and accurately identified. Identification would either be through standard clinical microbiology or the observation of symptoms unique to a pathogen. If multiple antibiotics targeting a specific pathogen could be developed, they could be used as part of a cocktail therapy to help extend the lifetime of the drugs, similar to the treatment regimens for TB.

Partly because the pharmaceutical industry preferentially develops (and doctors prescribe) the most broad-spectrum agents possible, the targets of current antibiotics are ubiquitous in the domain bacteria. Antibiotics that target a single Gram stain grouping share the same targets and are only selective due to the presence of other cellular features like the outer membrane of most Gram-negatives and efflux pumps that prevent the accumulation of therapeutic concentrations inside the cell. To develop an antibiotic against a specific species, unique targets that are not only essential to cell survival but also do not exist or can be compensated for in other bacteria are required. The presence or absence of important enzymes, unique cellular components, and key metabolic pathways are possible future drug targets. Additionally, specific protein folds or 3D structures unique to a single species in the otherwise common targets of other antibiotics could also be investigated. Prodrugs that are only activated once taken up by a specific pathogen would also impart specificity to compounds that become toxic when metabolized, such as with isoniazid in the treatment of TB.[116] Designing antibiotics around a specific target in one pathogen may even prove to be advantageous in that drug leads would not be required to have activity against distantly related homologs in other species. However, targets would need to be chosen carefully to ensure that sufficient differences between homologs exist and it remains to be seen if this is going to be a viable strategy.

A bountiful source of pathogen-specific drug leads may remain to be discovered in the realm of natural products. Some of these yet-to-be-discovered compounds may specifically interact with ecologically neighboring species while leaving others unaffected, either as signaling molecules or as mechanisms to fend off competitors. A compound that acts as a signaling molecule for a specific species

at environmentally relevant concentrations may kill the recipient outright while still maintaining specificity at higher concentrations. Even in the absence of co-evolution of species, microbial natural products are endowed with properties that render them better able to enter bacterial cells and interact with efficacious target(s).[117, 118] In addition to whole-cell screening with new compounds, pathogen-specific antibiotic discovery can also occur through screening against unique targets either *in vitro* or *in silico*, though serious pitfalls exist for these methods as discussed in section 1 and reviewed by Payne et al.[16] To date, few pathogen-specific antibiotics have been reported and still fewer have been followed up on to any significant extent. The scarcity stems in part from the limited antibiotic testing performed on many new natural products. Often, only common pathogens such as *S. aureus* and *Escherichia coli* are screened during testing so select activity against rarer pathogens is undetected (although the financial realities of developing such a compound would likely be limiting). Conversely, if only a small handful of organisms are used, a compound with activity against only one tested strain may also have activity against other untested pathogens or against the multitude of species in the human microbiome. As always, the researcher can only detect what he/she screens for.

Outside of TB-specific drugs, the only pathogen-specific antibiotic in use with FDA approval is fidaxomicin for the treatment of *C. difficile* (Figure 1.3),[95] which is ironic given that the explosion in *C. difficile* cases is the direct consequence of widespread use of broad-spectrum antibiotics. Although fidaxomicin does display some activity against other species, the minimum inhibitory concentrations in these cases tends to be 10-100 fold higher than for *C. difficile*,[119] allowing for the specific treatment of that pathogen with minimal effects on the gut microbiome.[76] Fidaxomicin has been shown to reduce recurrence rates of *C. difficile* in comparison to vancomycin, likely by allowing the gut microbiome to recover.[120, 121] The deployment of fidaxomicin to treat *C. difficile* infections represents an important milestone; however, the cure rates are still significantly lower than those achieved by fecal transplant.[121, 122]

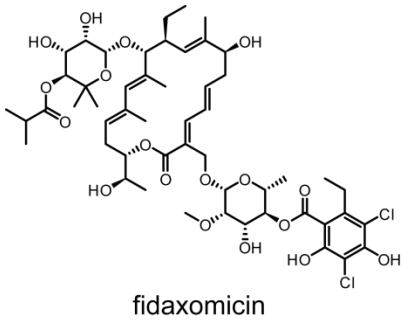


Figure 1.3. Structure of fidaxomicin.

A number of other pathogen-specific antibiotics have been reported in the literature but have not yet found clinical use. The following examples are not meant to be an exhaustive list of pathogen-specific agents and additional examples can be found in other reviews.[123, 124] Microcins are ribosomal peptide antibiotics, some of which are extensively post-translationally modified and display exquisitely selective activity.[125] A notable example is microcin B17 (Figure 1.4), exerting activity against only a handful of related gamma-proteobacteria that lack the associated immunity gene that protects the producing strains.[125, 126] The microcin B17 peptidic framework is decorated with a number of side chain-derived thiazole and oxazole rings[125] that rigidify the conformation of the peptide and provide an interaction surface to inhibit DNA gyrase.[127, 128] While microcin B17 does not exhibit the type of small molecule structure typically associated as being "drug-like", other peptides (and even larger biologics) have found clinical utility, such as the HIV fusion inhibitor enfuvirtide (Fuzeon) or insulin for diabetes.[129] Several other ribosomal peptides have been recently reported with pathogen-specific activity including the natural product plantazolicin (Figure 1.4), which, as another thiazole/oxazole-containing peptide, is biosynthesized in a similar manner to microcin B17 by *Bacillus* a*myloliquefaciens* and *Bacillus pumilus*.[130, 131] Plantazolicin has selective activity against *Bacillus anthracis* but the biological target has not yet been reported.[132, 133] The semisynthetic lanthipeptide NVB302 (from the natural product deoxyactagardine B) is another example of a ribosomal peptide-derived, pathogen-specific antibiotic (Figure 1.4);[134] it has recently completed a phase I clinical trial for the selective treatment of *C. difficile*.[135] Looking to more traditional, small molecule antibiotics, promysalin (Figure 1.4) is a novel

type of amphipathic salicylic acid-containing antibiotic produced by *Pseudomonas putida* with exquisitely selective activity against only other members of the *Pseudomonas* genus including *P. aeruginosa*, an ESKAPE pathogen.[136, 137] Yet another example is the pyloricidins (Figure 1.4), a family of peptide-like small molecules produced by *Bacillus* sp. HC-70 and HC-72.[138, 139] These compounds were discovered by screening specifically for activity against *Helicobacter pylori*, a traditionally difficult to treat gastric pathogen.[140] Like the example provided by *M. tuberculosis*, the selective targeting of other traditionally difficult pathogens may be the ultimate niche for further pathogen-specific antibiotic development.

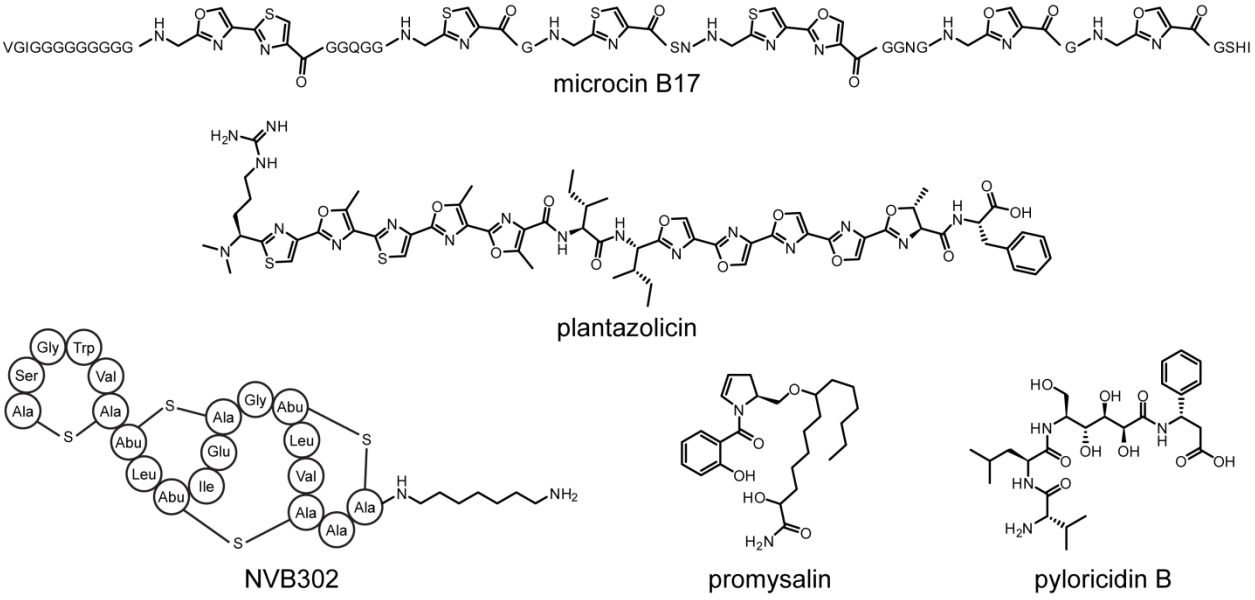


Figure 1.4. Structures of pathogen-specific antibiotics.

Most of the above examples demonstrate selectivity and potency *in vitro* which does not necessarily mean they would translate well into drugs. Additional studies are required, including determination of safety and pharmacokinetics, and the compounds (or derivatives) would need to meet all the stringent requirements for development into a drug. Whether this will happen is questionable given the high cost of drug development and the relatively small market for an antibiotic specific against rarer pathogens. A serious impediment in the development of pathogen-specific antibiotics is recruitment of sufficient patient populations known to be infected with the targeted pathogen, adding complexity to the

clinical trials. Granting orphan drug status to these compounds may help but additional financial incentives would likely also be required for pharmaceutical companies to proceed with development. Governmental agencies may have an interest in providing incentives for pathogens of homeland security concern, like *B. anthracis*. In the absence of subsidies however, it is probable that only pathogen-specific antibiotics against ubiquitous pathogens like methicillin-resistant *S. aureus* (MRSA) will attract interest from industry, analogous to the case of fidaxomicin for *C. difficile*.

1.2.4 Anti-virulence agents

An alternative strategy for treating disease caused by a particular bacterial pathogen is to interfere directly with pathogenesis. Such an anti-virulence strategy is anticipated to reduce the pathogen's ability to cause disease, rather than be growth suppressive. This entails targeting molecular entities that are not essential for the survival of the pathogen *in vitro* but are required for invasion and/or survival in the host.[3, 96, 97] A number of virulence targets have been explored, including inhibition or over-activation of quorum sensing,[141, 142] inhibition of bacterial adhesion,[143, 144] and inhibition of toxin production or delivery.[96] Since virulence factors are often unique to a single or small set of pathogens, agents that target them will intrinsically be pathogen-specific, although there are certainly cases in which members of the microbiome have similar factors that help them stably colonize a host without causing disease.[145] Targeting virulence rather than viability has also been postulated to cause slower resistance development due to the fact that the agent is not growth suppressive and thus elicits less selective pressure for acquiring resistance.[3, 96, 146] However, this prediction has not been sufficiently investigated and recent studies involving quorum sensing inhibition have cast some doubt on the idea.[147] We postulate that the propensity to develop resistance with anti-virulence agents will depend strongly on a number of factors, including how critical the virulence factor is to the pathogen for maintaining fitness in the host. If the virulence factor is essential to survival in the host, one would expect resistance to rise at approximately the same rate as if the cell wall or the ribosome were being targeted. Although it remains to be tested, anti-virulence agents also may not deliver clinical efficacy if supplied to patients as a monotherapy.[86] Formulating anti-

virulence agents with immune-stimulating drugs,[148, 149] or even a more conventional antibiotic, could prove to be the best future strategy for the patient and for the community.

A number of compounds that interfere with virulence through a wide array of pathways have been reported in the last decade. One of the first notable examples was virstatin (Figure 1.5).[150] It was found to transcriptionally prevent cholera toxin and pili production in *Vibrio cholerae* without causing any growth effects.[150] An *in vivo* infection model in mice demonstrated that virstatin treatment achieved a significant decrease in the bacterial burden.[150] Interestingly, virstatin was also recently found to inhibit biofilm formation in the ESKAPE pathogen *A. baumannii* through inhibition of pili biogenesis.[151] Another example of an anti-virulence agent that showed efficacy *in vivo* is BPH-652 (Figure 1.5), which was originally designed to inhibit cholesterol biosynthesis by targeting human squalene synthase. BPH-652 was also found to block a homologous enzyme in *S. aureus*, dehydrosqualene synthase, leading to inhibition of the oxidative stress-protective pigment staphyloxanthin.[152] Without the protection of staphyloxanthin, *S. aureus* was much more susceptible to the reactive oxygen species delivered by host immune cells and was cleared more efficiently in a mouse model of infection.[152] In addition to compounds that inhibit pathogenesis, there have been efforts to develop strategies to counteract toxins such as those produced by *B. anthracis* and *Clostridium botulinum*,[153, 154] similar to existing antibody based anti-toxin therapies (e.g. raxibacumab for anthrax toxin).[155] These types of compounds would not necessarily prevent or clear an infection alone but would rather reduce the damaging effects of the toxins and increase the chances of patient recovery. This is particular important for the pathogenic mechanism of diseases like anthrax and botulinum, which can cause mortality even if the causative pathogen is eradicated.[156, 157] It may also be possible to develop pseudo broad-spectrum anti-virulence antibiotics by purposefully targeting virulence factors that are employed by multiple different pathogens, such as the streptolysin S family of cytolytic toxins.[158] We found that the HIV protease inhibitor nelfinavir (Figure 1.5) blocked the proteolytic maturation of streptolysin S in *Streptococcus pyogenes* as well as related cytolysins from other pathogenic Firmicutes.[159] Such an approach may addresses the niche marketability

issue associated with pathogen-specific antibiotics while still providing the benefit of not disturbing the human microbiome.

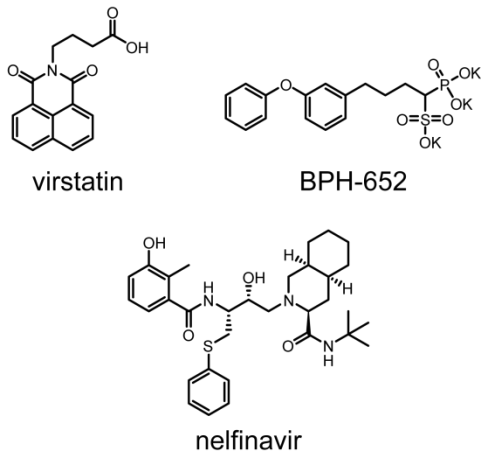


Figure 1.5. Structures of anti-virulence agents.

Despite the potential benefits of anti-virulence antibiotics, the fact that they don't suppress growth directly could also prove to be a major pitfall. Some pathogens deploy large arsenals of virulence factors[160-162] and inhibiting a single one may not sufficiently reduce their pathogenicity. Not all virulence factors are important throughout the course of infection either, with some required only for initial invasion or for dissemination to other body sites. This could prove to be an issue if treatment isn't started until symptoms appear and the infection is well underway, although utilization in prophylaxis for immune compromised individuals or surgery patients could still be helpful. A further complication is that many pathogens have variant strains that don't produce specific virulence factors and yet still cause disease; virulent strains of *V. cholerae* that do not produce cholera toxin and non-hemolytic strains of *S. pyogenes* that lack streptolysin S are examples of this.[150, 163, 164] If advances in diagnostics can overcome these challenges, however, anti-virulence agents could be an excellent avenue for treating bacterial infectious disease. Additionally, coupling anti-virulence drugs with conventional agents in a multicomponent cocktail, as is done in HIV, cancer, and TB therapy, could serve to significantly improve patient outcomes by directly reducing bacterial counts with the survivors more vulnerable to the host's immune system.

An additional benefit stemming from the development of anti-virulence agents are opportunities to use these agents as tools to study pathogenesis. Traditional methods for determining the roles of virulence factors include reconstitution of the factor's activity *in vitro* and generation of genetic deletions followed by *in vivo* infection models.[165, 166] While reconstitution of the activity of a virulence factor *in vitro* can provide a wealth of knowledge regarding the interaction of the factor with its cellular target or its effect on a single cell in culture, the role of the factor in the context of the array of other virulence factors and their global effects on the host are lost. Genetic deletion allows these effects to be more effectively investigated in certain situations but is a heavy-handed approach to a complex question; the importance of the virulence factor at different time points during the infection is impossible to explore and the quantity produced cannot be modulated beyond a simple on or off. Additionally, deletion of some cellular components that are recognized as important for virulence also causes detrimental effects on the pathogen when grown *in vitro*, such as with modulins.[165] Inhibitors of virulence factor production or action offer additional tools that help address these short comings. Anti-virulence agents can be used *in vitro* and *in vivo* to offer more temporal control over virulence factor action and provide a method to tune the activity of virulence factors through concentration-dependent inhibition. Anti-virulence compounds can also be useful in elucidating the biosynthesis of virulence factors, as is demonstrated in the second chapter of this thesis. Anti-virulence agents are simply one additional tool, however, and their use alongside other methods for probing pathogenesis is likely the best approach to providing useful insights into the mechanisms of infection.

1.2.5 Alternative narrow-spectrum therapeutic approaches

Another unconventional method for designing narrow-spectrum antibiotics is to attach broad-spectrum antibiotics, or even broadly toxic compounds, to a narrowly specific targeting domain.[167] This strategy has been employed in cancer therapy with several FDA-approved antibody-drug conjugates [brentuximab vedotin (Adcetris), trastuzumab emtansine (Kadcyla)].[168] Examples of targeting domains include peptides,[169, 170] antibodies,[171] and phage.[172] The conjugates are intended to bind specifically to the pathogen of interest and thus create a locally high concentration of the antibiotic that is sufficient to kill

the target. This strategy theoretically prevents non-targeted bacteria (and host cells in the case of general toxins) from experiencing harmful concentrations of the toxic payload. Oral administration of these antibiotic conjugates would likely be infeasible due to stability and bioavailability issues. The requirement of successfully targeting the pathogen and then delivering a sufficiently toxic payload also presents two different routes for resistance generation. Attempting to fight the pathogen outright with antibodies or bacteriophage may be more realistic. Both strategies have historical precedent and have been reviewed extensively elsewhere.[173-175] The use of bacteriophage to deliver CRISPR-Cas (CRISPR, clustered, regularly interspaced, short palindromic repeats; Cas, CRISPR-associated protein) systems for selective bacterial killing is also beginning to be explored.[176] Cas proteins are nucleases capable of cleaving DNA in a site-specific manner dependent on targeting by ~20-40 nucleotide RNA guides (CRISPR units).[177, 178] Cleavage of the bacterial genome by Cas nucleases has been shown to be lethal[179] and antimicrobials utilizing this strategy are currently in development by Eligo Bioscience.

1.2.6 Challenges facing narrow-spectrum therapy

If narrow-spectrum antibiotics manage to overcome the technological and financial hurdles to their development, they still face significant challenges in the clinic. An accurate and sensitive diagnosis would be required to rule out the possibility of a polymicrobial infection, otherwise a patient's condition may only worsen as an undetected pathogen flourishes during treatment of the known one. Additionally, immunocompromised patients are more susceptible to secondary infections that could be suppressed by a prophylactic broad-spectrum antibiotic but would be unaffected by a narrow-spectrum agent.

One of the intended benefits of narrow-spectrum antibiotics is that they would be less likely to disturb the human microbiome. However, there are an estimated 800 or more species in the gut[180] and new antibiotics are unlikely to be tested against a vast majority of these, especially given the difficulty in culturing many species in the microbiome. Even a pathogen-specific antibiotic may have unanticipated effects on some of these species, which could end up impacting the entire microbiome with deleterious effects. The argument that resistance may be slower to develop due to less possibility for HGT also loses some of its relevance if this is the case. The spectrum of activity of antibiotics against the gut microbiome

could be tested indirectly through metagenome analysis of fecal samples however, and this approach would also provide a more realistic picture of the effect of narrow-spectrum antibiotics in the context of the host. Regardless, resistance will always develop as bacteria eventually find a way to survive. Given the difficulty in finding novel antibiotics, strategies that provide new avenues to discovery or slow the spread of resistance are welcome and will hopefully be successfully combined in the future with other methods of fighting infections, such as boosting the human immune system or improved sterilization and sanitation in hospitals.

1.3 Reinvigorating natural product discovery as a source of new therapeutics

Natural products have historically served as important sources of drugs leads and as inspirations for synthetic or semisynthetic drugs.[34, 181, 182] Beyond the realm of antibiotics, drugs for other indications such as anticancer and immunosuppressive therapies have also had great success in coopting natural products.[183-185] The chemical diversity offered by collections of natural products has been recognized as an important factor in the hit rate of high-throughput screening campaigns for new drug leads[186] and compound libraries are starting to be designed to be more "natural product-like" in acknowledgement of this.[187-189] Natural products have also been extensively utilized as biological probes, aiding in the elucidation of enzyme activity and other cellular functions.[43, 44] For these reasons, continuing to expand our knowledge base of natural products is a worthwhile undertaking that will likely translate into medically and scientifically useful compounds.

Traditional methods for natural product discovery such as the Waksman platform for antibiotic discovery generally revolved around screening for a particular bioactivity from a culture or extract followed by isolation of the active compound by tracking the activity through purification steps.[15, 182, 190] While this strategy has resulted in the rapid discovery of a great many medically important natural products, there are significant flaws in the method that have become more apparent over time. The rediscovery of highly active compounds that are produced by a wide array of species results in wasted effort towards the isolation and characterization of known natural products.[191] For example, random screening of soil actinomycetes for antibiotic activity will yield streptomycin from roughly 1 in 100

strains.[13] Natural product discovery based on bioactivity screening will also only provide compounds that are active in the assay utilized. This can be a positive outcome in some cases, as any compounds found will have the desired activity, but also presents the drawback of missing any natural products with alternative activities that are not screened for. Compounds that are weakly active may also be missed due to insufficient abundance or interference from other components of the extract,[182] but these compounds could potentially be interesting as lead structures for further drug development. Therefore, novel methods for natural product discovery have been introduced and are continually being expanded.

Several strategies for natural product discovery that utilize bioactivity based screening but offer advantages such as more sensitive detection of low abundance compounds have been frequently employed. Prefractionation of extracts can remove many of the interfering or nuisance compounds that would otherwise interfere with the detection of weakly active or low abundance compounds in bioassays.[192, 193] Desired physicochemical properties that are known to make better drug leads can also be selected for during prefractionation, helping to eliminate active yet undesirable scaffolds.[194] Another method for improving discovery through bioactivity based screening is the use of novel cultivation techniques to stimulate robust natural product production or to successfully culture organisms that previously could not be cultivated in laboratory settings.[15, 195-197] Growing microorganisms in the context of their natural environment provides access to the myriad of factors normally present during growth and facilitates the culture of the roughly 99% of species that cannot be easily cultivated, many of which will hopefully provide new natural products.[42, 198]

Alternative strategies of natural product discovery that are independent of activity have also been developed. These methods often have the advantage of reduced rediscovery rates at the expense of potentially providing products with cryptic activity. Genomics has been extensively utilized in many of these methods to garner a sense of the biosynthetic potential of organisms and to provide initial clues as the structure of a natural product prior to isolation and characterization.[199, 200] Access to information about the natural product biosynthetic gene cluster of interest also allows for targeted strategies to stimulate production or facilitate isolation. Silent clusters can be activated through manipulation of promoters,

utilization of signaling molecules such as γ-butyrolactones, or expression in a heterologous host.[201, 202] Mass spectrometry guided methods have also been used not only for the detection of natural products with structural motifs predicted from genome mining but also to aid in final structural elucidation of the products.[203, 204] Another strategy is to employ chemoselective enrichment for functionality suspected to be present on the natural product of interest.[205] Similar to chemoselective enrichment, unique and reactive functional groups can be used as handles for tagging natural products as a method of reactivity-based natural product discovery.[206] The discovery of natural products with aldehyde or ketone functional groups via reactivity-based screening with an aminooxy probe is the focus of the third chapter in this thesis.

1.4 Summary and outlook

The golden age of antibiotics has ended and new strategies to develop antibiotics and fight resistance are needed. However, academia has fervently taken up the call and industry is starting to follow suite with prodding from public concern and government incentives. While the studies discussed in this thesis may not directly have an impact on the development of a new antibiotic, they help build the groundwork for future research in these areas. Developing a better understanding of virulence in *S. pyogenes* will be beneficial towards future endeavors to develop treatment strategies against this common pathogen, whether they are the design of new antibiotics or vaccines or simply more intelligent utilization of our current drugs.

History tells us that natural products are one of the best sources of new drugs, especially antibiotics, and continuing to launch new discovery efforts is a worthwhile pursuit. Reactivity-based natural product discovery is a relatively new endeavor so its utility towards large scale discovery is as yet untested. However, it offers another potentially useful set of tools for natural product screening. As additional reactivity-based probes are developed, they will hopefully allow the majority of natural products to be labeled selectively by at least one probe, adding in isolation and also potentially providing useful structural information.

1.5 References

1. Arias, E. (2014) United States life tables, 2010, *Natl. Vital Stat. Rep. 63*, 1-63.

2. Palumbi, S. R. (2001) Humans as the world's greatest evolutionary force, *Science 293*, 1786-1790.

3. Clatworthy, A. E., Pierson, E., and Hung, D. T. (2007) Targeting virulence: a new paradigm for antimicrobial therapy, *Nat. Chem. Biol. 3*, 541-548.

4. Walsh, C. T., and Wencewicz, T. A. (2014) Prospects for new antibiotics: a molecule-centered perspective, *J. Antibiot. 67*, 7-22.

5. Tsiodras, S., Gold, H. S., Sakoulas, G., Eliopoulos, G. M., Wennersten, C., Venkataraman, L., Moellering, R. C., and Ferraro, M. J. (2001) Linezolid resistance in a clinical isolate of Staphylococcus aureus, *Lancet 358*, 207-208.

6. Sabol, K., Patterson, J. E., Lewis, J. S., Owens, A., Cadena, J., and Jorgensen, J. H. (2005) Emergence of daptomycin resistance in Enterococcus faecium during daptomycin therapy, *Antimicrob. Agents Chemother. 49*, 1664-1665.

7. Gentry, D. R., McCloskey, L., Gwynn, M. N., Rittenhouse, S. F., Scangarella, N., Shawar, R., and Holmes, D. J. (2008) Genetic characterization of Vga ABC proteins conferring reduced susceptibility to pleuromutilins in Staphylococcus aureus, *Antimicrob. Agents Chemother. 52*, 4507-4509.

8. (2013) Antibiotic resistance threats in the United States, 2013, Centers for Disease Control and Prevention.

9. Rice, L. B. (2008) Federal funding for the study of antimicrobial resistance in nosocomial pathogens: No ESKAPE, *J. Infect. Dis. 197*, 1079-1081.

10. Boucher, H. W., Talbot, G. H., Bradley, J. S., Edwards, J. E., Gilbert, D., Rice, L. B., Scheld, M., Spellberg, B., and Bartlett, J. (2009) Bad Bugs, No Drugs: No ESKAPE! An Update from the Infectious Diseases Society of America, *Clin. Infect. Dis. 48*, 1-12.

11. Barbosa, T. M., and Levy, S. B. (2000) The impact of antibiotic use on resistance development and persistence, *Drug Resist. Updat. 3*, 303-311.

12. von Nussbaum, F., Brands, M., Hinzen, B., Weigand, S., and Habich, D. (2006) Antibacterial natural products in medicinal chemistry--exodus or revival?, *Angew. Chem. Int. Ed. 45*, 5072-5129.

13. Baltz, R. H. (2006) Marcel Faber Roundtable: Is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration?, *J. Ind. Microbiol. Biotechnol. 33*, 507-513.

14. Frantz, S. (2004) Better antibiotics through chemistry, *Nat. Rev. Drug Discov. 3*, 900-901.

15. Lewis, K. (2013) Platforms for antibiotic discovery, *Nat. Rev. Drug Discov. 12*, 371-387.

16. Payne, D. J., Gwynn, M. N., Holmes, D. J., and Pompliano, D. L. (2007) Drugs for bad bugs: confronting the challenges of antibacterial discovery, *Nat. Rev. Drug Discov. 6*, 29-40.

17. Spellberg, B. (2013) New antibiotic development: barriers and opportunities in 2012, *APUA Newsletter 30*, 8-10.

18. Clardy, J., Fischbach, M. A., and Walsh, C. T. (2006) New antibiotics from bacterial natural products, *Nat. Biotechnol. 24*, 1541-1550.

19. Nathan, C. (2004) Antibiotics at the crossroads, *Nature 431*, 899-902.

20. Shlaes, D. M. (2015) Research and Development of Antibiotics: The Next Battleground, *ACS Infect. Dis. 1*, 232-233.

21. Outterson, K. (2014) New Business Models for Sustainable Antibiotics, Centre on Global Health Security Working Group Papers, Chatham House.

22. Shlaes, D. M., Sahm, D., Opiela, C., and Spellberg, B. (2013) The FDA reboot of antibiotic development, *Antimicrob. Agents Chemother. 57*, 4605-4607.

23. Boucher, H. W., Talbot, G. H., Benjamin, D. K., Bradley, J., Guidos, R. J., Jones, R. N., Murray, B. E., Bonomo, R. A., Gilbert, D., and Amer, I. D. S. (2013) 10 x '20 Progress-Development of New Drugs Active Against Gram-Negative Bacilli: An Update From the Infectious Diseases Society of America, *Clin. Infect. Dis. 56*, 1685-1694.

24. Spellberg, B., Guidos, R., Gilbert, D., Bradley, J., Boucher, H. W., Scheld, W. M., Bartlett, J. G., Edwards, J., and Amer, I. D. S. (2008) The epidemic of antibiotic-resistant infections: A call to action for the medical community from the Infectious Diseases Society of America, *Clin. Infect. Dis. 46*, 155-164.

25. Gould, I. M. (2009) Antibiotic resistance: the perfect storm, *Int. J. Antimicrob. Agents 34 Suppl 3*, S2-5.

26. Cooper, M. A., and Shlaes, D. (2011) Fix the antibiotics pipeline, *Nature 472*, 32.

27. Outterson, K., Powers, J. H., Seoane-Vazquez, E., Rodriguez-Monguio, R., and Kesselheim, A. S. (2013) Approval and withdrawal of new antibiotics and other antiinfectives in the U.S., 1980-2009, *J. Law Med. Ethics 41*, 688-696.

28. Countries, J. o. I. i. D. (2011) Generating Antibiotic Incentives Now Act of 2011, *112th Cong., H.R. 2182/S. 1734: Generating Antibiotic Incentives Now Act of 2011*.

29. Brown, E. D. (2013) Is the GAIN Act a turning point in new antibiotic discovery?, *Can. J. Microbiol. 59*, 153-156.

30. (2014) Combating Antibiotic-Resistant Bacteria, *Executive Order 13676: Combating Antibiotic-Resistant Bacteria*.

31. (2015) National action plan for combating antibiotic-resistant bacteria, *National Action Plan for Combating Antibiotic-Resistant Bacteria*.

32. Hwang, T. J., Carpenter, D., and Kesselheim, A. S. (2015) Paying for innovation: reimbursement incentives for antibiotics, *Sci. Transl. Med. 7*, 276fs279.

33. Hwang, T. J., Powers, J. H., Carpenter, D., and Kesselheim, A. S. (2015) Accelerating innovation in rapid diagnostics and targeted antibacterials, *Nat. Biotechnol. 33*, 589-590.

34. Newman, D. J., and Cragg, G. M. (2012) Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010, *J. Nat. Prod. 75*, 311-335.

35. Rutledge, P. J., and Challis, G. L. (2015) Discovery of microbial natural products by activation of silent biosynthetic gene clusters, *Nat. Rev. Microbiol. 13*, 509-523.

36. Weber, T., Charusanti, P., Musiol-Kroll, E. M., Jiang, X., Tong, Y., Kim, H. U., and Lee, S. Y. (2015) Metabolic engineering of antibiotic factories: new tools for antibiotic production in actinomycetes, *Trends Biotechnol. 33*, 15-26.

37. Zerikly, M., and Challis, G. L. (2009) Strategies for the discovery of new natural products by genome mining, *ChemBioChem 10*, 625-633.

38. Olano, C., Garcia, I., Gonzalez, A., Rodriguez, M., Rozas, D., Rubio, J., Sanchez-Hidalgo, M., Brana, A. F., Mendez, C., and Salas, J. A. (2014) Activation and identification of five clusters for secondary metabolites in Streptomyces albus J1074, *Microb. Biotechnol. 7*, 242-256.

39. Spohn, M., Kirchner, N., Kulik, A., Jochim, A., Wolf, F., Muenzer, P., Borst, O., Gross, H., Wohlleben, W., and Stegmann, E. (2014) Overproduction of Ristomycin A by activation of a silent gene cluster in Amycolatopsis japonicum MG417-CF17, *Antimicrob. Agents Chemother. 58*, 6185-6196.

40. Pham, V. H., and Kim, J. (2012) Cultivation of unculturable soil bacteria, *Trends Biotechnol. 30*, 475-484.

41. Xiong, Z. Q., Wang, J. F., Hao, Y. Y., and Wang, Y. (2013) Recent advances in the discovery and development of marine microbial natural products, *Mar. Drugs 11*, 700-717.

42. Ling, L. L., Schneider, T., Peoples, A. J., Spoering, A. L., Engels, I., Conlon, B. P., Mueller, A., Schaberle, T. F., Hughes, D. E., Epstein, S., Jones, M., Lazarides, L., Steadman, V. A., Cohen, D. R., Felix, C. R., Fetterman, K. A., Millett, W. P., Nitti, A. G., Zullo, A. M., Chen, C., and Lewis, K. (2015) A new antibiotic kills pathogens without detectable resistance, *Nature 517*, 455-459.

43. Crews, C. M., and Splittgerber, U. (1999) Chemical genetics: exploring and controlling cellular processes with chemical probes, *Trends Biochem. Sci. 24*, 317-320.

44. Carlson, E. E. (2010) Natural Products as Chemical Probes, *ACS Chem. Biol. 5*, 639-653.

45. Koehn, F. E. (2008) New strategies and methods in the discovery of natural product anti-infective agents: the mannopeptimycins, *J. Med. Chem. 51*, 2613-2617.

46. Ashburn, T. T., and Thor, K. B. (2004) Drug repositioning: identifying and developing new uses for existing drugs, *Nat. Rev. Drug Discov. 3*, 673-683.

47. Rangel-Vega, A., Bernstein, L. R., Mandujano-Tinoco, E. A., Garcia-Contreras, S. J., and Garcia-Contreras, R. (2015) Drug repurposing as an alternative for the treatment of recalcitrant bacterial infections, *Front. Microbiol. 6*, 237.

48. (2011) Antibiotic Policies: Controlling Hospital Acquired Infection, *Antibiotic Policies: Controlling Hospital Acquired Infection*, 1-209.

49. May, A. K., Fleming, S. B., Carpenter, R. O., Diaz, J. J., Guillamondegui, O. D., Deppen, S. A., Miller, R. S., Talbot, T. R., and Morris, J. A. (2006) Influence of broad-spectrum antibiotic prophylaxis on intracranial pressure monitor infections and subsequent infectious complications in head-injured patients, *Surg. Infect. 7*, 409-417.

50. de Man, P., Verhoeven, B. A. N., Verbrugh, H. A., Vos, M. C., and van den Anker, J. N. (2000) An antibiotic policy to prevent emergence of resistant bacilli, *Lancet 355*, 973-978.

51. Dial, S., Kezouh, A., Dascal, A., Barkun, A., and Suissa, S. (2008) Patterns of antibiotic use and risk of hospital admission because of Clostridium difficile infection, *CMAJ 179*, 767-772.

52. Deshpande, A., Pasupuleti, V., Thota, P., Pant, C., Rolston, D. D., Sferra, T. J., Hernandez, A. V., and Donskey, C. J. (2013) Community-associated Clostridium difficile infection and antibiotics: a meta-analysis, *J. Antimicrob. Chemother. 68*, 1951-1961.

53. Lemon, K. P., Armitage, G. C., Relman, D. A., and Fischbach, M. A. (2012) Microbiota-Targeted Therapies: An Ecological Perspective, *Sci. Transl. Med. 4*, 137rv135.

54. Dellinger, R. P., Levy, M. M., Rhodes, A., Annane, D., Gerlach, H., Opal, S. M., Sevransky, J. E., Sprung, C. L., Douglas, I. S., Jaeschke, R., Osborn, T. M., Nunnally, M. E., Townsend, S. R., Reinhart, K., Kleinpell, R. M., Angus, D. C., Deutschman, C. S., Machado, F. R., Rubenfeld, G. D., Webb, S. A., Beale, R. J., Vincent, J. L., Moreno, R., and Surviving Sepsis Campaign Guidelines Committee including the Pediatric, S. (2013) Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012, *Crit. Care Med. 41*, 580-637.

55. Mancini, N., Carletti, S., Ghidoli, N., Cichero, P., Burioni, R., and Clementi, M. (2010) The Era of Molecular and Other Non-Culture-Based Methods in Diagnosis of Sepsis, *Clin. Microbiol. Rev. 23*, 235-251.

56. Paolucci, M., Landini, M. P., and Sambri, V. (2010) Conventional and molecular techniques for the early diagnosis of bacteraemia, *Int. J. Antimicrob. Agents 36*, S6-S16.

57. Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E., and Crook, D. W. (2012) Transforming clinical microbiology with bacterial genome sequencing, *Nat. Rev. Genet. 13*, 601-612.

58. Casadevall, A. (2009) The case for pathogen-specific therapy, *Expert Opin. Pharmacother. 10*, 1699-1703.

59. Kothari, A., Morgan, M., and Haake, D. A. (2014) Emerging Technologies for Rapid Identification of Bloodstream Pathogens, *Clin. Infect. Dis. 59*, 272-278.

60. Olsen, D., and Jorgensen, J. T. (2014) Companion diagnostics for targeted cancer drugs - clinical and regulatory aspects, *Front. Oncol. 4*, 105.

61. Acar, J. (1997) Broad- and narrow-spectrum antibiotics: an unhelpful categorization, *Clin. Microbiol. Infect. 3*, 395-396.

62. Fu, L. M., and Fu-Liu, C. S. (2002) Is Mycobacterium tuberculosis a closer relative to Gram-positive or Gram-negative bacterial pathogens?, *Tuberculosis 82*, 85-90.

63. Hersh, A. L., Shapiro, D. J., Pavia, A. T., and Shah, S. S. (2011) Antibiotic prescribing in ambulatory pediatrics in the United States, *Pediatrics 128*, 1053-1061.

64. Sarpong, E. M., and Miller, G. E. (2015) Narrow- and Broad-Spectrum Antibiotic Use among US Children, *Health Serv. Res. 50*, 830-846.

65. Page, M. G. (2004) Cephalosporins in clinical development, *Expert Opin. Invest. Drugs 13*, 973-985.

66. Kollef, M. H. (2008) Broad-spectrum antimicrobials and the treatment of serious bacterial infections: getting it right up front, *Clin. Infect. Dis. 47 Suppl 1*, S3-13.

67. Walsh, C. (2003) Where will new antibiotics come from?, *Nat. Rev. Microbiol. 1*, 65-70.

68. Goossens, H., Ferech, M., Stichele, R. V., Elseviers, M., and Grp, E. P. (2005) Outpatient antibiotic use in Europe and association with resistance: a cross-national database study., *Lancet 365*, 579-587.

69. Dortch, M. J., Fleming, S. B., Kauffmann, R. M., Dossett, L. A., Talbot, T. R., and May, A. K. (2011) Infection reduction strategies including antibiotic stewardship protocols in surgical and trauma intensive care units are associated with reduced resistant gram-negative healthcare-associated infections, *Surg. Infect. 12*, 15-25.

70. Salyers, A. A., Gupta, A., and Wang, Y. P. (2004) Human intestinal bacteria as reservoirs for antibiotic resistance genes, *Trends Microbiol. 12*, 412-416.

71. Salyers, A. A., Moon, K., and Schlesinger, D. (2007) The Human Intestinal Tract - a Hotbed of Resistance Gene Transfer? Part I, *Clin. Microbiol. Newsletter 29*, 17-21.

72. Salyers, A. A., Moon, K., and Schlesinger, D. (2007) The Human Intestinal Tract - a Hotbed of Resistance Gene Transfer? Part II, *Clin. Microbiol. Newsletter 29*, 25-30.

73. Jernberg, C., Lofmark, S., Edlund, C., and Jansson, J. K. (2010) Long-term impacts of antibiotic exposure on the human intestinal microbiota, *Microbiology 156*, 3216-3223.

74. Cotter, P. D., Stanton, C., Ross, R. P., and Hill, C. (2012) The Impact of Antibiotics on the Gut Microbiota as Revealed by High Throughput DNA Sequencing, *Discov. Med. 70*, 193-199.

75. Sullivan, A., Edlund, C., and Nord, C. E. (2001) Effect of antimicrobial agents on the ecological balance of human microflora, *Lancet Infect. Dis. 1*, 101-114.

76. Rashid, M. U., Weintraub, A., and Nord, C. E. (2012) Effect of new antimicrobial agents on the ecological balance of human microflora, *Anaerobe 18*, 249-253.

77. Bartlett, J. G. (2010) Clostridium difficile: progress and challenges, *Ann. N. Y. Acad. Sci. 1213*, 62-69.

78. Kelly, C. P., and LaMont, J. T. (2008) Clostridium difficile - More difficult than ever, *New Engl. J. Med. 359*, 1932-1940.

79. Kachrimanidou, M., and Malisiovas, N. (2011) Clostridium difficile infection: a comprehensive review, *Crit. Rev. Microbiol. 37*, 178-187.

80. Schubert, A. M., Sinani, H., and Schloss, P. D. (2015) Antibiotic-Induced Alterations of the Murine Gut Microbiota and Subsequent Effects on Colonization Resistance against Clostridium difficile, *mBio 6*, e00974.

81. Blaser, M. (2011) Stop the killing of beneficial bacteria, *Nature 476*, 393-394.

82. Moloney, R. D., Desbonnet, L., Clarke, G., Dinan, T. G., and Cryan, J. F. (2014) The microbiome: stress, health and disease, *Mamm. Genome 25*, 49-74.

83. Nylund, L., Satokari, R., Salminen, S., and de Vos, W. M. (2014) Intestinal microbiota during early life - impact on health and disease, *Proc. Nutr. Soc. 73*, 457-469.

84. Matamoros, S., Gras-Leguen, C., Le Vacon, F., Potel, G., and de La Cochetiere, M. F. (2013) Development of intestinal microbiota in infants and its impact on health, *Trends Microbiol. 21*, 167-173.

85. Quigley, E. M. (2013) Gut bacteria in health and disease, *Gastroenterol. Hepatol. 9*, 560-569.

86. Then, R. L., and Sahl, H. G. (2010) Anti-Infective Strategies of the Future: Is there Room for Species-Specific Antibacterial Agents?, *Curr. Pharm. Des. 16*, 555-566.

87. Torfoss, D., Hoiby, E. A., Holte, H., and Kvaloy, S. (2012) The Norwegian experience with penicillin G plus an aminoglycoside as initial empiric therapy in febrile neutropenia; a review, *Acta Oncol. 51*, 433-440.

88. Bergan, T. (2001) Antibiotic usage in Nordic countries, *Int. J. Antimicrob. Agents 18*, 279-282.

89. Cars, O., Molstad, S., and Melander, A. (2001) Variation in antibiotic use in the European Union, *Lancet 357*, 1851-1853.

90. Bailey, L. C., Forrest, C. B., Zhang, P., Richards, T. M., Livshits, A., and DeRusso, P. A. (2014) Association of antibiotics in infancy with early childhood obesity, *JAMA Pediatr. 168*, 1063-1069.

91. Kronman, M. P., Hersh, A. L., Feng, R., Huang, Y. S., Lee, G. E., and Shah, S. S. (2011) Ambulatory visit rates and antibiotic prescribing for children with pneumonia, 1994-2007, *Pediatrics 127*, 411-418.

92. Taylor, S. P., and Taylor, B. T. (2013) Health care-associated pneumonia in haemodialysis patients: clinical outcomes in patients treated with narrow versus broad spectrum antibiotic therapy, *Respirology 18*, 364-368.

93. Palmer, D. L., Pett, S. B., and Akl, B. F. (1995) Bacterial Wound Colonization after Broad-Spectrum Versus Narrow-Spectrum Antibiotics, *Ann. Thorac. Surg. 59*, 626-631.

94. Vuori-Holopainen, E., Peltola, H., Kallio, M. J. T., and Grp, S. T. S. (2000) Narrow- versus broad-spectrum parenteral antimicrobials against common infections of childhood: a prospective and randomised comparison between penicillin and cefuroxime, *Eur. J. Pediatr. 159*, 878-884.

95. Chahine, E. B., Sucher, A. J., and Mantei, K. (2014) Fidaxomicin: a novel macrolide antibiotic for Clostridium difficile infection, *Consult. Pharm. 29*, 614-624.

96. Rasko, D. A., and Sperandio, V. (2010) Anti-virulence strategies to combat bacteria-mediated disease, *Nat. Rev. Drug Discov. 9*, 117-128.

97. Heras, B., Scanlon, M. J., and Martin, J. L. (2015) Targeting virulence not viability in the search for future antibacterials, *Br. J. Clin. Pharmacol. 79*, 208-215.

98. Yacoby, I., and Benhar, I. (2007) Targeted anti bacterial therapy, *Infect. Disord. Drug Targets 7*, 221-229.

99. Butler, C. C., Kinnersley, P., Prout, H., Rollnick, S., Edwards, A., and Elwyn, G. (2001) Antibiotics and shared decision-making in primary care, *J. Antimicrob. Chemother. 48*, 435-440.

100. McDonnell Norms, G. (2008) Antibiotic overuse: the influence of social norms, *J. Am. Coll. Surg. 207*, 265-275.

101. Delcour, A. H. (2009) Outer membrane permeability and antibiotic resistance, *Biochim. Biophys. Acta 1794*, 808-816.

102. Page, M. G. (2012) The role of the outer membrane of Gram-negative bacteria in antibiotic resistance: Ajax' shield or Achilles' heel?, *Handb. Exp. Pharmacol.*, 67-86.

103. Coates, A. R., Halls, G., and Hu, Y. (2011) Novel classes of antibiotics or more of the same?, *Br. J. Pharmacol. 163*, 184-194.

104. Brickner, S. J. (1996) Oxazolidinone antibacterial agents, *Curr. Pharm. Des. 2*, 175-194.

105. Leach, K. L., Brickner, S. J., Noe, M. C., and Miller, P. F. (2011) Linezolid, the first oxazolidinone antibacterial agent, *Ann. N. Y. Acad. Sci. 1222*, 49-54.

106. Eisenstein, B. I., Oleson, F. B., and Baltz, R. H. (2010) Daptomycin: From the Mountain to the Clinic, with Essential Help from Francis Tally, MD, *Clin. Infect. Dis. 50*, S10-S15.

107. Rea, M. C., Dobson, A., O'Sullivan, O., Crispie, F., Fouhy, F., Cotter, P. D., Shanahan, F., Kiely, B., Hill, C., and Ross, R. P. (2011) Effect of broad- and narrow-spectrum antimicrobials on Clostridium difficile and microbial diversity in a model of the distal colon, *Proc. Natl. Acad. Sci. U. S. A. 108 Suppl 1*, 4639-4644.

108. Edlund, C., Barkholt, L., Olsson-Liljequist, B., and Nord, C. E. (1997) Effect of vancomycin on intestinal flora of patients who previously received antimicrobial therapy, *Clin. Infect. Dis. 25*, 729-732.

109. Sutter, V. L., Kwok, Y. Y., and Finegold, S. M. (1973) Susceptibility of Bacteroides-Fragilis to 6 Antibiotics Determined by Standardized Antimicrobial Disk Susceptibility Testing, *Antimicrob. Agents Chemother. 3*, 188-193.

110. Gonzales, M., Pepin, J., Frost, E. H., Carrier, J. C., Sirard, S., Fortier, L. C., and Valiquette, L. (2010) Faecal pharmacokinetics of orally administered vancomycin in patients with suspected Clostridium difficile infection, *BMC Infect. Dis. 10*.

111. Citron, D. M., Tyrrell, K. L., Merriam, C. V., and Goldstein, E. J. C. (2012) In Vitro Activities of CB-183,315, Vancomycin, and Metronidazole against 556 Strains of Clostridium difficile, 445 Other Intestinal Anaerobes, and 56 Enterobacteriaceae Species, *Antimicrob. Agents Chemother. 56*, 1613-1615.

112. Louie, T. J., Emery, J., Krulicki, W., Byrne, B., and Mah, M. (2009) OPT-80 Eliminates Clostridium difficile and Is Sparing of Bacteroides Species during Treatment of C-difficile Infection, *Antimicrob. Agents Chemother. 53*, 261-263.

113. Efferth, T. (2010) Personalized cancer medicine: from molecular diagnostics to targeted therapy with natural products, *Planta Med. 76*, 1143-1154.

114. Zumla, A., Nahid, P., and Cole, S. T. (2013) Advances in the development of new tuberculosis drugs and treatment regimens, *Nat. Rev. Drug Discov. 12*, 388-404.

115. Hett, E. C., and Rubin, E. J. (2008) Bacterial growth and cell division: a mycobacterial perspective, *Microbiol. Mol. Biol. Rev. 72*, 126-156, table of contents.

116. Timmins, G. S., and Deretic, V. (2006) Mechanisms of action of isoniazid, *Mol. Microbiol. 62*, 1220-1227.

117. Wright, G. D. (2014) Something old, something new: revisiting natural products in antibiotic drug discovery, *Can. J. Microbiol. 60*, 147-154.

118. Butler, M. S., and Buss, A. D. (2006) Natural products--the future scaffolds for novel antibiotics?, *Biochem. Pharmacol. 71*, 919-929.

119. Goldstein, E. J., Babakhani, F., and Citron, D. M. (2012) Antimicrobial activities of fidaxomicin, *Clin. Infect. Dis. 55 Suppl 2*, S143-148.

120. Crook, D., Weiss, K., Cornely, O. A., Miller, M., R., E., and Gorbach, S. (2010) Randomized clinical trial in Clostridium difficile infection confirms equivalent cure rate and lower recurrence rate of fidaxomicin vs. vancomycin, In *20th European Congress of Clinical Microbiology and Infectious Diseases, Abstract LB2401.*, Vienna, Austria.

121. Chaparro-Rojas, F., and Mullane, K. M. (2013) Emerging therapies for Clostridium difficile infection - focus on fidaxomicin, *Infect. Drug Resist. 6*, 41-53.

122. Rohlke, F., and Stollman, N. (2012) Fecal microbiota transplantation in relapsing Clostridium difficile infection, *Therap. Adv. Gastroenterol. 5*, 403-420.

123. Saleem, M., Nazir, M., Ali, M. S., Hussain, H., Lee, Y. S., Riaz, N., and Jabbar, A. (2010) Antimicrobial natural products: an update on future antibiotic drug candidates, *Nat. Prod. Rep. 27*, 238-254.

124. Wietz, M., Mansson, M., Vynne, N. G., and Gram, L. (2013) *Small-Molecule Antibiotics from Marine Bacteria and Strategies to Prevent Rediscovery of Known Compounds, in Marine Microbiology: Bioactive Compounds and Biotechnological Applications*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.

125. Duquesne, S., Destoumieux-Garzon, D., Peduzzi, J., and Rebuffat, S. (2007) Microcins, gene-encoded antibacterial peptides from enterobacteria, *Nat. Prod. Rep. 24*, 708-734.

126. Asensio, C., and Perez-Diaz, J. C. (1976) A new family of low molecular weight antibiotics from enterobacteria, *Biochem. Biophys. Res. Commun. 69*, 7-14.

127. Vizan, J. L., Hernandezchico, C., Delcastillo, I., and Moreno, F. (1991) The Peptide Antibiotic Microcin B17 Induces Double-Strand Cleavage of DNA Mediated by Escherichia-Coli DNA Gyrase, *Embo J. 10*, 467-476.

128. Heddle, J. G., Blance, S. J., Zamble, D. B., Hollfelder, F., Miller, D. A., Wentzell, L. M., Walsh, C. T., and Maxwell, A. (2001) The antibiotic microcin B17 is a DNA gyrase poison: Characterisation of the mode of inhibition, *J. Mol. Biol. 307*, 1223-1234.

129. Craik, D. J., Fairlie, D. P., Liras, S., and Price, D. (2013) The future of peptide-based drugs, *Chem. Biol. Drug Des. 81*, 136-147.

130. Scholz, R., Molohon, K. J., Nachtigall, J., Vater, J., Markley, A. L., Sussmuth, R. D., Mitchell, D. A., and Borriss, R. (2011) Plantazolicin, a novel microcin B17/streptolysin S-like natural product from Bacillus amyloliquefaciens FZB42, *J. Bacteriol. 193*, 215-224.

131. Kalyon, B., Helaly, S. E., Scholz, R., Nachtigall, J., Vater, J., Borriss, R., and Sussmuth, R. D. (2011) Plantazolicin A and B: structure elucidation of ribosomally synthesized thiazole/oxazole peptides from Bacillus amyloliquefaciens FZB42, *Org. Lett. 13*, 2996-2999.

132. Molohon, K. J., Melby, J. O., Lee, J., Evans, B. S., Dunbar, K. L., Bumpus, S. B., Kelleher, N. L., and Mitchell, D. A. (2011) Structure determination and interception of biosynthetic intermediates for the plantazolicin class of highly discriminating antibiotics, *ACS Chem. Biol. 6*, 1307-1313.

133. Hao, Y., Blair, P. M., Sharma, A., Mitchell, D. A., and Nair, S. K. (2015) Insights into Methyltransferase Specificity and Bioactivity of Derivatives of the Antibiotic Plantazolicin, *ACS Chem. Biol. 10*, 1209-1216.

134. Boakes, S., and Dawson, M. J. (2014) *Discovery and Development of NVB302, a Semisynthetic Antibiotic for Treatment of Clostridium difficile Infection, in Natural Products: Discourse, Diversity, and Design*, John Wiley & Sons, Inc., Hoboken, NJ, USA.

135. Sandiford, S. K. (2015) Perspectives on lantibiotic discovery - where have we failed and what improvements are required?, *Expert Opin. Drug Discov. 10*, 315-320.

136. Li, W., Estrada-de los Santos, P., Matthijs, S., Xie, G. L., Busson, R., Cornelis, P., Rozenski, J., and De Mot, R. (2011) Promysalin, a salicylate-containing Pseudomonas putida antibiotic,

promotes surface colonization and selectively targets other Pseudomonas, *Chem. Biol. 18*, 1320-1330.

137. Steele, A. D., Knouse, K. W., Keohane, C. E., and Wuest, W. M. (2015) Total Synthesis and Biological Investigation of (-)-Promysalin, *J. Am. Chem. Soc. 137*, 7314–7317.

138. Nakao, M., Miyagaw, K., Nakano, Y., Sakane, T., Tada, M., Nishimura, O., and Fujino, M. (2001) Pyloricidins, novel anti-helicobacterpylori antibiotics produced by Bacillus sp. I. Taxonomy, fermentation and biological activity, *J. Antibiot. 54*, 926-933.

139. Nagano, Y., Ikedo, K., Fujishima, A., Izawa, M., Tsubotani, S., Nishimura, O., and Fujino, M. (2001) Pyloricidins, novel anti-Helicobacter pylori antibiotics produced by bacillus sp. II. Isolation and structure elucidation, *J. Antibiot. 54*, 934-947.

140. Shiota, S., and Yamaoka, Y. (2014) Strategy for the treatment of Helicobacter pylori infection, *Curr. Pharm. Des. 20*, 4489-4500.

141. Tay, S. B., and Yew, W. S. (2013) Development of quorum-based anti-virulence therapeutics targeting Gram-negative bacterial pathogens, *Int. J. Mol. Sci. 14*, 16570-16599.

142. LaSarre, B., and Federle, M. J. (2013) Exploiting Quorum Sensing To Confuse Bacterial Pathogens, *Microbiol. Mol. Biol. Rev. 77*, 73-111.

143. Krachler, A. M., and Orth, K. (2013) Targeting the bacteria-host interface: strategies in anti-adhesion therapy, *Virulence 4*, 284-294.

144. Klemm, P., Hancock, V., Kvist, M., and Schembri, M. A. (2007) Candidate targets for new antivirulence drugs: selected cases of bacterial adhesion and biofilm formation, *Future Microbiol. 2*, 643-653.

145. Niu, C., Yu, D., Wang, Y., Ren, H., Jin, Y., Zhou, W., Li, B., Cheng, Y., Yue, J., Gao, Z., and Liang, L. (2013) Common and pathogen-specific virulence factors are different in function and structure, *Virulence 4*, 473-482.

146. Allen, R. C., Popat, R., Diggle, S. P., and Brown, S. P. (2014) Targeting virulence: can we make evolution-proof drugs?, *Nat. Rev. Microbiol. 12*, 300-308.

147. Kalia, V. C., Wood, T. K., and Kumar, P. (2014) Evolution of Resistance to Quorum-Sensing Inhibitors, *Microb. Ecol. 68*, 13-23.

148. Finlay, B. B., and Hancock, R. E. W. (2004) Opinion - Can innate immunity be enhanced to treat microbial infections?, *Nat. Rev. Microbiol. 2*, 497-504.

149. Kimura, H. J., Suzuki, K., Landek-Salgado, M. A., Caturegli, P., Jounai, N., Kobiyama, K., and Takeshita, F. (2011) Application of innate immune molecules for a new class of drugs: infection, inflammation and beyond, *Endocr. Metab. Immune Disord. Drug Targets 11*, 68-75.

150. Hung, D. T., Shakhnovich, E. A., Pierson, E., and Mekalanos, J. J. (2005) Small-molecule inhibitor of Vibrio cholerae virulence and intestinal colonization, *Science 310*, 670-674.

151. Chabane, Y. N., Ben Mlouka, M., Alexandre, S., Nicol, M., Marti, S., Pestel-Caron, M., Vila, J., Jouenne, T., and De, E. (2014) Virstatin inhibits biofilm formation and motility of Acinetobacter baumannii, *BMC Microbiol. 14*.

152. Liu, C. I., Liu, G. Y., Song, Y. C., Yin, F. L., Hensler, M. E., Jeng, W. Y., Nizet, V., Wang, A. H. J., and Oldfield, E. (2008) A cholesterol biosynthesis inhibitor blocks Staphylococcus aureus virulence, *Science 319*, 1391-1394.

153. Nestorovich, E. M., and Bezrukov, S. M. (2014) Designing inhibitors of anthrax toxin, *Expert Opin. Drug Discov. 9*, 299-318.

154. Pang, Y. P., Vummenthala, A., Mishra, R. K., Park, J. G., Wang, S., Davis, J., Millard, C. B., and Schmidt, J. J. (2009) Potent new small-molecule inhibitor of botulinum neurotoxin serotype A endopeptidase developed by synthesis-based computer-aided molecular design, *PLoS One 4*, e7730.

155. Kummerfeldt, C. E. (2014) Raxibacumab: potential role in the treatment of inhalational anthrax, *Infect. Drug Resist. 7*, 101-109.

156. Schneemann, A., and Manchester, M. (2009) Anti-toxin antibodies in prophylaxis and treatment of inhalation anthrax, *Future Microbiol. 4*, 35-43.

157. Rosow, L. K., and Strober, J. B. (2015) Infant botulism: review and clinical update, *Pediatr. Neurol. 52*, 487-492.

158. Molloy, E. M., Cotter, P. D., Hill, C., Mitchell, D. A., and Ross, R. P. (2011) Streptolysin S-like virulence factors: the continuing sagA, *Nat. Rev. Microbiol. 9*, 670-681.

159. Maxson, T., Deane, C. D., Molloy, E. M., Cox, C. L., Markley, A. L., Lee, S. W., and Mitchell, D. A. (2015) HIV Protease Inhibitors Block Streptolysin S Production, *ACS Chem. Biol. 10*, 1217-1226.

160. Nizet, V. (2007) Understanding how leading bacterial pathogens subvert innate immunity to reveal novel therapeutic targets, *J. Allergy Clin. Immun. 120*, 13-22.

161. Ballok, A. E., and O'Toole, G. A. (2013) Pouring Salt on a Wound: Pseudomonas aeruginosa Virulence Factors Alter Na+ and Cl- Flux in the Lung, *J. Bacteriol. 195*, 4013-4019.

162. Jedrzejas, M. J. (2001) Pneumococcal virulence factors: structure and function, *Microbiol. Mol. Biol. Rev. 65*, 187-207.

163. Shakhnovich, E. A., Sturtevant, D., and Mekalanos, J. J. (2007) Molecular mechanisms of virstatin resistance by non-O1/non-O139 strains of Vibrio cholerae, *Mol. Microbiol. 66*, 1331-1341.

164. Yoshino, M., Murayama, S. Y., Sunaoshi, K., Wajima, T., Takahashi, M., Masaki, J., Kurokawa, I., and Ubukata, K. (2010) Nonhemolytic Streptococcus pyogenes Isolates That Lack Large Regions of the sag Operon Mediating Streptolysin S Production, *J. Clin. Microbiol. 48*, 635-638.

165. Casadevall, A., and Pirofski, L. A. (2009) Virulence factors and their mechanisms of action: the view from a damage-response framework, *J. Water Health 7*, S2-S18.

166. Wu, H. J., Wang, A. H. J., and Jennings, M. P. (2008) Discovery of virulence factors of pathogenic bacteria, *Curr Opin Chem Biol 12*, 93-101.

167. Yacoby, I., Bar, H., and Benhar, I. (2007) Targeted drug-carrying bacteriophages as antibacterial nanomedicines, *Antimicrob. Agents Chemother. 51*, 2156-2163.

168. Sassoon, I., and Blanc, V. (2013) Antibody-drug conjugate (ADC) clinical pipeline: a review, *Methods Mol. Biol. 1045*, 1-27.

169. Domenyuk, V., Loskutov, A., Johnston, S. A., and Diehnelt, C. W. (2013) A technology for developing synbodies with antibacterial activity, *PLoS One 8*, e54162.

170. Eckert, R., Qi, F., Yarbrough, D. K., He, J., Anderson, M. H., and Shi, W. (2006) Adding selectivity to antimicrobial peptides: rational design of a multidomain peptide against Pseudomonas spp, *Antimicrob. Agents Chemother. 50*, 1480-1488.

171. Szynol, A., de Haard, J. J., Veerman, E. C., de Soet, J. J., and van Nieuw Amerongen, A. V. (2006) Design of a peptibody consisting of the antimicrobial peptide dhvar5 and a llama variable heavy-chain antibody fragment, *Chem. Biol. Drug Des. 67*, 425-431.

172. Yacoby, I., Shamis, M., Bar, H., Shabat, D., and Benhar, I. (2006) Targeting antibacterial agents by using drug-carrying filamentous bacteriophages, *Antimicrob. Agents Chemother. 50*, 2087-2097.

173. Oleksiewicz, M. B., Nagy, G., and Nagy, E. (2012) Anti-bacterial monoclonal antibodies: Back to the future?, *Arch. Biochem. Biophys. 526*, 124-131.

174. Golkar, Z., Bagasra, O., and Pace, D. G. (2014) Bacteriophage therapy: a potential solution for the antibiotic resistance crisis, *J. Infect. Dev. Ctries. 8*, 129-136.

175. Casadevall, A. (2006) The third age of antimicrobial therapy, *Clin. Infect. Dis. 42*, 1414-1416.

176. Bikard, D., Euler, C. W., Jiang, W. Y., Nussenzweig, P. M., Goldberg, G. W., Duportet, X., Fischetti, V. A., and Marraffini, L. A. (2014) Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials, *Nat. Biotechnol. 32*, 1146-1150.

177. Horvath, P., and Barrangou, R. (2010) CRISPR/Cas, the immune system of bacteria and archaea, *Science 327*, 167-170.

178. Sorek, R., Kunin, V., and Hugenholtz, P. (2008) CRISPR - a widespread system that provides acquired resistance against phages in bacteria and archaea, *Nat. Rev. Microbiol. 6*, 181-186.

179. Gomaa, A. A., Klumpe, H. E., Luo, M. L., Selle, K., Barrangou, R., and Beisel, C. L. (2014) Programmable Removal of Bacterial Strains by Use of Genome-Targeting CRISPR-Cas Systems, *mBio 5*, e00928.

180. Backhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A., and Gordon, J. I. (2005) Host-bacterial mutualism in the human intestine, *Science 307*, 1915-1920.

181. Dias, D. A., Urban, S., and Roessner, U. (2012) A historical overview of natural products in drug discovery, *Metabolites 2*, 303-336.

182. Harvey, A. L., Edrada-Ebel, R., and Quinn, R. J. (2015) The re-emergence of natural products for drug discovery in the genomics era, *Nat. Rev. Drug Discov. 14*, 111-129.

183. Mishra, B. B., and Tiwari, V. K. (2011) Natural products: an evolving role in future drug discovery, *Eur. J. Med. Chem. 46*, 4769-4807.

184. Cragg, G. M., and Newman, D. J. (2013) Natural products: a continuing source of novel drug leads, *Biochim. Biophys. Acta 1830*, 3670-3695.

185. Butler, M. S., Blaskovich, M. A., and Cooper, M. A. (2013) Antibiotics in the clinical pipeline in 2013, *J. Antibiot. 66*, 571-591.

186. Sukuru, S. C. K., Jenkins, J. L., Beckwith, R. E. J., Scheiber, J., Bender, A., Mikhailov, D., Davies, J. W., and Glick, M. (2009) Plate-Based Diversity Selection Based on Empirical HTS Data to Enhance the Number of Hits and Their Chemical Diversity, *J. Biomol. Screen. 14*, 690-699.

187. Wetzel, S., Bon, R. S., Kumar, K., and Waldmann, H. (2011) Biology-oriented synthesis, *Angew. Chem. Int. Ed. 50*, 10800-10826.

188. Grabowski, K., Baringhaus, K. H., and Schneider, G. (2008) Scaffold diversity of natural products: inspiration for combinatorial library design, *Nat. Prod. Rep. 25*, 892-904.

189. Huigens, R. W., 3rd, Morrison, K. C., Hicklin, R. W., Flood, T. A., Jr., Richter, M. F., and Hergenrother, P. J. (2013) A ring-distortion strategy to construct stereochemically complex and structurally diverse compounds from natural products, *Nat. Chem. 5*, 195-202.

190. Schatz, A., Bugie, E., and Waksman, S. A. (2005) Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria. 1944, *Clin. Orthop. Relat. Res.*, 3-6.

191. Sashidhara, K. V., and Rosaiah, J. N. (2007) Various dereplication strategies using LC-MS for rapid natural product lead identification and drug discovery, *Nat. Prod. Commun. 2*, 193-202.

192. Appleton, D. R., Buss, A. D., and Butler, M. S. (2007) A simple method for high-throughput extract prefractionation for biological screening, *Chimia 61*, 327-331.

193. Wagenaar, M. M. (2008) Pre-fractionated microbial samples - The second generation natural products library at Wyeth, *Molecules 13*, 1406-1426.

194. Camp, D., Davis, R. A., Campitelli, M., Ebdon, J., and Quinn, R. J. (2012) Drug-like properties: guiding principles for the design of natural product libraries, *J. Nat. Prod. 75*, 72-81.

195. Stewart, E. J. (2012) Growing Unculturable Bacteria, *J. Bacteriol. 194*, 4151-4160.

196. VanderMolen, K. M., Raja, H. A., El-Elimat, T., and Oberlies, N. H. (2013) Evaluation of culture media for the production of secondary metabolites in a natural products screening program, *AMB Express 3*.

197. Zazopoulos, E., Huang, K. X., Staffa, A., Liu, W., Bachmann, B. O., Nonaka, K., Ahlert, J., Thorson, J. S., Shen, B., and Farnet, C. M. (2003) A genomics-guided approach for discovering and expressing cryptic metabolic pathways, *Nat. Biotechnol. 21*, 187-190.

198. Kaeberlein, T., Lewis, K., and Epstein, S. S. (2002) Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment, *Science 296*, 1127-1129.

199. Doroghazi, J. R., Albright, J. C., Goering, A. W., Ju, K. S., Haines, R. R., Tchalukov, K. A., Labeda, D. P., Kelleher, N. L., and Metcalf, W. W. (2014) A roadmap for natural product discovery based on large-scale genomics and metabolomics, *Nat. Chem. Biol. 10*, 963-968.

200. Jensen, P. R., Chavarria, K. L., Fenical, W., Moore, B. S., and Ziemert, N. (2014) Challenges and triumphs to genomics-based natural product discovery, *J. Ind. Microbiol. Biotechnol. 41*, 203-209.

201. Scherlach, K., and Hertweck, C. (2009) Triggering cryptic natural product biosynthesis in microorganisms, *Org. Biomol. Chem. 7*, 1753-1760.

202. Rutledge, P. J., and Challis, G. L. (2015) Discovery of microbial natural products by activation of silent biosynthetic gene clusters, *Nat. Rev. Microbiol. 13*, 509-523.

203. Kersten, R. D., Yang, Y. L., Xu, Y., Cimermancic, P., Nam, S. J., Fenical, W., Fischbach, M. A., Moore, B. S., and Dorrestein, P. C. (2011) A mass spectrometry-guided genome mining approach for natural product peptidogenomics, *Nat. Chem. Biol. 7*, 794-802.

204. Medema, M. H., and Fischbach, M. A. (2015) Computational approaches to natural product discovery, *Nat. Chem. Biol. 11*, 639-648.

205. Odendaal, A. Y., Trader, D. J., and Carlson, E. E. (2011) Chemoselective enrichment for natural products discovery, *Chem. Sci. 2*, 760-764.

206. Cox, C. L., Tietz, J. I., Sokolowski, K., Melby, J. O., Doroghazi, J. R., and Mitchell, D. A. (2014) Nucleophilic 1,4-additions for natural product discovery, *ACS Chem. Biol. 9*, 2014-2022.

207. Maxson, T., and Mitchell, D. A. (2015) Targeted treatment for bacterial infections: prospects for pathogen-specific antibiotics coupled with rapid diagnostics, *Tetrahedron*, in press doi:10.1016/j.tet.2015.09.069

Chapter 2: HIV protease inhibitors block streptolysin S production

This chapter is taken in part from Maxson et al.[1] and is reproduced with permission from ACS Chemical Biology.

Douglas Mitchell generated ^{35}S-labeled SagA, performed membrane cleavage assays, and tested general mechanism based protease inhibitors. Andrew Markley and Shaun Lee (Lee lab) performed *E. coli* pETDuet experiments. Evelyn Molloy tested nelfinavir in *L. monocytogenes*. Caitlin Deane tested nelfinavir in *B. amyloliquefaciens*.

Abstract

Streptolysin S (SLS) is a post-translationally modified peptide cytolysin that is produced by the human pathogen *Streptococcus pyogenes.* SLS belongs to a large family of azole-containing natural products that are biosynthesized via an evolutionarily conserved pathway. SLS is an important virulence factor during *S. pyogenes* infections, but despite an extensive history of study, further investigations are needed to clarify several steps of its biosynthesis. To this end, chemical inhibitors of SLS biosynthesis would be valuable tools to interrogate the various maturation steps of both SLS and biosynthetically-related natural products. Such chemical inhibitors could also potentially serve as anti-virulence therapeutics, which in theory may alleviate the spread of antibiotic resistance. In this work, we demonstrate that FDA-approved HIV protease inhibitors, especially nelfinavir, block a key proteolytic processing step during SLS production. This inhibition was demonstrated in live *S. pyogenes* cells and through *in vitro* protease inhibition assays. A panel of 57 nelfinavir analogs was synthesized, leading to a series of compounds with improved anti-SLS activity while illuminating structure-activity relationships. Nelfinavir was also found to inhibit the maturation of other azole-containing natural products, namely those involved in listeriolysin S, clostridiolysin S, and plantazolicin production. The use of nelfinavir analogs as inhibitors of SLS production has allowed us to begin examining the proteolysis event in SLS maturation and will aid in further investigations of the biosynthesis of SLS and related natural products.

2.1 Introduction

The ribosomally synthesized and post-translationally modified peptides (RiPPs) comprise a rapidly expanding class of natural products that includes a wide variety of structural modifications.[2] These modifications impart RiPPs with diverse activities, giving rise to a range of products from

antibacterials[3-5] to anticancer agents.[6] The installation of azole and/or azoline heterocycles is one such modification common to many RiPPs, forming a sub-class of natural products called the thiazole/oxazole-modified microcins (TOMMs).[7] The azoles are biosynthesized by the cyclodehydration and subsequent dehydrogenation of cysteine, serine, and threonine residues to form thiazole and (methyl)oxazole rings on the C-terminal portion, or "core", of a ribosomally produced precursor peptide.[7] The azole-containing peptides will often undergo further processing, including the proteolytic removal of the N-terminal "leader" portion of the peptide and export of the mature product.[8] Although recent discoveries have shed light on the mechanism of azole formation,[9-11] the proteolytic processing step of most TOMMs has yet to be investigated.

Streptolysin S (SLS), a key virulence factor of *Streptococcus pyogenes*, is one such TOMM whose biosynthesis is incompletely understood.[12] *S. pyogenes* is the causative agent of diseases ranging in severity from pharyngitis to necrotizing fasciitis[13] and is a major global health burden, causing over 600 million infections and 500,000 deaths annually.[14] SLS is the cytolytic toxin responsible for the classic β-hemolytic phenotype when *S. pyogenes* is grown on blood agar[15] and has been shown to be critical to pathogenesis in mammalian infection models.[16-18] Although a few strains of non-β-hemolytic, pathogenic *S. pyogenes* have been described, such as the Lowry strain,[19] the vast majority of *S. pyogenes* isolates produce SLS.[20] The toxin is biosynthesized by a 9-gene biosynthetic operon that encodes the precursor peptide (*sagA*), cyclodehydratase and dehydrogenase enzymes (*sagBCD*), a putative leader peptidase (*sagE*), a multicomponent ABC-transporter (*sagGHI*), and a protein of unknown function (*sagF*) (Figure 2.1A).[15] The SagBCD heterocycle synthetase is known to install azol(in)e rings on the core region of the precursor peptide,[21, 22] which is followed by proteolytic removal of the leader peptide prior to export of the mature, bioactive natural product (Figure 2.1B,C). The final molecular weight of SLS has been inferred from classic gel filtration studies to be 2.8 kDa,[23] which is consistent with the bioinformatic prediction of the scissile bond being C-terminal to a Gly-Gly motif based on the similarity to bacteriocins from other Gram-positive bacteria.[24] Additionally, proteolysis following small residues is common in RiPPs with known cleavage sites.[8] Although SLS was first defined nearly 80 years ago,[25] with the β-

hemolytic phenotype being known since the late 1800s,[26] a detailed mechanism of SLS biosynthesis and the structure of the mature toxin remain elusive. Thus, effective SLS biosynthetic inhibitors could serve as powerful chemical tools to shed more light on the biochemistry of SLS and the infection biology of *S. pyogenes*.

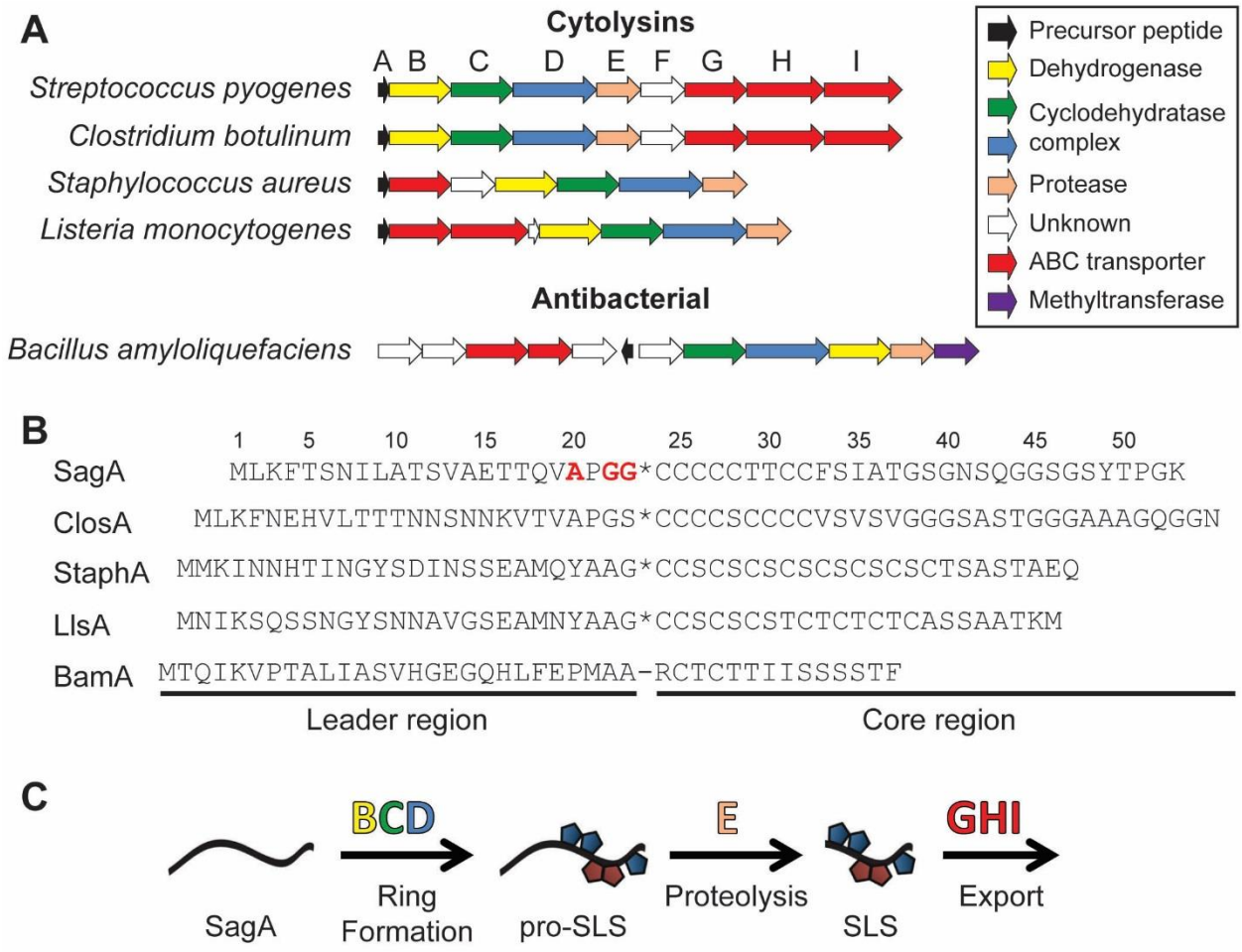


Figure 2.1. TOMM gene clusters and the biosynthetic pathway. (A) Open reading frame diagram showing the organization of several TOMM gene clusters, grouped by function of the mature TOMM. The letters over each gene correspond to the name of the *sag* genes in *S. pyogenes*. The function of each gene is color-coded in the legend. (B) Sequences of the precursor peptides from the clusters shown in panel A. The putative (*) or known (-) cleavage sites are shown. The residues substituted in SagA to generate SagA-VLPLL are shown in red. SagA, from *S. pyogenes*; ClosA, from *C. botulinum*; StaphA, from *S. aureus*; LlsA, from *L. monocytogenes*; BamA, from *B. amyloliquefaciens*. (C**)** Generalized mechanism of TOMM maturation with SLS as the example. The proteins putatively responsible for each step are shown above each arrow.

Gaining a better understanding of how pathogens employ their various virulence factors also aids in the development of selective treatment strategies that could help to increase the lifespan of clinically important antibiotics. Unlike traditional broad-spectrum antibiotic treatment, specifically targeting virulence would retain the human microbiota, helping to eradicate secondary infections and, in some cases, could theoretically reduce the evolutionary pressure for the development of resistance.[27] Previous approaches to targeting virulence have included disruption of quorum sensing, which often regulates virulence factor expression, blocking toxin delivery or function, and inhibition of bacterial adhesion.[27] Several compounds designed around these approaches have been efficacious *in vivo* and have prompted further study,[27, 28] but the selective nature of targeting virulence requires a tailored therapy for each pathogen, which when developed, would stimulate vast improvements in clinical diagnostics. SLS is an interesting anti-virulence target, as it plays a major role in paracellular invasion, immune evasion, and host-metabolism manipulation.[12, 29] Furthermore, additional roles of SLS in iron acquisition have been suggested but require further confirmation.[12] Due to the importance of SLS in a multitude of pathogenic processes, chemical inhibitors developed to probe the biosynthesis of SLS may find future roles in virulence attenuation strategies that protect the vital microbiota and limit the spread of antibiotic resistance in other pathogens that employ SLS-like toxins (*i.e. S. pyogenes* and specific strains of *Listeria monocytogenes*, *Clostridium botulinum*, and *Staphylococcus aureus*).

In this study, we identified inhibitors SLS biosynthesis in *S. pyogenes* by searching for compounds that block an essential proteolytic maturation step. This proteolysis event has been proposed to be performed by SagE, a putative peptidase with homology to a large family of proteases referred to as the CaaX proteases and bacteriocin-processing enzymes (CPBP; often confusingly annotated as abortive infection proteins, Abi), which includes the eukaryotic type II CaaX proteases as well as prokaryotic proteins with putative bacteriocin-related functions.[30, 31] The type II CaaX proteases are involved in the processing of a number of C-terminally prenylated proteins in eukaryotes and have been much more thoroughly studied than their prokaryotic counterparts.[32-34] Type II CaaX proteases were initially believed to be cysteine proteases,[35] but the presumed catalytic cysteine was later proven to be unnecessary for

activity.[34] In contrast, conserved glutamate and histidine residues were shown to be important for activity, leading to the hypothesis that type II CaaX proteases were metalloproteases.[34] Many of the prokaryotic members of the CPBP family, including SagE, have been annotated as immunity proteins due to the role in bacteriocin self-immunity of the family members in *Lactobacillus plantarum* and *L. sakei*.[36] However, the production of viable allelic exchange mutants of *sagE* and the homolog in *Listeria monocytogenes* (*llsP*) indicates that SagE may not be serving an immunity role or may be redundant with other uncharacterized immunity mechanisms.[15, 37] Thus, it is plausible that the prokaryotic CPBPs actually act as proteases and that this function is exploited to provide self-immunity in certain cases. The results presented herein support a role for SagE as a protease through the discovery and characterization of a family of small molecule inhibitors of SLS biosynthesis. Appealingly, these inhibitors were identified through the repurposing of existing drugs by an examination of known off-target effects. This approach facilitated the rapid identification of lead compounds without the need to perform expensive and laborious high-throughput screens, as well as aiding in the synthesis of analogs by leveraging previous work on the compounds.[38-40]

2.2 Evidence for the role of SagE as a protease

Reconstitution of SagE *in vitro* would be the most direct means of testing for the predicted protease activity and for the screening of inhibitors. Unfortunately, numerous attempts to heterologously express SagE proved unsuccessful and alternative assays were developed to address this issue. We attribute much of the difficulty in expressing SagE to the predicted transmembrane nature of the protein; topology modeling of SagE with SPOCTOPUS[41] predicted five transmembrane helixes and an N-terminal signal peptide (Figure 2.2). With this in mind, crude membranes from *S. pyogenes* were prepared from total cellular lysates by ultracentrifugation. The resultant samples were then assessed for proteolytic activity towards ^{35}S-labeled SagA, generated through *in vitro* transcription/translation. Robust proteolysis of SagA to the predicted molecular weight[23] was observed after treatment with the membrane fraction (Figure 2.3A). SagA processing in the whole cell lysate and supernatant fractions was much less extensive and often not observable (Figure 2.4). To test the substrate specificity of proteolysis, a mutant

version of SagA with residues Ala20, Gly22, and Gly23 mutated to leucines (SagA-VLPLL) was generated (Figure 2.1B). These residues are directly N-terminal to the predicted leader peptide cleavage site and were expected to be important for protease recognition. Pro21 was left intact to avoid inducing a drastic structural change on the peptide. When treated with isolated *S. pyogenes* membranes, SagA-VLPLL was processed at a rate slower than wild-type SagA (Figure 2.3A), suggesting that the cleavage was performed by a membrane protease specifically recognizing the SagA cleavage site (other membrane-bound proteases may also be minor contributors to the observed proteolysis).

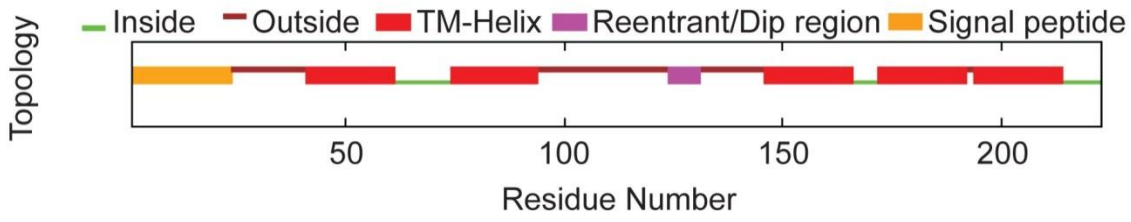


Figure 2.2. Hydropathy plot of SagE. The membrane topology of SagE was modeled using the SPOCTOPUS program[41] and is predicted to be an integral membrane protease with 5 transmembrane helices. The eukaryotic homologue Rce1 from *Methanococcus maripaludis* is known to also be an integral membrane protein with eight transmembrane α-helices (PDB ID 4CAD).[42]

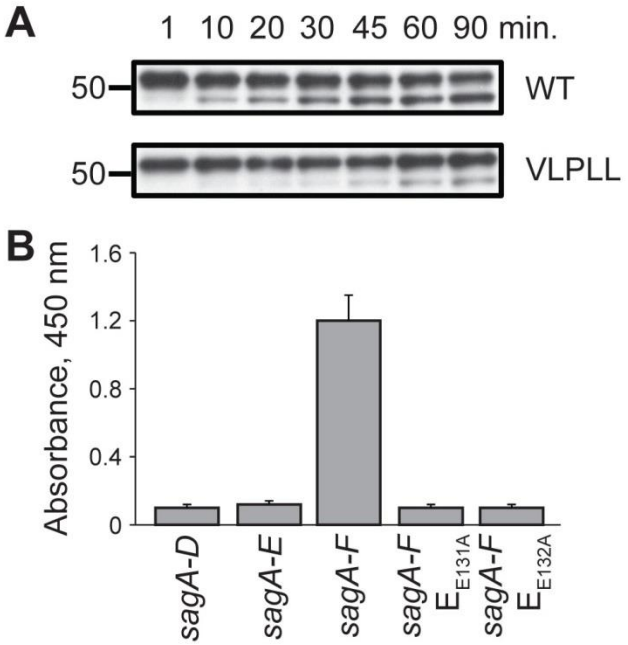


Figure 2.3. The role of SagE in the processing of pre-SLS. (A) Purified *S. pyogenes* membranes display proteolytic activity towards SagA. Both wild type (WT) SagA, and to a reduced extent, the predicted cut site mutant SagA-VLPLL, function as cleavage substrates. (B**)** Lytic activity of *E. coli* expressing the SLS biosynthetic machinery after a 2 h induction. The hemolytic activity of extracted SLS on erythrocytes is measured by a colorimetric readout of hemoglobin release.

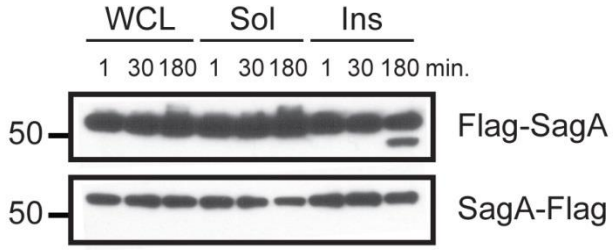


Figure 2.4. Cleavage of Flag-tagged SagA with the *S. pyogenes* membrane fraction. Proteins present within *S. pyogenes* membranes harbor proteolytic activity towards MBP-Flag-SagA (Flag tag is on the N-terminus of SagA). A cleavage band is not seen for MBP-SagA-Flag as the Flag tag here is appended to the much smaller C-terminal core peptide, which migrates with the dye front. WCL, whole cell lysate; Sol, soluble fraction; Ins, insoluble (crude membrane) fraction.

To provide additional evidence for the role of SagE, a previously reported multi-plasmid-based expression system for the generation of SLS in *Escherichia coli* was used. In that study, a lytic entity was generated by the expression of a maltose-binding protein (MBP)-tagged SagA in conjunction with SagB-D from pETDuet vectors after extended induction times.[43] SLS produced in this system could be extracted with bovine serum albumin (BSA) and applied to blood using procedures long established for *S.*

pyogenes.[44] The heterologously produced SLS was presumably exported by generalized transporters after non-specific proteolysis of the MBP tag or was released via *E. coli* cell death following buildup of the toxin. Using this system, no lytic activity was observable after a considerably shorter induction time (2 h) from a strain expressing only SagA-D; however, *E. coli* expressing SagA-F was highly lytic after 2 h (Figure 2.3B), indicating that SagEF expedite SLS maturation. The additional inclusion of SagF was required for this effect, although its functional role is unknown. SagF shares no homology to any known proteins but has previously been demonstrated to be necessary for SLS biosynthesis[15] and may aid in the folding or localization of the other modification machinery. To support the role of SagE as a protease involved in SLS maturation, two highly conserved glutamine residues (Glu131 and Glu132) that are critical for activity in mammalian CPBP family members[31, 34] were mutated to alanine, resulting in the complete loss of the observed lytic activity.

2.3 Aspartyl protease inhibitors block SagA proteolysis

A small panel of general mechanism-based protease inhibitors was screened for inhibition of SagA leader proteolysis using the membrane cleavage assay with *S. pyogenes* membranes. This assay was preferred over the *E. coli* multi-plasmid system to avoid potential issues with membrane penetration or general toxicity of the inhibitors. From the panel, only the aspartyl protease inhibitor pepstatin displayed inhibitory activity (Figure 2.5), which was unexpected given that SagE bears no similarity to known aspartyl proteases and that members of the type II CaaX protease family were previously hypothesized to function as zinc-dependent metalloproteases.[34] However, inhibition of a metalloprotease by aspartyl protease inhibitors has literature precedent, as several inhibitors of HIV protease were found to also inhibit the type I CaaX protease ZMPSTE24, which is a known zinc metalloprotease.[45-47] Lipodystrophy is a possible side effect of treatment with certain HIV protease inhibitors and also occurs from genetic deficiencies in ZMPSTE24, leading to the discovery of this off-target effect for these drugs.[45] Although type I and II CaaX proteases do not share sequence similarity, they redundantly process some of the same substrates and share similar substrate-binding site architectures.[42, 48] Given this, we reasoned that HIV protease inhibitors might be repurposed as inhibitors of SLS production.

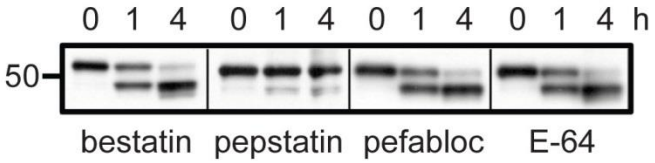


Figure 2.5. Inhibition of SagA proteolysis by general-mechanism based inhibitors. Proteolysis of WT [^{35}S]-MBP-SagA by *S. pyogenes* membranes treated with general mechanism-based protease inhibitors. Bestatin, metalloproteases; pepstatin, aspartyl proteases; pefabloc, serine proteases; E-64, cysteine proteases.

2.4 HIV protease inhibitors block SLS production

A whole-cell assay in *S. pyogenes* based on the extraction of SLS with BSA was used to screen a panel of nine FDA-approved HIV protease inhibitors (Figure 2.6). Nelfinavir, ritonavir, saquinavir, and lopinavir were found to inhibit the production of SLS when tested at 50 μM, while indinavir, amprenavir, atazanavir, and darunavir did not inhibit SLS production (Figure 2.7A). Interestingly, tipranavir caused significant growth suppression in *S. pyogenes*, and treated cultures never reached late exponential phase (Figure 2.8), which is when SLS becomes detectable *in vitro*.[49] The efficacies of the HIV protease inhibitors shown to inhibit SLS production in the initial screen were evaluated by determining 50% inhibition concentration (IC_{50}) values. The IC_{50} of nelfinavir (6 μM) was much lower than those of ritonavir (35 μM), saquinavir (25 μM), and lopinavir (25 μM). Owing to its greater potency, nelfinavir (Figure 2.7B) was selected for further study.

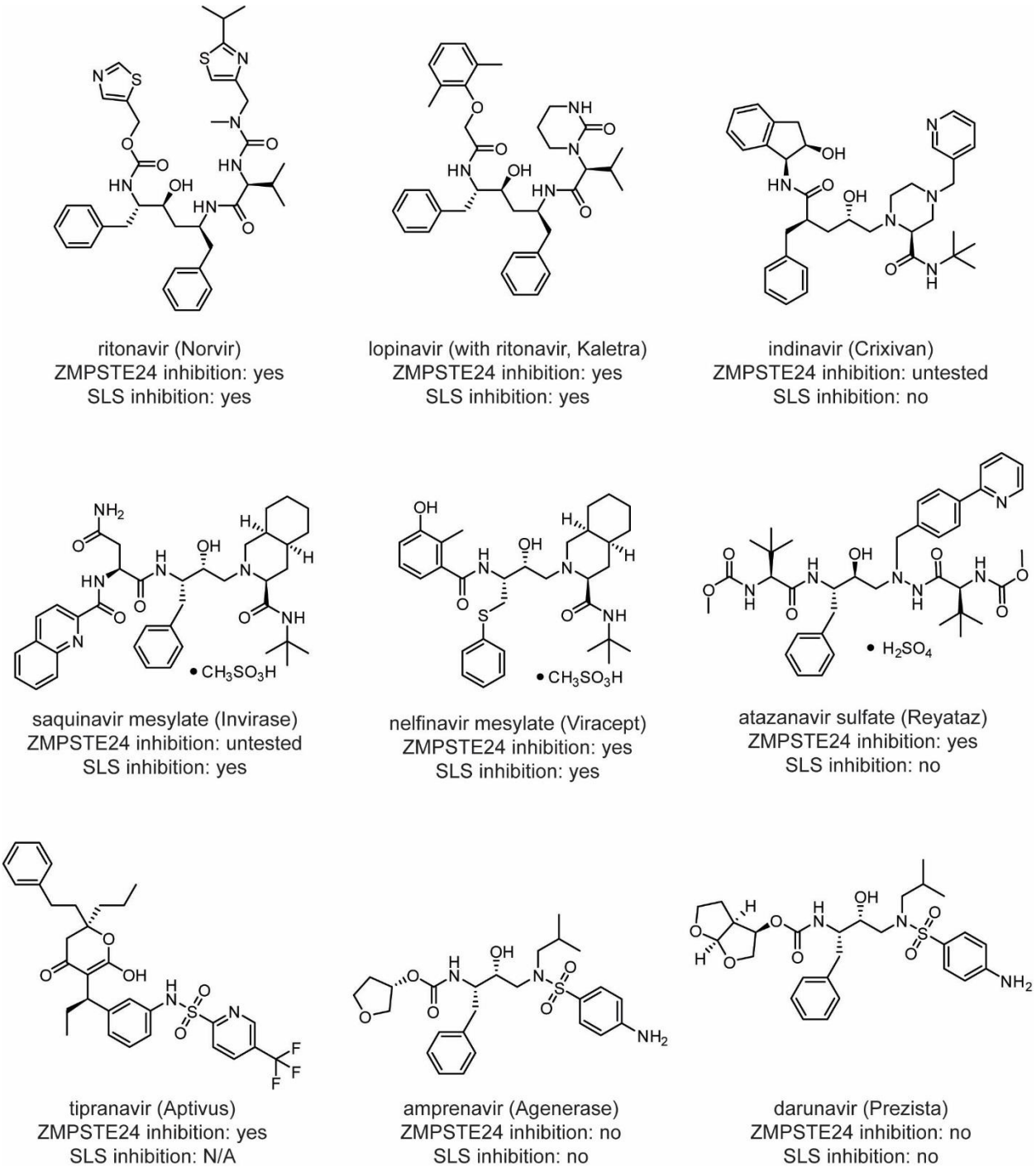


Figure 2.6. Structures of the FDA-approved HIV protease inhibitors. The structures of all FDA-approved HIV protease inhibitors are shown (U.S. brand name in parentheses) with the exception of fosamprenavir calcium (Lexiva), a pro-drug of amprenavir. Any previously reported activity for each drug on the CaaX protease ZMPSTE24 is listed,[46, 47] along with the activity for SLS production inhibition.

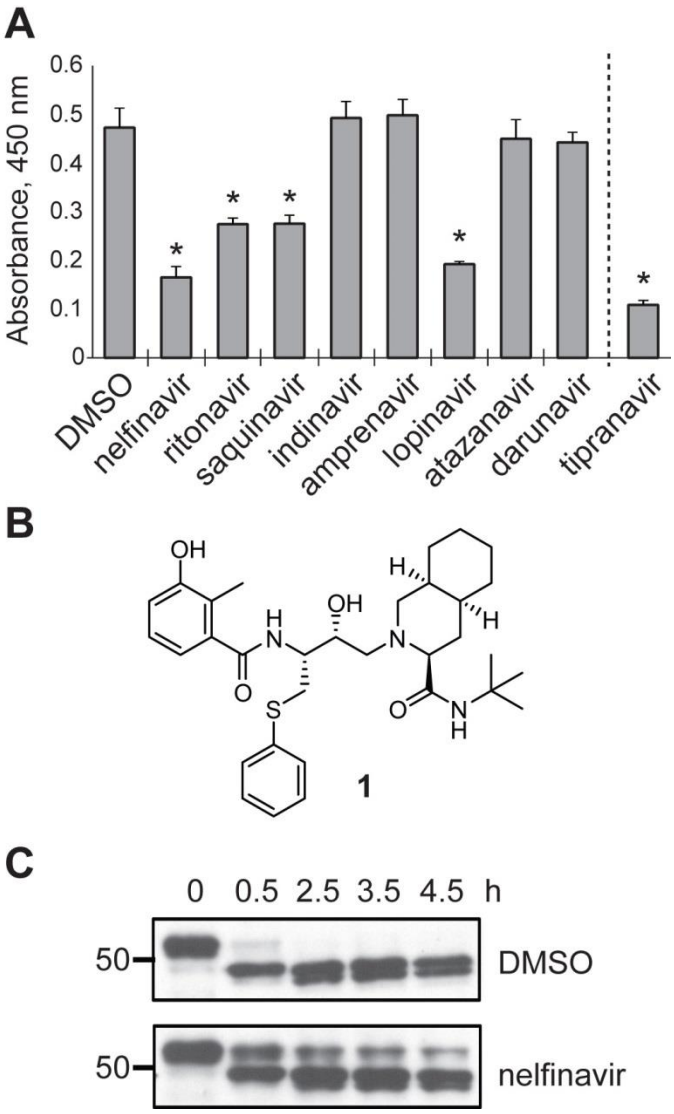


Figure 2.7. Inhibition of SLS production. **(**A) Hemolytic activity of SLS extracts from *S. pyogenes* treated with HIV protease inhibitors. Tipranavir is shown separated by a dashed line due to significant growth effects during treatment (Figure 2.8). Asterisks indicate a P-value < 0.01 relative to the DMSO control. (B) The structure of nelfinavir (**1**). (C) The proteolytic effect of *S. pyogenes* membranes treated with nelfinavir on MBP-SagA, relative to a DMSO-treated control.

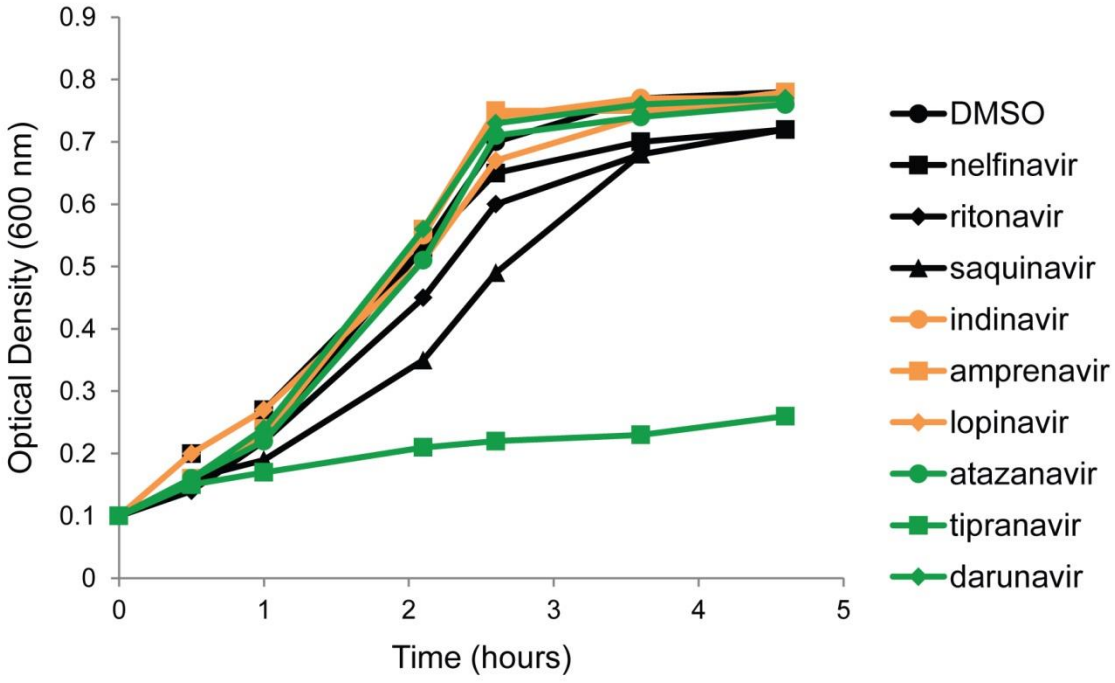


Figure 2.8. Growth effects of HIV protease inhibitors on *S. pyogenes*. *S. pyogenes* M1 5448 was grown in the presence of the indicated HIV protease inhibitor at 50 μM and the optical density at 600 nm was recorded. Tipranavir was found to significantly inhibit growth.

Nelfinavir was evaluated in the membrane proteolysis assay to determine if the observed loss of β-hemolysis was due to inhibition of the proteolytic processing step of SLS maturation. As expected, treatment with nelfinavir greatly reduced the proteolytic activity toward SagA contained within *S. pyogenes* membranes (Figure 2.7C, the membrane fraction used in this experiment was prepared on a different day than the fraction used in Figure 2.3 and displayed much higher activity). Evidence that nelfinavir was not drastically perturbing normal cellular function came from the observation that the growth rates of *S. pyogenes* treated with nelfinavir were identical to the DMSO control (Figure 2.8). Additionally, minimum inhibitory concentration (MIC) testing revealed no growth inhibition up to the highest concentration tested (64 μM) for a range of bacterial species (Table 2.1). Transmission electron microscopy (TEM) was used to examine the morphology of *S. pyogenes* treated with nelfinavir, with no apparent changes compared to the control sample (Figure 2.9). The transcription levels of a panel of virulence factor genes, as assessed by qRT-PCR, were also not significantly impacted (Table 2.2).

Notably, the levels of *sagA* and *sagB* expression were unchanged. These data indicate that nelfinavir inhibited peptide processing directly, rather than through transcriptional regulation or by significantly perturbing other cellular processes.

Bacterial strain	**MIC (μM)**
Streptococcus pyogenes M1 5448	>64
Listeria monocytogenes 4b F2365	>64
Clostridium sporogenes ATCC 19404	>64
Staphylococcus aureus USA300	>64
Enterococcus faecium U503	>64
Bacillus subtilis 168	>64
Klebsiella pneumoniae ATCC 27736	>64
Pseudomonas aeruginosa PA01	>64
Escherichia coli MC4100	>64

Table 2.1. The minimum inhibitory concentrations of nelfinavir against a range of bacterial species, including Gram-positive and Gram-negative organisms, are shown. No growth inhibition was seen for any species up to the solubility limit of nelfinavir after 16 h of growth.

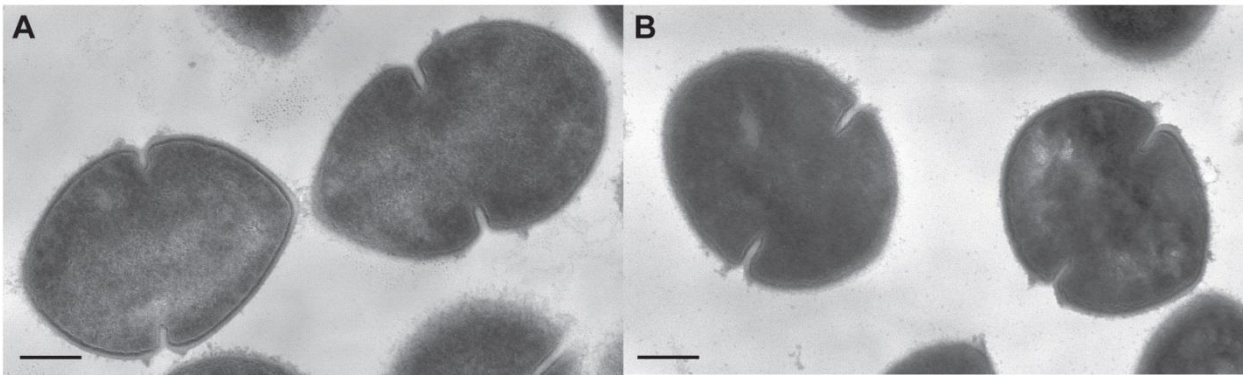


Figure 2.9. Transmission electron microscopy (TEM) of nelfinavir-treated *S. pyogenes*. TEM images of *S. pyogenes* M1 5448 treated with (A) DMSO or (B) 25 μM nelfinavir. Scale bars are 100 nm. These images are representative of the entire sample, where more than 100 cells were closely examined.

Gene	Log_2 ratio (treated/untreated)	Virulence factor
emm49	-1.98 ± 0.77	M protein
slo	0.02 ± 1.22	streptolysin O (SLO)
sagA	0.32 ± 0.20	streptolysin S (precursor peptide)
sagB	0.10 ± 0.70	streptolysin S (dehydrogenase)
scpA	-0.86 ± 0.26	C5A peptidase
ska	-1.23 ± 0.83	streptokinase
speB	1.38 ± 0.92	streptococcal pyrogenic exotoxin B (SpeB)
nga	-1.30 ± 0.90	NAD glycohydrolase
spd3	-0.83 ± 0.28	streptodornase

Table 2.2. Changes in virulence factor expression as measured by qRT-PCR with *S. pyogenes* M49 NZ131 during treatment with nelfinavir versus a DMSO control. Values are from three biological replicates averaged from three technical replicates each. The 16S rRNA gene was used as an internal control gene and relative gene expression was calculated by a comparative C_T method.[50] Positive and negative values represent up- and down-regulation upon nelfinavir treatment, respectively. The primers used are given in the supplemental methods. A transcription change was not considered significant if the log_2 ratio was < 1.5 fold or if the P value was > 0.01. Based on these criteria, none of the genes were significantly altered upon nelfinavir treatment. Similar results were obtained for an additional strain, *S. pyogenes* M1 5448.

2.5 Structure-activity relationships

A series of nelfinavir analogs was synthesized to gain a better understanding of the structure-activity relationships (SAR) for inhibition of SLS maturation. If nelfinavir were to interact with the SLS leader peptidase in a manner analogous to HIV protease, the secondary hydroxyl group would function as a tetrahedral intermediate mimic, as is the case for many aspartyl and metalloprotease inhibitors (Figure 2.10). Synthetic routes to nelfinavir have been thoroughly explored and a route that allowed facile derivatization was adapted for this study (Scheme 2.1).[38] The efficacy of each analog at 25 μM was evaluated using the hemolysis assay (Table 2.3 and Table 2.4).

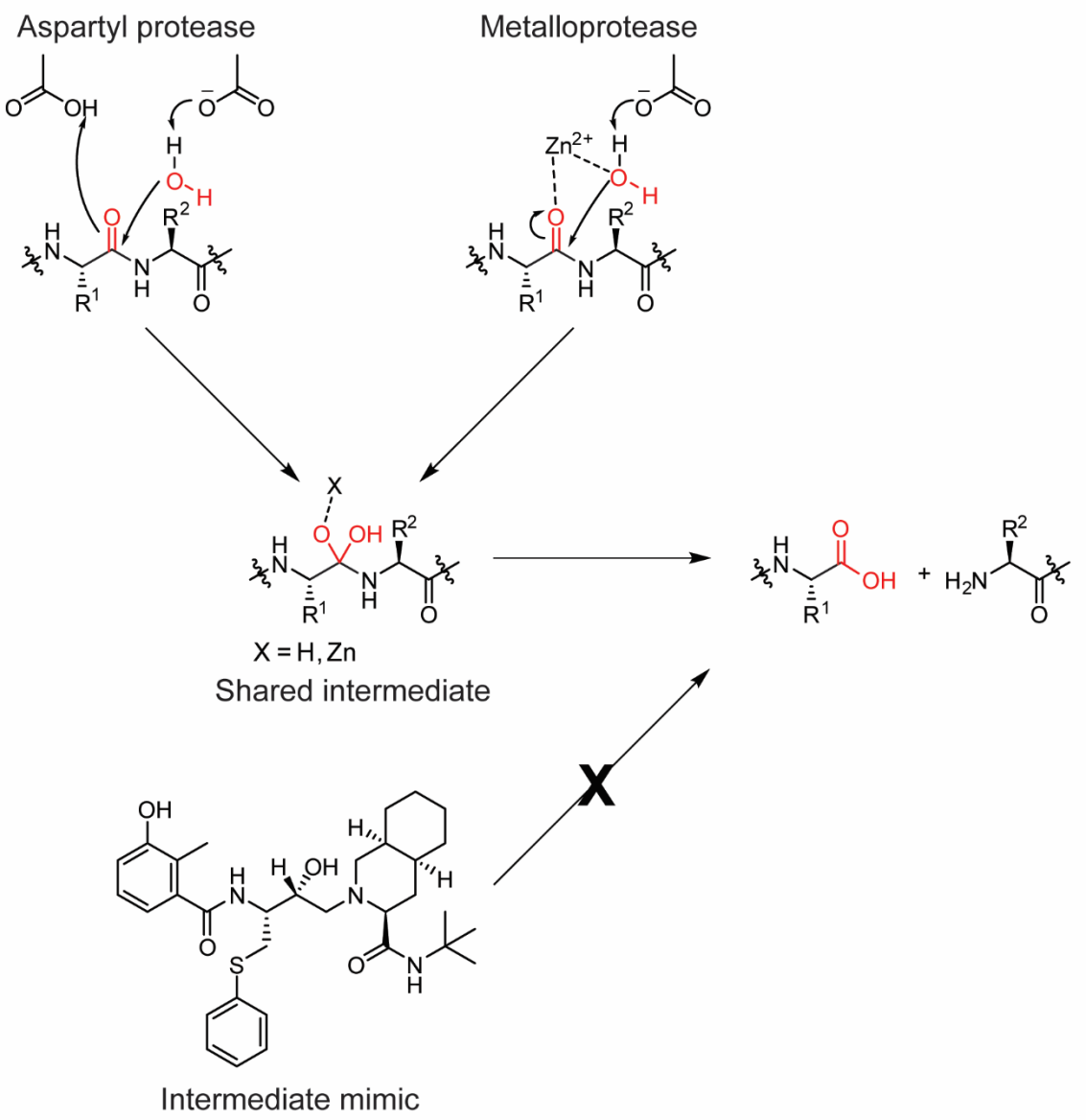
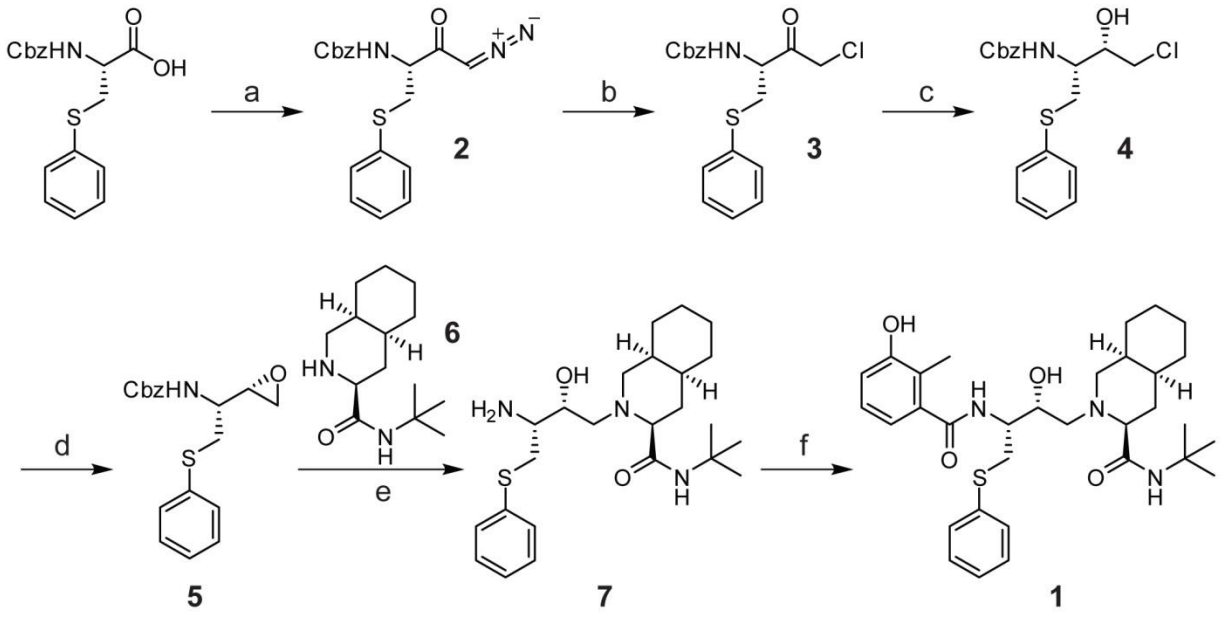


Figure 2.10. Aspartyl and metalloproteases share a common intermediate. Both types of protease activate water, which performs a nucleophilic attack on the scissile amide bond. This attack results in a *gem*-diol intermediate shared between the two mechanisms, unlike the covalent acyl-enzyme intermediate formed by serine, threonine, and cysteine proteases. Nelfinavir and other aspartyl and metalloprotease inhibitors usually mimic the intermediate with the non-hydrolyzable secondary hydroxyl group. However, our data suggest that nelfinavir does not inhibit the SLS leader peptidase in this fashion.

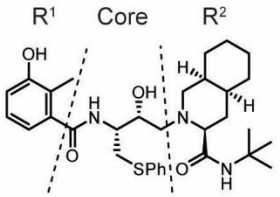


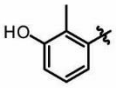
Scheme 2.1. Synthesis of nelfinavir. **a**, *i*-BuOCOCl, Et_3N, THF; CH_2N_2, Et_2O (62%); **b**, HCl, Et_2O; **c**, $NaBH_4$, THF (62% over two steps); **d**, KOH, EtOH (91%); **e**, **6**, KOH, IPA, 80 °C (84%); **f**, 3-hydroxy-2-methylbenzoic acid, EDC, HOBt, THF (66%).

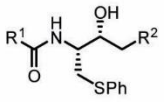
Nelfinavir has peptidomimetic features, including an *S*-phenyl-ethyl group intended to replace the side chain of phenylalanine commonly found in the P1 position of the HIV protease substrate. Since the putative P1 position of SagA is a much smaller glycine residue (Figure 2.1B, Gly23), we surmised that smaller substituents at this location on nelfinavir might improve the observed anti-SLS activity. Instead, removal of the side chain (**8**) abolished detectable activity. It is possible that the cleavage site is closer to the N-terminus, however, and that Pro21 or an aliphatic residue preceding it resides in the P1 position. Accordingly, analogs mimicking these amino acids (**9, 10**) were synthesized, but these compounds did not exhibit detectable inhibitory activity. Even retention of the phenyl ring in a phenylalanine mimic (**11**) was insufficient to maintain activity. Conversely, a tryptophan mimic (**12**) was inhibitory, albeit weakly, suggesting that a relatively large group in that position may be necessary for activity.

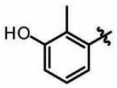
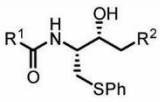
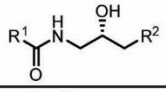
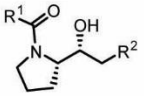
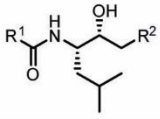
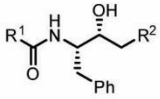
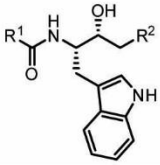
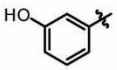
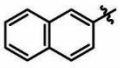
The lack of detectable activity with the series of P1 position analogs prevented the rigorous establishment of SAR. Modifications to other portions of the molecule were prepared in order to enhance the inhibitory activity to address this pitfall. As installation of the benzamide is the final step of the

synthesis (Scheme 2.1), preparation of analogs at this location was convenient. Initial analogs revealed that neither the hydroxyl nor the methyl groups (**13–15**) were important for activity but that the ring itself was necessary (**16**). Replacement of the ring with bulkier naphthyl and cyclohexyl groups (**17, 18**) provided >5-fold increases in potency, indicating that this part of the pharmacophore may reside in a hydrophobic pocket not fully occupied by the single planar ring. However, increasing the size of this moiety to an anthracene (**19**) greatly reduced activity. In general, electron-deficient rings had improved activity relative to electron-rich rings (Table 2.4), although hydrophobicity appeared to be a more significant contributing factor towards potency.



R^1 = A = 

Core = B = 

Compound	R^1	Core	Relative Activity[a]
nelfinavir (**1**)	 A	 B	++
8	A		-
9	A		-
10	A		-
11	A		-
12	A		+
13		B	++
14		B	++
15		B	++
16		B	+
17		B	+++
18		B	+++

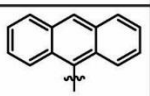
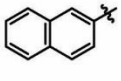
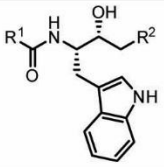
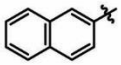
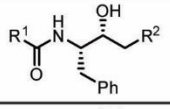
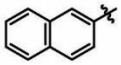
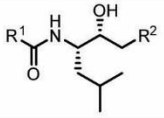
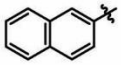
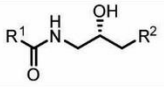
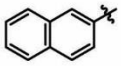
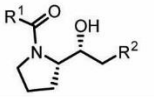
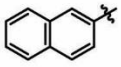
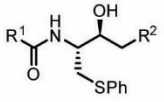
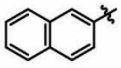
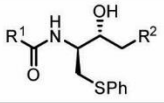
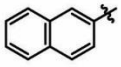
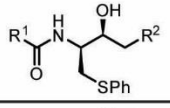
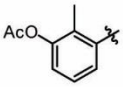
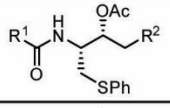
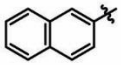
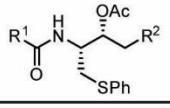
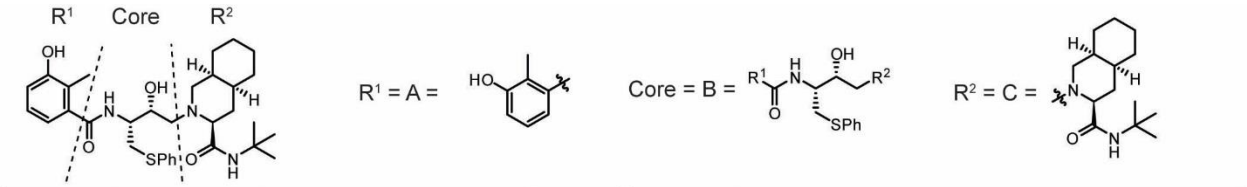
Compound	R^1	Core	Relative Activity[a]
19		B	+
20			+++
21			++
22			++
23			+
24			-
25			++
26			+++
27			+++
28			+++
29			++

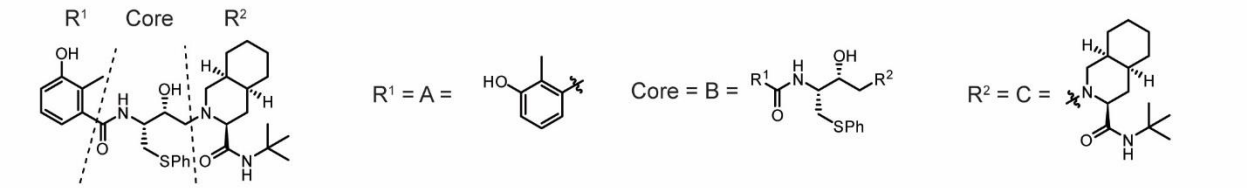
Table 2.3. Relative activity of nelfinavir analogs for SLS inhibition.
[a]The activity of each analog is reported qualitatively relative to nelfinavir due to variability in commercial blood lots and extraction effectiveness. Inhibitory activity that is >3-fold than nelfinavir is designated as (+++); activity that is equal to nelfinavir is (++); detectable activity that is <3-fold than nelfinavir is (+); non-detectable activity is denoted (-).

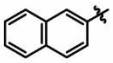
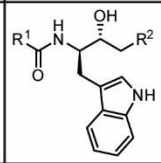
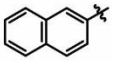
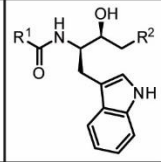
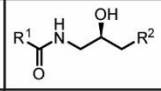
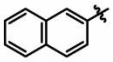
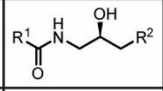
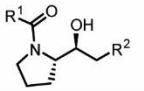
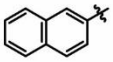
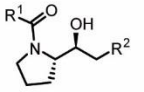
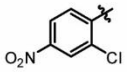
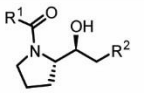
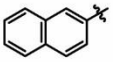
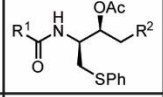
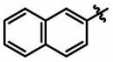
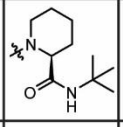
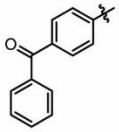


Compound	R^1	Core	R^2	Relative Activity
nelfinavir (1)	A	B	C	++
74		B	C	++
75		B	C	+++
76		B	C	+++
77		B	C	++
78		B	C	++
79		B	C	+++
80		B	C	+++
81		B	C	+++
82		B	C	++
83		B	C	+
84		B	C	++
85		B	C	++
86		B	C	++
87		B	C	+

Compound	R^1	Core	R^2	Relative Activity
88		B	C	+
89			C	+
90			C	-
91	A		C	-
94	A		C	-
95			C	++
102	A		C	-
103			C	+++
106	A		C	-
107			C	+
114			C	++
121	A		C	-

Table 2.4. Relative activity of additional nelfinavir analogs for SLS inhibition. The activity of each analog is reported qualitatively relative to nelfinavir due to lot variation in the commercial sources of blood and the BSA extraction efficiency. Activity >3-fold that of nelfinavir is designated as (+++); activity that is equal to nelfinavir is (++); detectable activity that is <3-fold that of nelfinavir is (+); non-detectable activity is denoted (-).



Compound	R1	Core	R2	Relative Activity
122			C	+++
125			C	-
126	A		C	-
127			C	+
131	A		C	-
132			C	-
133			C	-
134			C	+
139		B		++
140		B	C	-

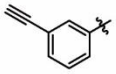
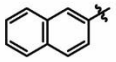
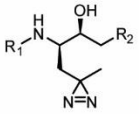
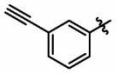
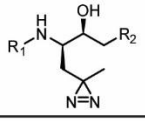
Compound	R1	Core	R2	Relative Activity
141		B	C	++
149			C	++
150			C	+

Table 2.4 (cont.)

Given the enhanced activity of the naphthylamide-bearing compound (**17**), this group was incorporated into the collection of P1 position analogs (Table 2.3; Table 2.4), increasing the activity of

these analogs to detectable levels. The tryptophan mimic with the naphthylamide (**20**) displayed considerably increased potency relative to the 3-hydroxy-2-methylbenzamide analog (**12**). Within the naphthylamide series, the phenylalanine and leucine mimics (**21, 22**) were equivalently active to nelfinavir, while the glycine mimic (**23**) displayed weak SLS-inhibitory activity. The proline mimic (**24**) remained devoid of activity, possibly due to structural perturbations enforced by the ring. These data support a trend of larger substituents imparting higher activity that was foreshadowed by the initial SAR.

To further probe the pseudo-P1 position of nelfinavir, the stereochemical configurations of both the side chain and the secondary hydroxyl group were varied. Surprisingly, inverting stereocenters with this series of analogs (**25**–**27**) did not result in large changes in activity. Many of the derivatives were still highly potent, suggesting that nelfinavir may not inhibit SagE through mimicking the proteolytic tetrahedral intermediate (Figure 2.10). This is further supported by the retention of activity in acetylated derivatives (**28, 29**). Overall, the sum of the SAR analysis resulted in the development of an analog with significantly improved potency (**17**, IC_{50} = 1 μM). Additionally, the activity trends led us to believe that nelfinavir might serve as an inhibitor for the protease in other TOMM natural product biosynthetic clusters in which the precursor peptides do not always contain a predicted Gly-Gly cleavage motif (Figure 2.1B).

2.6 Biosynthesis inhibition of other TOMMs

SLS is the best-studied member of a group of related cytolysins produced by a number of bacteria, including pathogens such as *L. monocytogenes* and *C. botulinum*.[12, 21] The SLS-like biosynthetic gene clusters in these strains are highly similar to that of *S. pyogenes* and include orthologs of *sagE* (Figure 2.1A, Figure 2.11). The SLS-like toxin from *L. monocytogenes*, listeriolysin S (LLS), is known to be expressed during oxidative stress; therefore, a strain with LLS under the control of a constitutive promoter was used in the blood lysis assay to determine if nelfinavir could also inhibit LLS production.[51] The strain was deficient in production of an unrelated cytolysin, listeriolysin O (LLO), to ensure any hemolysis observed derived from LLS production. When treated with nelfinavir, this strain produced

significantly less LLS (Figure 2.12A). The LLO$^-$/LLS$^-$ strain of *L. monocytogenes* was included as a negative control to demonstrate the observed hemolysis was indeed LLS-dependent (Figure 2.12A).

```
SagE        --MPLSIQCLNLCFLLVTFCPSIPMQA--IFGKEDSGYAFNLI-GFLRATLIYDILA---  52
LlsP        ---------MHNIFFLITLFPGMLLLL----TKWIP-VLSRKS-TFFQYLL-----C---  37
StaphE      MIFHILWHSLIGLMFFLTCFPGILLIC----IRKVP-YFKYLT-EFHKYIL-----V---  46
ClosE       --MNIRTQCLDLLILLVTFDPIVLAQL--FLTKENS-YNESLERSFKALAIIYAI-I---  51
BamE        -MMEL-RACIFIGILIHLMGPYVSKLTVMYQIKHRT-KEYKLT-RHITPQRSYDIITVFG  56
                      :   :::    * :         :            .

SagE        ---LVSIYVLSPQITLSLESIDSKTFFMGLVFCVLIVLI--ELIFLHGLRCWQK-----K 102
LlsP        ---LFLITIMN------------SLFF-RQQFVVVLSLI--CILFLPFILFFVEYIFVER 79
StaphE      ---LIIVLIID------------MCLF-LSYINIYLCVF--VIIITPVIIYITEKSLL--  86
ClosE       ---LVGIFIISK------------EKFYINIDSNYHLYIM--AIIFVPITICWEI-----I 90
BamE        GIMCCSIFLID------------DNFFLIHSDNAYGVLLFSALGFFFPYV----------- 94
                  : ::.              :         : ::    ::

SagE        QWL-----------PATFSFVGTTNDWSKIGYPLLLALF---EETIYRFLWFNI-LAFQW 147
LlsP        QWKKLLT-------------IYKKNKIF-IQSIVWFPVL---EEIIFRFFIYQYCELFDF 122
StaphE      FNKNLFD-------------FYNKNSKFILSVLIAYPIL---EELNFRYFLYNFSLIYHY 130
ClosE       AGYIMLGVSNKAIPKFKISLVFTKTNINVIILSVFIGIC---EEFAYRQLWFNI-LFKDF 146
BamE        IWK-----------GKNKSVVFVKRKVNKLDTLIFLPLCAVSEELIWR-----ICIPFLF 138
                           .  . .   :   :   :    **  :*          :

SagE        HLPTI----IVLIVTSFCYALNHLLMGKSIFYAKLVTGIIYGSIYMLTSQLWLVV-IMHV 202
LlsP        SNIQY----ILLATFS--FVIAHIFYQGVSSIVKILFSFILSILFLLTLNIFLTI-IIHC 175
StaphE      HYIQY----IILSVLT--FVFSHFLYQGKSSIIKIFFATMQTLIYLCFRDIFINI-IVHK 183
ClosE       KMNIA----IVLIITSFVYALNHIFLGKQVFFQKMVAGLIYGALFYFSGFNIIVPIITHC 202
BamE        KLYFFDSVLLCTCISSIGFLLLHIPLGGVKSAPYMTVFTVIAVIA--AQFDFLSAGAFHI 196
                      :     :  : : *:          :    :   :        :   . *

SagE        GGNLLVECLSHLQTKKKKEVT--------      223
LlsP        IFNFLVYIVRTSKYENHRNW---------      195
StaphE      IFNILVFSSQMKKINYSIITK--------      204
ClosE       LENIVILLKG-------------------      212
BamE        VHNLTIQYFRPVSKKKINTSSPTMSQAEW      225
              *: :
```

Figure 2.11. Amino acid sequence alignment of select bacterial CaaX proteases and bacteriocin-processing enzymes (CPBPs). A Clustal Omega alignment of SagE and related proteases in other natural product gene clusters is shown. Conserved residues shown to be important for catalysis in eukaryotic homologues are highlighted in yellow and their proposed role in catalysis is shown in Figure 2.18.[34] The C-terminal 35 residues were manually aligned. Gene names and producing organisms: SagE, *S. pyogenes*; LlsP, *L. monocytogenes*; StaphE, *S. aureus*; ClosE, *C. sporogenes/botulinum*; BamE, *B. amyloliquefaciens*.

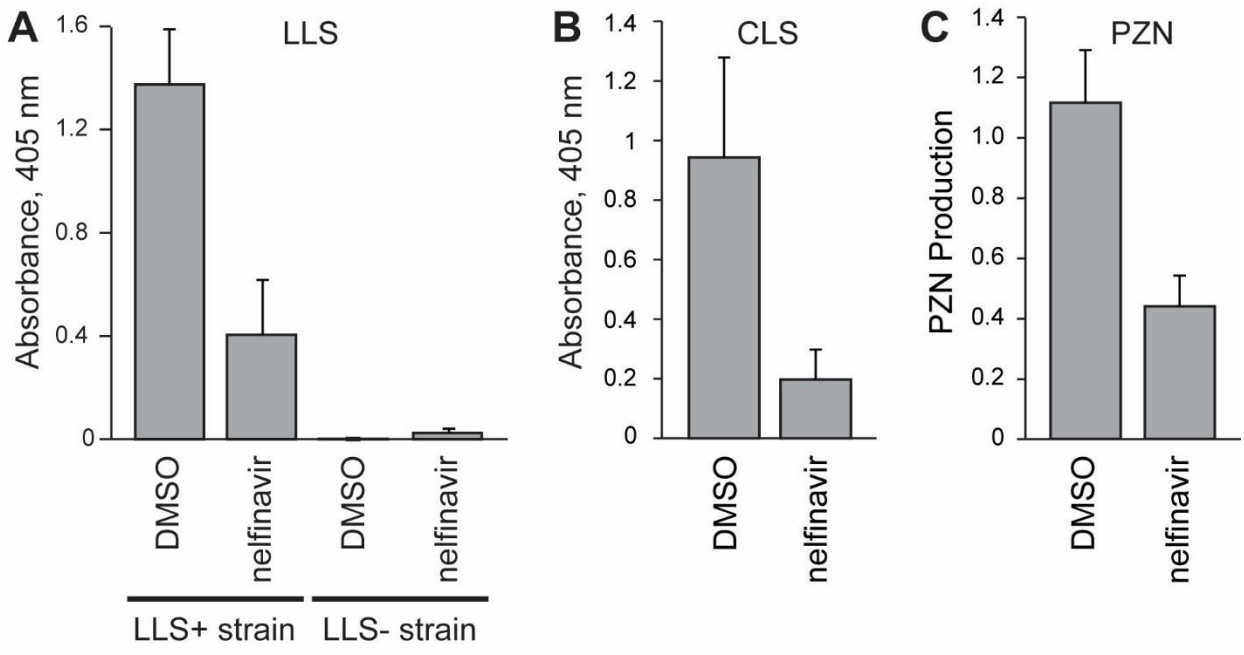


Figure 2.12. Inhibition of the biosynthesis of other TOMMs by nelfinavir. (A) Lytic activity of extracts from an LLS-producing strain (LLS+) of *L. monocytogenes* treated with nelfinavir relative to a DMSO control (n = 3). Extracts from a separate strain with *llsA* deleted (LLS-) were also used as a control. (B) Lytic activity of extracts from a CLS-producing strain of *C. sporogenes* treated with nelfinavir relative to a DMSO control (n = 3). (C) Production of PZN by *B. amyloliquefaciens* treated with nelfinavir relative to a DMSO control (n = 4). Nelfinavir was used at 50 μM in all cases. Excluding LLS- negative control, P-values < 0.05 were obtained for all nelfinavir-treated samples relative to the corresponding DMSO positive controls.

Similar to SLS and LLS, clostridiolysin S (CLS) is a hemolysin from *C. botulinum* as well as from certain strains of *C. sporogenes*, which are nearly identical to *C. botulinum* but do not produce botulinum toxin.[22] The presence of the CLS cluster in an unsequenced strain of *C. sporogenes* (ATCC 19404) known to be hemolytic on blood agar was confirmed by PCR amplification of *closC* and *closD* (the cyclodehydratase genes). When *C. sporogenes* was grown in the presence of nelfinavir, the production of CLS was significantly reduced (Figure 2.12B). The inhibition of not only SLS but also LLS and CLS production suggests that nelfinavir would likely inhibit the production of additional TOMM cytolysins.

TOMM biosynthetic gene clusters are wide-spread among bacteria and archaea, and the products of the vast majority of these clusters have not been structurally or functional characterized.[7] A

bioinformatic analysis of TOMM clusters revealed that 22% (328 out of 1520 as of October 2014) contained a CPBP family member within the cluster, many of which are predicted or known to have non-cytolytic products. The presence of a CPBP member in these clusters led us to postulate that nelfinavir would also inhibit TOMM production in these cases. Plantazolicin (PZN) is a TOMM produced by *Bacillus amyloliquefaciens* FZB42 with highly selective antibacterial activity against *Bacillus anthracis* and has a *sagE*-like gene (*bamE*) in its biosynthetic gene cluster (Figure 2.1A, Figure 2.11).[52] The structure of PZN has been determined; thus the precise cleavage site is known (after Ala27, Figure 2.1B), although no biochemical studies have directly linked BamE to leader peptide cleavage. Methanolic surface extracts from *B. amyloliquefaciens* were analyzed by liquid chromatography-mass spectrometry and the nelfinavir-treated cultures were found to produce significantly less PZN than a DMSO control (Figure 2.12C). Unlike the cytolysins, PZN can be readily observed and quantified by mass spectrometry, which permitted confirmation of the nelfinavir dose-dependent inhibition (Figure 2.13). The nelfinavir-dependent inhibition in divergent organisms of additional TOMM cytolysins, as well as a functionally distinct antimicrobial TOMM, not only supports the assignment of the CPBP protein being responsible for leader peptide removal during maturation, but also suggests that nelfinavir, and analogs thereof, could be generally useful for inhibition of TOMM production in a large number of hitherto uncharacterized clusters.

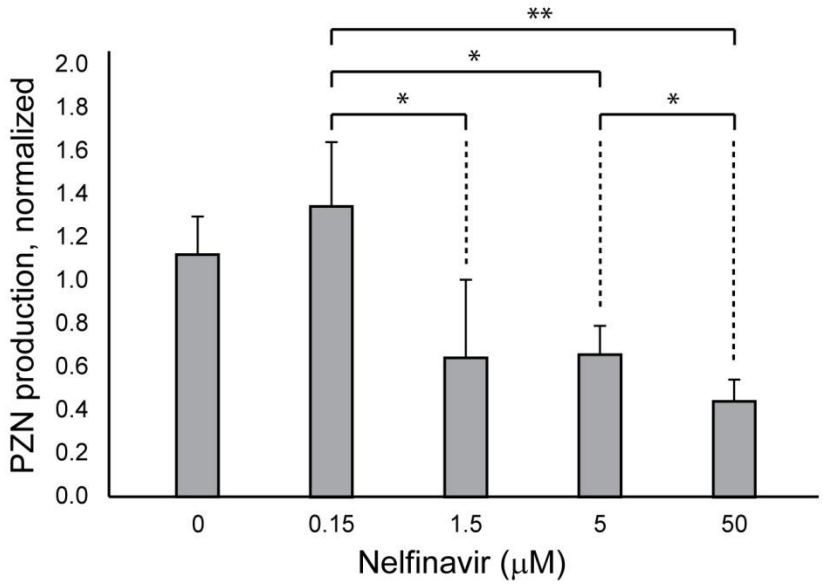


Figure 2.13. Dose-dependent inhibition of PZN by nelfinavir. The production of plantazolicin by *B. amyloliquefaciens* is reduced in a dose-dependent manner when treated with nelfinavir, relative to a DMSO control. *P<0.05, **P<0.005.

Additional CPBP family members are also present in bacteria beyond TOMM biosynthetic gene clusters. One such family member is PrsW from *Bacillus subtilis*, which is involved in cleaving the peptide RsiW (anti-σ factor) during a signaling cascade following cell envelop stress.[53] It is believed that PrsW is responsible for sensing antimicrobial peptides and other cell envelop stressing agents.[53] Since PrsW is a more divergent CPBP family member and is not expected to be involved in TOMM biosynthesis, we were interested in determining if it would also be inhibited by nelfinavir. Previously reported strains of *B. subtilis* containing an IPTG inducible, FLAG-tagged *rsiW* with and without *prsW* genetically deleted[53] were grown in the presence or absence of nelfinavir. After induction of *prsW*, the cells were treated with small amounts sodium hydroxide as a model cell envelop stressing agent. Rather than observing nelfinavir-dependent inhibition of PrsW, however, nelfinavir appeared to activate the protease even in the absence of sodium hydroxide (Figure 2.14). While this could indicate that nelfinavir is binding to the protease but resulting in activation instead of inhibition, it is also possible that nelfinavir is simply acting as a mild cell envelop stressing agent. This would result in non-specific activation of PrsW, complicating any attempts to determine the interaction between nelfinavir and the protease.

Nevertheless, it appears that the strong inhibitory activity of nelfinavir towards CPBP family members likely does not extend significantly beyond members in TOMM biosynthetic gene clusters.

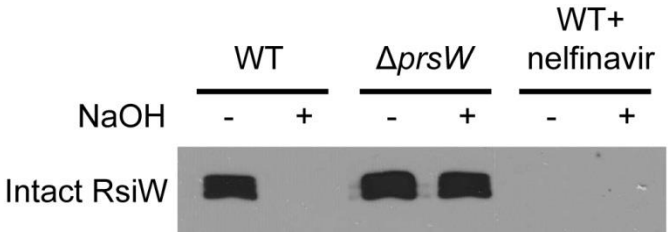


Figure 2.14. Cleavage of RsiW by PrsW. The CPBP family member PrsW cleaves RsiW under conditions of cell envelop stress, created here by sodium hydroxide. The WT strain is CDE743, containing a FLAG-tagged *rsiW* under IPTG controlled induction. The Δ*prsW* strain is CDE744, a modified version of CDE743 with *prsW* genetically deleted.

2.7 Target validation studies

Although the previous data strongly support SagE as the target of nelfinavir in SLS biosynthesis inhibition, we sought to further validate this hypothesis through affinity pulldown with a photoactivatable crosslinker probe. The previously described synthesis (Scheme 2.1) was utilized to install photoactivatable moieties as well as terminal alkynes for subsequent azide-alkyne cycloaddition click chemistry. As the most facile position to modify, replacement of the benzamide was initially investigated. The installation of a photoactivatable benzophone moiety at this position unfortunately yielded a barely active compound (**140)**, but installation of the smaller aryl alkyne did not significantly affect activity (**141**). Thus, as leucine-mimic analogs of nelfinavir had been determined to be active (**22, 89, 114**), an analog derived from D-photo-leucine containing a photoactivatable diazirine moiety was synthesized with the aryl alkyne in the benzamide position. The D-form of photo-leucine was chosen rather than the L-form as head-to-head comparisons of **22** and **114** consistently indicated that the R stereochemistry at the side chain of the P1 position was more active. The probe (**150**, Figure 2.15A) was determined to retain inhibitory activity via the hemolysis assay, albeit weakly, and was thus used for subsequent affinity pulldown studies. Cultures of *S. pyogenes* were treated with **150** before irradiation with UV light, cell lysis, and conjugation to biotin via copper mediated azide-alkyne cycloaddition. The lysates were then passed over streptavidin resin for affinity pulldown or were visualized directly by western blotting.

Unfortunately, it became quickly apparent that copious non-specific crosslinking was occurring due to the presence of numerous bands by western blot (Figure 2.15B), none of which were diminished in competition experiments with nelfinavir. As SagE was shown to be active in *E. coli*, a strain overexpressing SagE was utilized for labeling and also displaying non-specific crosslinking (Figure 2.15C). A proteomics analysis of the labeled proteins revealed that several of the most abundant cellular proteins were crosslinking to the probe (**150**), including EF-tu and protein S9. These abundant housekeeping proteins overwhelmed the analysis and no proteins that would reasonably be expected to be involved in SLS biosynthesis were detected. It is probable that **150**, as well as nelfinavir, non-specifically stick to many different cellular components, likely contributing to the range of known off-target activities of nelfinavir.

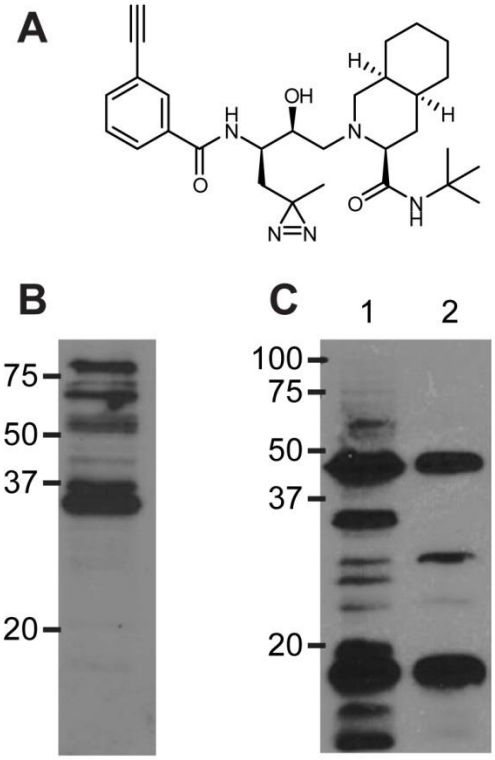


Figure 2.15. Cell extracts photocrosslinked with **150**. (A) Structure of probe **150**. (B) Western blot of the insoluble fraction of lysates from *S. pyogenes* M1 WT cells labeled with probe **150**. (C) Western blot of total lysate (lane 1) and the insoluble fraction (lane 2) from *E. coli* labeled with probe **150**.

Although initial attempts to heterologously express and reconstitute SagE were unsuccessful, we believed it might be possible to generate properly folded SagE in the presence of lipid nanodiscs as the protein is expected to normally be membrane bound (Figure 2.2). Lipid nanodiscs are portions of lipid bilayer solubilized by the action of membrane scaffolding proteins that wrap around the hydrophobic tails of the lipids.[54] The nanodiscs have been shown to be effective platforms for stably reconstituting otherwise intransigent membrane proteins *in vitro*. SagE was generated through a coupled *in vitro* transcription/translation system in the presence of lipid nanodiscs. The reactions were then tested for their ability to cleave an MBP-SagA substrate. Although intact SagE could be visualized by western blotting, no cleavage activity could be detected (Figure 2.16).

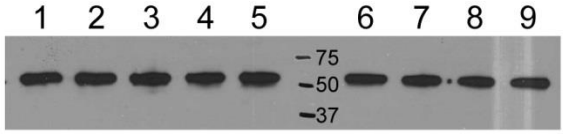


Figure 2.16. Representative example of the lack of MBP-SagA cleavage by SagE expressed *in vitro* with lipid nanodiscs. Each lane of the western blot represents a different reaction condition for the *in vitro* transcription/translation (TNT) of SagE or for the cleavage of MBP-SagA. (1) no DNA, 30 °C TNT, 37 °C cleavage; (2) pET28-Mistic-SagE, 30 °C TNT, 37 °C cleavage; (3) pET28-Mistic-SagE, 18 °C TNT, 37 °C cleavage; (4) pCOLA- SagE-SagF-Stag, 30 °C TNT, 37 °C cleavage; (5) pCOLA- SagE-SagF-Stag, 18 °C TNT, 37 °C cleavage; (6) pET28-Mistic-SagE, 30 °C TNT, 18 °C cleavage; (7) pET28-Mistic-SagE, 18 °C TNT, 18 °C cleavage; (8) pCOLA- SagE-SagF-Stag, 30 °C TNT, 18 °C cleavage; (9) pCOLA- SagE-SagF-Stag, 18 °C TNT, 18 °C cleavage. No cleavage is seen in any condition.

In a final effort to validate SagE as the target of nelfinavir, *sagE* (with and without *sagF*) was cloned into the *S. pyogenes* overexpression vector pIB184 under constitutive control.[55] If inhibition of SagE by nelfinavir causes a bottleneck in SLS biosynthesis, overexpression of SagE may abrogate the inhibition. The amount of SLS produced by strains containing the *sagE* vectors was compared to an empty vector control in the presence or absence of nelfinavir via the hemolysis assay. No difference in the amount of SLS produced between *sagE*-containing or empty vectors could be detected in any instance (Figure 2.17). Taking the opposite approach, overexpression vectors containing antisense sequences to *sagE* were created, transformed into *S. pyogenes*, and tested in the hemolysis assay. Again, however, no synergistic inhibition of SLS biosynthesis was observed (Figure 2.17). qRT-PCR analysis of strains

containing the overexpression vectors against empty vector controls indicated that the presence of the vectors was resulting in a several fold increase in *sagE* RNA quantities (Table 2.5). Therefore, it is likely that the levels of SagE were being effectively modulated but that this was having no effect on SLS biosynthesis in the presence of nelfinavir. It is possible that SagE requires other cellular components to function that are more limiting than the quantity of SagE. If nelfinavir only binds to SagE in the context of this interaction, modulating the expression of SagE would not be expected to have a significant effect on nelfinavir inhibition. However, it is also possible that nelfinavir does not directly inhibit SagE or that the inhibition of nelfinavir biosynthesis is achieved through the interaction of nelfinavir and a different target. Even in this latter scenario, it is still possible that nelfinavir inhibits proteolysis via SagE, but that this simply isn't a limiting factor in the biosynthesis of the mature cytolysin.

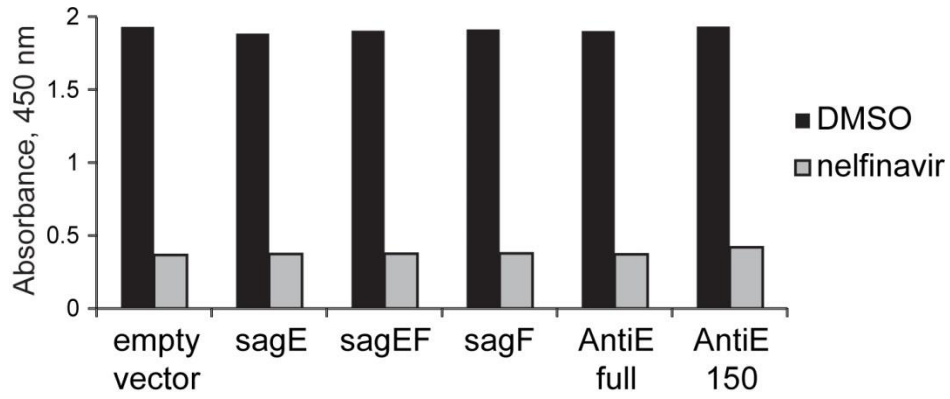


Figure 2.17. Over- and under-expression of SagE does not affect SLS biosynthesis inhibition by nelfinavir. Lytic activity of extracts from *S. pyogenes* containing pIB184 vectors with *sagE*/*F* or antisense *sagE* treated with nelfinavir relative to DMSO controls (n = 2).This subset of data is representative of the overall trend.

Vector	Gene	Log_2 ratio (treated/untreated)
pIB184-*sagE*	*sagE*	4.54 ± 1.32
pIB184-*sagEF*	*sagE*	3.11*
pIB184-*sagEF*	*sagF*	2.55*
pIB184-*sagF*	*sagF*	4.63 ± 0.72
pIB184-AntiEfull	*sagE*	6.07 ± 0.41

Table 2.5. Changes in SagE/F expression as measured by qRT-PCR with *S. pyogenes* M1 5448 transformed with pIB184 vectors. Values are from three biological replicates averaged from three technical replicates each, except for the pIB184-*sagEF* vector which was only one biological replicate (indicated by a *). The *sagB* gene was used as an internal control gene and relative gene expression was calculated by a comparative C_T method.[50] Positive values represent up-regulation of the genes. *sagE* appears to be upregulated with the antisense vector as the primers used cannot differentiate between normal *sagE* RNA and antisense *sagE* RNA. The primers used are given in the supplemental methods.

2.8 Discussion

In this work, the FDA-approved HIV protease inhibitor nelfinavir was repurposed as the first small molecule inhibitor of SLS production in *S. pyogenes*, displaying low micromolar activity. Nelfinavir was identified as a lead compound by leveraging the extensive basic and clinical research data accumulated on the effects of the drug. Lipodystrophy, a known side effect of nelfinavir and several other HIV protease inhibitors, had been previously linked to the off-target inhibition of the CaaX protease ZMPSTE24. We surmised that the HIV protease inhibitors would also inhibit SagE due to its homology with CaaX proteases, allowing us to rapidly identify a lead compound for the inhibition of SLS production. This strategy for lead identification negated the need for high-throughput screening or for a crystal structure of the target for *in silico* and structure-based design. Utilizing a drug with synthetic routes that have been thoroughly explored also greatly accelerated the creation of analogs for SAR efforts that yielded compound **17**, with an improved IC_{50} value of 1 μM.

Nelfinavir and related compounds are inhibitors of SLS biosynthesis, most likely through inhibition of proteolytic cleavage of the protoxin by the CPBP family member SagE. Although *in vitro* reconstitution was unsuccessful, the necessity for SagE during SLS production in the multi-plasmid expression system in *E. coli* provides considerable evidence for its role in proteolytic processing. Like

many CPBP members, SagE is commonly referred to as an immunity protein in the literature,[15, 16] but inhibition by nelfinavir did not have any effect on the growth of *S. pyogenes*, providing evidence that SagE is not involved in self-immunity. Alternatively, compensatory mutations that abolish SLS production may arise when SagE is inactivated, as has been previously suggested.[16] The original annotation of SagE as an immunity protein stems from its similarity to PlnP from *Lactobacillus plantarum*. PlnP and several related proteins are found downstream of bacteriocin structural genes in *L. plantarum* and have been shown to provide immunity to the antibacterial effect of the respective bacteriocins.[36] Yet, unlike these bacteriocins, SLS has not been demonstrated to possess any antibacterial activity against intact *S. pyogenes* cells. A large buildup of intracellular SLS might result in toxicity, but this would be expected to affect any *S. pyogenes* strains in which the transport machinery is inactivated as well, which has not been observed.[15] Furthermore, treatment of SagA with the cyclization machinery (SagBCD) *in vitro* results in a lytic entity without cleavage of the leader peptide, indicating that while proteolysis is required for cellular export, it is unnecessary for lytic activity.[21] These observations lead us to conclude that the principal function of SagE is to proteolytically mature SLS.

In addition to experimentally supporting a biochemical role for SagE, we have also addressed the mechanism of proteolysis and the probable inhibition by nelfinavir. CPBP family members have been postulated to function through a zinc metalloprotease mechanism based on the presence of two glutamates, two histidines, and an asparagine residue that are conserved across the family (Figure 2.11).[34] Mutation of these residues in the eukaryotic type II CaaX protease Ras-converting enzyme (Rce1) has also demonstrated that these residues are critical for activity.[34] We found that Glu131 and Glu132 were critical for the activity of SagE. Thus, it was initially unexpected that proteolysis was inhibited by pepstatin, a general aspartyl protease inhibitor, but not by bestatin, a metalloprotease inhibitor (Figure 2.5). However, a recent report detailing the crystal structure of Rce1 from *Methanococcus maripaludis* provides compelling evidence that this family of proteases actually functions through a novel glutamate-dependent mechanism (Figure 2.18).[42] The authors hypothesized that a glutamate residue extending into the active site is responsible for deprotonating a water molecule, activating it for nucleophilic attack.

Given the similarities between this proposed mechanism and the mechanism of aspartyl proteases, it is perhaps unsurprising that a CaaX protease homolog would be inhibited by aspartyl protease inhibitors, including nelfinavir.

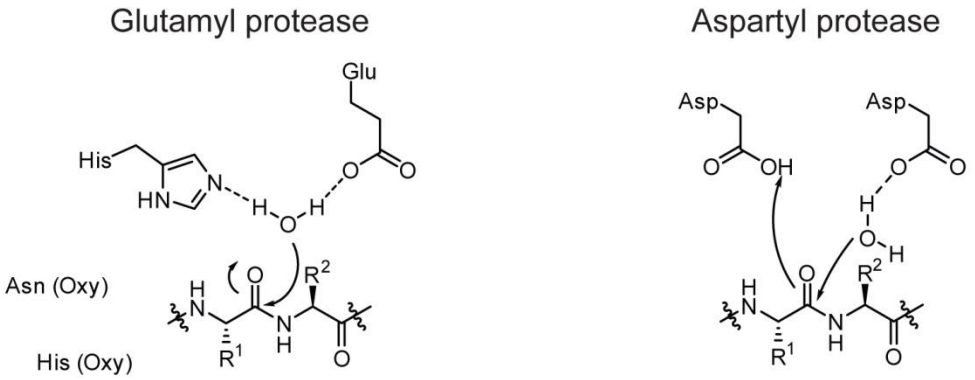


Figure 2.18. Proposed mechanism for glutamate-dependent proteases. A potential mechanism for glutamate dependent proteases based on the crystal structure of Rce1 is shown.[42] Glutamate and histidine activate water for nucleophilic attack while a second histidine residue and an asparagine residue labeled "Oxy" comprise the oxyanion hole. A second conserved glutamate near the active site is thought to be structurally important. A general mechanism for aspartyl proteases is shown for comparison.

In addition to increasing the potency of inhibition with compound **17**, our synthetic effort also yielded information on how nelfinavir may interact with SagE. The SAR analysis revealed that a rather large side chain in the pseudo P1 position of the structure was required for potent activity, with the original *S*-phenyl-ethyl group displaying the greatest inhibition. This result was not anticipated, given that the P1 residue in the SagA substrate is putatively glycine. One possible explanation for this discrepancy is that the catalytic site architecture of SagE is highly conserved with the type II CaaX proteases, which normally cleave substrates with a prenylated cysteine in the P1 position. In this case, nelfinavir would be an ideal fit for the active site, as the core of the molecule closely resembles a cysteine with a hydrophobic group appended. An alternative explanation is that nelfinavir does not bind in the active site in the expected fashion (with the secondary hydroxyl interacting with the catalytic residues, Figure 2.10) or does not bind in the active site at all. These possibilities are supported by the potent activity of several analogs with different stereochemical configurations, as these molecules would likely be forced into different conformations that do not allow the same favorable interactions with active site residues. In this scenario,

binding of nelfinavir to the membrane protease may be driven by hydrophobic interactions. This explanation would also account for the attenuated activity of saquinavir, which is nearly identical to nelfinavir except for the presence of a more hydrophilic group at the benzamide position (Figure 2.6, XLogP3 values from PubChem for nelfinavir and saquinavir are 5.7 and 4.2, respectively). Comparisons to the other HIV protease inhibitors do not yield much additional information due to their low similarity to nelfinavir, as reflected in the Tanimoto similarity coefficients (Table 2.6).

	ampr	ataz	daru	indi	**lopi**	**nelf**	**rito**	**saqu**	tipr
ampr	1.00	0.24	0.67	0.14	0.21	0.14	0.21	0.20	0.12
ataz	0.24	1.00	0.22	0.15	0.20	0.13	0.21	0.20	0.11
daru	0.67	0.22	1.00	0.13	0.20	0.13	0.20	0.19	0.11
indi	0.14	0.15	0.13	1.00	0.15	0.18	0.14	0.22	0.10
lopi	0.21	0.20	0.20	0.15	1.00	0.14	0.28	0.19	0.11
nelf	0.14	0.13	0.13	0.18	0.14	1.00	0.10	**0.40**	0.09
rito	0.21	0.21	0.20	0.14	0.28	0.10	1.00	0.17	0.10
saqu	0.20	0.20	0.19	0.22	0.19	**0.40**	0.17	1.00	0.10
tipr	0.12	0.11	0.11	0.10	0.11	0.09	0.10	0.10	1.00

Table 2.6. The Tanimoto similarity coefficients for the FDA-approved HIV protease inhibitors. Higher Tanimoto coefficients indicate higher similarity. Compounds that inhibited the production of SLS are highlighted in green. Abbreviations: ampr, amprenavir; ataz, atazanavir; daru, darunavir; indi, indinavir; lopi, lopinavir; nelf, nelfinavir; rito, ritonavir; saqu, saquinavir; tipr, tipranavir.

Unequivocal confirmation of the bacterial target of nelfinavir (and analogs) was unfortunately not possible in the present study, which can be attributed to the technical challenges inherent to the study of integral membrane proteins. Nelfinavir also likely interacts with multiple targets in *S. pyogenes*, as the drug is known to have multiple off-target effects in humans.[45] A significant body of work in mammalian cell lines has demonstrated that nelfinavir displays promiscuous activity, such as interruption of Akt signaling and inhibition of the proteasome.[56] Target promiscuity likely also exists in bacteria, so possible interactions of nelfinavir with additional targets cannot be ruled out. Thus, nelfinavir may be inhibiting

additional participants that indirectly result in the inhibition of SLS production in a mechanism unrelated to proteolysis. Further targets of nelfinavir may also exist that do not result in observable phenotypes. However, nelfinavir also inhibited the biosynthesis of other natural products that include a CPBP family member in the gene cluster (*i.e.* LLS, CLS, and PZN). The fact that this inhibition occurred in a range of disparate bacterial species decreases the probability that another protease is responsible and provides substantial, albeit indirect, support that SagE is the primary target of nelfinavir in *S. pyogenes*. Finally, the HIV protease inhibitors found to inhibit SLS production parallel those capable of inhibiting the human CaaX protease ZMPSTE24 (Figure 2.6),[46, 47] providing additional evidence that nelfinavir and its analogs inhibit SLS production by blocking the action of SagE.

2.9 Summary and outlook

Despite their prevalence, prokaryotic members of the CPBP family have not yet been thoroughly investigated. Many of the family members are incorrectly annotated or have a predicted function based solely on distant homology. The discovery of nelfinavir as an inhibitor of CPBPs has provided evidence that SagE functions as a protease and will aid in the assignment of functions to other family members, including other human pathogens such as *S. aureus*. Additionally, nelfinavir is the first reported inhibitor of the production of SLS and related toxins. Nelfinavir and improved analogs will provide a new tool to investigate toxin function without the need to create genetic deletions while also allowing for temporal control over toxin production. Reversible control of SLS production with nelfinavir analogs will also help to clarify the precise contribution of SLS to virulence in *in vivo* models of infection and may open the door to the development of virulence-targeting strategies for the control of *S. pyogenes* infections. Finally, the chemical knockdown effect of nelfinavir can be utilized for the discovery of natural products from the 22% of TOMM gene clusters that contain a CPBP family member, potentially accelerating the structural and functional characterization of these compounds.

2.10 Experimental

2.10.1 Materials

All chemicals were purchased from Sigma-Aldrich, VWR, Fisher Scientific, or Oakwood Products and used without further purification unless otherwise specified. HIV protease inhibitors were obtained from the NIH AIDS Research and Reference Program, Division of Aids, NIAID, NIH. Nelfinavir, ritonavir, saquinavir, indinavir sulfate, amprenavir, lopinavir, atazanavir sulfate, tipranavir, and darunavir were received through distribution by Thermo Fisher.

2.10.2 Plasmid construction

The genes encoding Flag-SagA and SagA-Flag were generated in pDCerm via PCR utilizing primers specific to *sagA* and containing the Flag tag sequence and XbaI (Flag-sagA) or BamHI (sagA-Flag) restriction sites. For heterologous expression of these peptides in *Escherichia coli*, the genes of interest were subcloned into a pET28 vector containing an N-terminal maltose binding protein (MBP) fusion as previously described.[21] The construction of the Duet vectors pETDuet-1-sagB-sagC and pACYCDuet-1-MBP-sagA-sagD has been previously described.[43] The Duet vector containing *sagE* and *sagF* was constructed using the pCOLADuet-1 vector (EMD Millipore) in an analogous manner.[43] Point mutations of SagA and SagE were introduced by the Quikchange site-directed mutagenesis kit (Agilent), following the manufacturer's instructions. The genes encoding SagE, SagEF, and SagF were inserted into pIB184 using Gibson cloning following the manufacturer's instructions (New England Biolabs) with the EcoRI cut site to generate pIB184-sagE, pIB184-sagEF, and pIB184-sagF respectively. Antisense RNA comprising the entire sequence of *sagE* and the first 150 bp of *sagE* was inserted into pIB184 in a similar manner to generate pIB184-AntiEfull and pIB184-AntiE150 respectively.

2.10.3 Generation of SagA substrates

^{35}S-Methionine-labeled MBP-SagA was generated as previously described (from rabbit reticulocyte, T7-coupled *in vitro* transcription/translation, Promega).[57] MBP-Flag-SagA and MBP-SagA-Flag were generated by heterologous expression and purification as previously described.[21]

2.10.4 Preparation of S. pyogenes membranes

An overnight culture (5 mL) of *S. pyogenes* M1 5448 Δ*sagA*[16, 21] was grown in Todd-Hewitt Broth (THB) at 37 °C and used to inoculate a 50 mL culture of THB. This culture was grown to an OD_{600} of 0.6 and harvested at 4,000 x g (5 min, 4 °C), and the cell pellet was transferred to a 1.7 mL microfuge tube using 1 mL of cell lysis buffer [50 mM MOPS (pH 7.6), 50 mM NaCl, 5% (v/v) glycerol, 1500 U mutanolysin, and 5 mg lysozyme]. The cell wall was digested for 1 h at 37 °C with agitation before cooling to 4 °C, directly prior to cell disruption by sonication (3 rounds for 30 s each, 30% power using a microprobe). The sample was then subjected to 3 rounds of freeze/thaw cycling using liquid nitrogen with cap venting upon warming to room temperature. A final round of sonication was then applied and any remaining intact cells were removed by a low-speed centrifugation step (500 x g, 3 min, 4 °C). Omission of either the sonication or freeze/thaw cycling steps in this procedure gave poor membrane fraction preparations. The supernatant of the low-speed spin is referred to as the whole-cell lysate fraction. The majority of this fraction was transferred to a thick-walled microfuge tube prior to a subsequent high-speed spin (50,000 x g, 45 min, 4 °C), in which the membrane fraction was separated from the whole-cell lysate. The supernatant of this spin, referred to as the soluble fraction, was transferred to a clean microfuge tube. Proteins from the harvested membrane fraction were resolubilized in 300 µL of membrane extraction buffer [50 mM HEPES (pH 7.4), 125 mM NaCl, 5% (v/v) glycerol, and 1% (w/w) dodecyl maltoside] with light sonication (2 rounds for 1 s each, 10% power using a microprobe). The extraction was allowed to proceed for 1 h at 4 °C with rocking. Protein concentration in each sample (whole-cell lysate, soluble, and membrane fractions) were quantified using the Bradford assay, adjusted to the same concentration (3.8 mg/mL) and would typically produce a total of 1-2 mg of protein for each fraction.

2.10.5 Localization of SagA/SLS proteolytic activity

To *S. pyogenes* fractions (whole-cell lysate, soluble, and membrane, 13.2 µL, 50 µg protein each), was added TCEP (1 µL of 50 mM), purified ^{35}S-methionine-labeled MBP-SagA, MBP-Flag-SagA, or MBP-SagA-Flag protein pre-diluted in membrane extraction buffer (1 µL of 90 µg/mL), and 14.8 µL of membrane extraction buffer for a total reaction volume of 30 µL. Aliquots were removed at specific

times of reaction at 25 °C and quenched with SDS-PAGE loading buffer. Proteins were separated on a 10% SDS-PAGE gel with 25 ng of MBP-SagA protein loaded per lane. For samples utilizing ^{35}S-methionine-labeled MBP-SagA, the gels were dried and visualized by autoradiography (typically requiring ~10 h of exposure at -80 °C using a Kodak BioMax low-energy isotope intensifying screen). For samples utilizing MBP-Flag-SagA or MBP-SagA-Flag, protein was transferred to nitrocellulose membranes (Pierce) by electroblot. Membranes were blocked for 30 min at 25 °C in 4% non-fat milk in TBST [25 mM Tris (pH 7.6), 150 mM NaCl, 0.1% (v/v) Tween 20] before anti-Flag antibodies (M2, mouse monoclonal, Sigma-Aldrich) were added at a 1:5,000 dilution directly to the buffer and allowed to bind for 1 h at 25 °C. The buffer was discarded and the membranes were washed 3 times for 5 min each with TBST. The membranes were then treated with anti-mouse horseradish peroxidase (HRP, GE Healthcare) at a 1:5,000 dilution for 30 min at 25 °C followed by 3 washes for 5 min each with TBST. The membranes were then treated with the SuperSignal West Pico Chemiluminescent substrate (Pierce) prior to film imaging.

2.10.6 Expression of SLS in E. coli

The three vectors comprising SagABCDEF were co-transformed into BL21-DE3 cells (Invitrogen) and plated on LB agar plates with 100 μg/mL ampicillin, 35 μg/mL chloramphenicol, and 30 μg/mL kanamycin. The resulting colonies were then chosen and tested for their ability to produce hemolytic activity. To perform hemolytic tests, 10 mL LB cultures with 100 μg/mL ampicilln, 35 μg/mL chloramphenicol, and 30 μg/mL kanamycin were grown overnight at 37°C with shaking. Separate 10 mL LB cultures were inoculated 1:50 with each of these starter cultures and grown to OD_{600} 0.7. Cultures were induced with 0.1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and bovine serum albumin (BSA, 10 mg/ml) was added to the induced cultures. An aliquot was removed at 2 and 8 h for hemolytic assay testing. Aliquots were centrifuged at 4,000 x g for 15 min and supernatants were collected and directly added to washed, defibrinated sheep blood.

2.10.7 Erythrocyte lysis assay

Defibrinated sheep blood (Hemostat Laboratories) was rinsed by resuspension in phosphate buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4 \cdot 2H_2O$, 2 mM KH_2PO_4, pH 7.4) followed by centrifugation at 300 x g for 10 min at 4 °C. The supernatant was discarded and the pellet was rinsed again, repeating until the supernatant was nearly colorless. The blood was resuspended in fresh PBS to a concentration within the linear absorbance range at 405 or 450 nm when treated with 0.1% Triton X-100. SLS extracts were applied to the blood in a 3:4 ratio and incubated at 37 °C for 30 to 120 min. Lysed cells were removed by centrifugation at 300 x g for 10 min at 4 °C and hemoglobin release was measured by A_{405} (FilterMax F5 plate reader) or A_{450} (Cary 4000 UV-Vis).

2.10.8 Membrane proteolysis inhibition

Purified membranes from *S. pyogenes* M1 5448 were used to digest [^{35}S]-MBP-SagA in the presence of nelfinavir or mechanism-based protease inhibitors. A typical reaction (25 µL total volume) consisted of 18 µL of *S. pyogenes* membranes (5.0 mg/mL protein concentration), [^{35}S]-MBP-SagA, protease inhibitor, and TCEP (2 mM). The inhibitors tested were pepstatin A (for aspartyl proteases, 60 µM), bestatin (for metalloproteases, 200 µM), E-64 (for cysteine proteases, 200 µM), and pefabloc (for serine proteases, 400 µM). Nelfinavir was also tested in subsequent assays. Aliquots were removed at specific times of reaction at 25 °C, quenched with SDS-PAGE loading buffer, and analyzed by SDS-PAGE as described above.

2.10.9 Effect of FDA-approved HIV protease inhibitors on S. pyogenes

An overnight culture (5 mL) of *S. pyogenes* M1 5448 was grown in THB without shaking at 37 °C and used to inoculate multiple, identical 50 mL cultures of THB. Upon reaching an OD_{600} of 0.2, the cultures were diluted two-fold using an equal volume of THB containing the desired concentration of protease inhibitor (initial screening was at 50 µM; nelfinavir analog screening was at 25 µM) from 10 mM stock solutions in DMSO. An equal volume of DMSO was used as a negative control for these assays. The OD_{600} of these cultures was measured in triplicate every 30 min using a Cary 4000 UV-Vis

spectrophotometer. For the cultures that reached an OD_{600} of 0.6, any SLS present on the cell surface was isolated using a BSA extraction procedure. SLS activity was then quantified by erythrocyte lysis assay.

2.10.10 BSA extraction of SLS

S. pyogenes M1 5448 was grown in THB at 37 °C as described above to an OD_{600} of 0.6. Cultures were harvested by centrifugation at 4000 x g for 10 min at 4 °C and the supernatant was discarded. The pellets were resuspended in fresh THB containing 10 mg/mL BSA and incubated at 37 °C for 1 h. The cultures were harvested by centrifugation at 4000 x g for 10 min at 4 °C and the supernatant was subsequently used in the erythrocyte lysis assay either undiluted or at a 1:3 dilution in fresh THB containing 10 mg/mL BSA.

2.10.11 Determination of IC50 values

S. pyogenes M1 5448 was grown in the presence of inhibitors with a concentration range of 0.2 to 64 μM. SLS extracts were applied to the erythrocyte lysis assay alongside a DMSO treated control. Samples were centrifuged at 300 x g for 10 min at 4 °C and the supernatants were assayed for hemoglobin release when DMSO control treated blood had reached roughly 90% lysis. IC_{50} values were defined as the concentration of inhibitor needed to prevent 50% of the hemolysis relative to the DMSO control over the given time course.

2.10.12 Minimum inhibitory concentration (MIC) testing

Cultures of *Bacillus subtilis* 168 and *E. coli* MC4100 were grown in Luria-Bertani (LB) broth; cultures of *Listeria monocytogenes* 4b F2365, *Staphylococcus aureus* USA300 (MRSA), and *Enterococcus faecium* U503 (VRE) were grown in brain heart infusion (BHI) broth; cultures of *S. pyogenes* M1 5448 were grown in THB; cultures of *C. sporogenes* ATCC 19404 were grown in anaerobically prepared THB supplemented with 5 mM L-cysteine; cultures of *Bacillus amyloliquefaciens* RS6, *Klebsiella pneumoniae* ATCC 27736, and *Pseudomonas aeruginosa* PA01 were grown in Mueller Hinton broth. All strains were started as overnight cultures, were incubated at 37 °C, and were used at 1:100 dilutions to inoculate fresh cultures in the same media. The cultures were grown at 37 °C to mid-

exponential phase prior to MIC testing. Nelfinavir was added to the cultures at 1 to 64 µM from a 10 mM DMSO stock. The treated cultures were incubated for 16 h at 37 °C and growth was visually examined.

2.10.13 Electron microscopy

An overnight culture of *S. pyogenes* M1 5448 was grown in THB at 37 °C and used to inoculate 5 mL cultures of THB containing 25 µM nelfinavir (from a 10 mM stock) or an equivalent amount of DMSO at a 1:20 dilution. The cultures were grown to an OD_{600} of 0.6, harvested by centrifugation (5,000 x g, 10 min, 25 °C), and fixed with a Karnovsky's fixative in phosphate buffered 2% glutaraldehyde and 2.5% paraformaldehyde. Microwave fixation was used with this primary fixative and the tissue was then washed in cacodylate buffer with no further additives. Microwave fixation was also used with the secondary 2% osmium tetroxide fixative, followed by the addition of 3% potassium ferricyanide for 30 min. After washing with water, saturated uranyl acetate was added for enbloc staining. The sample was dehydrated in a series of increasing concentrations of ethanol. Acetonitrile was used as the transition fluid between ethanol and the epoxy. Infiltration series were done with an epoxy mixture using the epon substitute Lx112. The resulting blocks were polymerized at 90 °C overnight, trimmed, and ultrathin sectioned with diamond knives. Sections were stained with uranyl acetate and lead citrate and photographed with a Hitachi H600 Transmission Electron Microscope.

2.10.14 qRT-PCR of virulence gene expression

An overnight culture of *S. pyogenes* M49 NZ131 was grown in THB at 37 °C and used to inoculate 2 mL cultures of THB containing 25 µM nelfinavir (from a 10 mM stock) or an equivalent amount of DMSO at a 1:20 dilution. The cultures were grown to an OD_{600} of 0.8 before RNA was isolated using an RNeasy Protect Bacteria Mini Kit (Qiagen) following the manufacturer's instructions, except for the lysis step, which was extended to 1 h with the addition of 250 U mutanolysin. The absence of DNA contamination was confirmed by PCR with primers to the 16S rRNA and *sagA* genes. Isolated RNA was used to generate cDNA using the ImProm II kit (Promega) following the manufacturer's instructions, running reactions both with and without reverse transcriptase (RT). RT negative reactions again confirmed the absence of DNA contamination. qRT-PCR reactions were initiated with 2 uL of cDNA

(~2.5 ng/µL), 2 µL of 10 µM forward primer, 2 µL of 10 µM reverse primer, 10 µL of iTaq Universal SYBR Green Supermix (Bio-Rad), and 4 µL of water and were run on a Lightcycler 480 (Roche). The fold-change in gene expression was calculated using a comparative C_T method.[50] We defined a significant change in gene expression to be genes that were either up- or down-regulated by more than 1.5 fold (log_2) with P-values <0.01.

2.10.15 Inhibition of LLS production

Overnight cultures (5 mL) of *L. monocytogenes* were grown in BHI media at 37 °C and used to inoculate a 50 mL culture of BHI at a 1:20 dilution. Upon reaching an OD_{600} of 0.2, the culture was used to create multiple, identical cultures diluted by twofold using an equal volume of BHI containing 50 µM nelfinavir from a stock solution of 10 mM in DMSO. An equal volume of DMSO was used as a negative control for these assays. *L. monocytogenes* F2365LLS^C Δ*hly* (LLO genetically deleted, LLS constitutively expressed) was used as the "+LLS" strain and *L. monocytogenes* F2365LLS^C Δ*hly*Δ*llsB* (LLO and LLS genetically deleted) was used as the "-LLS" strain[51]. LLS was extracted in a manner analogous to SLS and applied to the erythrocyte lysis assay.

2.10.16 Inhibition of CLS production

Overnight cultures (5 mL) of *Clostridium sporogenes* ATCC 19404 were grown in anaerobically prepared THB containing 5 mM L-cysteine at 37 °C in an anaerobic chamber (Coy Lab Products) and used to inoculate 5 mL cultures at a 1:40 dilution of anaerobically prepared THB containing 5 mM L-cysteine. The cultures were incubated at 37 °C until an OD_{600} of 0.15, at which point 50 µM nelfinavir from a stock solution of 10 mM in DMSO was added. An equal volume of DMSO was used as a negative control. CLS was extracted in a manner analogous to SLS and applied to the erythrocyte lysis assay.

2.10.17 Inhibition of plantazolicin production

10 mL cultures of *B. amyloliquefaciens* RS6 (*sfp*::*ermAM bac*::*cmR*, deficient in lipopeptides, polyketides, and bacilysin)[58] were grown in LB broth at 37 °C to an OD_{600} of 0.2 or 0.6 (early- to mid-log phase, respectively), at which point nelfinavir or an equivalent volume of DMSO was added to the liquid cultures. Following inhibitor addition, cultures were grown with shaking at 37 °C for 24 h. Cells were

harvested by centrifugation (4,000 x g, 10 min, 4 °C), washed with Tris-buffered saline (TBS, pH 8), and harvested again (8,000 x g, 10 min, 4 °C). PZN was obtained by a non-lytic, methanolic cell surface extraction, as described previously.[52, 59] Briefly, cells were resuspended in MeOH (1 mL), agitated by vortex for 30 s, and equilibrated for 15 min at 22 °C. Cells were harvested by centrifugation (8,000 x g, 10 min, 4 °C), and the MeOH supernatant was evaporated in a speed-vac concentrator with minimal heating. Dried, crude extracts were resuspended in 100 µL of MeOH and filtered through a 0.22 µm filter in preparation for analysis by liquid chromatography-mass spectrometry (LC-MS) using an Agilent 1200 HPLC system coupled to a G1956B quadrupole mass spectrometer with an electrospray ionization (ESI) source. LC used a 4.6 mm x 250 mm Thermo Scientific BetaSil C18 column (100 Å, 5 µm) with a gradient of 40-90% MeCN (with 10 mM NH_4HCO_3 in the aqueous phase) over 25 min, with the analytes eluted directly into the MS. PZN was quantified by integrating the total area under the curve for the single-ion monitoring (SIM) trace of *m/z* 668 ($[M+H]^{2+}$), normalized to a negative control sample included in each set.

2.10.18 Bioinformatics analysis of CaaX-like proteins in divergent TOMM clusters

All YcaO superfamily proteins (InterPro IPR003776, SagD-like) were obtained from UniProt on October 28, 2014. Genome clusters were defined by using regions of 1000 bp on either side of the YcaO-like protein. Gene cluster families were created using MultiGeneBlast[11] with a cutoff of total score above 10. TOMM clusters were defined as containing both a YcaO-like protein and a SagC or SagA like ortholog. All gene clusters were manually examined with special attention toward stand-alone YcaO clusters such as bottromycin (*i.e.* lacking the C-protein). CaaX-like proteins were defined as being orthologous to the SagE protein and identified using a Hidden Markov Model (HMM) created with SagE orthologs from *Clostridium* and *Bacillus*.

2.10.19 PrsW cleavage of RsiW

B. subtilis strains CDE743 (*B. subtilis* PY79 *amyE*::*Phyperspank-3xflag-rsiW*) and CDE744 (*B. subtilis* PY79 *amyE*::*Phyperspank-3xflag-rsiW psrW*::*erm*) were obtained from Craig Ellermeier (University of Iowa, Iowa City). Overnight cultures of CDE743 and CDE744 were grown in LB

containing 5 μg/mL kanamycin (for maintenance of the *rsiW*-containing vector) at 37 °C with shaking. An additional 25 μg/mL erythromycin was added to all cultures of CDE744. The overnight cultures were used to inoculate fresh 2 ml cultures in LB containing 5 μg/mL kanamycin at a 1:100 dilution. The cultures were grown at 37 °C with shaking to an OD_{600} of 0.1 before the addition of 1 mM IPTG and either 50 μM nelfinavir or an equivalent amount of DMSO. The cultures were then grown to OD_{600} of 0.7 before 24 mM NaOH from a 100x stock was added to half of the cultures. The cultures were incubated at 37 °C with shaking for an additional 1 h. The cultures were pelleted and the supernatant was discarded. The pellets were resuspended in 375 μL of TAE buffer (40 mM Tris HCl, 20 mM acetic acid, 1mM EDTA, pH 8.0) with 10 mg/mL lysozyme. After a 10 min incubation at room temperature, the samples were mixed with 4x reducing loading dye and heated at 65 °C for 10 min. The samples were run on 12% SDS-PAGE gels and protein was transferred to PVDF membranes (Pierce) by electroblot. Membranes were blocked overnight at 4 °C in 5% non-fat milk in TBST before horse radish peroxidase (HRP) conjugate anti-Flag antibodies (M2, mouse monoclonal, Sigma-Aldrich) were added at a 1:5,000 dilution directly to the buffer and allowed to bind for 1 h at 4 °C. The buffer was discarded and the membranes were washed 5 times for 5 min each with TBST. The membranes were then treated with the SuperSignal West Pico Chemiluminescent substrate (Pierce) prior to film imaging.

2.10.20 Affinity labeling with crosslinking probe

S. pyogenes M1 WT and *E. coli* BL21-DE3 were grown in THB and LB, respectively, as above. Overnight cultures of each species were used to inoculate fresh 50 mL cultures in the same media. The cultures were grown to an OD_{600} of 0.9 at 37 °C and then pelleted at 4,000 x g for 10 min. The supernatants were discarded and the pellets were washed with 2 mL water before being resuspended in 1 mL of water. Probe **150** was added at 50 μM to each sample and the suspensions were incubated at 37 °C with shaking for 30 min. The samples were transferred to a 6-well plate and exposed to 365 nm light from a 4 W UV wand placed directly on the well for 20 min. The samples were transferred back to an Eppendorf tube and pelleted at 2,000 x g for 10 min. The pellets were resuspended in MOPS lysis buffer (50 mM MOPS, 50 mM NaCl, 5% glycerol, pH 7.0) containing 15 μL/mL protease inhibitor cocktail (2

μM leupeptin, 200 μM PMSF, 2 mM benzamidine, 2 μM E64) and either 200 U/mL mutanolysin (for *S. pyogenes*) or 15 mg/mL (for *E. coli*). For *S. pyogenes* samples, cell lysis was carried out by a one hour incubation at room temperature followed by freezing, thawing, and sonicating the cells for 30s, with the freeze/thaw/sonicate cycle being repeated 4 times total. For *E. coli* samples, cell lysis was carried out by a 30 min incubation at room temperature followed by four rounds of sonication for 30 s each. Biotin-peg3-azide (Jena Bioscience, Jena, Germany; 100 μM), copper sulfate (10 mM), and sodium ascorbate (20 mM) were added to each sample. The samples were incubated for 1 h and were then used as total lysate or were centrifuged at 17,000 x g for 10 min. The supernatant was used as the soluble portion of the lysate while the pellet was used as the insoluble portion. The insoluble portion was resuspended in 100 μL MOPS lysis buffer with 1% CHAPS with a reducing loading dye before application to 12% SDS-PAGE gels. Protein was transferred to PVDF membranes (Pierce) by electroblot. Membranes were blocked overnight at 4 °C in 5% BSA in TBST before neutrAvidin-HRP (Pierce) was added at a 1:30,000 dilution directly to the buffer and allowed to bind for 1 h at 4 °C. The buffer was discarded and the membranes were washed 5 times for 5 min each with TBST. The membranes were then treated with the SuperSignal West Pico Chemiluminescent substrate (Pierce) prior to film imaging.

2.10.21 Expression of SagE in lipid nanodiscs

POPC lipid nanodiscs (disc concentration at 3uM, in 20mM tris, 0.1% NaCl, 15% glyercol) were obtained from Steven Sligar (University of Illinois, Urbana, IL). Coupled transcription/translation reactions were carried out with the Promega TNT T7 Coupled Rabbit Reticulocyte Lysate System (Progema product number L4610) following the manufacturer's guidelines. Reactions contained 12.5 μL rabbit reticulocyte lysate, 1 μL TNT reaction buffer (supplied with kit), 0.5 μL T7 polymerase, 0.25 μL amino acid mix minus Met, 0.25 μL amino acid mix minus Leu, 0.5 μL RNasin, 6.5 μL POPC nanodiscs, and 1 μg DNA. The DNA used for each reaction was either pET28-Mistic-SagE or pCOLA-SagE-SagF-Stag. Generation of the pCOLA is described above while the pET28 vector was generated in the Dixon lab (University of California, San Diego (UCSD), La Jolla, CA) by cloning from *S. pyogenes* M1 5448. The reactions were incubated at either 18 or 30 °C for 90 min before 2 μL of each reaction was mixed

with 3 µL of 33.3 ng/µL MBP-SagA in various cleavage buffers (50 mM tris, 50-125 mM NaCl, 0-20 mM $MgCl_2$, 0-50 µM $ZnCl_2$, 2.5 mM TCEP or 10 mM DTT, 0-2.5% glycerol, 0-3 mM ATP, pH 7.5). The cleavage reactions were incubated at 18 °C or 37 °C for 2 h before being mixed with a reducing loading dye and applied to 12% SDS-PAGE gels. Protein was transferred to PVDF membranes (Pierce) by electroblot. Membranes were blocked overnight at 4 °C in 5% non-fat milk in TBST before anti-MBP primary antibodies (rabbit polyclonal, generated in the Dixon lab, UCSD) were added at a 1:2,000 dilution directly to the buffer and allowed to bind for 2 h at 4 °C. The buffer was discarded and the membranes washed twice with TBST for 30 s each. The membranes were placed in 10 mL fresh TBST and 1.5 µL of anti-rabbit-HRP goat secondary antibodies were added to the buffer. The membranes were incubated for 45 min at 4 °C. The buffer was discarded and the membranes were washed 5 times for 5 min each with TBST. The membranes were then treated with the SuperSignal West Pico Chemiluminescent substrate (Pierce) prior to film imaging.

2.10.22 qRT-PCR of strains containing pIB184-sagE/F vectors

Overnight cultures of *S. pyogenes* M1 5448 transformed with pIB184-*sagE*, pIB184-*sagEF*, pIB184-*sagF*, or pIB184-AntiEfull were grown in THB at 37 °C with 50 µg/mL kanamycin and used to inoculate 2 mL cultures of THB. The cultures were grown to an OD_{600} of 0.65 before RNA was isolated using an RNeasy Protect Bacteria Mini Kit (Qiagen) following the manufacturer's instructions, except for the lysis step, which was extended to 1 h with the addition of 250 U mutanolysin. The absence of DNA contamination was confirmed by PCR with primers to the 16S rRNA and *sagE* genes. Isolated RNA was used to generate cDNA using the ImProm II kit (Promega) following the manufacturer's instructions, running reactions both with and without reverse transcriptase (RT). RT negative reactions again confirmed the absence of DNA contamination. qRT-PCR reactions were initiated with 2 uL of cDNA (~2.5 ng/µL), 2 µL of 10 µM forward primer, 2 µL of 10 µM reverse primer, 10 µL of iTaq Universal SYBR Green Supermix (Bio-Rad), and 4 µL of water and were run on a Lightcycler 480 (Roche). The fold-change in gene expression was calculated using a comparative C_T method.[50]

2.10.23 Tanimoto similarity analysis

The structure of each HIV protease inhibitor in ChemDraw (PerkinElmer) was converted to .sdf format. Tanimoto similarity coefficients were calculated in Discovery Study Client 2.5 (Accelrys) using the .sdf files as input reference ligands for the Library Analysis protocol "Find Similar Molecules by Fingerprints". ECFP_6 was used and the minimum similarity was set at zero. The compounds were all compared pair-wise.

2.10.24 List of primers

Primer	Sequence (5' to 3')	Description
Construct Generation		
pDC-Flag-SagA-fwd	GAGTCTAGAATGGATTACAAGGATGACGACGATAAGTTCAAATTTACTTCAAATATTTTA	Forward primer to append an N-terminal Flag tag to *sagA*
pDC-SagA-rev	ACAGGATCCTTATTTACCTGGCGTATAACTTCC	Reverse primer to append an N-terminal Flag tag to *sagA*
pDC-SagA-fwd	GAGTCTAGAATGTTAAAATTTACTTCAAATATTTTAG	Forward primer to append a C-terminal Flag tag to *sagA*
pDC-SagA-Flag-rev	ACAGGATCCTTACTTGTCGTCGTCATCCTTGTAGTCTTTACCTGGCGTATAACTTCC	Reverse primer to append a C-terminal Flag tag to *sagA*
Flag-SagA-fwd	AAAAGGATCCATGGACTACAAGGATGACGAC	Forward primer to subclone Flag-*sagA* from pDCerm construct
Flag-SagA-rev	AAAAGCGGCCGCTTATTTACCTGGCGTATAACTTCCG	Reverse primer to subclone Flag-*sagA* from pDCerm construct
SagA-Flag-fwd	AAAAGGATCCATGTTAAAATTTACTTCAAATATTTTAGCTACTAGTGT	Forward primer to subclone *sagA*-Flag from pDCerm construct
SagA-Flag-rev	AAAAGCGGCCGCTTACTTGTCGTCGTCATCCTTGTA	Reverse primer to subclone *sagA*-Flag from pDCerm construct
MBP-fwd	ATGAAGCCCTGAAAGACG	Sequencing primer for pET28-MBP constructs
SagE-fwd	AAAACCATGGCGCCTTTGTCCATCCAATGC	Forward primer to clone *sagE* into pCOLADuet-1
SagE-rev	AAAGCGGCCGCTCATGTCACCTCCTTCTTCTTTTTTG	Reverse primer to clone *sagE* into pCOLADuet-1
SagF-fwd	AAAAAAAACATATGATGCTATTGGTTTTGCTGTCG	Forward primer to clone *sagF* into pCOLADuet-1
SagF-rev	AAAACTCGAGCTAATACTCTTTGCAACTAATCATCAAATAAGTC	Reverse primer to clone *sagF* into pCOLADuet-1
SagE-E131A-fwd	GGCTATCCTTTATTACTAGCTTTATTTgcaGAGACGATTTATCG	Forward primer for incorporation of alanine at position E131 in *sagE*
SagE-E131A-rev	CGATAAATCGTCTCtgcAAATAAAGCTAGTAATAAAGGATAGCC	Reverse primer for incorporation of alanine at position E131 in *sagE*
SagE-E132A-fwd	CTAGCTTTATTTGAAgcgACGATTTATCGTTTTTTGTGG	Forward primer for incorporation of alanine at position E132 in *sagE*
SagE-E132A-rev	CCACAAAAAACGATAAATCGTcgcTTCAAATAAAGCTAG	Reverse primer for incorporation of alanine at position E132 in *sagE*
SagA VAPGG-VLPLL-fwd	GCTACTAGTGTAGCTGAAACAACTCAAGTTcttCCTctactcTGCTGTTGCTGCTGTAC	Forward primer for incorporation of leucine at positions A20, G22, and G23 in *sagA*
SagA VAPGG-VLPLL-rev	GTACAGCAGCAACAGCAgagtagAGGaagAACTTGAGTTGTTTCAGCTACACTAGTAGC	Reverse primer for incorporation of leucine at positions A20, G22, and G23 in *sagA*
G184-SagE-fwd	CCCGCGGTACCCGGGATGCCTTTGTCCATCCAATG	Forward primer to clone *sagE* (and *sagEF*) into pIB184
G184-SagE-rev	AGATCTCGAGCTCTAGTCATGTCACCTCCTTCTTC	Reverse primer to clone *sagE* into pIB184
G184-SagF-fwd	CCCGCGGTACCCGGGATGATGCTATTGGTTTTGC	Forward primer to clone *sagF* into pIB184
G184-SagF-rev	AGATCTCGAGCTCTAGCTAATACTCTTTGCAACTAATCATC	Reverse primer to clone *sagF* (and *sagEF*) into pIB184

2.10.24 List of primers (cont.)

G184-AntiE-fwd	CCCGCGGTACCCGGGTCATGTCACCTCCTTCTTC	Forward primer to clone antisense *sagE* into pIB184
G184-AntiE-rev	agatctcgagctctagATGCCTTTGTCCATCCAATG	Reverse primer to clone antisense *sagE*-full into pIB184
G184-AntiE150-rev	CCCGCGGTACCCGGGAATGTCATAAATCAGAGTTGCTC	Reverse primer to clone antisense *sagE*-150 into pIB184
qRT-PCR		
q16S-fwd	GAGAGTTTGATCCTGGC	Forward primer for qRT-PCR on the 16S rRNA gene
q16S-rev	TTGCCGAAGATTCCCTA	Reverse primer for qRT-PCR on the 16S rRNA gene
qM49-fwd	TTAGTTTTCTTCTTTGCGTTTTAGA	Forward primer for qRT-PCR on *emm49*
qM49-rev	CGAAGCTAAGAAAAAAGTAGAAG	Reverse primer for qRT-PCR on *emm49*
qSLO2-fwd	GATGTGTTTGATAAATCAGTGAC	Forward primer for qRT-PCR on *slo*
qSLO2-rev	TCAGTTCTGTTATTGACACC	Reverse primer for qRT-PCR on *slo*
qsagA-fwd	ATGTTAAAATTTACTTCAAATATTTTAGCT	Forward primer for qRT-PCR on *sagA*
qsagA-rev	TATTTACCTGGCGTATAACTTC	Reverse primer for qRT-PCR on *sagA*
qsagB-fwd	ATGTCATTTTTTACAAAGGAACAA	Forward primer for qRT-PCR on *sagB*
qsagB-rev	ATTGACGATGACTTCTTCG	Reverse primer for qRT-PCR on *sagB*
qscpA-fwd	TTAGAAGATCGTTTCTCTAGAGTA	Forward primer for qRT-PCR on *scpA*
qscpA-rev	TACTGTTCCATTGAAAATGTCA	Reverse primer for qRT-PCR on *scpA*
qska-fwd	ATGAAAAATTACTTATCTTTTGGGAT	Forward primer for qRT-PCR on *ska*
qska-rev	CTTGCAAAATCAATGACCTC	Reverse primer for qRT-PCR on *ska*
qspeB-fwd	CTAAGGTTTGATGCCTACAA	Forward primer for qRT-PCR on *speB*
qspeB-rev	TGTTGGTATTTCAGTAGACATG	Reverse primer for qRT-PCR on *speB*
qnga-fwd	AAACAAAAAAGTAACATTAGCTCAT	Forward primer for qRT-PCR on *nga*
qnga-rev	GTATTTAACATCAGCCTTTGC	Reverse primer for qRT-PCR on *nga*
qspd3-fwd	ATGTCTAAATCAAATCGTCGT	Forward primer for qRT-PCR on *spd3*
qspd3-rev	GTCATTTAAGCCGCTAAATTG	Reverse primer for qRT-PCR on *spd3*
qsagE-fwd	GCCTTTGTCCATCCAAT	Forward primer for qRT-PCR on *sagE*
qsagE-rev	AATCGTCTCTTCAAATAAAGCT	Reverse primer for qRT-PCR on *sagE*
qsagF-fwd	TGCTATTGGTTTTGCTGT	Forward primer for qRT-PCR on *sagF*
qsagF-rev	ATACTTAAGACTGGTTTATCTTGT	Reverse primer for qRT-PCR on *sagF*

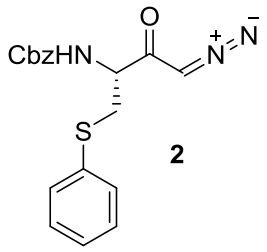
2.10.24 List of primers (cont.)

Gene Verification		
ClosC-fwd	ATGGAAAATAATACTATATACAAATTAAGCAATAATTTAA	Forward primer for confirming the presence of *closC*
ClosC-rev	TTAAGAATTAATATCTTCTAAAATTTCATCCACAA	Reverse primer for confirming the presence of *closC*
ClosD-fwd	ATGATAAAATTTTACCCAAGTTTTAATAATATATTAGAG	Forward primer for confirming the presence of *closD*
ClosD-rev	TTAAGGCATTGGGTGCG	Reverse primer for confirming the presence of *closD*

Primers used in this study. Restriction endonuclease recognition sites are underlined. Lower-case letters indicate bases targeted for mutagenesis. The function of each gene analyzed by qRT-PCR is given in Table 2.2.

2.10.25 Compound Syntheses

^{1}H, ^{13}C, and ^{19}F NMR spectra were collected on Varian Inova 400 and 500 MHz spectrometers. All ^{1}H and ^{13}C spectra collected in $CDCl_3$ were referenced to an internal tetramethylsilane standard. All ^{1}H and ^{13}C spectra collected in $(CD_3)_2SO$ were referenced to the solvent peak. All 19F spectra were referenced to an external $CFCl_3$ standard. High-resolution mass spectrometry (HRMS) data were obtained on a Micromass Q-TOF Ultima tandem quadrupole mass-spectrometer at the University of Illinois at Urbana-Champaign Mass Spectrometry Laboratory.

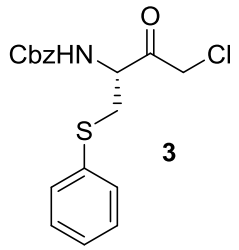


2

(R)-Benzyl (4-diazo-3-oxo-1-(phenylthio)butan-2-yl)carbamate (2)

To a stirred solution of N-Z-S-phenyl-L-cysteine (5.00 g, 15.1 mmol) in anhydrous tetrahydrofuran (60 mL) under nitrogen at -78 °C was added anhydrous triethylamine (3.14 mL, 22.7 mmol) followed by isobutyl chloroformate (2.95 mL, 22.7 mmol). The reaction was stirred cold for 1 h and then warmed to room temperature for 1.5 h. The reaction was cooled to -78 °C before an ethereal

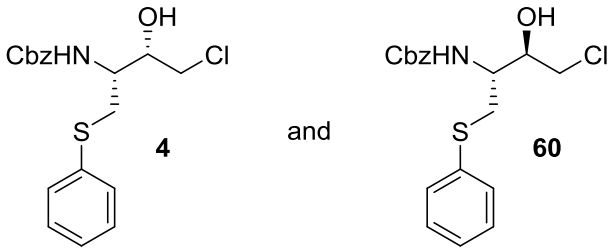
solution of diazomethane (100 mL, 32.4 mmol) was added in one portion and the reaction was then warmed to room temperature with stirring for 1 h. The reaction was evaporated to dryness, resulting in a yellow solid. The solid was taken up in ethyl acetate, washed twice with water, twice with a saturated solution of sodium bicarbonate, and once with brine. The organic phase was dried over sodium sulfate and concentrated by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 0-4% EtOAc in $CHCl_3$) to yield **2** as a yellow oil (3.33 g, 62%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.31 (d, *J* = 7.5 Hz, 2H), 7.29-7.17 (m, 6H), 7.13 (t, *J* = 7.5 Hz, 2H), 5.79 (d, *J* = 7.5 Hz, 1H), 5.39 (s, 1H), 5.01 (s, 2H), 4.32 (br s, 1H), 3.20 (m, 2H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 130.05, 128.98, 128.38, 128.09, 127.92, 126.79, 67.39, 60.64, 57.35, 55.15, 36.38, 21.38, 14.44. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{18}N_3O_3S$, 356.1069; observed, 356.1072.



3

(R)-Benzyl (4-chloro-3-oxo-1-(phenylthio)butan-2-yl)carbamate (3)

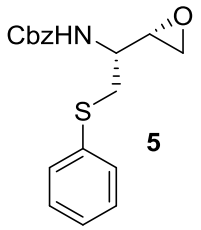
To a solution of **2** (3.33 g, 9.37 mmol) in anhydrous tetrahydrofuran under nitrogen at 0 °C was added 1 M HCl in diethyl ether (14.1 mL, 14.1 mmol) dropwise. The reaction was stirred at 0 °C for 30 min after complete addition. The solution was dried under reduced pressure to yield **3** as a slightly yellow solid (3.4 g). No further purification was made. ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.40-7.21 (m, 10H), 5.77 (d, *J* = 7.5 Hz, 1H), 5.08 (s, 2H), 4.69 (q, *J* = 7 Hz, 1H), 4.24-4.08 (m, 2H), 3.35 (m, 2H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 199.52, 155.65, 135.76, 133.83, 130.64, 129.25, 128.46, 128.23, 127.99, 127.30, 67.24, 57.03, 46.93, 35.42. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{19}NO_3SCl$, 364.0774; observed, 364.0771.



Benzyl ((2R,3S)-4-chloro-3-hydroxy-1-(phenylthio)butan-2-yl)carbamate (4) and benzyl ((2R,3R)-4-chloro-3-hydroxy-1-(phenylthio)butan-2-yl)carbamate (60)

To a solution of **3** (3.34 g, 11.06 mmol) in dichloromethane (25 mL) and methanol (15 mL) under nitrogen at 0 °C was added sodium borohydride (293 mg, 7.74 mmol). The reaction was stirred at 0 °C for 1 h, then at room temperature for 1 h. The reaction was cooled to 0 °C and quenched with slow addition of 2 M HCl (6 mL). The reaction was evaporated to dryness and the solid was taken up in dichloromethane. The solution was washed twice with water followed by brine, dried over sodium sulfate, and concentrated by rotary evaporation. TLC analysis showed two diastereomers with the lower Rf compound being the (R,S)-isomer. The (R,S)-isomer was recrystallized in hexanes/ethyl acetate. The crystals were collected and washed with cold 2:1 hexanes:ethyl acetate. The filtrate was collected, dried to a solid, and recrystallized again in the same manner. Both diastereomers were then separated further by silica flash column chromatography (gradient of 0-40% EtOAc in hexanes) to yield (R,S)-isomer (**4**) as a white solid (2.51 g, 62%) and (R,R)-isomer (**60**) as a colorless oil (439 mg, 11%) over two steps. Characterization of the (R,S)-isomer (**4**): ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.41-7.15 (complex, 10H), 5.15 (br s, 1H), 5.07 (s, 2H), 3.93 (m, 2H), 3.68 (dd, J_1 = 3.5 Hz, J_2 = 11.5 Hz, 1H), 3.60 (dd, J_1 = 7 Hz, J_2 = 11.5 Hz, 1H), 3.30 (m, 2H), 2.81 (br s, 1H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 154.16, 130.11, 129.17, 128.57, 128.29, 128.10, 127.92, 126.84, 120.32, 72.85, 67.17, 53.33, 47.45, 35.26. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{21}NO_3SCl$, 366.0931; observed, 366.0928. Characterization of the (R,R)-isomer (**60**): ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.38-7.09 (complex, 10H), 5.52 (d, J = 9 Hz, 1H), 5.06 (dd, J_1 = 12.5 Hz, J_2 = 16 Hz, 2H), 4.10 (m, 1H), 3.87 (m, 1H), 3.61 (d, J = 4 Hz, 1H), 3.46 (m, 2H), 3.18 (dd, J_1 = 6.5 Hz, J_2 = 13.5 Hz, 1H), 3.04 (dd, J_1 = 8 Hz, J_2 = 13.5 Hz, 1H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm

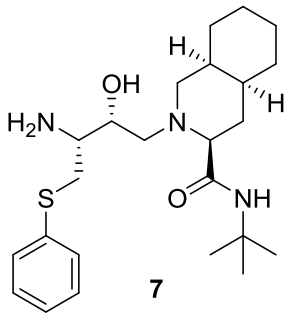
156.33, 135.90, 134.98, 129.26, 128.96, 128.39, 128.08, 127.88, 126.31, 70.71, 66.98, 52.04, 46.89, 35.35. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{21}NO_3SCl$, 366.0931; observed, 366.0933.



5

Benzyl ((R)-1-((S)-oxiran-2-yl)-2-(phenylthio)ethyl)carbamate (5)

To a solution of **4** (1.40 g, 3.83 mmol) in ethanol (60 mL) at 0 °C was added solid potassium hydroxide (237 mg, 4.22 mmol). The reaction was stirred cold for 15 min then warmed to room temperature and stirred for an additional 3 h. The reaction was evaporated to dryness and the solid was purified by silica flash column chromatography (gradient of 0-2.5% EtOAc in DCM) to yield **5** as a viscous yellow oil (1.15 g, 91%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.41-7.31 (m, 7H), 7.28 (t, J = 7.5 Hz, 2H), 7.21 (t, J = 7.5 Hz), 5.36 (br s, 1H), 5.09 (dd, J_1 = 12.5 Hz, J_2 = 16 Hz, 2H), 3.72 (br s, 1H), 3.20 (s, 2H), 2.99 (s, 1H), 2.76 (br s, 2H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 155.67, 136.02, 135.02, 129.69, 128.95, 128.36, 128.03, 127.93, 126.53, 66.78, 52.51, 52.07, 46.55, 35.76. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{20}NO_3S$, 330.1164; observed, 330.1159.



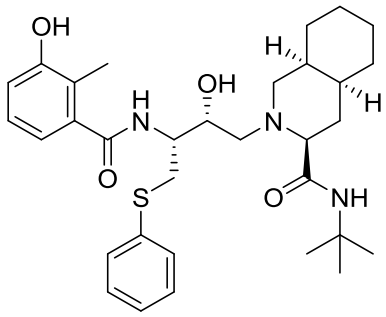
7

(3S,4aS,8aS)-2-((2R,3R)-3-amino-2-hydroxy-4-(phenylthio)butyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (7)

To a solution of **5** (775 mg, 2.35 mmol) in isopropyl alcohol (15 mL) was added (3S,4aS,8aS)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (**6**, 617 mg, 2.59 mmol). The solution was heated to reflux and stirred. After 5 h, 2 M KOH (5.9 mL, 11.8 mmol) was added and the reaction was stirred at reflux for an additional 14 h. The reaction was diluted with toluene (30 mL) and extracted with 1 M HCl (10 mL) three times. The aqueous fractions were combined and washed with toluene (15 mL) three times. The aqueous phase was adjusted to pH 12 with 5 M KOH and was extracted with toluene (20 mL) three times. The combined toluene fractions were washed with brine (15 mL), dried over sodium sulfate, and concentrated by rotary evaporation to yield **6** as a white solid (859 mg, 84%). No further purification was made. ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.39 (d, J = 7.5 Hz, 2H), 7.27 (t, J = 7.5 Hz, 2H), 7.17 (t, J = 7.5 Hz, 1H), 6.63 (br s, 1H), 6.22 (s, 1H), 3.90 (q, J = 6 Hz, 1H), 3.49 (dd, J_1 = 3 Hz, J_2 = 14 Hz, 1H), 3.11-2.98 (m, 3H), 2.72 (dd, J_1 = 3 Hz, J_2 = 6 Hz, 1H), 2.59 (dd, J_1 = 8 Hz, J_2 = 14 Hz, 1H), 2.46 (dd, J_1 = 5 Hz, J_2 = 14 Hz, 1H), 2.29 (d, J = 11.5 Hz, 1H), 1.93 (q, J = 13 Hz, 1H), 1.85-1.14 (complex, 19H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.50, 134.80, 129.67, 129.06, 126.38, 69.89, 60.06, 59.20, 54.91, 51.14, 36.41, 35.78, 33.03, 30.97, 30.39, 28.72, 28.61, 26.15, 25.60, 20.45. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{24}H_{40}N_3O_2S$, 434.2841; observed, 434.2841.

General procedure for EDC coupling to compound 7

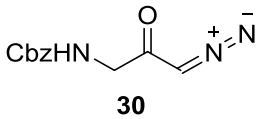
To a solution of **7** (1.0 eq) in THF (0.01-0.05 M) was added carboxylic acid (1.05 eq), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (1.05 eq), and hydroxybenzo-triazole hydrate (1.05 eq). The solution was stirred at room temperature for 16-48 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography.



1

(3S,4aS,8aS)-N-(tert-butyl)-2-((2R,3R)-2-hydroxy-3-(3-hydroxy-2-methylbenzamido)-4-(phenylthio)butyl)decahydroisoquinoline-3-carboxamide (1)

This reaction was performed according to the general EDC coupling procedure (0.083 mmol 3-hydroxy-2-methylbenzoic acid for 16 h). Purification by flash column chromatography (gradient of 2-5% MeOH in DCM) yielded **1** as a white solid (31 mg, 66%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.43 (d, J = 8 Hz, 2H), 7.30-7.7.14 (m, 4H), 7.02 (d, J = 4.5 Hz, 2H), 6.82 (t, J = 4.5 Hz, 1H), 5.87 (br s, 1H), 5.56 (s, 1H), 4.45 (m, 1H), 4.07 (m, 1H), 3.86 (br s, 1H), 3.74 (dd, J_1 = 9 Hz, J_2 = 13.5 Hz, 1H), 3.41 (dd, J_1 = 5 Hz, J_2 = 13.5 Hz, 1H), 2.92 (d, J = 11.5 Hz, 1H), 2.56 (dd, J_1 = 9 Hz, J_2 = 12.5 Hz, 1H), 2.44 (d, J = 11 Hz, 1H), 2.28 (s, 3H), 2.24-2.17 (m, 2H), 1.97 (q, J = 12.5 Hz, 1H), 1.80-1.14 (complex, 11H), 1.09 (s, 9H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 174.04, 171.76, 154.51, 137.42, 135.65, 130.14, 129.05, 127.92, 126.53, 122.65, 119.96, 116.88, 70.66, 70.48, 59.76, 58.74, 54.60, 51.26, 36.08, 35.41, 33.72, 30.99, 30.89, 28.41, 26.28, 25.98, 20.51, 12.50. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{32}H_{46}N_3O_4S$, 568.3205; observed, 568.3209.

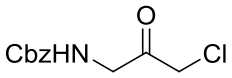


30

Benzyl (3-diazo-2-oxopropyl)carbamate (30)

To a stirred solution of Z-glycine (720 mg, 3.44 mmol) in anhydrous tetrahydrofuran (12 mL) under nitrogen at -78 °C was added anhydrous triethylamine (716 μL, 5.16 mmol) followed by isobutyl

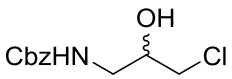
chloroformate (672 μL, 5.16 mmol). The reaction was stirred cold for 1 h and then warmed to room temperature for 1.5 h. The reaction was cooled to -78 °C before an ethereal solution of diazomethane (25 mL, 8.1 mmol) was added in one portion and the reaction was then warmed to room temperature with stirring for 1 h. The reaction was evaporated to dryness, resulting in a yellow solid. The solid was taken up in ethyl acetate, washed twice with water, twice with a saturated solution of sodium bicarbonate, and once with brine. The organic phase was dried over sodium sulfate and concentrated by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 0-8% EtOAc in $CHCl_3$) to yield **30** as a yellow solid (565 mg, 70%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.32-7.28 (complex, 5H), 5.90 (m, 1H), 5.37 (m, 1H), 5.09 (s, 2H), 3.89 (bs s, 2H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 190.56, 156.25, 136.01, 128.30, 127.97, 127.81, 66.79, 53.32, 47.93. HRMS (*m/z*): $[M + Na]^+$ calc. for $C_{11}H_{11}N_3O_3Na$, 256.0698; observed, 256.0706.



31

Benzyl (3-chloro-2-oxopropyl)carbamate (31)

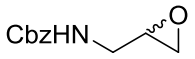
To a solution of **30** (565 mg, 2.43 mmol) in anhydrous tetrahydrofuran (8 mL) under nitrogen at 0 °C was added 1 M HCl in diethyl ether (3.65 mL, 3.65 mmol) dropwise. The reaction was stirred at 0 °C for 30 min after complete addition. The solution was dried under reduced pressure to yield **31** as orange solid (571 mg). No further purification was made. ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.35-7.28 (complex, 5H), 5.64 (br s, 1H), 5.09 (s, 2H), 4.20 (d, J = 4.5 Hz, 2H), 4.10 (s, 2H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 198.61, 156.17, 135.97, 128.40, 128.10, 127.91, 67.01, 48.43, 46.03. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{11}H_{13}NO_3Cl$, 242.0584; observed, 242.0585.



32

Racemic benzyl (3-chloro-2-hydroxypropyl)carbamate (32)

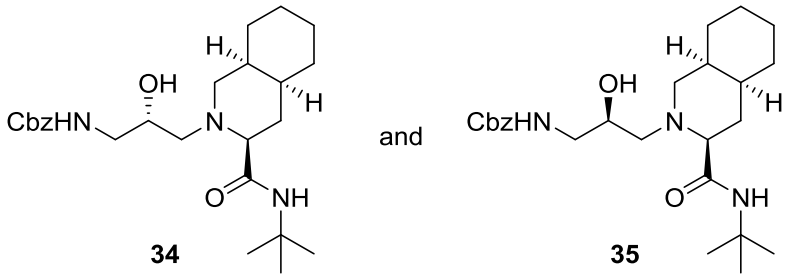
To a solution of **31** (571 mg, 2.36 mmol) in dichloromethane (7 mL) and methanol (4 mL) under nitrogen at 0 °C was added sodium borohydride (63 mg, 1.65 mmol). The reaction was stirred at 0 °C for 1 h, and then quenched with slow addition of 2 M HCl (2 mL). The reaction was evaporated to dryness and the solid was taken up in ethyl acetate. The solution was washed twice with water followed by brine, dried over sodium sulfate, and concentrated by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 0-8% EtOAc in DCM) to yield racemic **32** as a colorless oil (423 mg, 74% over two steps). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.33-7.28 (complex, 5H), 5.63 (t, J = 6 Hz, 1H), 5.06 (s, 2H), 3.88 (s, 2H), 3.48 (m, 2H), 3.39 (m, 1H), 3.22 (m, 1H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 157.23, 135.93, 128.36, 128.06, 127.91, 70.52, 66.90, 46.44, 43.96. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{11}H_{15}NO_3Cl$, 244.0740; observed, 244.0741.



33

Racemic benzyl (oxiran-2-ylmethyl)carbamate (33)

To a solution of **32** (333 mg, 1.37 mmol) in ethanol (15 mL) at 0 °C was added solid potassium hydroxide (84 mg, 1.50 mmol). The reaction was stirred cold for 15 min then warmed to room temperature and stirred for an additional 3 h. The reaction was evaporated to dryness and the solid was purified by silica flash column chromatography (gradient of 0-6% EtOAc in DCM) to yield **33** as a colorless oil (237 mg, 83%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.35-7.28 (complex, 5H), 5.29 (br s, 1H), 5.09 (s, 2H), 3.57 (dq, J_1 = 3 Hz, J_2 = 14.5 Hz, 1H), 3.22 (dt, J_1 = 6 Hz, J_2 = 14.5 Hz, 1H), 3.07 (m, 1H), 2.74 (t, J = 4 Hz, 1H), 2.56 (dd, J_1 = 2.5 Hz, J_2 = 4 Hz, 1H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 156.39, 136.23, 128.35, 127.99, 127.90, 66.66, 50.51, 44.87, 42.05. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{11}H_{14}NO_3$, 208.0974; observed, 208.0982.

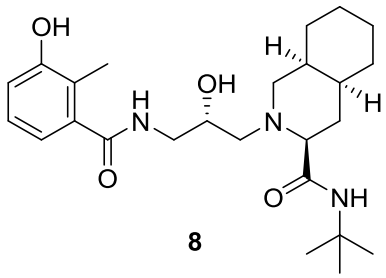


34 **35**

Benzyl ((S)-3-((3S,4aS,8aS)-3-(tert-butylcarbamoyl)octahydroisoquinolin-2(1H)-yl)-2-hydroxypropyl)carbamate (34) and benzyl ((R)-3-((3S,4aS,8aS)-3-(tert-butylcarbamoyl)octahydroisoquinolin-2(1H)-yl)-2-hydroxypropyl)carbamate (35)

To a solution of **33** (237 mg, 1.14 mmol) in isopropyl alcohol (15 mL) was added (3S,4aS,8aS)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (**6**, 327 mg, 1.37 mmol). The solution was heated to reflux and stirred. After 5 h, the reaction was cooled and dried to a solid. TLC analysis showed two diastereomers with the lower Rf compound being the S-isomer. The products were purified by silica flash column chromatography (gradient of 0-4% MeOH in DCM) to yield the S-isomer (**34**) as a white solid (205 mg, 40%) and R-isomer (**35**) as a white solid (205 mg, 40%). Characterization of the S-isomer (**34**): ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.35-7.25 (complex, 5H), 6.25 (m, 1H), 6.03 (s, 1H), 5.08 (m, 2H), 3.85 (d, J = 3.5 Hz, 1H), 3.74 (s, 1H), 3.56 (m, 1H), 3.19 (m, 1H), 2.79 (d, J = 11.5 Hz, 1H), 2.52 (dd, J_1 = 2.5 Hz, J_2 = 11.5 Hz, 1H), 2.44 (dd, J_1 = 9 Hz, J_2 = 12.5 Hz, 1H), 2.17 (m, 2H), 1.93 (q, J = 12 Hz, 1H), 1.84-1.14 (complex, 20H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.98, 157.68, 136.46, 128.24, 127.75, 127.72, 70.14, 68.27, 66.46, 58.99, 58.87, 50.92, 44.98, 35.71, 33.14, 30.71, 30.52, 28.53, 26.09, 25.44, 20.25. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{25}H_{40}N_3O_4$, 446.3019; observed, 446.3023. Characterization of the R-isomer (**35**): ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.35-7.28 (complex, 5H), 6.34 (s, 1H), 5.61 (t, J = 5.5 Hz, 1H), 5.06 (dd, J_1 = 12.5 Hz, J_2 = 14.5 Hz, 2H), 3.83 (s, 2H), 3.35 (dq, J_1 = 3 Hz, J_2 = 14 Hz, 1H), 3.03 (m, 1H), 2.83 (d, J = 11.5 Hz, 1H), 2.62 (dd, J_1 = 3 Hz, J_2 = 11 Hz, 1H), 2.47 (t, J = 11 Hz, 1H), 2.14 (dd, J_1 = 3 Hz, J_2 = 11 Hz, 1H), 2.01 (dd, J_1 = 2 Hz, J_2 = 13.5 Hz, 1H), 1.91-1.17 (complex, 21H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.48, 156.77, 136.33, 128.31, 127.90, 127.88, 69.54, 66.55, 66.39,

58.73, 58.05, 50.69, 44.70, 35.66, 33.09, 30.76, 30.73, 28.45, 26.16, 25.35, 20.12. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{25}H_{40}N_3O_4$, 446.3019; observed, 446.3021.

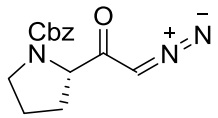


8

(3S,4aS,8aS)-N-(tert-butyl)-2-((S)-2-hydroxy-3-(3-hydroxy-2-methylbenzamido)propyl)decahydroisoquinoline-3-carboxamide (8)

A solution of **34** (44 mg, 0.099 mmol) and 10% palladium on carbon (32 mg, 0.030 mmol) in methanol (10 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite and washed with excess methanol. The reaction was then dried to a solid and redissolved in THF (3 mL). 3-hydroxy-2-methylbenzoic acid (17 mg, 0.109 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (21 mg, 0.109 mmol), and hydroxybenzotriazole hydrate (18 mg, 0.109 mmol) were added to the reaction and the solution was stirred at room temperature for 18 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-4% MeOH in DCM) to yield **8** as a white solid (24 mg, 54%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.45 (t, J = 5 Hz, 1H), 6.96 (t, J = 7.5 Hz, 1H), 6.91 (d, J = 6.5 Hz, 1H), 6.80 (d, J = 7.5 Hz, 1H), 5.83 (s, 1H), 3.94 (m, 2H), 3.39 (dt, J_1 = 5 Hz, J_2 = 14 Hz, 1H), 2.76 (d, J = 10 Hz, 1H), 2.51 (dd, J_1 = 2.5 Hz, J_2 = 11 Hz, 1H), 2.41 (dd, J_1 = 9 Hz, J_2 = 12.5 Hz, 1H), 2.18 (m, 4H), 2.00 (q, J = 12 Hz, 1H), 1.86 (qd, J_1 = 3 Hz, J_2 = 12.5 Hz, 1H), 1.70 (m, 2H), 1.59-1.14 (complex, 18H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 174.19, 172.84, 154.81, 137.62, 126.32, 122.40, 119.02, 116.66, 70.42,

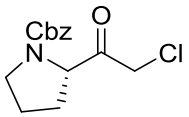
69.17, 59.46, 58.07, 51.35, 43.48, 35.93, 33.45, 30.90, 30.79, 28.45, 26.22, 25.50, 20.31, 12.51. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{25}H_{40}N_3O_4$, 446.3019; observed, 446.3021.



36

(S)-benzyl 2-(2-diazoacetyl)pyrrolidine-1-carboxylate (36)

To a stirred solution of Z-L-proline (1.02 g, 4.10 mmol) in anhydrous tetrahydrofuran (12 mL) under nitrogen at -78 °C was added anhydrous triethylamine (853 μL, 6.15 mmol) followed by isobutyl chloroformate (800 μL, 6.15 mmol). The reaction was stirred cold for 1 h and then warmed to room temperature for 1.5 h. The reaction was cooled to -78 °C before an ethereal solution of diazomethane (25 mL, 8.1 mmol) was added in one portion and the reaction was then warmed to room temperature with stirring for 1 h. The reaction was evaporated to dryness, resulting in a yellow solid. The solid was taken up in ethyl acetate, washed twice with water, twice with a saturated solution of sodium bicarbonate, and once with brine. The organic phase was dried over sodium sulfate and concentrated by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 3-4% EtOAc in $CHCl_3$) to yield **36** as a yellow oil (894 mg, 80%). ^{1}H NMR (500 MHz, $CDCl_3$, mixture of cis and trans isomers) δ ppm 7.38-7.26 (m, 10H), 5.50 (br s, 1H), 5.29 (br s, 1H), 5.19-5.05 (m, 4H), 4.33 (br s, 1H), 4.28 (br s, 1H), 3.56-3.47 (m, 4H), 2.17 (m, 1H), 2.10-1.83 (m, 7H). ^{13}C NMR (500 MHz, $CDCl_3$, mixture of cis and trans isomers) δ ppm 194.95, 194.18, 154.92, 154.24, 136.30, 136.14, 128.24, 128.18, 127.80, 127.69, 127.62, 66.88, 63.75, 53.02, 52.42, 47.09, 46.66, 30.93, 29.43, 24.09, 23.30. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{14}H_{16}N_3O_3$, 274.1192; observed, 274.1202.



37

(S)-benzyl 2-(2-chloroacetyl)pyrrolidine-1-carboxylate (37)

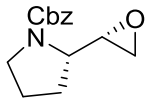
To a solution of **36** (894 mg, 3.27 mmol) in anhydrous tetrahydrofuran (8 mL) under nitrogen at 0 °C was added 1 M HCl in diethyl ether (4.91 mL, 4.91 mmol) dropwise. The reaction was stirred at 0 °C for 30 min after complete addition. The solution was dried under reduced pressure to yield **37** as a slightly yellow oil (860 mg). No further purification was made. ^{1}H NMR (500 MHz, $CDCl_3$, mixture of cis and trans isomers) δ ppm 7.35-7.25 (complex, 10H), 5.15-5.01 (complex, 4H), 4.74 (dd, J_1 = 4 Hz, J_2 = 8.5 Hz, 1H), 4.53 (dd, J_1 = 4 Hz, J_2 = 9 Hz, 1H), 4.31 (dd, J_1 = 16 Hz, J_2 = 21 Hz, 2H), 4.05 (dd, J_1 = 16 Hz, J_2 = 21 Hz, 2H), 3.59-3.48 (complex, 4H), 2.26-2.12 (complex, 2H), 1.99-1.81 (complex, 6H). ^{13}C NMR (500 MHz, $CDCl_3$, mixture of cis and trans isomers) δ ppm 201.36, 201.30, 154.79, 153.61, 136.12, 135.72, 128.29, 128.21, 127.99, 127.81, 127.79, 127.50, 67.14, 66.92, 62.80, 62.50, 46.96, 46.94, 46.48, 46.21, 30.08, 29.04, 24.32, 23.42. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{14}H_{17}NO_3Cl$, 282.0897; observed, 282.0894.

38 **39**

(S)-benzyl 2-((S)-2-chloro-1-hydroxyethyl)pyrrolidine-1-carboxylate (38) and (S)-benzyl 2-((R)-2-chloro-1-hydroxyethyl)pyrrolidine-1-carboxylate (39)

To a solution of **37** (860 mg, 3.05 mmol) in dichloromethane (7 mL) and methanol (4 mL) under nitrogen at 0 °C was added sodium borohydride (81 mg, 2.14 mmol). The reaction was stirred at 0 °C for 1 h, and then quenched with slow addition of 2 M HCl (2 mL). The reaction was evaporated to dryness and the solid was taken up in ethyl acetate. The solution was washed twice with water followed by brine, dried over sodium sulfate, and concentrated by rotary evaporation. TLC analysis showed two diastereomers with the lower Rf compound being the (S,S) product. Both diastereomers were purified by silica flash column chromatography (gradient of 0-10% EtOAc in DCM) to yield the (S,S)-isomer (**38**) as a white solid (279 mg, 32%) and (S,R)-isomer (**39**) as a white solid (429 mg, 50%) over two steps.

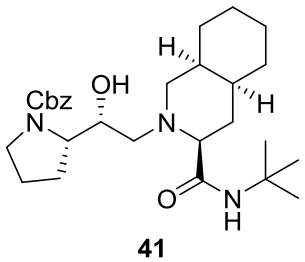
Characterization of the (S,S)-isomer (**38**): ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.36-7.28 (complex, 5H), 5.12 (dd, J_1 = 13 Hz, J_2 = 23.5 Hz, 2H), 4.04-3.95 (complex, 3H), 3.60-3.29 (complex, 4H), 2.08-1.74 (complex, 4H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 156.01, 136.27, 128.32, 127.89, 127.67, 72.52, 66.98, 61.04, 47.15, 46.84, 26.62, 23.92. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{14}H_{19}NO_3Cl$, 284.1053; observed, 284.1056. Characterization of the (S,R)-isomer (**39**): ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.37-7.28 (complex, 5H), 5.13 (dd, J_1 = 12.5 Hz, J_2 = 16 Hz, 2H), 4.95 (d, J = 2 Hz, 1H), 4.11 (m, 1H), 3.81 (br s, 1H), 3.72 (d, J = 11 Hz, 1H), 3.55 (m, 2H), 3.37 (m, 1H), 2.03 (m, 1H), 1.89 (m, 1H), 1.81 (m, 1H), 1.72 (m, 1H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 157.52, 136.04, 128.27, 127.87, 127.65, 74.69, 67.22, 60.57, 47.91, 47.05, 28.12, 23.94. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{14}H_{19}NO_3Cl$, 284.1053; observed, 284.1055.



40

(S)-benzyl 2-((S)-oxiran-2-yl)pyrrolidine-1-carboxylate (40)

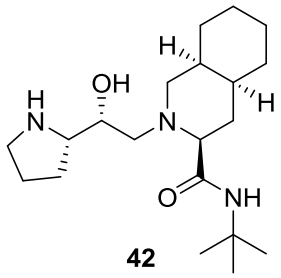
To a solution of **38** (297 mg, 0.983 mmol) in ethanol (15 mL) at 0 °C was added solid potassium hydroxide (61 mg, 1.08 mmol). The reaction was stirred cold for 15 min then warmed to room temperature and stirred for an additional 3 h. The reaction was evaporated to dryness and the solid was purified by silica flash column chromatography (gradient of 0-5% EtOAc in DCM) to yield **40** as a colorless oil (224 mg, 92%). ^{1}H NMR (500 MHz, $CDCl_3$, mixture of cis and trans isomers) δ ppm 7.38-7.28 (complex, 10H), 5.17-5.06 (complex, 4H), 3.72 (d, J = 3.5 Hz, 1H), 3.59 (br s, 1H), 3.49 (br s, 3H), 3.42 (m, 1H), 3.03 (s, 1H), 2.88 (s, 1H), 2.83 (t, J = 4.5 Hz, 1H), 2.78 (s, 1H), 2.65 (s, 1H), 2.49 (s, 1H), 2.09-1.83 (complex, 8H). ^{13}C NMR (500 MHz, $CDCl_3$, mixture of cis and trans isomers) δ ppm 154.96, 154.77, 136.57, 136.27, 128.28, 128.25, 128.01, 127.94, 127.72, 127.58, 66.89, 66.50, 58.87, 58.25, 52.67, 52.26, 47.38, 47.18, 46.96, 46.52, 28.39, 27.07, 23.84, 22.95. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{14}H_{18}NO_3$, 248.1287; observed, 248.1290.



41

(S)-benzyl 2-((R)-2-((3S,4aS,8aS)-3-(tert-butylcarbamoyl)octahydroisoquinolin-2(1H)-yl)-1-hydroxyethyl)pyrrolidine-1-carboxylate (41)

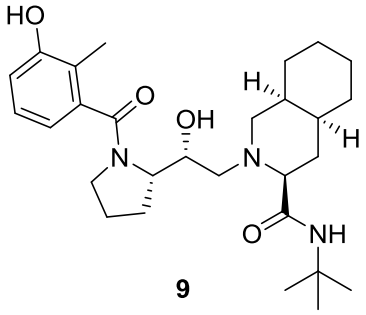
To a solution of **40** (233 mg, 0.902 mmol) in isopropyl alcohol (12 mL) was added (3S,4aS,8aS)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (**6**, 258 mg, 1.08 mmol). The solution was heated to reflux and stirred. After 5 h, the reaction was cooled and dried to a solid. The product was purified by silica flash column chromatography (gradient of 0-7% MeOH in DCM) to yield **41** as a white solid (211 mg, 48%). ^{1}H NMR (500 MHz, $CDCl_3$, predominately one cis/trans isomer) δ ppm 7.38-7.29 (complex, 5H), 6.31 (s, 1H), 5.12 (m, 2H), 4.18 (m, 1H), 4.02-3.87 (complex, 2H), 3.56 (m, 2H), 3.36 (m, 2H), 3.22 (d, J = 11.5 Hz, 1H), 2.56-2.41 (complex, 3H), 2.22-1.00 (complex, 25H). ^{13}C NMR (500 MHz, $CDCl_3$, predominately one cis/trans isomer) δ ppm 174.04, 155.82, 136.46, 128.37, 127.87, 127.65, 71.05, 70.11, 66.82, 61.80, 59.02, 58.73, 50.37, 47.43, 35.60, 32.88, 30.51, 28.57, 26.34, 26.08, 25.60, 25.21, 24.32, 20.49. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{28}H_{44}N_3O_4$, 486.3332; observed, 486.3325.



42

(3S,4aS,8aS)-N-(tert-butyl)-2-((R)-2-hydroxy-2-((S)-pyrrolidin-2-yl)ethyl)decahydroisoquinoline-3-carboxamide (42)

A solution of **41** (211 mg, 0.434 mmol) and 10% palladium on carbon (140 mg, 0.13 mmol) in methanol (16 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite, washed with excess methanol, and dried to a solid. The solid was taken up in ethyl acetate and extracted with 1 M HCl three times. The aqueous fractions were combined, adjusted to pH 12 with 10 M NaOH, and extracted with ethyl acetate three times. The combined ethyl acetate fractions were washed with brine, dried over sodium sulfate, and concentrated by rotary evaporation. The solid was further purified by silica flash column chromatography (gradient of 4-10% MeOH in DCM with 1% triethylamine) to yield **42** as a yellow solid (136 mg). The product was used without further purification.

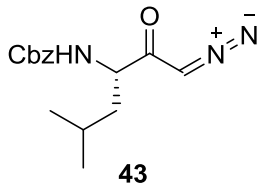


9

(3S,4aS,8aS)-N-(tert-butyl)-2-((R)-2-hydroxy-2-((S)-1-(3-hydroxy-2-methylbenzoyl)pyrrolidin-2-yl)ethyl)decahydroisoquinoline-3-carboxamide (9)

To a solution of **42** (24 mg) in THF (3 mL) was added 3-hydroxy-2-methylbenzoic acid (11 mg, 0.075 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (14 mg, 0.075 mmol), and hydroxybenzotriazole hydrate (12 mg, 0.075 mmol). The solution was stirred at room temperature for 16 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-5% MeOH in DCM) to yield **9** as a white solid (27 mg, 73% over two steps). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.00 (t, *J* = 7.5 Hz, 1H), 6.81 (d, *J* = 7.5 Hz, 1H), 6.66 (d, *J* = 7.5 Hz, 1H), 6.50 (s, 1H), 4.31 (t, *J* = 7.5 Hz, 1H), 4.16 (d,

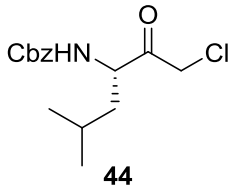
J = 5 Hz, 1H), 3.48 (s, 2H), 3.31 (d, *J* = 11 Hz, 1H), 3.24 (t, *J* = 9 Hz, 1H), 3.12 (br s, 1H), 2.64 (m, 2H), 2.26 (m, 2H), 2.12 (m, 4H), 1.89-1.21 (complex, 24H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 174.67, 171.81, 162.54, 155.3, 138.36, 126.93, 115.81, 72.06, 70.15, 62.76, 59.41, 59.12, 50.75, 50.72, 35.77, 32.88, 30.71, 28.71, 27.45, 26.23, 25.63, 24.70, 20.54, 12.57. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{28}H_{44}N_3O_4$, 486.3332; observed, 486.3317.



43

(S)-benzyl (1-diazo-5-methyl-2-oxohexan-3-yl)carbamate (43)

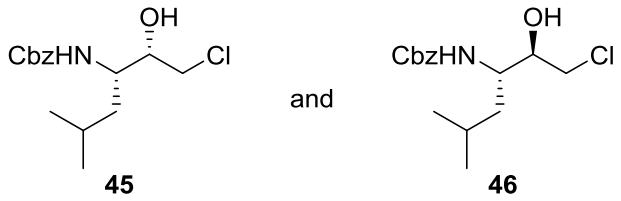
To a stirred solution of Z-L-leucine (1.02 g, 3.83 mmol) in anhydrous tetrahydrofuran (12 mL) under nitrogen at -78 °C was added anhydrous triethylamine (796 μL, 5.74 mmol) followed by isobutyl chloroformate (747 μL, 5.74 mmol). The reaction was stirred cold for 1 h and then warmed to room temperature for 1.5 h. The reaction was cooled to -78 °C before an ethereal solution of diazomethane (25 mL, 8.1 mmol) was added in one portion and the reaction was then warmed to room temperature with stirring for 1 h. The reaction was evaporated to dryness, resulting in a yellow solid. The solid was taken up in ethyl acetate, washed twice with water, twice with a saturated solution of sodium bicarbonate, and once with brine. The organic phase was dried over sodium sulfate and concentrated by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 0-6% EtOAc in $CHCl_3$) to yield **43** as a yellow solid (951 mg, 86%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.34-7.27 (complex, 5H), 5.77 (d, *J* = 8.5 Hz, 1H), 5.48 (s, 1H), 5.08 (dd, J_1 = 12.5 Hz, J_2 = 17.5 Hz, 2H), 4.25 (br s, 1H), 1.68 (m, 1H), 1.55 (m, 1H), 1.46 (m, 1H), 0.93-0.84 (complex, 6H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 194.39, 155.92, 136.05, 128.26, 127.91, 127.76, 66.68, 56.29, 53.53, 40.94, 24.45, 22.83, 21.50. HRMS (*m/z*): $[M + Na]^+$ calc. for $C_{15}H_{19}N_3O_3Na$, 312.1324; observed, 312.1322.



44

(S)-benzyl (1-chloro-5-methyl-2-oxohexan-3-yl)carbamate (44)

To a solution of **43** (951 mg, 3.29 mmol) in anhydrous tetrahydrofuran (8 mL) under nitrogen at 0 °C was added 1 M HCl in diethyl ether (4.93 mL, 4.93 mmol) dropwise. The reaction was stirred at 0 °C for 30 min after complete addition. The solution was dried under reduced pressure to yield **44** as a yellow oil (951 mg). No further purification was made. ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.34-7.28 (complex, 5H), 5.60 (d, J = 8.5 Hz, 1H), 5.07 (dd, J_1 = 12 Hz, J_2 = 15.5 Hz, 2H), 4.57 (m, 1H), 4.26 (dd, J_1 = 16 Hz, J_2 = 26.5 Hz, 2H), 1.69 (m, 1H), 1.57 (m, 1H), 1.42 (m, 1H), 0.96-0.86 (complex, 6H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 201.86, 156.04, 135.87, 128.35, 128.05, 127.83, 66.95, 56.15, 46.53, 39.88, 24.58, 22.99, 21.23. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{15}H_{21}NO_3Cl$, 298.1210; observed, 298.1201.

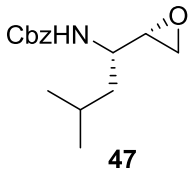


45 **46**

Benzyl ((2S,3S)-1-chloro-2-hydroxy-5-methylhexan-3-yl)carbamate (45) and benzyl ((2R,3S)-1-chloro-2-hydroxy-5-methylhexan-3-yl)carbamate (46)

To a solution of **44** (951 mg, 3.19 mmol) in dichloromethane (7 mL) and methanol (4 mL) under nitrogen at 0 °C was added sodium borohydride (85 mg, 2.24 mmol). The reaction was stirred at 0 °C for 1 h, and then quenched with slow addition of 2 M HCl (2 mL). The reaction was evaporated to dryness and the solid was taken up in ethyl acetate. The solution was washed twice with water followed by brine, dried over sodium sulfate, and concentrated by rotary evaporation. TLC analysis showed two diastereomers with the lower Rf compound being the (S,S)-isomer. Both diastereomers were purified by silica flash column chromatography (gradient of 0-5% EtOAc in DCM) to yield the (S,S)-isomer (**45**) as a

white solid (569 mg, 59%) and the (R,S)-isomer (**46**) as a white solid (62 mg, 6%) over two steps. Characterization of the (S,S)-isomer (**45**): ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.36-7.28 (complex, 5H), 5.08 (m, 3H), 3.81 (m, 2H), 3.57 (dd, J_1 = 4 Hz, J_2 = 11.5 Hz, 1H), 3.50 (dd, J_1 = 8 Hz, J_2 = 11.5 Hz, 1H), 3.27 (d, J = 4 Hz, 1H), 1.64 (m, 1H), 1.41 (m, 1H), 1.30 (m, 1H), 0.91 (t, J = 7 Hz, 6H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 156.51, 136.12, 128.46, 128.13, 127.97, 74.45, 66.92, 51.99, 46.79, 38.35, 24.49, 23.59, 21.33. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{15}H_{23}NO_3Cl$, 300.1366; observed, 300.1374. Characterization of the (R,S)-isomer (**46**): ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.38-7.30 (complex, 5H), 5.09 (s, 2H), 5.02 (d, J = 9.5 Hz, 1H), 3.83 (m, 1H), 3.76 (m, 1H), 3.62 (dd, J_1 = 4 Hz, J_2 = 11.5 Hz, 1H), 3.49 (dd, J_1 = 9 Hz, J_2 = 11.5 Hz, 1H), 2.87 (d, J = 4.5 Hz, 1H), 1.61 (m, 2H), 1.35 (m, 1H), 0.93 (dd, J_1 = 3 Hz, J_2 = 6 Hz, 6H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 156.64, 136.25, 128.53, 128.19, 128.00, 73.58, 66.95, 51.02, 47.84, 41.55, 24.62, 22.99, 22.04. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{15}H_{23}NO_3Cl$, 300.1366; observed, 300.1360.

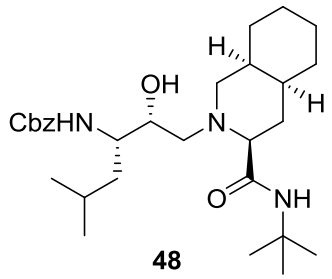


47

Benzyl ((S)-3-methyl-1-((S)-oxiran-2-yl)butyl)carbamate (47)

To a solution of **45** (569 mg, 1.90 mmol) in ethanol (15 mL) at 0 °C was added solid potassium hydroxide (117 mg, 2.09 mmol). The reaction was stirred cold for 15 min then warmed to room temperature and stirred for an additional 3 h. The reaction was evaporated to dryness and the solid was purified by silica flash column chromatography (gradient of 0-5% EtOAc in DCM) to yield **47** as a colorless oil (455 mg, 91%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.33-7.26 (complex, 5H), 5.21 (d, J = 8.5 Hz, 1H), 5.06 (s, 2H), 3.57 (m, 1H), 2.82 (br s, 1H), 2.72 (s, 1H), 2.68 (m, 1H), 1.70 (m, 1H), 1.38 (t, J = 7 Hz, 2H), 0.90 (dd, J_1 = 6.5 Hz, J_2 = 12.5 Hz, 6H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 155.87,

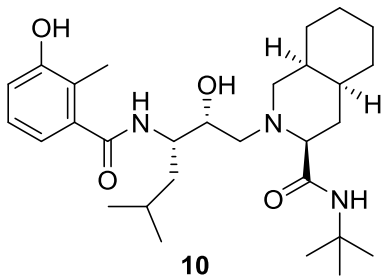
136.21, 128.21, 127.82, 127.72, 66.39, 54.02, 50.56, 45.58, 40.39, 24.12, 23.04, 21.46. HRMS (*m/z*): [M + H]$^+$ calc. for $C_{15}H_{22}NO_3$, 264.1600; observed, 264.1597.



48

Benzyl ((2R,3S)-1-((3S,4aS,8aS)-3-(tert-butylcarbamoyl)octahydroisoquinolin-2(1H)-yl)-2-hydroxy-5-methylhexan-3-yl)carbamate (48)

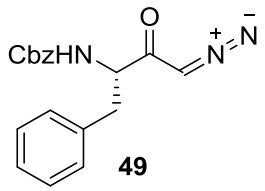
To a solution of **47** (455 mg, 1.73 mmol) in isopropyl alcohol (15 mL) was added (3S,4aS,8aS)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (**6**, 494 mg, 2.07 mmol). The solution was heated to reflux and stirred. After 5 h, the reaction was cooled and dried to a solid. The product was purified by silica flash column chromatography (gradient of 0-4% MeOH in DCM) to yield **48** as a white solid (764 mg, 88%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.33-7.27 (complex, 5H), 6.14 (s, 1H), 5.61 (d, *J* = 9 Hz, 1H), 5.10 (d, *J* = 12 Hz, 1H), 5.04 (d, *J* = 12 Hz, 1H), 3.08 (m, 2H), 3.55 (s, 1H), 3.00 (dd, J_1 = 2 Hz, J_2 = 12 Hz, 1H), 2.55 (m, 2H), 2.18 (m, 3H), 1.88 (q, *J* = 12 Hz, 1H), 1.70-1.17 (complex, 22H), 0.91 (d, *J* = 6.5 Hz, 6H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.89, 156.91, 136.47, 128.26, 127.80, 72.77, 70.04, 66.47, 64.05, 59.05, 58.85, 53.61, 50.72, 38.59, 35.61, 33.18, 30.55, 30.43, 28.57, 25.99, 25.80, 25.18, 24.60, 23.54, 21.40, 20.61. HRMS (*m/z*): [M + H]$^+$ calc. for $C_{29}H_{48}N_3O_4$, 502.3645; observed, 502.3637.



10

(3S,4aS,8aS)-N-(tert-butyl)-2-((2R,3S)-2-hydroxy-3-(3-hydroxy-2-methylbenzamido)-5-methylhexyl)decahydroisoquinoline-3-carboxamide (10)

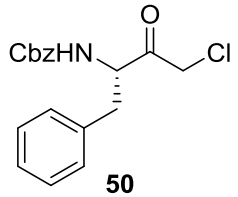
A solution of **48** (53 mg, 0.106 mmol) and 10% palladium on carbon (34 mg, 0.032 mmol) in methanol (6 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite and washed with excess methanol. The reaction was then dried to a solid and redissolved in THF (3 mL). 3-hydroxy-2-methylbenzoic acid (18 mg, 0.117 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (22 mg, 0.117 mmol), and hydroxybenzotriazole hydrate (19 mg, 0.117 mmol) were added to the reaction and the solution was stirred at room temperature for 18 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-5% MeOH in DCM) to yield **10** as a white solid (39 mg, 73%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.40 (br s, 1H), 6.97 (t, J = 7.5 Hz, 1H), 6.87 (d, J = 7.5 Hz, 1H), 6.83 (d, J = 8 Hz, 1H), 6.62 (d, J = 8 Hz, 1H), 5.99 (s, 1H), 4.28 (m, 1H), 3.89 (m, 1H), 3.00 (dd, J_1 = 2.5 Hz, J_2 = 12 Hz, 1H), 2.55 (dd, J_1 = 6 Hz, J_2 = 13 Hz, 1H), 2.50 (dd, J_1 = 3.5 Hz, J_2 = 11 Hz, 1H), 2.22 (m, 4H), 1.91 (q, J = 11.5 Hz, 1H), 1.85-1.19 (complex, 25H), 0.97 (d, J = 6 Hz, 6H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 174.27, 171.55, 155.03, 137.79, 126.34, 122.54, 118.70, 116.75, 72.90, 70.24, 64.43, 59.46, 59.20, 52.82, 51.04, 38.96, 35.86, 33.36, 30.70, 28.54, 26.11, 25.94, 25.31, 23.59, 21.50, 20.67, 12.54. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{29}H_{48}N_3O_4$, 502.3645; observed, 502.3635.



49

(S)-benzyl (4-diazo-3-oxo-1-phenylbutan-2-yl)carbamate (49)

To a stirred solution of Z-L-phenylalanine (971 mg, 3.25 mmol) in anhydrous tetrahydrofuran (12 mL) under nitrogen at -78 °C was added anhydrous triethylamine (675 μL, 4.87 mmol) followed by isobutyl chloroformate (633 μL, 4.87 mmol). The reaction was stirred cold for 1 h and then warmed to room temperature for 1.5 h. The reaction was cooled to -78 °C before an ethereal solution of diazomethane (25 mL, 8.1 mmol) was added in one portion and the reaction was then warmed to room temperature with stirring for 1 h. The reaction was evaporated to dryness, resulting in a yellow solid. The solid was taken up in ethyl acetate, washed twice with water, twice with a saturated solution of sodium bicarbonate, and once with brine. The organic phase was dried over sodium sulfate and concentrated by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 2-3% EtOAc in $CHCl_3$) to yield **49** as a yellowish foam (642 mg, 61%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.35-7.21 (complex, 8H), 7.16 (d, *J* = 7 Hz, 2H), 5.52 (d, *J* = 7.5 Hz, 1H), 5.22 (s, 1H), 5.06 (m, 2H) 4.47 (d, *J* = 6.5 Hz, 1H), 3.07-2.98 (m, 2H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 192.73, 155.65, 136.06, 135.97, 129.23, 128.57, 128.44, 128.12, 127.95, 126.97, 66.90, 58.77, 54.50, 38.29. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{18}N_3O_3$, 324.1348; observed, 324.1355.

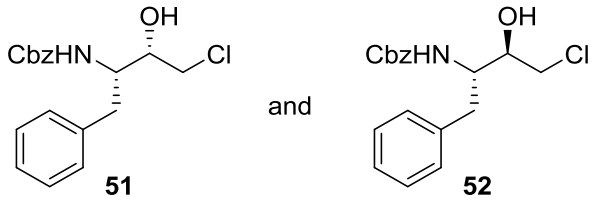


50

(S)-benzyl (4-chloro-3-oxo-1-phenylbutan-2-yl)carbamate (50)

To a solution of **49** (642 mg, 1.99 mmol) in anhydrous tetrahydrofuran (8 mL) under nitrogen at 0 °C was added 1 M HCl in diethyl ether (2.99 mL, 2.99 mmol) dropwise. The reaction was stirred at 0 °C for 30 min after complete addition. The solution was dried under reduced pressure to yield **50** as a white solid (717 mg). No further purification was made. ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.35-7.22 (complex, 8H), 7.12 (d, *J* = 7 Hz, 2H), 5.56 (d, *J* = 7.5 Hz, 1H), 5.04 (dd, J_1 = 12.5 Hz, J_2 = 16 Hz, 2H), 4.72 (dd, J_1 = 7 Hz, J_2 = 14.5 Hz, 1H), 4.15 (d, *J* = 16 Hz, 1H), 3.97 (d, *J* = 16.5 Hz, 1H), 3.08 (dd, J_1 = 7

Hz, J_2 = 14 Hz, 1H), 2.97 (dd, J_1 = 7 Hz, J_2 = 14 Hz, 1H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 200.90, 155.72, 135.83, 135.26, 128.98, 128.79, 128.43, 128.16, 127.94, 127.24, 67.06, 58.65, 47.36, 37.35. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{19}NO_3Cl$, 332.1053; observed, 332.1052.

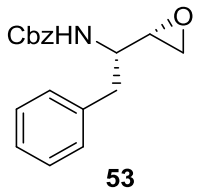


51 **52**

Benzyl ((2S,3S)-4-chloro-3-hydroxy-1-phenylbutan-2-yl)carbamate (51) and benzyl ((2S,3R)-4-chloro-3-hydroxy-1-phenylbutan-2-yl)carbamate (52)

To a solution of **50** (717 mg, 2.16 mmol) in dichloromethane (7 mL) and methanol (4 mL) under nitrogen at 0 °C was added sodium borohydride (57 mg, 1.51 mmol). The reaction was stirred at 0 °C for 1 h, and then quenched with slow addition of 2 M HCl (2 mL). The reaction was evaporated to dryness and the solid was taken up in ethyl acetate. The solution was washed twice with water followed by brine, dried over sodium sulfate, and concentrated by rotary evaporation. TLC analysis showed two diastereomers with the lower Rf compound being the (S,S)-isomer. The (S,S)-isomer was recrystallized in hexanes/ethyl acetate. The crystals were collected and washed with cold 2:1 hexanes:ethyl acetate. The filtrate was collected, dried to a solid, and recrystallized again in the same manner. Both diastereomers were then separated further by silica flash column chromatography (gradient of 0-5% EtOAc in DCM) to yield the (S,S)-isomer (**51**) as a white solid (452 mg, 63%) and the (S,R)-isomer (**52**) as a white solid (115 mg, 16%) over two steps. Characterization of the (S,S)-isomer (**51**): ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.35-7.16 (complex, 10H), 5.02 (s, 2H), 4.92 (d, J = 9 Hz, 1H), 3.99 (m, 1H), 3.84 (m, 1H), 3.63 (dd, J_1 = 3.5 Hz, J_2 = 11.5 Hz, 1H), 3.55 (dd, J_1 = 7.5 Hz, J_2 = 11 Hz, 1H), 3.12 (d, J = 4 Hz, 1H), 2.97 (dd, J_1 = 4.5 Hz, J_2 = 14 Hz, 1H), 2.90 (dd, J_1 = 8.5 Hz, J_2 = 13.5 Hz, 1H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 156.22, 137.04, 136.12, 129.37, 128.57, 128.49, 128.15, 127.92, 126.67, 73.16, 66.87, 54.60, 47.46, 35.52. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{21}NO_3Cl$, 334.1210; observed, 334.1213. Characterization of

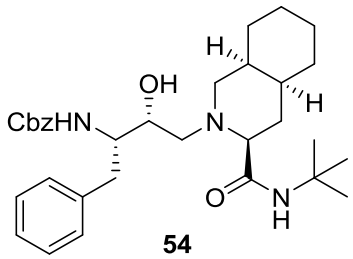
the (S,R)-isomer (**52**): ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.35-7.19 (complex, 10H), 5.32 (d, J = 9 Hz, 1H), 5.05 (dd, J_1 = 12.5 Hz, J_2 = 16 Hz, 2H), 3.95 (dd, J_1 = 8 Hz, J_2 = 16 Hz, 1H), 3.73 (s, 1H), 3.49 (d, J = 6.5 Hz, 2H), 3.26 (s, 1H), 2.98-2.87 (complex, 2H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 156.49, 137.44, 136.12, 129.20, 128.54, 128.47, 128.13, 127.92, 126.60, 71.13, 66.89, 54.23, 47.55, 38.42. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{21}NO_3Cl$, 334.1210; observed, 334.1207.



53

Benzyl ((S)-1-((S)-oxiran-2-yl)-2-phenylethyl)carbamate (53)

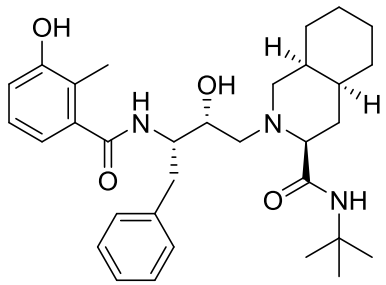
To a solution of **51** (452 mg, 1.35 mmol) in ethanol (15 mL) and dichloromethane (5 mL) at 0 °C was added solid potassium hydroxide (83 mg, 1.49 mmol). The reaction was stirred cold for 15 min then warmed to room temperature and stirred for an additional 3 h. The reaction was evaporated to dryness and the solid was purified by silica flash column chromatography (gradient of 0-3% EtOAc in DCM) to yield **53** as a white solid (361 mg, 90%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.32-7.14 (complex, 10H), 5.00 (m, 3H), 3.76 (br s, 1H), 2.96-2.72 (complex, 5H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 155.71, 136.40, 136.12, 129.17, 128.40, 128.30, 127.94, 127.80, 126.56, 66.52, 52.96, 46.46, 37.27. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{20}NO_3$, 298.1443; observed, 298.1447.



54

Benzyl ((2S,3R)-4-((3S,4aS,8aS)-3-(tert-butylcarbamoyl)octahydroisoquinolin-2(1H)-yl)-3-hydroxy-1-phenylbutan-2-yl)carbamate (54)

To a solution of **53** (360 mg, 1.21 mmol) in isopropyl alcohol (10 mL) was added (3S,4aS,8aS)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (**6**, 346 mg, 1.45 mmol). The solution was heated to reflux and stirred. After 5 h, the reaction was cooled and dried to a solid. The product was purified by silica flash column chromatography (gradient of 0-4% MeOH in DCM) to yield **54** as a white solid (606 mg, 93%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.28-7.14 (complex, 10H), 6.19 (s, 1H), 5.74 (d, J = 8.5 Hz, 1H), 4.99 (d, J = 12.5 Hz, 1H), 4.90 (d, J = 12.5 Hz, 1H), 3.99 (m, 3H), 2.95 (m, 3H), 2.66 (dd, J_1 = 6 Hz, J_2 = 12.5 Hz, 1H), 2.57 (d, J = 9.5 Hz, 1H), 2.27 (dd, J_1 = 5.5 Hz, J_2 = 12.5 Hz, 1H), 2.19 (d, J = 10 Hz, 1H), 1.94-1.15 (complex, 21H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.80, 156.50, 138.22, 136.38, 129.07, 128.10, 127.57, 127.45, 125.98, 71.07, 70.07, 66.10, 58.94, 58.79, 55.90, 50.67, 35.55, 35.10, 33.08, 30.47, 30.40, 28.43, 25.95, 25.59, 20.47. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{32}H_{46}N_3O_4$, 536.3488; observed, 536.3486.

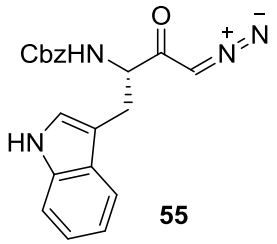


11

(3S,4aS,8aS)-N-(tert-butyl)-2-((2R,3S)-2-hydroxy-3-(3-hydroxy-2-methylbenzamido)-4-phenylbutyl)decahydroisoquinoline-3-carboxamide (11)

A solution of **54** (37 mg, 0.069 mmol) and 10% palladium on carbon (22 mg, 0.021 mmol) in methanol (10 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite and washed with excess methanol. The reaction was then dried to a solid and redissolved in THF (3 mL). 3-hydroxy-2-methylbenzoic acid (12 mg, 0.076 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (15 mg, 0.076 mmol), and hydroxybenzotriazole hydrate (12 mg, 0.076 mmol) were added

to the reaction and the solution was stirred at room temperature for 18 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-4% MeOH in DCM) to yield **11** as a white solid (21 mg, 57%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.29 (m, 4H), 7.22 (m, 1H), 6.86 (t, J = 7.5 Hz, 1H), 6.74 (d, J = 7.5 Hz, 1H), 6.52 (m, 2H), 5.88 (s, 1H), 4.52 (m, 1H), 4.00 (m, 1H), 3.27 (dd, J_1 = 11.5 Hz, J_2 = 14.5 Hz, 1H), 2.98 (m, 2H), 2.63 (dd, J_1 = 7 Hz, J_2 = 13 Hz, 1H), 2.54 (dd, J_1 = 3 Hz, J_2 = 11 Hz, 1H), 2.30 (dd, J_1 = 6.5 Hz, J_2 = 13 Hz, 1H), 2.25 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 1.97 (m, 4H), 1.85 (qd, J_1 = 3 Hz, J_2 = 13.5 Hz, 1H), 1.76-1.15 (complex, 21H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 174.07, 171.41, 154.69, 138.49, 137.58, 129.26, 128.53, 126.47, 126.29, 122.41, 118.84, 116.66, 71.78, 70.41, 59.70, 59.07, 55.61, 51.13, 35.97, 35.47, 33.43, 30.76, 28.52, 26.18, 25.86, 20.61, 12.22. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{32}H_{46}N_3O_4$, 536.3488; observed, 536.3487.

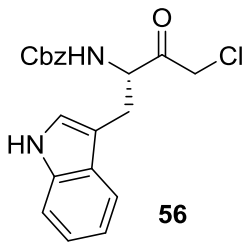


55

(S)-benzyl (4-diazo-1-(1H-indol-3-yl)-3-oxobutan-2-yl)carbamate (55)

To a stirred solution of Z-L-tryptophan (1.01 g, 2.97 mmol) in anhydrous tetrahydrofuran (12 mL) under nitrogen at -78 °C was added anhydrous triethylamine (618 μL, 4.46 mmol) followed by isobutyl chloroformate (580 μL, 4.46 mmol). The reaction was stirred cold for 1 h and then warmed to room temperature for 1.5 h. The reaction was cooled to -78 °C before an ethereal solution of diazomethane (25 mL, 8.1 mmol) was added in one portion and the reaction was then warmed to room temperature with stirring for 1 h. The reaction was evaporated to dryness, resulting in a yellow solid. The solid was taken up in ethyl acetate, washed twice with water, twice with a saturated solution of sodium

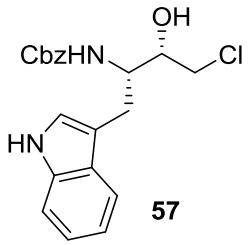
bicarbonate, and once with brine. The organic phase was dried over sodium sulfate and concentrated by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 5-6% EtOAc in $CHCl_3$) to yield **55** as a yellowish foam (916 mg, 85%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.08 (s, 1H), 7.62 (d, *J* = 7.5 Hz, 1H), 7.38-7.28 (m, 6H), 7.20 (t, *J* = 7.5 Hz, 1H), 7.12 (t, *J* = 8 Hz, 1H), 7.00 (s, 1H), 5.45 (s, 1H), 5.16-5.07 (m, 3H), 4.59 (s, 1H), 3.27 (dd, J_1 = 5 Hz, J_2 = 14 Hz, 1H), 3.19 (dd, J_1 = 7.5 Hz, J_2 = 14 Hz, 1H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 193.51, 155.81, 136.15, 128.53, 128.21, 128.11, 127.39, 123.13, 123.04, 122.38, 119.88, 118.77, 111.23, 110.17, 67.01, 58.20, 54.45, 28.51. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{20}H_{19}N_4O_3$, 363.1457; observed, 363.1464.



56

(S)-benzyl (4-chloro-1-(1H-indol-3-yl)-3-oxobutan-2-yl)carbamate (56)

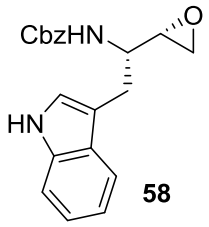
To a solution of **55** (916 mg, 2.53 mmol) in anhydrous tetrahydrofuran (8 mL) under nitrogen at 0 °C was added 1 M HCl in diethyl ether (3.80 mL, 3.80 mmol) dropwise. The reaction was stirred at 0 °C for 30 min after complete addition. The solution was dried under reduced pressure to yield **56** as a pink solid (886 mg). No further purification was made. ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.14 (s, 1H), 7.59 (d, *J* = 8 Hz, 1H), 7.37-7.31 (complex, 6H), 7.22 (t, *J* = 7.5 Hz, 1H), 7.14 (t, *J* = 7.5 Hz, 1H), 6.95 (s, 1H), 5.43 (d, *J* = 7Hz, 1H), 5.09 (m, 2H), 4.86 (dd, J_1 = 7 Hz, J_2 = 14 Hz, 1H), 4.09 (d, *J* = 16.5 Hz, 1H), 3.95 (d, *J* = 16.5 Hz, 1H), 3.25 (m, 2H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 201.63, 155.85, 136.17, 135.92, 128.56, 128.31, 128.15, 126.93, 122.95, 122.62, 120.08, 118.51, 111.42, 109.30, 67.22, 58.07, 47.70, 27.80. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{20}H_{20}N_2O_3Cl$, 371.1162; observed, 371.1167.



57

Benzyl ((2S,3S)-4-chloro-3-hydroxy-1-(1H-indol-3-yl)butan-2-yl)carbamate (57)

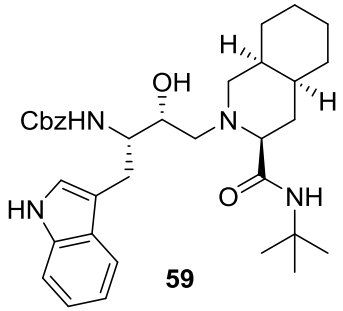
To a solution of **56** (886 mg, 2.38 mmol) in dichloromethane (7 mL) and methanol (4 mL) under nitrogen at 0 °C was added sodium borohydride (67 mg, 1.77 mmol). The reaction was stirred at 0 °C for 1 h, and then quenched with slow addition of 2 M HCl (2 mL). The reaction was evaporated to dryness and the solid was taken up in ethyl acetate. The solution was washed twice with water followed by brine, dried over sodium sulfate, and concentrated by rotary evaporation. TLC analysis showed two diastereomers with the lower Rf compound being the desired (S,S)-isomer. The product was purified by silica flash column chromatography (gradient of 0-2% MeOH in DCM) to yield **57** as a white solid (566 mg, 60% over two steps). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 10.78 (s, 1H), 7.51 (d, J = 8 Hz, 1H), 7.34-7.25 (complex, 3H), 7.21 (d, J = 6.5 Hz, 2H), 7.16 (d, J = 9 Hz, 1H), 7.07 (d, J = 1.5 Hz, 1H), 7.04 (t, J = 7 Hz, 1H), 6.93 (t, J = 7 Hz, 1H), 5.47 (d, J = 6.5 Hz, 1H), 4.92 (dd, J_1 = 12.5 Hz, J_2 = 20 Hz, 2H), 3.78-3.68 (complex, 2H), 3.64 (m, 1H), 3.52 (dd, J_1 = 7 Hz, J_2 = 11 Hz, 1H), 3.08 (dd, J_1 = 3 Hz, J_2 = 14.5 Hz, 1H), 2.76 (dd, J_1 = 9 Hz, J_2 = 14.5 Hz, 1H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 155.76, 137.25, 136.08, 128.28, 127.68, 127.60, 127.36, 123.29, 120.69, 120.58, 118.37, 118.11, 111.22, 72.41, 64.95, 54.30, 48.00, 25.18. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{20}H_{22}N_2O_3Cl$, 373.1319; observed, 373.1319.



58

Benzyl ((S)-2-(1H-indol-3-yl)-1-((S)-oxiran-2-yl)ethyl)carbamate (58)

To a solution of **57** (57 mg, 0.153 mmol) in ethanol (5 mL) and dichloromethane (5 mL) at 0 °C was added solid potassium hydroxide (10 mg, 0.168 mmol). The reaction was stirred cold for 15 min then warmed to room temperature and stirred for an additional 3 h. The reaction was evaporated to dryness and the solid was purified by silica flash column chromatography (gradient of 1-3% EtOAc in DCM) to yield **58** as a white solid (45 mg, 87%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.13 (s, 1H), 7.64 (d, J = 8 Hz, 1H), 7.35-7.28 (complex, 6H), 7.20 (t, J = 8 Hz, 1H), 7.12 (t, J = 8 Hz, 1H), 7.01 (s, 1H), 5.04 (s, 2H), 4.84 (m, 1H), 3.80 (br s, 1H), 3.11 (d, J = 5 Hz, 2H), 2.93 (br s, 1H), 2.79 (br s, 2H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 155.89, 136.22, 128.49, 128.13, 128.06, 122.96, 122.27, 119.76, 118.81, 111.17, 66.74, 53.13, 52.68, 47.15, 27.00. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{20}H_{21}N_2O_3$, 337.1552; observed, 337.1562.

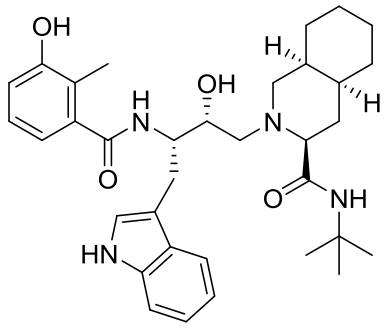


59

Benzyl ((2S,3R)-4-((3S,4aS,8aS)-3-(tert-butylcarbamoyl)octahydroisoquinolin-2(1H)-yl)-3-hydroxy-1-(1H-indol-3-yl)butan-2-yl)carbamate (59)

To a solution of **58** (41 mg, 0.122 mmol) in isopropyl alcohol (12 mL) was added (3S,4aS,8aS)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (**6**, 35 mg, 0.146 mmol). The solution was heated to reflux and stirred. After 5 h, the reaction was cooled and dried to a solid. The product was purified by silica flash column chromatography (gradient of 0-6% MeOH in DCM) to yield **59** as a white solid (70 mg, 100%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.28 (s, 1H), 7.63 (d, J = 8 Hz, 1H), 7.38-7.21 (complex, 6H), 7.16 (t, J = 7 Hz, 1H), 7.07 (t, J = 7.5 Hz, 1H), 7.01 (br s, 1H), 5.99 (s, 1H), 5.36 (d, J = 8 Hz, 1H), 5.00 (dd, J_1 = 12.5 Hz, J_2 = 16 Hz, 2H), 4.10 (m, 1H), 3.89 (m, 1H), 3.58 (br s, 1H), 3.20 (dd, J_1 = 9 Hz, J_2 = 15 Hz, 1H), 3.01 (m, 2H), 2.69 (dd, J_1 = 6 Hz, J_2 = 13.5 Hz, 1H), 2.58 (dd, J_1 = 2.5 Hz, J_2 = 11 Hz,

1H), 2.31 (dd, J_1 = 6 Hz, J_2 = 13.5 Hz, 1H), 2.23 (d, J = 11 Hz, 1H), 1.97-1.18 (complex, 21H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.75, 156.86, 136.44, 136.25, 128.35, 128.01, 127.89, 127.81, 122.65, 121.84, 119.27, 118.84, 111.73, 111.11, 71.02, 70.38, 66.53, 59.48, 58.98, 55.39, 50.87, 35.75, 33.25, 30.67, 30.62, 28.54, 26.08, 25.87, 25.49, 20.69. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{34}H_{47}N_4O_4$, 575.3597; observed, 575.3589.

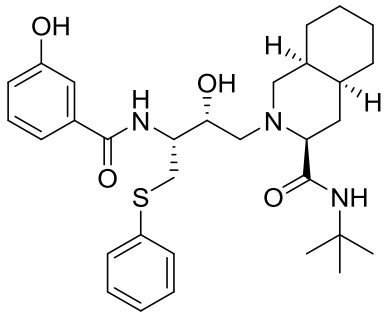


12

(3S,4aS,8aS)-N-(tert-butyl)-2-((2R,3S)-2-hydroxy-3-(3-hydroxy-2-methylbenzamido)-4-(1H-indol-3-yl)butyl)decahydroisoquinoline-3-carboxamide (12)

A solution of **59** (37 mg, 0.064 mmol) and 10% palladium on carbon (20 mg, 0.019 mmol) in methanol (6 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite and washed with excess methanol. The reaction was then dried to a solid and redissolved in THF (3 mL). 3-hydroxy-2-methylbenzoic acid (11 mg, 0.070 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (13 mg, 0.070 mmol), and hydroxybenzotriazole hydrate (11 mg, 0.070 mmol) were added to the reaction and the solution was stirred at room temperature for 18 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-7% MeOH in DCM) to yield **12** as a white solid (22 mg, 59%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.29 (s, 1H), 7.66 (d, J = 8 Hz, 1H), 7.32 (d, J = 8.5 Hz, 1H), 7.16 (t, J = 7.5 Hz, 1H), 7.08 (t, J = 7.5 Hz, 1H), 7.03 (s, 1H), 6.82 (t, J = 8 Hz, 1H), 6.73 (d, J = 8 Hz, 1H), 6.60 (d, J

= 8.5 Hz, 1H), 6.53 (d, J = 7.5 Hz, 1H), 5.92 (s, 1H), 4.62 (m, 1H), 4.04 (m, 2H), 3.44 (dd, J_1 = 11 Hz, J_2 = 15.5 Hz, 1H), 3.08 (m, 2H), 2.69 (dd, J_1 = 7.5 Hz, J_2 = 13.5 Hz, 1H), 2.57 (dd, J_1 = 2 Hz, J_2 = 10.5 Hz, 1H), 2.33 (dd, J_1 = 6.5 Hz, J_2 = 13 Hz, 1H), 2.27 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 2.03-1.12 (complex, 25H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 174.00, 171.55, 154.53, 137.60, 136.33, 127.72, 126.39, 122.33, 122.29, 122.00, 119.36, 119.15, 118.87, 116.69, 112.37, 111.17, 71.89, 70.51, 59.75, 59.06, 55.02, 51.07, 35.96, 33.41, 30.85, 30.76, 28.47, 26.20, 26.01, 25.86, 25.35, 20.65, 12.17. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{34}H_{47}N_4O_4$, 575.3597; observed, 575.3600.

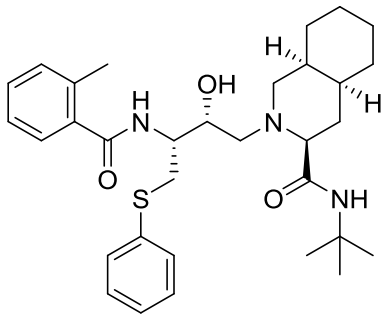


13

(3S,4aS,8aS)-N-(tert-butyl)-2-((2R,3R)-2-hydroxy-3-(3-hydroxybenzamido)-4-(phenylthio)butyl)decahydroisoquinoline-3-carboxamide (13)

This reaction was performed according to the general EDC coupling procedure (0.039 mmol 3-hydroxybenzoic acid for 48 h). Purification by flash column chromatography (gradient of 1-5% MeOH in DCM) yielded **13** as a white solid (11 mg, 54%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.58 (d, J = 8.5 Hz, 1H), 7.39 (d, J = 8 Hz, 2H), 7.33 (d, J = 8 Hz, 1H), 7.29 (s, 1H), 7.26-7.20 (m, 3H), 7.17-7.13 (m, 2H), 6.96 (dd, J_1 = 2.5 Hz, J_2 = 8 Hz, 1H), 5.61 (s, 1H), 4.43 (m, 1H), 4.12 (m, 1H), 3.78 (dd, J_1 = 9 Hz, J_2 = 13.5 Hz, 1H), 3.40 (dd, J_1 = 5.5 Hz, J_2 = 14 Hz, 1H), 2.93 (d, J = 11 Hz, 1H), 2.58 (dd, J_1 = 9 Hz, J_2 = 12.5 Hz, 1H), 2.45 (d, J = 9 Hz, 1H), 2.21 (dd, J_1 = 4.5 Hz, J_2 = 12.5 Hz, 1H), 2.17 (dd, J_1 = 2.5 Hz, J_2 = 11.5 Hz, 1H), 2.01 (q, J = 12 Hz, 1H), 1.82-1.17 (complex, 11H), 1.12 (s, 9H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 174.23, 169.80, 156.44, 135.04, 130.14, 129.64, 129.04, 127.92, 126.36, 120.32, 119.13,

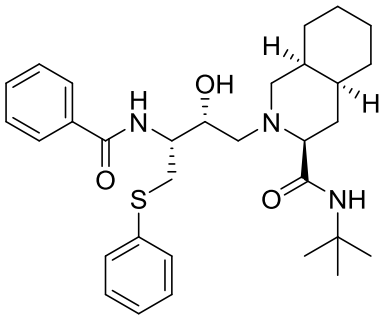
114.87, 70.77, 70.33, 59.53, 58.88, 55.48, 51.39, 36.05, 35.18, 33.73, 30.96, 30.93, 28.47, 26.33, 25.97, 20.47. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{31}H_{44}N_3O_4S$, 554.3053; observed, 554.3047.



14

(3S,4aS,8aS)-N-(tert-butyl)-2-((2R,3R)-2-hydroxy-3-(2-methylbenzamido)-4-(phenylthio)butyl)decahydroisoquinoline-3-carboxamide (14)

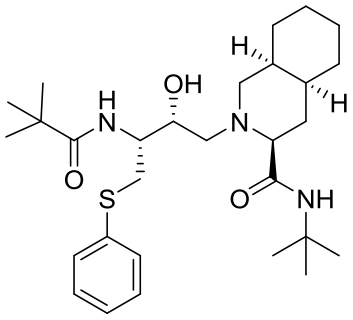
This reaction was performed according to the general EDC coupling procedure (0.048 mmol o-toluic acid for 48 h). Purification by flash column chromatography (gradient of 1-3% MeOH in DCM) yielded **14** as a white solid (22 mg, 83%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.50 (d, J = 8 Hz, 1H), 7.44 (dd, J_1 = 1 Hz, J_2 = 8 Hz, 2H), 7.30-7.23 (m, 3H), 7.21-7.15 (m, 3H), 5.47 (s, 1H), 4.49 (m, 1H), 3.83 (dd, J_1 = 9.5 Hz, J_2 = 14 Hz, 1H), 3.43 (dd, J_1 = 5 Hz, J_2 = 14 Hz, 1H), 2.92 (dd, J_1 = 2 Hz, J_2 = 12 Hz, 1H), 2.56 (dd, J_1 = 9.5 Hz, J_2 = 13 Hz, 1H), 2.48 (s, 3H), 2.42 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 2.22-2.15 (m, 2H), 2.98 (q, J = 11.5 Hz, 1H), 1.80-1.14 (complex, 11H), 1.02 (s, 9H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.96, 171.97, 136.93, 135.61, 135.36, 130.90, 130.17, 129.03, 128.01, 126.41, 125.66, 120.31, 70.74, 70.41, 59.67, 58.71, 54.89, 51.18, 36.08, 35.52, 33.78, 30.99, 30.93, 28.33, 26.30, 26.01, 20.46, 20.23. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{32}H_{46}N_3O_3S$, 552.3260; observed, 552.3254.



15

(3S,4aS,8aS)-2-((2R,3R)-3-benzamido-2-hydroxy-4-(phenylthio)butyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (15)

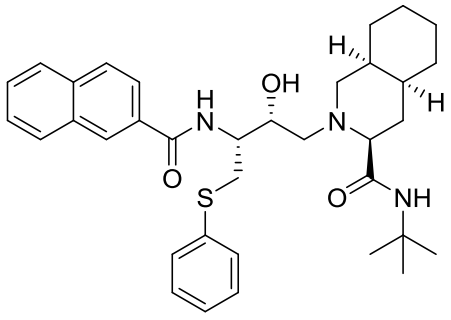
This reaction was performed according to the general EDC coupling procedure (0.042 mmol benzoic acid for 48 h). Purification by flash column chromatography (gradient of 1-3% MeOH in DCM) yielded **15** as a white solid (17 mg, 75%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.85 (d, *J* = 8 Hz, 2H), 7.66 (d, *J* = 8 Hz, 1H), 7.46 (t, *J* = 8 Hz, 1H), 7.44-7.37 (m, 4H), 7.29-7.24 (m, 2H), 7.20-7.15 (m, 2H), 5.46 (s, 1H), 4.48 (m, 1H), 4.13 (m, 1H), 3.88 (dd, J_1 = 9.5 Hz, J_2 = 14 Hz, 1H), 3.45 (dd, J_1 = 5.5 Hz, J_2 = 14 Hz, 1H), 2.95 (d, *J* = 11.5 Hz, 1H), 2.58 (m, 1H), 2.41 (d, *J* = 11.5 Hz, 1H), 2.21-2.13 (m, 2H), 2.03 (q, *J* = 12.5 Hz, 1H), 1.83-1.15 (complex, 11H), 1.05 (s, 9H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.95, 169.94, 135.45, 133.79, 131.64, 130.16, 129.04, 128.29, 128.06, 126.42, 70.91, 70.22, 59.53, 58.86, 55.66, 51.24, 36.10, 35.52, 33.88, 31.06, 30.97, 28.42, 26.35, 26.04, 20.44. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{31}H_{44}N_3O_3S$, 538.3103; observed, 538.3108.



16

(3S,4aS,8aS)-N-(tert-butyl)-2-((2R,3R)-2-hydroxy-4-(phenylthio)-3-pivalamidobutyl)decahydroisoquinoline-3-carboxamide (16)

This reaction was performed according to the general EDC coupling procedure (0.049 mmol pivalic acid for 48 h). Purification by flash column chromatography (gradient of 1-4% MeOH in DCM) yielded **16** as a white solid (13.6 mg, 56%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.38 (d, J = 7.5 Hz, 2H), 7.27 (t, J = 7.5 Hz, 2H), 7.18 (t, J = 7.5 Hz, 1H), 6.31 (d, J = 7.5 Hz, 1H), 5.73 (s, 1H), 4.16 (m, 1H), 3.97 (br s, 1H), 3.91 (m, 1H), 3.52 (dd, J_1 = 9 Hz, J_2 = 14 Hz, 1H), 3.27 (dd, J_1 = 4.5 Hz, J_2 = 14 Hz, 1H), 2.92 (dd, J_1 = 2.5 Hz, J_2 = 11.5 Hz, 1H), 2.61 (dd, J_1 = 7 Hz, J_2 = 13 Hz, 1H), 2.55 (dd, J_1 = 3 Hz, J_2 = 10.5 Hz, 1H), 2.24-2.19 (m, 2H), 1.94 (q, J = 11.5 Hz, 1H), 1.80-1.17 (complex, 20H), 1.14 (s, 9H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 180.12, 173.80, 135.77, 129.85, 129.29, 126.59, 71.09, 70.61, 59.94, 58.70, 53.56, 51.31, 39.05, 36.12, 34.71, 33.79, 31.33, 30.99, 29.11, 27.67, 26.38, 26.11, 20.88. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{29}H_{48}N_3O_3S$, 518.3416; observed, 518.3422.

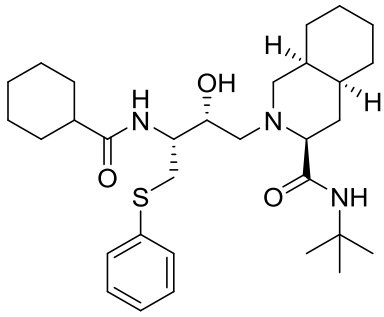


17

(3S,4aS,8aS)-2-((2R,3R)-3-(2-naphthamido)-2-hydroxy-4-(phenylthio)butyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (17)

This reaction was performed according to the general EDC coupling procedure (0.046 mmol 2-naphthoic acid for 48 h). Purification by flash column chromatography (gradient of 1-3% MeOH in DCM) yielded **17** as a white solid (19 mg, 70%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.35 (s, 1H), 7.92 (t, J = 9.5 Hz, 2H), 7.86-7.82 (m, 2H), 7.57-7.50 (m, 2H), 7.45 (d, J = 7.5 Hz, 2H), 7.29-7.7.24 (m, 2H), 7.20-7.15 (m, 2H), 5.40 (s, 1H), 4.56 (m, 1H), 4.16 (m, 1H), 3.93 (dd, J_1 = 9 Hz, J_2 = 13.5 Hz, 1H), 3.49

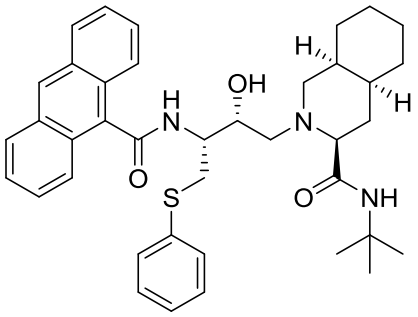
(dd, J_1 = 5.5 Hz, J_2 = 14 Hz, 1H), 3.97 (dd, J_1 = 2 Hz, J_2 = 11.5 Hz, 1H), 2.61 (dd, J_1 = 9.5 Hz, J_2 = 13 Hz, 1H), 2.41 (dd, J_1 = 3.5 Hz, J_2 = 12.5 Hz, 1H), 2.22-2.15 (m, 2H), 2.04 (q, J = 12.5 Hz, 1H), 1.85-1.17 (complex, 11H), 0.93 (s, 9H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.96, 169.96, 135.80, 134.97, 132.58, 131.02, 130.18, 129.20, 129.07, 128.64, 128.09, 127.92, 127.58, 126.45, 124.63, 121.21, 70.93, 70.34, 59.60, 58.91, 55.90, 51.21, 36.14, 35.50, 33.93, 31.16, 31.00, 28.30, 26.38, 26.07, 20.46. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{35}H_{46}N_3O_3S$, 588.3260; observed, 588.3262.



18

(3S,4aS,8aS)-N-(tert-butyl)-2-((2R,3R)-3-(cyclohexanecarboxamido)-2-hydroxy-4-(phenylthio)butyl)decahydroisoquinoline-3-carboxamide (18)

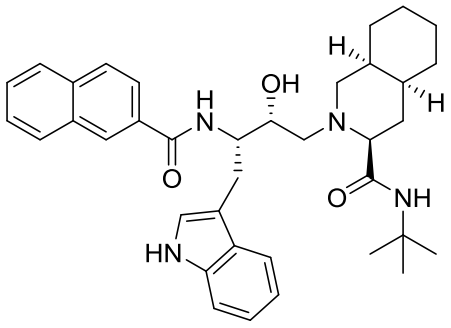
This reaction was performed according to the general EDC coupling procedure (0.056 mmol cyclohexanecarboxylic acid for 48 h). Purification by flash column chromatography (gradient of 1-3% MeOH in DCM) yielded **18** as a white solid (24 mg, 83%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.39 (d, J = 7.5 Hz, 2H), 7.27 (t, J = 7.5 Hz, 2H), 7.18 (t, J = 7.5 Hz, 1H), 6.66 (d, J = 8 Hz, 1H), 5.64 (s, 1H), 4.24 (m, 1H), 4.00 (m, 1H), 3.57 (dd, J_1 = 9 Hz, J_2 = 13.5 Hz, 1H), 3.29 (dd, J_1 = 5 Hz, J_2 = 13.5 Hz, 1H), 2.92 (dd, J_1 = 2 Hz, J_2 = 11.5 Hz, 1H), 2.57 (dd, J_1 = 8.5 Hz, J_2 = 13 Hz, 1H), 2.48 (dd, J_1 = 2.5 Hz, J_2 = 11 Hz, 1H), 2.18-2.09 (m, 2H) 1.99 (q, J = 11.5 Hz, 1H), 1.86 (d, J = 12.5 Hz, 1H), 1.80-1.14 (complex, 30H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 178.02, 173.94, 135.82, 129.75, 128.96, 126.25, 70.69, 70.11, 59.55, 58.70, 53.96, 51.26, 45.23, 36.00, 34.99, 33.76, 31.08, 30.88, 29.88, 29.67, 28.83, 26.24, 25.93, 25.79, 25.65, 25.52, 20.51. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{31}H_{50}N_3O_3S$, 544.3573; observed, 544.3562.



19

(3S,4aS,8aS)-2-((2R,3R)-3-(anthracene-9-carboxamido)-2-hydroxy-4-(phenylthio)butyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (19)

This reaction was performed according to the general EDC coupling procedure (0.048 mmol 9-anthracenecarboxylic acid for 48 h). Purification by flash column chromatography (gradient of 1-3% MeOH in DCM) yielded **19** as a light yellow solid (19.6 mg, 67%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.43 (s, 1H), 8.13 (br s, 1H), 7.96 (d, J = 7 Hz, 2H), 7.57 (d, J = 8.5 Hz, 1H), 7.52 (d, J = 7.5 Hz, 2H), 7.46 (br s, 4H), 7.31 (t, J = 7.5 Hz, 2H), 7.20 (t, J = 7.5 Hz, 1H), 5.35 (s, 1H), 4.88 (m, 1H), 4.22 (m, 1H), 3.83 (dd, J_1 = 9.5 Hz, J_2 = 13.5 Hz, 1H), 3.52 (dd, J_1 = 4 Hz, J_2 = 13.5 Hz, 1H), 2.97 (dd, J_1 = 2 Hz, J_2 = 11.5 Hz, 1H), 2.68 (dd, J_1 = 9 Hz, J_2 = 13 Hz, 1H), 2.40 (dd, J_1 = 3 Hz, J_2 = 11 Hz, 1H), 2.30 (dd, J_1 = 4.5 Hz, J_2 = 13 Hz, 1H), 2.19 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 1.88 (q, J = 11.5 Hz, 1H), 1.77-1.12 (complex, 11H), 0.67 (s, 9H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.65, 171.22, 135.91, 131.70, 131.13, 129.91, 129.05, 128.28, 128.12, 126.61, 126.35, 125.97, 125.35, 119.23, 70.81, 70.31, 59.73, 58.35, 54.63, 50.92, 36.02, 35.36, 33.76, 31.05, 30.84, 28.01, 26.23, 26.04, 20.52. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{39}H_{48}N_3O_3S$, 638.3416; observed, 638.3414.

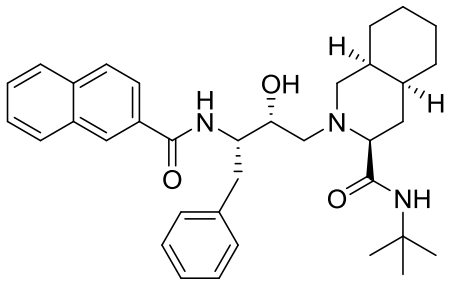


20

(3S,4aS,8aS)-2-((2R,3S)-3-(2-naphthamido)-2-hydroxy-4-(1H-indol-3-yl)butyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (20)

A solution of **59** (30 mg, 0.052 mmol) and 10% palladium on carbon (17 mg, 0.016 mmol) in methanol (6 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite and washed with excess methanol. The reaction was then dried to a solid and redissolved in THF (3 mL). 2-naphthoic acid (10 mg, 0.057 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (11 mg, 0.057 mmol), and hydroxybenzotriazole hydrate (9 mg, 0.057 mmol) were added to the reaction and the solution was stirred at room temperature for 18 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-5% MeOH in DCM) to yield **20** as a white solid (18 mg, 58%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.10 (br s, 1H), 8.03 (s, 1H), 7.80 (dd, J_1 = 4.5 Hz, J_2 = 8 Hz, 2H), 7.76 (d, J = 9 Hz, 2H), 7.72 (dd, J_1 = 2 Hz, J_2 = 9 Hz, 1H), 7.51 (td, J_1 = 1 Hz, J_2 = 7 Hz, 1H), 7.47 (td, J_1 = 1 Hz, J_2 = 7.5 Hz, 1H), 7.35 (m, 2H), 7.22 (td, J_1 = 1 Hz, J_2 = 8 Hz, 2H), 7.15 (t, J = 7 Hz, 1H), 5.92 (s, 1H), 4.71 (m, 1H), 4.17 (m, 1H), 3.81 (dd, J_1 = 10 Hz, J_2 = 15 Hz, 1H), 3.14 (m, 2H), 2.78 (dd, J_1 = 8 Hz, J_2 = 13 Hz, 1H), 2.57 (dd, J_1 = 2 Hz, J_2 = 11 Hz, 1H), 2.39 (dd, J_1 = 6 Hz, J_2 = 13 Hz, 1H), 2.30 (dd, J_1 = 3 Hz, J_2 = 12 Hz, 1H), 2.11 (q, J = 12, 1H), 1.99 (qd, J_1 = 3 Hz, J_2 = 13 Hz, 1H), 1.79-1.65 (complex, 3H), 1.55-1.20 (complex, 8H), 0.97 (s, 9H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.05, 169.23, 136.26, 134.77, 132.51, 131.16, 129.03, 128.11, 128.07, 127.52, 127.48, 126.39, 124.23, 122.39, 122.36, 122.13, 119.60,

119.02, 112.89, 111.20, 71.37, 71.08, 60.11, 59.40, 56.38, 51.02, 36.09, 33.71, 30.90, 30.83, 28.31, 26.35, 26.31, 26.10, 20.63. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{37}H_{47}N_4O_3$, 595.3648; observed, 595.3649.

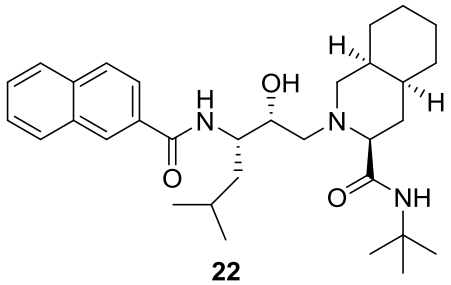


21

(3S,4aS,8aS)-2-((2R,3S)-3-(2-naphthamido)-2-hydroxy-4-phenylbutyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (21)

A solution of **54** (42 mg, 0.078 mmol) and 10% palladium on carbon (25 mg, 0.024 mmol) in methanol (15 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite and washed with excess methanol. The reaction was then dried to a solid and redissolved in THF (3 mL). 2-naphthoic acid (15 mg, 0.086 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (16 mg, 0.086 mmol), and hydroxybenzotriazole hydrate (14 mg, 0.086 mmol) were added to the reaction and the solution was stirred at room temperature for 18 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-3% MeOH in DCM) to yield **21** as a white solid (25 mg, 58%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.14 (s, 1H), 7.85 (d, J = 8 Hz, 1H), 7.80 (m, 2H), 7.72 (dd, J_1 = 1.5 Hz, J_2 = 9 Hz, 1H), 7.50 (m, 2H), 7.41 (m, 3H), 7.33 (t, J = 7.5 Hz, 2H), 7.21 (t, J = 7.5 Hz, 1H), 5.61 (s, 1H), 4.62 (m, 1H), 4.49 (br s, 1H), 4.10 (m, 1H), 3.70 (dd, J_1 = 11 Hz, J_2 = 14 Hz, 1H), 3.02 (m, 2H), 2.67 (dd, J_1 = 9 Hz, J_2 = 13 Hz, 1H), 2.47 (dd, J_1 = 3 Hz, J_2 = 11 Hz, 1H), 2.27 (dd, J_1 = 5 Hz, J_2 = 13 Hz, 1H), 2.21 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 2.06 (q, J = 12 Hz, 1H), 1.89 (qd, J_1 = 3.5 Hz, J_2 = 13.5 Hz, 1H), 1.76-0.95

(complex, 19H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.91, 169.65, 139.15, 134.81, 132.52, 131.34, 129.27, 129.01, 128.54, 128.29, 128.08, 127.55, 127.48, 126.40, 124.24, 71.63, 70.90, 59.50, 59.16, 57.41, 51.04, 36.53, 36.15, 33.76, 30.97, 28.38, 26.34, 26.03, 20.52. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{35}H_{46}N_3O_3$, 556.3539; observed, 556.3536.

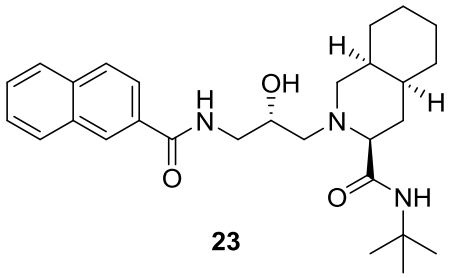


22

(3S,4aS,8aS)-2-((2R,3S)-3-(2-naphthamido)-2-hydroxy-5-methylhexyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (22)

A solution of **48** (107 mg, 0.213 mmol) and 10% palladium on carbon (68 mg, 0.064 mmol) in methanol (6 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite and washed with excess methanol. The reaction was then dried to a solid and redissolved in THF (5 mL). 2-naphthoic acid (40 mg, 0.234 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (45 mg, 0.234 mmol), and hydroxybenzotriazole hydrate (38 mg, 0.234 mmol) were added to the reaction and the solution was stirred at room temperature for 18 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-4% MeOH in DCM) to yield **22** as a white solid (75 mg, 67%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.41 (s, 1H), 7.93 (m, 2H), 7.84 (dd, J_1 = 3 Hz, J_2 = 8 Hz, 2H), 7.53 (m, 3H), 5.81 (s, 1H), 4.46 (m, 1H), 3.97 (m, 1H), 3.00 (dd, J_1 = 2 Hz, J_2 = 11.5 Hz, 1H), 2.59 (dd, J_1 = 8 Hz, J_2 = 12.5 Hz, 1H), 2.43 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 2.16 (m, 3H), 1.96 (q, J = 12 Hz, 1H), 1.82 (m, 2H), 1.69 (m, 2H), 1.59 (d, J = 12.5 Hz, 1H), 1.51-0.96 (complex, 23H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm

174.04, 169.38, 134.83, 132.58, 131.31, 129.04, 128.29, 128.06, 127.54, 127.45, 126.39, 124.40, 72.68, 70.66, 59.31, 59.18, 54.11, 50.93, 39.86, 36.00, 33.60, 30.82, 30.75, 28.36, 26.24, 26.02, 25.52, 23.48, 21.63, 20.48. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{32}H_{48}N_3O_3$, 522.3696; observed, 522.3700.

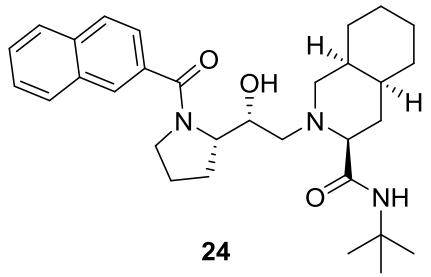


23

(3S,4aS,8aS)-2-((S)-3-(2-naphthamido)-2-hydroxypropyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (23)

A solution of **34** (50 mg, 0.112 mmol) and 10% palladium on carbon (36 mg, 0.034 mmol) in methanol (10 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite and washed with excess methanol. The reaction was then dried to a solid and redissolved in THF (3 mL). 3-hydroxy-2-methylbenzoic acid (21 mg, 0.123 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (24 mg, 0.123 mmol), and hydroxybenzotriazole hydrate (20 mg, 0.123 mmol) were added to the reaction and the solution was stirred at room temperature for 18 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-3% MeOH in DCM) to yield **23** as a white solid (36 mg, 66%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.52 (s, 1H), 8.33 (t, J = 5.5 Hz, 1H), 8.04 (dd, J_1 = 1.5 Hz, J_2 = 8.5 Hz, 1H), 7.93 (d, J = 8 Hz, 1H), 7.85 (m, 2H), 7.52 (m, 2H), 5.63 (s, 1H), 4.04 (m, 2H), 3.64 (dt, J_1 = 4 Hz, J_2 = 14 Hz, 1H), 2.78 (dd, J_1 = 2 Hz, J_2 = 11.5 Hz, 1H), 2.52 (dd, J_1 = 2.5 Hz, J_2 = 11.5 Hz, 1H), 2.44 (dd, J_1 = 10 Hz, J_2 = 12.5 Hz, 1H), 2.19 (m, 2H), 2.08 (q, J = 12 Hz, 1H), 1.85 (qd, J_1 = 3.5 Hz, J_2 = 13 Hz, 1H), 1.73 (m, 1H), 1.63 (m, 1H), 1.56 (m, 1H), 1.48 (m, 2H), 1.39-1.05 (complex, 14H). ^{13}C NMR (500 MHz,

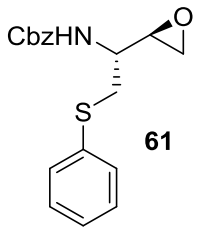
$CDCl_3$) δ ppm 174.06, 170.19, 134.83, 132.59, 131.16, 129.10, 128.43, 127.93, 127.55, 127.39, 126.30, 124.54, 70.58, 69.19, 59.24, 57.71, 51.28, 43.70, 35.93, 33.69, 31.05, 30.97, 28.41, 26.21, 25.57, 20.22. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{28}H_{40}N_3O_3$, 466.3070; observed, 466.3065.



24

(3S,4aS,8aS)-2-((R)-2-((S)-1-(2-naphthoyl)pyrrolidin-2-yl)-2-hydroxyethyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (24)

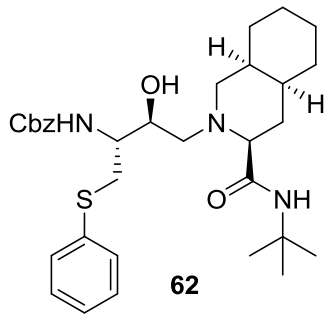
To a solution of **42** (30 mg, 0.085 mmol) in THF (3 mL) was added 2-naphthoic acid (16 mg, 0.094 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (18 mg, 0.094 mmol), and hydroxybenzotriazole hydrate (15 mg, 0.094 mmol). The solution was stirred at room temperature for 16 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-4% MeOH in DCM) to yield **24** as a white solid (33 mg, 68% over two steps). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.02 (s, 1H), 7.87 (m, 3H), 7.61 (d, J = 8.5 Hz, 1H), 7.54 (m, 1H), 6.45 (s, 1H), 4.42 (t, J = 8 Hz, 1H), 4.25 (br s, 1H), 3.58 (m, 2H), 3.47 (s, 2H), 3.33 (d, J = 12 Hz, 1H), 2.71 (dd, J_1 = 3.5 Hz, J_2 = 13 Hz, 1H), 2.65 (dd, J_1 = 2 Hz, J_2 = 10.5 Hz, 1H), 2.26 (m, 2H), 2.13 (m, 1H), 2.03-1.20 (complex, 24H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 174.09, 171.38, 133.91, 133.72, 132.40, 128.54, 128.13, 127.72, 127.32, 127.27, 126.66, 124.28, 71.56, 70.32, 62.82, 59.17, 58.87, 51.81, 50.58, 35.74, 32.95, 30.71, 30.59, 28.70, 26.61, 26.21, 25.60, 25.27, 20.52. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{31}H_{44}N_3O_3$, 506.3383; observed, 506.3389.



61

Benzyl ((R)-1-((R)-oxiran-2-yl)-2-(phenylthio)ethyl)carbamate (61)

To a solution of **60** (439 mg, 1.33 mmol) in ethanol (25 mL) at 0 °C was added solid potassium hydroxide (82 mg, 1.47 mmol). The reaction was stirred cold for 15 min then warmed to room temperature and stirred for an additional 2 h. The reaction was evaporated to dryness and the solid was purified by silica flash column chromatography (gradient of 0-4% EtOAc in DCM) to yield **61** as a white solid (285 mg, 65%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.33 (m, 9H), 7.17 (t, J = 7 Hz, 1H), 5.06 (s, 3H), 4.06 (d, J = 6.5 Hz, 1H), 3.23 (m, 2H), 3.07 (dd, J_1 = 7.5 Hz, J_2 = 13 Hz, 1H), 2.69 (t, J = 4 Hz, 1H), 2.58 (s, 1H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 155.77, 136.08, 135.12, 129.42, 128.95, 128.38, 128.04, 127.89, 126.38, 66.77, 51.81, 51.75, 49.44, 44.45, 36.63. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{20}NO_3S$, 330.1164; observed, 330.1169.

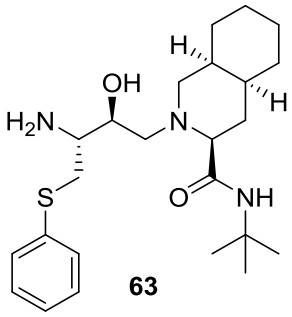


62

Benzyl ((2R,3S)-4-((3S,4aS,8aS)-3-(tert-butylcarbamoyl)octahydroisoquinolin-2(1H)-yl)-3-hydroxy-1-(phenylthio)butan-2-yl)carbamate (62)

To a solution of **61** (285 mg, 0.865 mmol) in isopropyl alcohol (7 mL) was added (3S,4aS,8aS)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (**6**, 248 mg, 1.04 mmol). The solution was heated to reflux and stirred for 5 h, then dried to a solid. The product was purified by silica flash column

chromatography (gradient of 0-4% MeOH in DCM) to yield **62** as a white solid (490 mg, 99%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.42 (d, J = 7.5 Hz, 2H), 7.32 (m, 7H), 7.16 (t, J = 7.5 Hz, 1H), 5.83 (s, 1H), 5.35 (d, J = 9 Hz, 1H), 5.28 (s, 1H), 5.12 (d, J = 12.5 Hz, 1H), 5.00 (d, J = 12.5 Hz, 1H), 4.09 (dd, J_1 = 2 Hz, J_2 = 11 Hz, 1H), 3.80 (s, 1H), 3.60 (m, 1H), 3.25 (dd, J_1 = 5.5 Hz, J_2 = 13.5 Hz, 1H), 3.10 (dd, J_1 = 9 Hz, J_2 = 13.5 Hz, 1H), 2.79 (dd, J_1 = 1.5 Hz, J_2 = 11.5 Hz, 1H), 2.62 (dd, J_1 = 3 Hz, J_2 = 11 Hz, 1H), 2.56 (dd, J_1 = 11 Hz, J_2 = 12.5 Hz, 1H), 2.19 (dd, J_1 = 3 Hz, J_2 = 12 Hz, 1H), 1.98 (dd, J_1 = 3 Hz, J_2 = 13 Hz, 1H), 1.93 (q, J = 12 Hz, 1H), 1.74 (m, 3H), 1.61-1.19 (complex, 19H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.05, 156.06, 136.19, 135.82, 128.92, 128.66, 128.34, 127.95, 127.89, 125.88, 69.55, 66.67, 65.33, 58.33, 58.01, 52.09, 50.88, 35.71, 35.56, 33.18, 30.82, 30.78, 28.51, 26.18, 25.32, 20.15. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{32}H_{46}N_3O_4S$, 568.3209; observed, 568.3196.

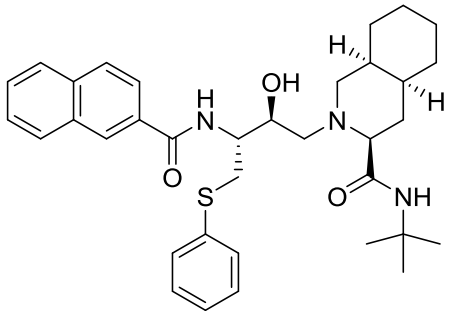


63

(3S,4aS,8aS)-2-((2S,3R)-3-amino-2-hydroxy-4-(phenylthio)butyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (63)

A solution of **62** (113 mg, 0.119 mmol) in isopropyl alcohol (10 mL) was heated to reflux with stirring and 2 M NaOH (0.50 mL, 1.0 mmol) was added. The solution was refluxed for 16 h, then dried to a solid. The product was purified by silica flash column chromatography (gradient of 0-6% MeOH in DCM) to yield **63** as a white solid (53 mg, 61%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.35 (dd, J_1 = 1 Hz, J_2 = 8 Hz, 2H), 7.28 (t, J = 8 Hz, 2H), 7.19 (t, J = 7.5 Hz, 1H), 6.26 (s, 1H), 3.71 (dt, J_1 = 3 Hz, J_2 = 10.5 Hz, 1H), 3.16 (dd, J_1 = 4.5 Hz, J_2 = 13.5 Hz, 1H), 2.89 (dd, J_1 = 8.5 Hz, J_2 = 13.5 Hz, 1H), 2.84 (dd, J_1 = 1.5 Hz, J_2 = 11.5 Hz, 1H), 2.67 (m, 3H), 2.14 (dd, J_1 = 3.5 Hz, J_2 = 11.5 Hz, 1H), 2.00 (dd, J_1 = 2.5 Hz, J_2

= 12.5 Hz, 1H), 1.88 (q, *J* = 12.5 Hz, 1H), 1.77 (m, 3H), 1.63-1.20 (complex, 19H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.38, 135.59, 129.76, 129.00, 126.37, 69.94, 68.52, 59.35, 58.21, 52.89, 50.70, 39.63, 35.79, 33.22, 30.92, 30.83, 28.59, 26.28, 25.49, 20.21. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{24}H_{40}N_3O_2S$, 434.2841; observed, 434.2837.

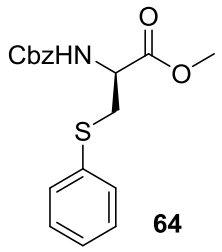


25

(3S,4aS,8aS)-2-((2S,3R)-3-(2-naphthamido)-2-hydroxy-4-(phenylthio)butyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (25)

To a solution of **63** (25 mg, 0.058 mmol) in THF (3 mL) was added 2-naphthoic acid (11 mg, 0.063 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (12 mg, 0.063 mmol), and hydroxybenzotriazole hydrate (10 mg, 0.063 mmol). The solution was stirred at room temperature for 20 h. The reaction was taken up in ethyl acetate and washed once with water, once with saturated sodium bicarbonate, and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-3% MeOH in DCM) to yield **25** as a white solid (33 mg, 97%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.27 (s, 1H), 7.91 (d, *J* = 7.5 Hz, 1H), 7.84 (m, 3H), 7.52 (m, 4H), 7.30 (t, *J* = 7.5 Hz, 2H), 7.15 (t, *J* = 7.5 Hz, 1H), 7.02 (d, *J* = 8.5 Hz, 1H), 5.45 (s, 1H), 4.38 (br s, 1H), 4.21 (dd, J_1 = 2 Hz, J_2 = 11 Hz, 1H), 4.13 (m, 1H), 3.47 (dd, J_1 = 6 Hz, J_2 = 14 Hz, 1H), 3.23 (dd, J_1 = 9 Hz, J_2 = 14 Hz, 1H), 2.84 (dd, J_1 = 1.5 Hz, J_2 = 11.5 Hz, 1H), 2.61 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 2.51 (dd, J_1 = 11 Hz, J_2 = 12.5 Hz, 1H), 2.24 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 2.09 (dd, J_1 = 3 Hz, J_2 = 13 Hz, 1H), 2.01 (q, *J* = 13 Hz, 1H), 1.74 (m, 2H), 1.62-1.15 (complex, 9H), 1.00 (s, 9H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.03, 167.05, 136.08, 134.77, 132.54,

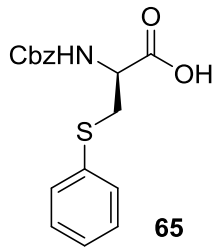
131.08, 129.01, 128.61, 128.27, 127.61, 127.58, 127.56, 126.57, 125.84, 123.69, 69.26, 65.79, 58.33, 58.05, 50.99, 50.94, 35.85, 35.34, 33.41, 31.07, 30.97, 28.37, 26.25, 25.35, 20.20. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{35}H_{46}N_3O_3S$, 588.3260; observed, 588.3254.



64

Methyl N-((benzyloxy)carbonyl)-S-phenyl-D-cysteinate (64)

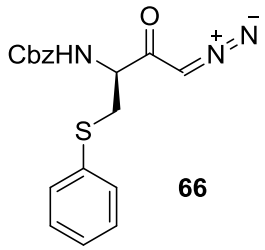
To a stirred solution of triphenylphosphine (2.43 g, 9.25 mmol) in anhydrous tetrahydrofuran (20 mL) under nitrogen at 0 °C was added diisopropyl azodicarboxylate (1.82 mL, 9.25 mmol) dropwise over 15 min. The reaction was stirred cold for an additional 40 min before Z-D-serine methyl ester (1.95 g, 7.71 mmol) in anhydrous tetrahydrofuran (20 mL) was added slowly over 15min. Immediately after addition, thiophenol (1.03 mL, 10.02 mmol) was added dropwise over 1 min. The reaction was warmed to room temperature and stirred for 18 h. Took up the reaction in ethyl acetate and washed once with water, once with sodium bicarbonate, and once with brine. The organic phase was dried over sodium sulfate and concentrated by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 0-20% EtOAc in hexanes) to yield **64** as a colorless oil (699 mg, 26%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.32 (m, 9H), 7.20 (t, J = 7.5 Hz, 1H), 5.61 (d, J = 7.5 Hz, 1H), 5.07 (dd, J_1 = 12 Hz, J_2 = 16 Hz, 2H), 4.63 (m, 1H), 3.53 (s, 3H), 3.40 (m, 2H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 170.59, 155.48, 136.08, 134.42, 131.15, 128.99, 128.50, 128.17, 128.05, 127.12, 67.03, 53.61, 52.37, 37.07. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{20}NO_4S$, 346.1113; observed, 346.1120.



65

N-((benzyloxy)carbonyl)-S-phenyl-D-cysteine (65)

To a stirred solution of **64** (521 mg, 1.51 mmol) in acetonitrile (4 mL) was added 1M aq. sodium carbonate (8 mL). The reaction was stirred vigorously for 18 h, then heated to 50 °C and stirred for an additional 6 h. The pH of the reaction was adjusted to 1 with 6M HCl and extracted three times with ethyl acetate. The combined organic fraction was washed once with brine, dried over sodium sulfate, and concentrated by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 2-10% MeOH in DCM) to yield **65** as an off-white solid (249 mg, 50%). ^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 10.18 (br s, 1H), 7.30 (m, 10H), 5.72 (s, 1H), 5.00 (m, 2H), 4.53 (m, 1H), 3.42 (m, 2H). ^{13}C NMR (400 MHz, $CDCl_3$) δ ppm 175.03, 155.88, 135.98, 134.79, 130.46, 128.94, 128.54, 128.42, 128.17, 128.02, 126.78, 67.06, 54.04, 36.47. HRMS (*m/z*): $[M + Na]^+$ calc. for $C_{17}H_{17}NO_4NaS$, 354.0776; observed, 354.0782.

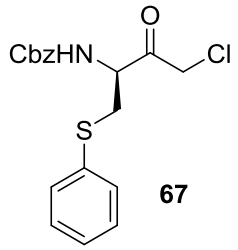


66

Benzyl (S)-(4-diazo-3-oxo-1-(phenylthio)butan-2-yl)carbamate (66)

To a stirred solution of **65** (334 mg, 1.01 mmol) in anhydrous tetrahydrofuran (4 mL) under nitrogen at -78 °C was added anhydrous triethylamine (210 μL, 1.51 mmol) followed by isobutyl chloroformate (196 μL, 1.51 mmol). The reaction was stirred cold for 1 h and then warmed to room temperature for 1.5 h. The reaction was cooled to -78 °C before an ethereal solution of diazomethane (10

mL, 3.24 mmol) was added in one portion and the reaction was then warmed to room temperature with stirring for 1 h. The reaction was evaporated to dryness, resulting in a yellow solid. The solid was taken up in ethyl acetate, washed twice with water, twice with a saturated solution of sodium bicarbonate, and once with brine. The organic phase was dried over sodium sulfate and concentrated by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 10-40% EtOAc in hexanes) to yield **66** as a yellow oil (202 mg, 56%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.32 (m, 9H), 7.19 (t, J = 7.5 Hz, 1H), 5.89 (d, J = 8 Hz, 1H), 5.47 (s, 1H), 5.08 (s, 2H), 4.39 (m, 1H), 3.30 (dd, J_1 = 6 Hz, J_2 = 14 Hz, 1H), 3.23 (dd, J_1 = 6 Hz, J_2 = 14 Hz, 1H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 191.59, 155.62, 135.89, 135.52, 130.05, 128.97, 128.37, 128.08, 127.92, 126.78, 69.74, 66.97, 56.92, 54.78, 35.93, 21.76. Ethyl acetate peaks are visible in the NMR. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{18}N_3O_3S$, 356.1069; observed, 356.1070.

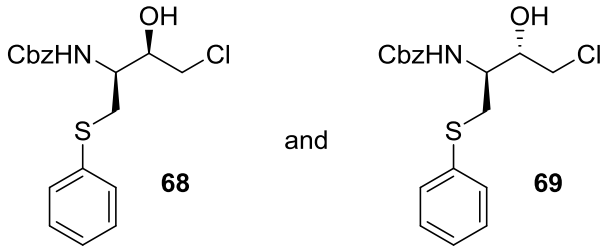


67

(S)-benzyl (4-chloro-3-oxo-1-(phenylthio)butan-2-yl)carbamate (67)

To a solution of **66** (202 mg, 0.57 mmol) in anhydrous tetrahydrofuran (8 mL) under nitrogen at 0 °C was added 1 M HCl in diethyl ether (0.85 mL, 0.85 mmol) dropwise. The reaction was stirred at 0 °C for 30 min after complete addition. The solution was dried under reduced pressure to yield **67** as an off-white solid (189 mg). No further purification was made. ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.30 (m, 10H), 5.75 (d, J = 7 Hz, 1H), 5.06 (s, 2H), 4.68 (q, J = 7 Hz, 1H), 4.18 (dd, J_1 = 16.5 Hz, J_2 = 26 Hz, 2H), 3.39 (dd, J_1 = 6 Hz, J_2 = 14.5 Hz, 1H), 3.28 (dd, J_1 = 6.5 Hz, J_2 = 14.5 Hz, 1H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 199.60, 155.67, 135.74, 133.80, 130.65, 129.23, 128.49, 128.26, 128.02, 127.32, 67.26,

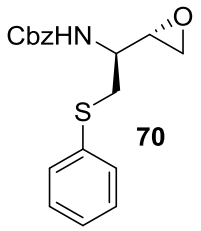
56.99, 46.98, 35.41, 21.84. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{19}NO_3SCl$, 364.0774; observed, 364.0772.



Benzyl ((2S,3R)-4-chloro-3-hydroxy-1-(phenylthio)butan-2-yl)carbamate (68) and benzyl ((2S,3S)-4-chloro-3-hydroxy-1-(phenylthio)butan-2-yl)carbamate (69)

To a solution of **67** (202 mg, 0.57 mmol) in dichloromethane (7 mL) and methanol (4 mL) under nitrogen at 0 °C was added sodium borohydride (15 mg, 0.40 mmol). The reaction was stirred at 0 °C for 1 h, and then quenched with slow addition of 2 M HCl (2 mL). The reaction was evaporated to dryness and the solid was taken up in ethyl acetate. The solution was washed twice with water followed by brine, dried over sodium sulfate, and concentrated by rotary evaporation. TLC analysis showed two diastereomers with the lower Rf compound being the (S,R)-isomer. Both diastereomers were purified by silica flash column chromatography (gradient of 0-4% EtOAc in DCM) to yield (S,R)-isomer (**68**) as a white solid (98 mg, 47%) and (S,S)-isomer (**69**) as a white solid (25 mg, 12%) over two steps. Characterization of the (S,R)-isomer (**68**): ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.30 (m, 10H), 5.19 (m, 1H), 5.06 (m, 2H), 3.91 (m, 2H), 3.66 (dd, J_1 = 3.5 Hz, J_2 = 11.5 Hz, 1H), 3.59 (dd, J_1 = 7 Hz, J_2 = 11.5 Hz, 1H), 3.28 (d, J = 4 Hz, 2H), 2.89 (s, 1H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 156.15, 136.03, 135.24, 130.06, 129.13, 128.53, 128.25, 128.07, 126.78, 72.86, 67.13, 53.33, 47.39, 35.21. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{21}NO_3SCl$, 366.0931; observed, 366.0927. Characterization of the (S,S)-isomer (**69**): ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.41 (d, J = 7.5 Hz, 2H), 7.33 (m, 7H), 7.20 (t, J = 7.5 Hz, 1H), 5.28 (d, J = 9 Hz, 1H), 5.10 (m, 2H), 4.18 (m, 1H), 3.87 (q, J = 7.5 Hz, 1H), 3.53 (m, 2H), 3.26 (dd, J_1 = 6 Hz, J_2 = 14 Hz, 1H), 3.11 (dd, J_1 = 7.5 Hz, J_2 = 14 Hz, 1H), 2.91 (d, J = 3.5 Hz, 1H). ^{13}C NMR (500 MHz, $CDCl_3$)

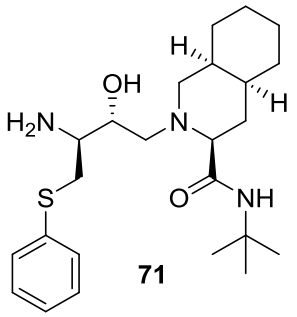
δ ppm 156.33, 136.07, 135.02, 129.51, 129.14, 128.57, 128.54, 128.26, 128.22, 128.12, 128.07, 126.56, 70.84, 67.15, 66.98, 52.27, 47.40, 35.58. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{21}NO_3SCl$, 366.0931; observed, 366.0930.



70

Benzyl ((S)-1-((S)-oxiran-2-yl)-2-(phenylthio)ethyl)carbamate (70)

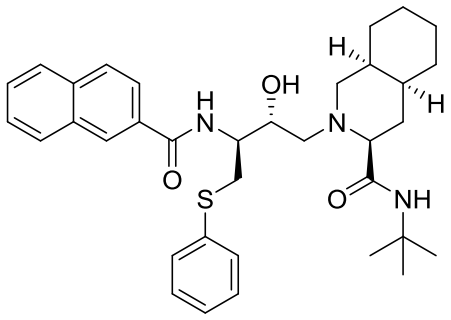
To a solution of **69** (25 mg, 0.068 mmol) in ethanol (5 mL) at 0 °C was added solid potassium hydroxide (4 mg, 0.075 mmol). The reaction was stirred cold for 15 min then warmed to room temperature and stirred for an additional 2 h. The reaction was evaporated to dryness and the solid was purified by silica flash column chromatography (gradient of 1-4% EtOAc in DCM) to yield **70** as a white solid (10 mg, 45%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.42 (d, J = 7 Hz, 2H), 7.34 (m, 7H), 7.20 (t, J = 7 Hz, 1H), 5.09 (s, 2H), 4.88 (d, J = 7 Hz, 1H), 4.11 (m, 1H), 3.29 (m, 2H), 3.11 (dd, J_1 = 8 Hz, J_2 = 13.5 Hz, 1H), 2.75 (t, J = 4.5 Hz, 1H), 2.62 (s, 1H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 155.89, 136.16, 135.20, 129.65, 129.13, 128.55, 128.22, 128.06, 126.60, 67.00, 51.77, 49.43, 44.52, 36.99. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{20}NO_3S$, 330.1164; observed, 330.1160.



71

(3S,4aS,8aS)-2-((2R,3S)-3-amino-2-hydroxy-4-(phenylthio)butyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (71)

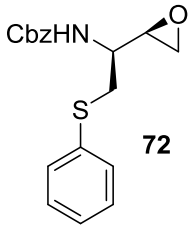
To a solution of **70** (10 mg, 0.030 mmol) in isopropyl alcohol (5 mL) was added (3S,4aS,8aS)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (**6**, 8 mg, 0.033 mmol). The solution was heated to reflux and stirred. After 5 h, 2 M NaOH (75 μL, 0.15 mmol) was added and the reaction was stirred at reflux for an additional 15 h. The reaction was diluted with toluene and extracted with 1 M HCl three times. The aqueous fractions were combined and washed with toluene twice. The aqueous phase was adjusted to pH 12 with 10 M NaOH and was extracted with toluene five times. The combined toluene fractions were washed with brine, dried over sodium sulfate, and concentrated to a solid by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 1-9% MeOH in DCM) to yield **71** as a white solid (9 mg, 69%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.42 (d, J = 7.5 Hz, 2H), 7.31 (t, J = 7.5 Hz, 2H), 7.21 (t, J = 7.5 Hz, 1H), 5.88 (s, 1H), 5.30 (s, 1H), 4.08 (m, 1H), 3.50 (dd, J_1 = 9 Hz, J_2 = 14 Hz, 1H), 3.40 (m, 1H), 3.34 (dd, J_1 = 6 Hz, J_2 = 14 Hz, 1H), 2.60 (dd, J_1 = 2.5 Hz, J_2 = 11.5 Hz, 2H), 2.48 (m, 2H), 2.16 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 1.90 (q, J = 12 Hz, 1H), 1.76-0.62 (complex, 22H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 174.82, 133.18, 130.09, 129.40, 127.04, 69.12, 66.84, 58.20, 56.72, 52.43, 52.16, 35.44, 33.31, 30.96, 30.80, 29.70, 28.72, 26.09, 25.05, 20.08. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{24}H_{40}N_3O_2S$, 434.2841; observed, 434.2835.



26

(3S,4aS,8aS)-2-((2R,3S)-3-(2-naphthamido)-2-hydroxy-4-(phenylthio)butyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (26)

To a solution of **71** (9 mg, 0.021 mmol) in THF (3 mL) was added 2-naphthoic acid (4 mg, 0.023 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (4 mg, 0.023 mmol), and hydroxybenzotriazole hydrate (4 mg, 0.023 mmol). The solution was stirred at room temperature for 20 h. The reaction was taken up in ethyl acetate and washed once with water, once with saturated sodium bicarbonate, and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-3% MeOH in DCM) to yield **26** as a white solid (5 mg, 41%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.39 (s, 1H), 7.92 (m, 2H), 7.86 (m, 2H), 7.79 (m, 1H), 7.54 (m, 3H), 7.45 (d, J = 8 Hz, 2H), 7.27 (t, J = 7.5 Hz, 2H), 7.13 (t, J = 7.5 Hz, 1H), 5.77 (br s, 1H), 4.25 (m, 2H), 3.61 (dd, J_1 = 6.5 Hz, J_2 = 14 Hz, 1H), 3.46 (dd, J_1 = 8 Hz, J_2 = 14 Hz, 1H), 2.83 (m, 1H), 2.60 (m, 2H), 2.29 (m, 3H), 2.09-0.81 (complex, 20H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 168.83, 135.96, 134.89, 132.58, 131.24, 129.09, 129.01, 128.73, 128.30, 128.16, 127.74, 127.63, 126.54, 125.86, 125.86, 124.17, 70.56, 68.82, 59.81, 59.00, 55.97, 53.97, 53.09, 51.25, 35.74, 33.68, 33.41, 30.70, 29.70, 28.42, 26.00, 25.44, 20.33. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{35}H_{46}N_3O_3S$, 588.3260; observed, 588.3256.

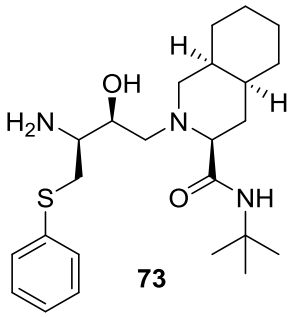


72

Benzyl ((S)-1-((R)-oxiran-2-yl)-2-(phenylthio)ethyl)carbamate (72)

To a solution of **68** (98 mg, 0.268 mmol) in ethanol (7 mL) at 0 °C was added solid potassium hydroxide (17 mg, 0.295 mmol). The reaction was stirred cold for 15 min then warmed to room temperature and stirred for an additional 2 h. The reaction was evaporated to dryness and the solid was purified by silica flash column chromatography (gradient of 1-4% EtOAc in DCM) to yield **72** as a white solid (71 mg, 80%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.34 (m, 7H), 7.26 (t, J = 8 Hz, 2H), 7.19 (t, J = 7.5 Hz, 1H), 5.18 (br s, 1H), 5.07 (dd, J_1 = 12 Hz, J_2 = 16 Hz, 2H), 3.69 (m, 1H), 3.20 (d, J = 5.5 Hz, 2H),

2.98 (s, 1H), 2.75 (s, 2H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 155.74, 136.11, 135.07, 129.92, 129.09, 128.49, 128.17, 128.05, 126.72, 66.95, 52.59, 52.20, 46.74, 35.99. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{20}NO_3S$, 330.1164; observed, 330.1166.

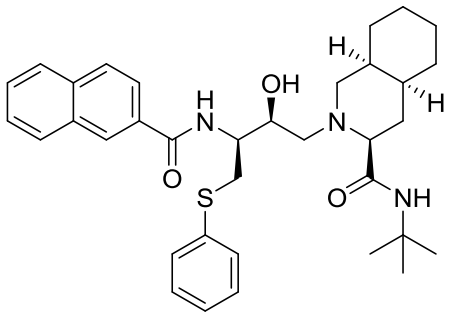


73

(3S,4aS,8aS)-2-((2S,3S)-3-amino-2-hydroxy-4-(phenylthio)butyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (73)

To a solution of **72** (71 mg, 0.216 mmol) in isopropyl alcohol (15 mL) was added (3S,4aS,8aS)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (**6**, 57 mg, 0.237 mmol). The solution was heated to reflux and stirred. After 5 h, 2 M NaOH (0.54 mL, 1.08 mmol) was added and the reaction was stirred at reflux for an additional 15 h. The reaction was diluted with toluene and extracted with 1 M HCl three times. The aqueous fractions were combined and washed with toluene twice. The aqueous phase was adjusted to pH 12 with 10 M NaOH and was extracted with toluene five times. The combined toluene fractions were washed with brine, dried over sodium sulfate, and concentrated to a solid by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 0-8% MeOH in DCM) to yield **73** as a white solid (47 mg, 50%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.36 (dd, J_1 = 1 Hz, J_2 = 8 Hz, 2H), 7.27 (t, J = 7.5 Hz, 2H), 7.19 (t, J = 7.5 Hz, 1H), 6.27 (s, 1H), 3.76 (m, 1H), 3.29 (dd, J_1 = 3 Hz, J_2 = 13.5 Hz, 1H), 2.91 (d, J = 12.5 Hz, 2H), 2.82 (br s, 2H), 2.81 (dd, J_1 = 9.5 Hz, J_2 = 13.5 Hz, 1H), 2.70 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 2.59 (dd, J_1 = 10 Hz, J_2 = 12.5 Hz, 1H), 2.18 (dd, J_1 = 1.5 Hz, J_2 = 12.5 Hz, 1H), 1.88 (q, J = 11.5 Hz, 1H), 1.75 (m, 3H), 1.63-1.15 (complex, 19H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.34, 135.18, 129.71, 129.02, 126.40, 69.95, 69.18, 58.31, 58.05, 53.47, 50.90,

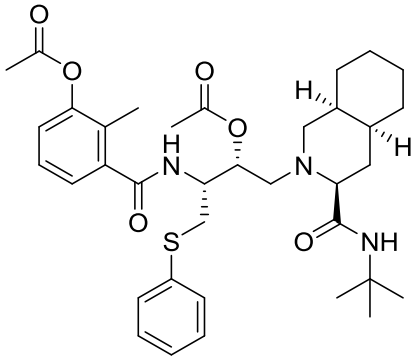
37.29, 35.75, 33.19, 30.89, 30.80, 28.59, 26.27, 25.45, 20.21. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{24}H_{40}N_3O_2S$, 434.2841; observed, 434.2841.



27

(3S,4aS,8aS)-2-((2S,3S)-3-(2-naphthamido)-2-hydroxy-4-(phenylthio)butyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (27)

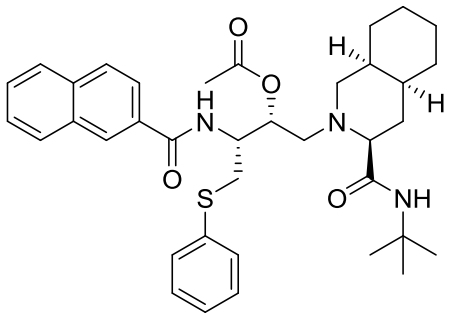
To a solution of **73** (47 mg, 0.108 mmol) in THF (5 mL) was added 2-naphthoic acid (21 mg, 0.119 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (23 mg, 0.119 mmol), and hydroxybenzotriazole hydrate (19 mg, 0.119 mmol). The solution was stirred at room temperature for 20 h. The reaction was taken up in ethyl acetate and washed once with water, once with saturated sodium bicarbonate, and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-3% MeOH in DCM) to yield **27** as a white solid (47 mg, 74%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.15 (s, 1H), 7.84 (m, 3H), 7.74 (dd, J_1 = 1.5 Hz, J_2 = 9 Hz, 1H), 7.54 (m, 2H), 7.40 (d, J = 7.5 Hz, 2H), 7.25 (t, J = 8 Hz, 2H), 7.16 (t, J = 7.5 Hz, 1H), 6.89 (d, J = 8 Hz, 1H), 6.18 (s, 1H), 4.30 (m, 1H), 3.99 (m, 1H), 3.39 (dd, J_1 = 8 Hz, J_2 = 14 Hz, 1H), 3.25 (dd, J_1 = 4.5 Hz, J_2 = 14 Hz, 1H), 2.80 (d, J = 10 Hz, 1H), 2.61 (m, 2H), 2.14 (m, 2H), 1.91 (q, J = 12 Hz, 1H), 1.79-1.66 (complex, 3H), 1.57-1.13 (complex, 18H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.51, 167.67, 135.68, 134.76, 132.49, 131.17, 129.78, 129.13, 128.95, 128.35, 127.68, 127.57, 126.68, 126.53, 123.63, 69.77, 69.17, 58.52, 58.34, 53.09, 50.98, 35.81, 34.37, 33.27, 31.02, 30.86, 28.65, 26.23, 25.40, 20.17. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{35}H_{46}N_3O_3S$, 588.3260; observed, 588.3266.



28

(2R,3R)-3-(3-acetoxy-2-methylbenzamido)-1-((3S,4aS,8aS)-3-(tert-butylcarbamoyl)octahydroisoquinolin-2(1H)-yl)-4-(phenylthio)butan-2-yl acetate (28)

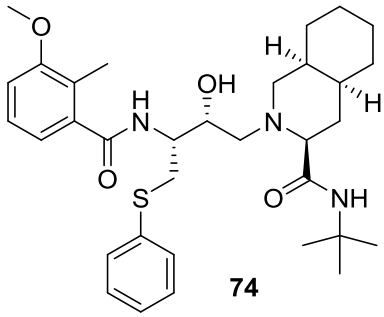
To a stirred solution of **1** (75 mg, 0.132 mmol) in anhydrous tetrahydrofuran (3 mL) under nitrogen at room temperature was added anhydrous triethylamine (37 μL, 0.264 mmol) followed by acetic anhydride (125 μL, 1.32 mmol). The reaction was stirred at room temperature for 16 h. Ethyl acetate was added to the reaction. The crude material was washed once with saturated sodium bicarbonate and once with brine. The organic phase was dried over sodium sulfate and concentrated by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 0-50% EtOAc in hexanes) to yield **28** as a white solid (41 mg, 48%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.44 (dd, J_1 = 1 Hz, J_2 = 7.5 Hz, 2H), 7.39 (d, J = 9.5 Hz, 1H), 7.34 (dd, J_1 = 1 Hz, J_2 = 7.5 Hz, 1H), 7.28 (t, J = 7.5 Hz, 2H), 7.19 (m, 2H), 7.04 (dd, J_1 = 0.5 Hz, J_2 = 8 Hz, 1H), 5.70 (s, 1H), 5.36 (m, 1H), 4.66 (m, 1H), 3.59 (dd, J_1 = 9 Hz, J_2 = 13.5 Hz, 1H), 3.44 (dd, J_1 = 5.5 Hz, J_2 = 13.5 Hz, 1H), 3.00 (dd, J_1 = 2 Hz, J_2 = 12 Hz, 1H), 2.64 (dd, J_1 = 10 Hz, J_2 = 12.5 Hz, 1H), 2.43 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 2.30 (s, 3H), 2.23 (s, 3H), 2.14 (m, 2H), 2.03 (s, 3H), 1.97 (q, J = 12 Hz, 1H), 1.75-1.10 (complex, 20H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.79, 170.22, 169.18, 168.95, 149.62, 138.74, 135.61, 129.87, 128.95, 128.50, 126.32, 124.85, 123.23, 70.72, 70.69, 59.58, 55.92, 51.28, 50.62, 35.95, 35.61, 33.72, 30.97, 30.91, 28.36, 26.29, 25.82, 21.18, 20.78, 20.35, 12.84. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{36}H_{50}N_3O_6S$, 652.3420; observed, 652.3414.



29

(2R,3R)-3-(2-naphthamido)-1-((3S,4aS,8aS)-3-(tert-butylcarbamoyl)octahydroisoquinolin-2(1H)-yl)-4-(phenylthio)butan-2-yl acetate (29)

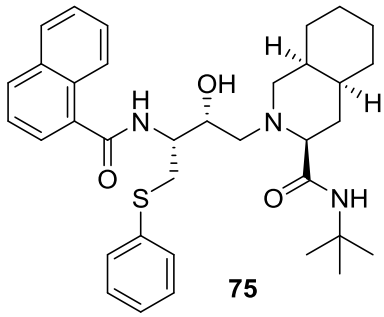
To a stirred solution of **17** (23.2 mg, 0.040 mmol) in anhydrous tetrahydrofuran (2 mL) under nitrogen at room temperature was added anhydrous triethylamine (10.9 μL, 0.079 mmol) followed by acetic anhydride (7.5 μL, 0.079 mmol). The reaction was stirred for 8 h, then additional acetic anhydride was added (15 μL, 0.16 mmol) and the reaction was stirred at room temperature for 12 h. Ethyl acetate was added to the reaction. The crude material was washed once with saturated sodium bicarbonate and once with brine. The organic phase was dried over sodium sulfate and concentrated by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 1-3% MeOH in DCM) to yield **29** as a white solid (9.8 mg, 39%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.33 (s, 1H), 7.92 (m, 2H), 7.84 (m, 2H), 7.64 (d, J = 9.5 Hz, 1H), 7.52 (m, 2H), 7.45 (dd, J_1 = 1 Hz, J_2 = 7.5 Hz, 2H), 7.26 (t, J = 7.5 Hz, 2H), 7.15 (*t*, J = 7.5 Hz, 1H), 5.73 (s, 1H), 5.45 (m, 1H), 4.77 (m, 1H), 3.62 (dd, J_1 = 7.5 Hz, J_2 = 13.5 Hz, 1H), 3.51 (dd, J_1 = 6 Hz, J_2 = 14 Hz, 1H), 3.06 (dd, J_1 = 2 Hz, J_2 = 12 Hz, 1H), 2.78 (dd, J_1 = 9.5 Hz, J_2 = 12.5 Hz, 1H), 2.47 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 2.21-2.16 (m, 2H), 2.07 (s, 3H), 2.02 (q, J = 12 Hz, 1H), 1.83-1.13 (complex, 20H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.74, 170.49, 167.77, 135.85, 132.60, 131.96, 129.80, 129.17, 128.99, 127.96, 127.56, 127.27, 126.29, 126.24, 124.45, 71.01, 70.88, 59.69, 56.27, 51.39, 51.25, 36.00, 35.74, 33.85, 31.24, 30.95, 28.56, 26.33, 25.87, 21.27, 20.44. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{37}H_{48}N_3O_4S$, 630.3366; observed, 630.3361.



74

(3S,4aS,8aS)-N-(tert-butyl)-2-((2R,3R)-2-hydroxy-3-(3-methoxy-2-methylbenzamido)-4-(phenylthio)butyl)decahydroisoquinoline-3-carboxamide (74)

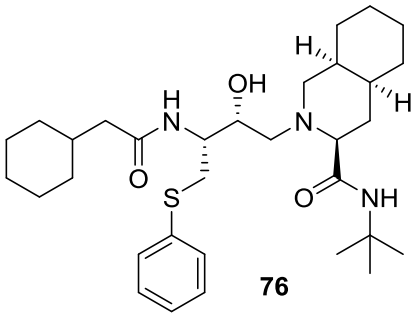
To a solution of **1** (41 mg, 0.074 mmol) in dry DMF (1 mL) under nitrogen was added potassium carbonate (20 mg, 0.15 mmol) and methyl iodide (9.1 μL). The solution was heated to 60 °C and stirred for 3 h. The reaction was cooled to room temperature and taken up in water. The aqueous phase was extracted three times with ethyl acetate. The combined organic phase was washed with water then brine, dried over sodium sulfate, and concentrated by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 2-3% MeOH in DCM) to yield **74** as a white solid (22.3 mg, 52%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.46 (d, J = 8 Hz, 2H), 7.30 (t, J = 8 Hz, 1H), 7.22 (t, J = 7.5 Hz, 1H), 7.18 (t, J = 8 Hz, 2H), 7.10 (d, J = 7.5 Hz, 1H), 6.89 (d, J = 8 Hz, 1H), 5.51 (s, 1H), 4.44 (m, 1H), 4.04 (m, 1H), 3.92 (br s, 1H), 3.84 (s, 3H), 3.78 (dd, J_1 = 9 Hz, J_2 = 13 Hz, 1H), 3.44 (dd, J_1 = 5 Hz, J_2 = 14 Hz, 1H), 2.94 (dd, J_1 = 2.5 Hz, J_2 = 12 Hz, 1H), 2.59 (dd, J_1 = 9 Hz, J_2 = 13 Hz, 1H), 2.46 (dd, J_1 = 2.5 Hz, J_2 = 11 Hz, 1H), 2.32 (s, 3H), 2.26-2.19 (m, 2H), 2.00 (q, J = 12 Hz, 1H), 1.80-1.18 (complex, 11H), 1.08 (s, 9H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.88, 171.79, 157.96, 137.24, 135.63, 130.10, 129.01, 127.91, 126.39, 125.45, 119.77, 111.74, 70.62, 70.38, 59.68, 58.69, 55.68, 54.57, 51.15, 36.03, 35.38, 33.71, 30.96, 30.89, 28.37, 26.26, 25.95, 20.48, 12.74. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{33}H_{48}N_3O_4S$, 582.3366; observed, 582.3361.



75

(3S,4aS,8aS)-2-((2R,3R)-3-(1-naphthamido)-2-hydroxy-4-(phenylthio)butyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (75)

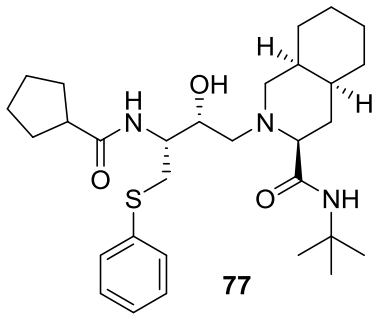
This reaction was performed according to the general EDC coupling procedure (0.046 mmol 1-naphthoic acid for 16 h). Purification by flash column chromatography (gradient of 1-3% MeOH in DCM) yielded **75** as a white solid (11.4 mg, 45%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.45 (d, J = 8 Hz, 1H), 7.90 (d, J = 8.5 Hz, 1H), 7.82 (d, J = 7.5 Hz, 1H), 7.77 (dd, J_1 = 1 Hz, J_2 = 7 Hz, 1H), 7.60 (br d, J = 8.5 Hz), 7.54-7.43 (complex, 6H), 7.29 (t, J = 7.5 Hz, 2H), 7.20 (t, J = 7.5 Hz, 1H), 5.36 (s, 1H), 4.61 (m, 1H), 4.14 (m, 1H), 3.93 (dd, J_1 = 9.5 Hz, J_2 = 13.5 Hz, 1H), 3.46 (dd, J_1 = 5 Hz, J_2 = 14 Hz, 1H), 2.94 (dd, J_1 = 2 Hz, J_2 = 12 Hz, 1H), 2.59 (dd, J_1 = 9.5 Hz, J_2 = 13.5 Hz, 1H), 2.38 (dd, J_1 = 3 Hz, J_2 = 11 Hz, 1H), 2.23 (dd, J_1 = 4.5 Hz, J_2 = 13 Hz, 1H), 2.17 (dd, J_1 = 3 Hz, J_2 = 12 Hz, 1H), 1.98 (q, J = 11.5 Hz, 1H), 1.81-1.16 (complex, 11H), 0.78 (s, 9H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 174.12, 171.78, 135.78, 133.95, 131.22, 130.92, 130.53, 129.30, 128.26, 127.29, 126.72, 126.63, 126.40, 126.26, 125.96, 124.91, 71.01, 70.76, 59.93, 59.02, 55.44, 51.31, 36.35, 35.89, 34.03, 31.22, 31.19, 28.32, 26.56, 26.28, 20.69. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{35}H_{46}N_3O_3S$, 588.3260; observed, 588.3259.



76

(3S,4aS,8aS)-N-(tert-butyl)-2-((2R,3R)-3-(2-cyclohexylacetamido)-2-hydroxy-4-(phenylthio)butyl)decahydroisoquinoline-3-carboxamide (76)

This reaction was performed according to the general EDC coupling procedure (0.039 mmol cyclohexaneacetic acid for 16 h). Purification by flash column chromatography (gradient of 1-3% MeOH in DCM) yielded **76** as a white solid (13.5 mg, 65%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.38 (dd, J_1 = 2 Hz, J_2 = 8 Hz, 2H), 7.27 (t, J = 8 Hz, 2H), 7.17 (t, J = 7.5 Hz, 1H), 6.80 (d, J = 8 Hz, 1H), 5.61 (s, 1H), 4.27 (m, 1H), 4.03 (m, 1H), 3.54 (dd, J_1 = 8.5 Hz, J_2 = 13.5 Hz, 1H), 3.33 (dd, J_1 = 6 Hz, J_2 = 14 Hz, 1H), 2.93 (dd, J_1 = 2.5 Hz, J_2 = 12 Hz, 1H), 2.57 (dd, J_1 = 9 Hz, J_2 = 12.5 Hz, 1H), 2.47 (dd, J_1 = 2.5 Hz, J_2 = 11 Hz, 1H), 2.19-2.10 (m, 3H), 2.05-1.96 (m, 2H), 1.82-1.07 (complex, 29H), 0.98-0.88 (m, 2H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 174.38, 173.99, 135.82, 129.55, 128.97, 126.19, 70.69, 70.01, 59.57, 58.79, 54.15, 51.32, 44.47, 36.00, 35.16, 34.88, 33.80, 33.04, 33.00, 31.10, 30.89, 28.91, 26.24, 26.04, 25.93, 20.51. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{32}H_{52}N_3O_3S$, 558.3729; observed, 558.3740.

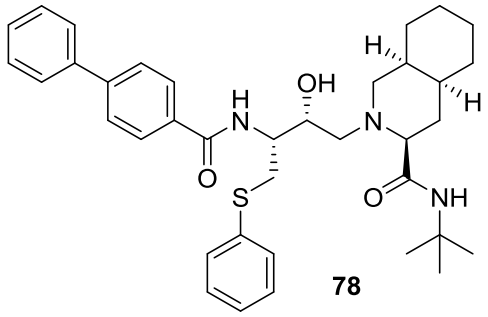


77

(3S,4aS,8aS)-N-(tert-butyl)-2-((2R,3R)-3-(cyclopentanecarboxamido)-2-hydroxy-4-(phenylthio)butyl)decahydroisoquinoline-3-carboxamide (77)

This reaction was performed according to the general EDC coupling procedure (0.063 mmol cyclopentanecarboxylic acid for 48 h). Purification by flash column chromatography (gradient of 1-4% MeOH in DCM) yielded **77** as a white solid (17.3 mg, 54%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.39 (d, J = 7 Hz, 2H), 7.28 (t, J = 7.5 Hz, 2H), 7.18 (t, J = 7.5 Hz, 1H), 6.66 (d, J = 8.5 Hz, 1H), 5.63 (s, 1H), 4.24 (m, 1H), 4.01 (m, 1H), 3.75 (br s, 1H), 3.55 (dd, J_1 = 9 Hz, J_2 = 13.5 Hz, 1H), 3.31 (dd, J_1 = 5.5 Hz, J_2 = 13.5 Hz, 1H), 2.92 (dd, J_1 = 2 Hz, J_2 = 11.5 Hz, 1H), 2.60-2.54 (m, 2H), 2.48 (dd, J_1 = 3 Hz, J_2 = 11

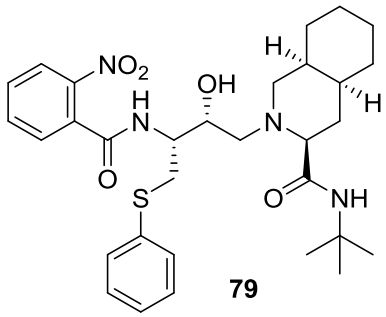
Hz, 1H), 2.20-2.15 (m, 2H), 1.99 (q, J = 11.5 Hz, 1H), 1.89 (m, 1H), 1.81-1.14 (complex, 27H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 178.44, 174.19, 136.02, 130.00, 129.23, 126.53, 70.97, 70.41, 59.83, 59.00, 54.40, 51.50, 45.87, 36.25, 35.18, 33.99, 31.31, 31.13, 30.38, 29.04, 26.49, 26.17, 26.13, 20.76. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{30}H_{48}N_3O_3S$, 530.3416; observed, 530.3422.



78

(3S,4aS,8aS)-2-((2R,3R)-3-([1,1'-biphenyl]-4-ylcarboxamido)-2-hydroxy-4-(phenylthio)butyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (78)

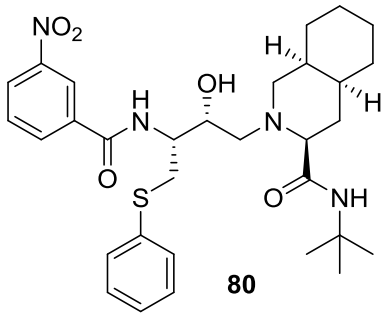
This reaction was performed according to the general EDC coupling procedure (0.058 mmol biphenyl-4-carboxylic acid for 16 h). Purification by flash column chromatography (gradient of 0.5-3% MeOH in DCM) yielded **78** as a white solid (25.3 mg, 75%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.92 (dd, J_1 = 2 Hz, J_2 = 7 Hz, 2H), 7.76 (d, J = 8.5 Hz, 1H), 7.61-7.58 (m, 4H), 7.47 (t, J = 8 Hz, 2H), 7.44 (dd, J_1 = 1 Hz, J_2 = 7 Hz, 2H), 7.39 (t, J = 7.5 Hz, 1H), 7.28 (t, J = 7.5 Hz, 2H), 7.19 (t, J = 7.5 Hz, 1H), 5.47 (s, 1H), 4.51 (m, 1H), 4.15 (m, 1H), 3.89 (dd, J_1 = 4 Hz, J_2 = 13.5 Hz, 1H), 3.46 (dd, J_1 = 5 Hz, J_2 = 13.5 Hz, 1H), 2.95 (dd, J_1 = 2 Hz, J_2 = 12 Hz, 1H), 2.59 (dd, J_1 = 10 Hz, J_2 = 12.5 Hz, 1H), 2.41 (dd, J_1 = 3 Hz, J_2 = 12 Hz, 1H), 2.19 (dd, J_1 = 4 Hz, J_2 = 12.5 Hz, 1H), 2.15 (dd, J_1 = 3.5 Hz, J_2 = 11.5 Hz, 1H), 2.04 (q, J = 11.5 Hz, 1H), 1.84-1.15 (complex, 11H), 1.05 (s, 9H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.99, 169.70, 144.43, 140.28, 135.50, 132.55, 130.16, 129.04, 128.88, 128.62, 127.88, 127.17, 126.96, 126.42, 70.92, 70.23, 59.53, 58.93, 55.77, 51.26, 36.12, 35.55, 33.90, 31.11, 30.99, 28.44, 26.37, 26.06, 20.45. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{37}H_{48}N_3O_3S$, 614.3416; observed, 614.3414.



79

(3S,4aS,8aS)-N-(tert-butyl)-2-((2R,3R)-2-hydroxy-3-(2-nitrobenzamido)-4-(phenylthio)butyl)decahydroisoquinoline-3-carboxamide (79)

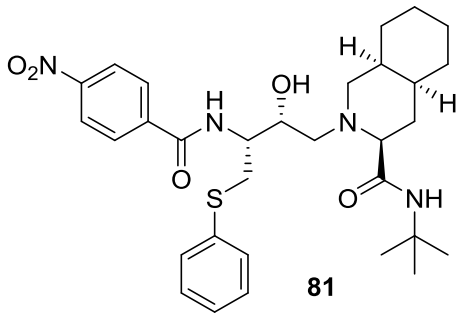
This reaction was performed according to the general EDC coupling procedure (0.036 mmol 2-nitrobenzoic acid for 16 h). Purification by flash column chromatography (gradient of 1-3% MeOH in DCM) yielded **79** as a white solid (15.8 mg, 80%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.96 (dd, J_1 = 1 Hz, J_2 = 8 Hz, 1H), 7.92 (d, J = 9 Hz, 1H), 7.64 (td, J_1 = 1 Hz, J_2 = 7.5 Hz, 1H), 7.59 (dd, J_1 = 1.5 Hz, J_2 = 7.5 Hz, 1H), 7.53 (m, 1H), 7.48 (dd, J_1 = 1.5 Hz, J_2 = 8.5 Hz, 2H), 7.30 (t, J = 7.5 Hz, 2H), 7.20 (tt, J_1 = 1.5 Hz, J_2 = 7.5 Hz, 1H), 5.45 (s, 1H), 4.59 (m, 1H), 4.16 (m, 1H), 3.81 (dd, J_1 = 9 Hz, J_2 = 13.5 Hz, 1H), 3.41 (dd, J_1 = 10.5 Hz, J_2 = 14 Hz, 1H), 2.95 (dd, J_1 = 2 Hz, J_2 = 12 Hz, 1H), 2.66 (dd, J_1 = 9.5 Hz, J_2 = 13 Hz, 1H), 2.43 (dd, J_1 = 2.5 Hz, J_2 = 11.5 Hz, 1H), 2.29 (dd, J_1 = 4.5 Hz, J_2 = 13 Hz, 1H), 2.18 (dd, J_1 = 3.5 Hz, J_2 = 11.5 Hz, 1H), 1.99 (q, J = 11.5 Hz, 1H), 1.79-1.12 (complex, 11H), 1.00 (s, 9H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 174.30, 167.86, 147.23, 135.64, 133.43, 133.23, 130.20, 129.97, 129.15, 129.06, 126.36, 124.29, 70.62, 69.65, 59.64, 59.03, 55.13, 51.21, 36.10, 35.39, 33.90, 31.21, 30.97, 28.28, 26.35, 26.00, 20.44. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{31}H_{43}N_4O_5S$, 583.2954; observed, 583.2951.



80

(3S,4aS,8aS)-N-(tert-butyl)-2-((2R,3R)-2-hydroxy-3-(3-nitrobenzamido)-4-(phenylthio)butyl)decahydroisoquinoline-3-carboxamide (80)

This reaction was performed according to the general EDC coupling procedure (0.065 mmol 3-nitrobenzoic acid for 40 h). Purification by flash column chromatography (gradient of 1-3% MeOH in DCM) yielded **80** as a white solid (18.8 mg, 61%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.72 (t, J = 2 Hz, 1H), 8.32 (m, 1H), 8.21 (d, J = 8 Hz, 1H), 8.18 (d, 8.5 Hz, 1H), 7.58 (t, J = 8 Hz, 1H), 7.43 (dd, J_1 = 1 Hz, J_2 = 8 Hz, 2H), 7.28 (t, J = 7.5 Hz, 2H), 7.18 (t, J = 7.5 Hz, 1H), 5.41 (s, 1H), 4.56 (m, 1H), 4.18 (m, 1H), 3.88 (dd, J_1 = 9 Hz, J_2 = 14 Hz, 1H), 3.63 (br s, 1H), 3.44 (dd, J_1 = 5 Hz, J_2 = 13.5 Hz, 1H), 2.93 (dd, J_1 = 2 Hz, J_2 = 11.5 Hz, 1H), 2.54 (dd, J_1 = 10 Hz, J_2 = 12.5 Hz, 1H), 2.41 (dd, J_1 = 2.5 Hz, J_2 = 11.5 Hz, 1H), 2.22-2.15 (m, 2H), 2.05 (q, J = 12 Hz, 1H), 1.81-1.15 (complex, 11H), 1.03 (s, 9H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 174.06, 167.38, 148.19, 135.80, 135.41, 134.03, 130.15, 129.32, 129.08, 126.51, 125.96, 123.35, 70.80, 69.81, 59.66, 58.94, 56.27, 51.29, 36.14, 35.66, 34.01, 31.26, 30.98, 28.33, 26.38, 26.12, 20.43. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{31}H_{43}N_4O_5S$, 583.2954; observed, 583.2951.

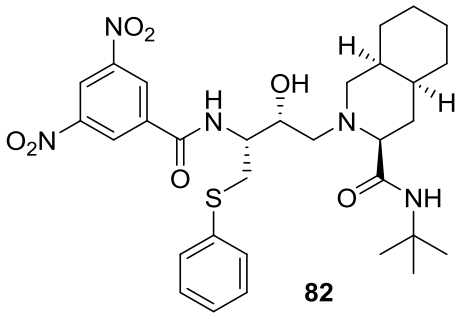


81

(3S,4aS,8aS)-N-(tert-butyl)-2-((2R,3R)-2-hydroxy-3-(4-nitrobenzamido)-4-(phenylthio)butyl)decahydroisoquinoline-3-carboxamide (81)

This reaction was performed according to the general EDC coupling procedure (0.065 mmol 4-nitrobenzoic acid for 40 h). Purification by flash column chromatography (gradient of 1-3% MeOH in DCM) yielded **81** as a white solid (19 mg, 53%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.23 (dd, J_1 = 2 Hz, J_2 = 7 Hz, 2H), 8.15 (d, J = 8.5 Hz, 1H), 8.02 (dd, J_1 = 2 Hz, J_2 = 7 Hz, 2H), 7.43 (dd, J_1 = 1.5 Hz, J_2 = 8

Hz, 2H), 7.28 (t, J = 7.5 Hz, 2H), 7.20 (tt, J_1 = 1.5 Hz, J_2 = 7.5 Hz, 1H), 5.40 (s, 1H), 4.56 (m, 1H), 4.17 (m, 1H), 3.92 (dd, J_1 = 9 Hz, J_2 = 13.5 Hz, 1H), 3.41 (dd, J_1 = 10 Hz, J_2 = 13.5 Hz, 1H), 2.93 (dd, J_1 = 2 Hz, J_2 = 11.5 Hz, 1H), 2.53 (dd, J_1 = 10 Hz, J_2 = 12.5 Hz, 1H), 2.41 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 2.21-2.14 (m, 2H), 2.40 (q, J = 11.5 Hz, 1H), 1.82-1.15 (complex, 11H), 1.04 (s, 9H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 174.14, 167.77, 149.60, 139.65, 135.38, 130.13, 129.41, 129.09, 126.51, 123.34, 70.80, 69.76, 59.60, 58.85, 56.39, 51.36, 36.12, 35.64, 34.00, 31.26, 30.98, 28.39, 26.38, 26.13, 20.41. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{31}H_{43}N_4O_5S$, 583.2954; observed, 583.2954.

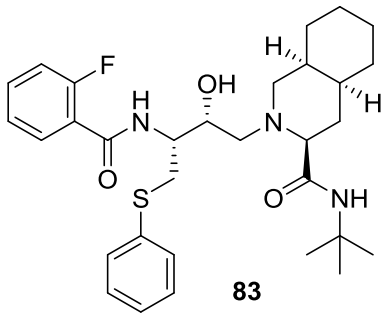


82

(3S,4aS,8aS)-N-(tert-butyl)-2-((2R,3R)-3-(3,5-dinitrobenzamido)-2-hydroxy-4-(phenylthio)butyl)decahydroisoquinoline-3-carboxamide (82)

This reaction was performed according to the general EDC coupling procedure (0.048 mmol 3,5-dinitrobenzoic acid for 40 h). Purification by flash column chromatography (gradient of 1-3% MeOH in DCM) yielded **82** as a slightly yellow solid (15 mg, 52%). ^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 9.13 (t, J = 2 Hz, 1H), 9.07 (d, J = 2 Hz, 2H), 8.79 (d, J = 8.8 Hz, 1H), 7.45 (dd, J_1 = 0.8 Hz, J_2 = 8.4 Hz, 2H), 7.27 (t, J = 8 Hz, 2H), 7.17 (t, J = 8 Hz, 1H), 5.36 (s, 1H), 4.61 (m, 1H), 4.24 (m, 1H), 3.86 (dd, J_1 = 8.8 Hz, J_2 = 14 Hz, 1H), 3.42 (dd, J_1 = 5.6 Hz, J_2 = 14 Hz, 1H), 3.09 (br s, 1H), 2.93 (d, J = 10.4, 1H), 2.52 (dd, J_1 = 10 Hz, J_2 = 12.8 Hz, 1H), 2.41 (dd, J_1 = 2.4 Hz, J_2 = 11.6 Hz, 1H), 2.23 (dd, J_1 = 4 Hz, J_2 = 12.8 Hz, 1H), 2.18 (dd, J_1 = 3.2 Hz, J_2 = 11.6 Hz, 1H), 2.07 (q, J = 12.4 Hz, 1H), 1.81-1.14 (complex, 11H), 1.01 (s, 9H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 174.14, 164.99, 148.29, 137.94, 135.24, 130.07, 129.13,

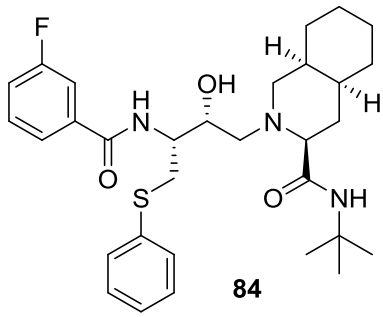
128.49, 126.59, 120.77, 70.72, 69.20, 59.80, 59.09, 56.93, 51.34, 36.10, 35.75, 34.05, 31.34, 30.94, 28.12, 26.39, 26.17, 20.38. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{31}H_{42}N_5O_7S$, 628.2805; observed, 628.2807.



83

(3S,4aS,8aS)-N-(tert-butyl)-2-((2R,3R)-3-(2-fluorobenzamido)-2-hydroxy-4-(phenylthio)butyl)decahydroisoquinoline-3-carboxamide (83)

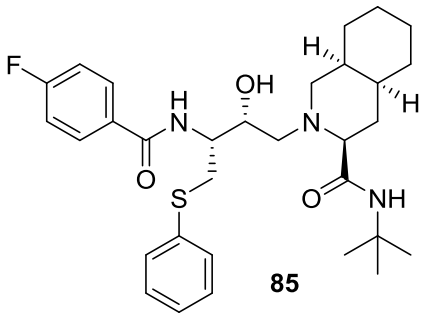
This reaction was performed according to the general EDC coupling procedure (0.046 mmol 2-fluorobenzoic acid for 48 h). Purification by flash column chromatography (gradient of 1-3% MeOH in DCM) yielded **83** as a white solid (13.7 mg, 56%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.93 (dt, J_1 = 2 Hz, J_2 = 8 Hz, 1H), 7.47-7.39 (m, 4H), 7.21 (t, J = 8 Hz, 2H), 7.19 (d, J = 1 Hz, 1H), 7.14-7.07 (m, 2H), 5.66 (s, 1H), 4.42 (m, 1H), 4.08 (m, 1H), 3.57 (dd, J_1 = 8 Hz, J_2 = 14 Hz, 1H), 3.45 (dd, J_1 = 5 Hz, J_2 = 14 Hz, 1H), 2.94 (dd, J_1 = 2.5 Hz, J_2 = 11.5 Hz, 1H), 2.62 (dd, J_1 = 7.5 Hz, J_2 = 13 Hz, 1H), 2.51 (dd, J_1 = 3 Hz, J_2 = 11 Hz, 1H), 2.27 (dd, J_1 = 5.5 Hz, J_2 = 13 Hz, 1H), 2.23 (dd, J_1 = 6.5 Hz, J_2 = 12 Hz, 1H), 1.98 (q, J = 13 Hz, 1H), 1.82-1.10 (complex, 20H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.62, 164.84, 135.58, 133.32, 133.24, 131.82, 130.04, 128.95, 126.37, 124.70, 116.22, 116.02, 70.60, 70.38, 59.73, 58.83, 54.50, 51.06, 35.98, 34.82, 33.56, 30.89, 30.81, 28.54, 26.19, 25.86, 20.59. ^{19}F NMR (500 MHz, $CDCl_3$) δ ppm -112.63. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{31}H_{43}N_3O_3SF$, 556.3009; observed, 556.3013.



84

(3S,4aS,8aS)-N-(tert-butyl)-2-((2R,3R)-3-(3-fluorobenzamido)-2-hydroxy-4-(phenylthio)butyl)decahydroisoquinoline-3-carboxamide (84)

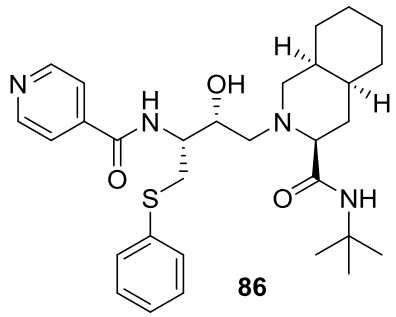
This reaction was performed according to the general EDC coupling procedure (0.046 mmol 3-fluorobenzoic acid for 48 h). Purification by flash column chromatography (gradient of 1-3% MeOH in DCM) yielded **84** as a white solid (23.4 mg, 96%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.82 (d, J = 8.5 Hz, 1H), 7.65 (d, J = 7.5 Hz, 1H), 7.56 (m, 1H), 7.42 (d, J = 7.5 Hz, 2H), 7.37 (m, 1H), 7.27 (t, J = 7.5 Hz, 2H), 7.18 (t, J = 7.5 Hz, 1H), 5.46 (s, 1H), 4.50 (m, 1H), 4.14 (m, 1H), 3.88 (dd, J_1 = 9 Hz, J_2 = 13.5 Hz, 1H), 3.43 (dd, J_1 = 5.5 Hz, J_2 = 13.5 Hz, 1H), 2.94 (dd, J_1 = 2 Hz, J_2 = 12 Hz, 1H), 2.56 (dd, J_1 = 9.5 Hz, J_2 = 12.5 Hz, 1H), 2.42 (dd, J_1 = 2.5 Hz, J_2 = 11.5 Hz, 1H), 2.20-2.14 (m, 2H), 2.04 (q, J = 11.5 Hz, 1H), 1.82-1.15 (complex, 11H), 1.09 (s, 9H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.99, 168.55, 130.17, 129.91, 129.85, 129.05, 126.47, 123.76, 118.61, 118.44, 115.33, 115.15, 70.86, 70.03, 59.59, 58.84, 55.91, 51.33, 36.12, 35.55, 33.95, 31.16, 30.99, 28.43, 26.37, 26.07, 20.44. ^{19}F NMR (500 MHz, $CDCl_3$) δ ppm -112.95. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{31}H_{43}N_3O_3SF$, 556.3009; observed, 556.3013.



85

(3S,4aS,8aS)-N-(tert-butyl)-2-((2R,3R)-3-(4-fluorobenzamido)-2-hydroxy-4-(phenylthio)butyl)decahydroisoquinoline-3-carboxamide (85)

This reaction was performed according to the general EDC coupling procedure (0.053 mmol 4-fluorobenzoic acid for 48 h). Purification by flash column chromatography (gradient of 1-3% MeOH in DCM) yielded **85** as a white solid (20.4 mg, 72%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.86 (dd, J_1 = 5.5 Hz, J_2 = 8.5 Hz, 2H), 7.73 (d, J = 8.5 Hz, 1H) 7.42 (d, J = 8 Hz, 2H), 7.27 (t, J = 7.5 Hz, 2H), 7.18 (t, J = 8 Hz, 1H), 7.06 (t, J = 7.5 Hz, 2H), 5.42 (s, 1H), 4.48 (m, 1H), 4.13 (m, 1H), 3.89 (dd, J_1 = 9.5 Hz, J_2 = 14 Hz, 1H), 3.42 (dd, J_1 = 5.5 Hz, J_2 = 14 Hz, 1H), 2.94 (d, J = 12 Hz, 1H), 2.55 (dd, J_1 = 10.5 Hz, J_2 = 12.5 Hz, 1H), 2.40 (dd, J_1 = 2.5 Hz, J_2 = 11 Hz, 1H), 2.19-2.13 (m, 2H), 2.04 (q, J = 12 Hz, 1H), 1.82-1.16 (complex, 11H), 1.07 (s, 9H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 174.04, 168.96, 135.47, 130.59, 130.52, 130.17, 129.05, 126.43, 115.31, 115.14, 70.95, 70.11, 59.53, 58.90, 56.02, 51.30, 36.13, 35.64, 33.95, 31.14, 31.00, 28.41, 26.38, 26.10, 20.43. ^{19}F NMR (500 MHz, $CDCl_3$) δ ppm -108.75. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{31}H_{43}N_3O_3SF$, 556.3009; observed, 556.3004.

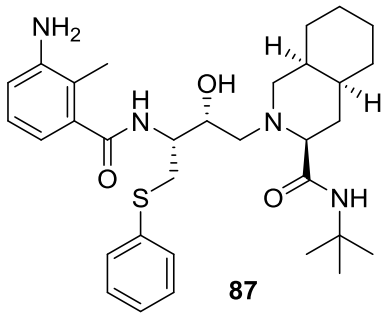


86

(3S,4aS,8aS)-N-(tert-butyl)-2-((2R,3R)-2-hydroxy-3-(isonicotinamido)-4-(phenylthio)butyl)decahydroisoquinoline-3-carboxamide (86)

This reaction was performed according to the general EDC coupling procedure (0.080 mmol isonicotinic acid for 48 h). Purification by flash column chromatography (gradient of 1-5% MeOH in DCM) yielded **86** as a white solid (32 mg, 78%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.68 (dd, J_1 = 1 Hz, J_2 = 4.5 Hz, 2H), 8.10 (d, J = 8.5 Hz, 1H), 7.71 (dd, J_1 = 1 Hz, J_2 = 4.5 Hz, 2H), 7.42 (d, J = 7.5 Hz, 2H),

7.28 (t, J = 7.5 Hz, 2H), 7.19 (t, J = 7.5 Hz, 1H), 5.51 (s, 1H), 4.55 (m, 1H), 4.17 (m, 1H), 3.87 (dd, J_1 = 9 Hz, J_2 = 13.5 Hz, 1H), 3.70 (br s, 1H), 3.42 (dd, J_1 = 5.5 Hz, J_2 = 13.5 Hz, 1H), 2.93 (dd, J_1 = 1.5 Hz, J_2 = 12 Hz, 1H), 2.54 (dd, J_1 = 10 Hz, J_2 = 9 Hz, 1H), 2.21-2.13 (m, 2H), 2.04 (q, J = 13 Hz, 1H), 1.82-1.15 (complex, 11H), 1.08 (s, 9H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 174.11, 167.80, 150.19, 141.10, 135.40, 130.05, 129.06, 126.46, 121.96, 70.72, 69.63, 59.53, 58.79, 55.94, 51.33, 36.07, 35.49, 33.94, 31.19, 30.95, 28.40, 26.34, 26.06, 20.39. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{30}H_{43}N_4O_3S$, 539.3056; observed, 539.3060.

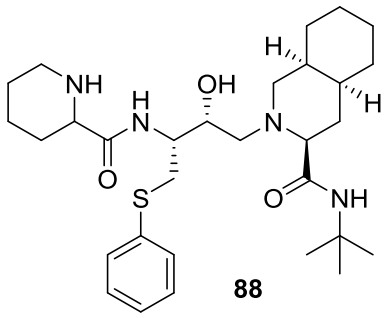


87

(3S,4aS,8aS)-2-((2R,3R)-3-(3-amino-2-methylbenzamido)-2-hydroxy-4-(phenylthio)butyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (87)

This reaction was performed according to the general EDC coupling procedure (0.069 mmol 3-amino-2-methylbenzoic acid for 48 h). Purification by flash column chromatography (gradient of 1-5% MeOH in DCM) yielded **87** as a white solid (24 mg, 65%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.43 (d, J = 7 Hz, 2H), 7.28 (t, J = 7.5 Hz, 2H), 7.19 (t, J = 7.5 Hz, 1H), 7.13 (d, J = 8.5 Hz, 1H), 7.01 (t, J = 7.5 Hz, 1H), 6.92 (d, J = 7.5 Hz, 1H), 6.71 (t, J = 8 Hz, 1H), 5.55 (s, 1H), 4.45 (m, 1H), 4.05 (m, 1H), 3.75 (dd, J_1 = 9 Hz, J_2 = 13.5 Hz, 1H), 3.66 (br s, 1H), 3.42 (dd, J_1 = 5 Hz, J_2 = 14 Hz, 1H), 2.93 (dd, J_1 = 2.5 Hz, J_2 = 12 Hz, 1H), 2.57 (dd, J_1 = 9 Hz, J_2 = 13 Hz, 1H), 2.45 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 2.24-2.17 (m, 3H), 1.97 (q, J = 11.5 Hz, 1H), 1.80-1.15 (complex, 11H), 1.09 (s, 9H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.88, 172.29, 145.26, 136.98, 135.63, 130.12, 129.01, 126.40, 120.51, 120.31, 118.13,

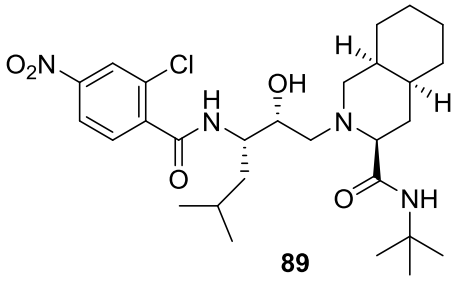
116.68, 70.65, 70.51, 59.68, 58.67, 54.49, 51.17, 36.04, 35.38, 33.70, 30.94, 30.89, 28.38, 26.26, 25.97, 20.50, 13.88. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{32}H_{47}N_4O_3S$, 567.3369; observed, 567.3364.



88

(3S,4aS,8aS)-N-(tert-butyl)-2-((2R,3R)-2-hydroxy-4-(phenylthio)-3-(piperidine-2-carboxamido)butyl)decahydroisoquinoline-3-carboxamide (88)

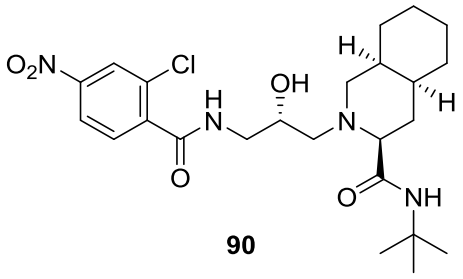
This reaction was performed according to the general EDC coupling procedure (0.061 mmol piperidine-2-carboxylic acid for 16 h). Purification by flash column chromatography (gradient of 3-10% MeOH in DCM) yielded **88** as a white solid (10.7 mg, 34%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.64 (d, J = 7.5 Hz, 1H), 7.39 (dd, J_1 = 1 Hz, J_2 = 8 Hz, 2H), 7.27 (t, J = 7.5 Hz, 2H), 7.17 (t, J = 7.5 Hz, 1H), 6.03 (s, 1H), 4.11 (m, 1H), 3.99 (m, 1H), 3.43 (br s, 1H), 3.36 (dd, J_1 = 9 Hz, J_2 = 14 Hz, 1H), 3.27-3.23 (m, 2H), 3.12 (d, J = 12.5 Hz, 1H), 2.99 (dd, J_1 = 2 Hz, J_2 = 12 Hz, 1H), 2.68-2.61 (m, 3H), 2.31-2.24 (m, 2H), 2.00-1.91 (m, 4H), 1.82-1.15 (complex, 23H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 186.55, 173.50, 136.25, 129.84, 129.21, 126.42, 70.57, 70.35, 60.08, 58.99, 53.94, 51.36, 45.55, 36.00, 34.09, 33.63, 31.04, 30.92, 28.95, 26.35, 25.97, 24.91, 24.85, 23.75, 20.86, 11.06. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{30}H_{49}N_4O_3S$, 545.3525; observed, 545.3525.



89

(3S,4aS,8aS)-N-(tert-butyl)-2-((2R,3S)-3-(2-chloro-4-nitrobenzamido)-2-hydroxy-5-methylhexyl)decahydroisoquinoline-3-carboxamide (89)

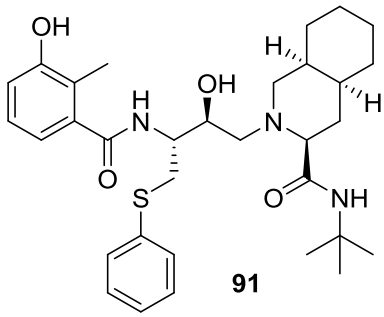
A solution of **48** (527 mg, 1.05 mmol) and 10% palladium on carbon (335 mg, 0.315 mmol) in methanol (15 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite and washed with excess methanol. The reaction was then dried to a solid and redissolved in THF (10 mL). 2-chloro-4-nitrobenzoic acid (234 mg, 1.16 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (222 mg, 1.16 mmol), and hydroxybenzotriazole hydrate (188 mg, 1.16 mmol) were added to the reaction and the solution was stirred at room temperature for 18 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-3% MeOH in DCM) to yield **89** as a yellow solid (99 mg, 17%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.23 (d, J = 2.5 Hz, 1H), 8.12 (dd, J_1 = 2 Hz, J_2 = 8.5 Hz, 1H), 7.73 (d, J = 8.5 Hz, 1H), 7.63 (d, J = 9 Hz, 1H), 5.76 (s, 1H), 4.21 (t, J = 10 Hz, 1H), 3.95 (t, J = 6 Hz, 1H), 2.95 (d, J = 12 Hz, 1H), 2.53 (dd, J_1 = 7.5 Hz, J_2 = 13 Hz, 1H), 2.47 (d, J = 9 Hz, 1H), 2.24 (dd, J_1 = 6 Hz, J_2 = 13 Hz, 1H), 2.19 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 2.01 (m, 1H), 1.90 (q, J = 12 Hz, 1H), 1.84-0.85 (complex, 28H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 174.13, 166.02, 148.44, 141.76, 132.56, 130.42, 125.09, 121.66, 71.78, 70.28, 59.34, 59.12, 53.66, 51.03, 39.52, 35.91, 33.60, 30.94, 30.75, 28.45, 26.15, 26.04, 25.24, 23.53, 21.55, 20.51. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{28}H_{44}N_4O_5Cl$, 551.3000; observed, 351.3005.



90

(3S,4aS,8aS)-N-(tert-butyl)-2-((S)-3-(2-chloro-4-nitrobenzamido)-2-hydroxypropyl)decahydroisoquinoline-3-carboxamide (90)

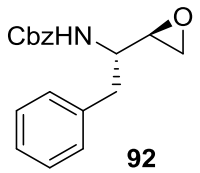
A solution of **34** (104 mg, 0.233 mmol) and 10% palladium on carbon (74 mg, 0.070 mmol) in methanol (15 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite and washed with excess methanol. The reaction was then dried to a solid and redissolved in THF (5 mL). 2-chloro-4-nitrobenzoic acid (52 mg, 0.256 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (49 mg, 0.256 mmol), and hydroxybenzotriazole hydrate (42 mg, 0.256 mmol) were added to the reaction and the solution was stirred at room temperature for 18 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-3% MeOH in DCM) to yield **90** as a yellow solid (77 mg, 67%). Crystals were grown for x-ray crystallography from an ethyl acetate/pentane solution. ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.41 (q, J = 4 Hz, 1H), 8.24 (d, J = 2 Hz, 1H), 8.13 (dd, J_1 = 2 Hz, J_2 = 8.5 Hz, 1H), 7.76 (d, J = 8.5 Hz, 1H), 5.60 (s, 1H), 4.04 (m, 2H), 3.47 (dt, J_1 = 4 Hz, J_2 = 14 Hz, 1H), 3.35 (br, 1H), 2.71 (dd, J_1 = 2 Hz, J_2 = 11.5 Hz, 1H), 2.49 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 2.36 (dd, J_1 = 10 Hz, J_2 = 12.5 Hz, 1H), 2.22 (dd, J_1 = 5 Hz, J_2 = 12.5 Hz, 1H), 2.18 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 1.95 (q, J = 12 Hz, 1H), 1.80-1.08 (complex, 20H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 174.16, 167.06, 148.39, 142.00, 132.80, 130.18, 124.96, 121.56, 70.40, 68.29, 59.09, 57.54, 51.27, 43.27, 35.83, 33.55, 31.02, 30.86, 28.39, 26.19, 25.52, 20.18. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{24}H_{36}N_4O_5Cl$, 495.2374; observed, 495.2376.



91

(3S,4aS,8aS)-N-(tert-butyl)-2-((2S,3R)-2-hydroxy-3-(3-hydroxy-2-methylbenzamido)-4-(phenylthio)butyl)decahydroisoquinoline-3-carboxamide (91)

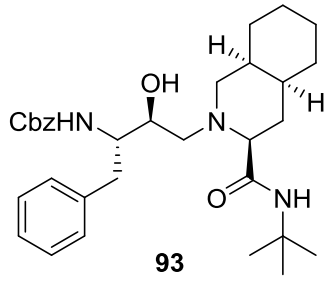
To a solution of **63** (25 mg, 0.058 mmol) in THF (3 mL) was added 3-hydroxy-2-methylbenzoic acid (10 mg, 0.063 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (12 mg, 0.063 mmol), and hydroxybenzotriazole hydrate (10 mg, 0.063 mmol). The solution was stirred at room temperature for 20 h. The reaction was taken up in ethyl acetate and washed once with water, once with saturated sodium bicarbonate, and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-4% MeOH in DCM) to yield **91** as a white solid (22 mg, 67%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.48 (d, J = 7.5 Hz, 2H), 7.29 (t, J = 7.5 Hz, 2H), 7.16 (t, J = 7.5 Hz, 1H), 7.13 (br s, 1H), 6.86 (m, 3H), 6.43 (d, J = 9 Hz, 1H), 5.73 (s, 1H), 4.16 (d, J = 10 Hz, 1H), 4.03 (q, J = 8 Hz, 1H), 3.37 (dd, J_1 = 6 Hz, J_2 = 13.5 Hz, 1H), 3.20 (dd, J_1 = 8.5 Hz, J_2 = 13.5 Hz, 1H), 2.85 (d, J = 10 Hz, 1H), 2.63 (dd, J_1 = 3 Hz, J_2 = 11 Hz, 1H), 2.54 (dd, J_1 = 11 Hz, J_2 = 12.5 Hz, 1H), 2.22 (s, 3H), 2.20 (dd, J_1 = 3 Hz, J_2 = 12.5 Hz, 1H), 2.07 (dd, J_1 = 2.5 Hz, J_2 = 12.5 Hz, 1H), 1.99 (q, J = 12.5 Hz, 1H), 1.88-1.18 (complex, 21H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm173.65, 170.34, 154.87, 137.56, 135.92, 129.04, 128.82, 126.35, 125.99, 122.24, 118.61, 116.64, 69.49, 66.15, 58.58, 58.04, 51.25, 50.93, 35.81, 35.43, 33.45, 31.14, 30.94, 28.61, 26.24, 25.32, 20.24, 12.50. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{32}H_{45}N_3O_4S$, 568.3209; observed, 568.3212.



92

Benzyl ((S)-1-((R)-oxiran-2-yl)-2-phenylethyl)carbamate (92)

To a solution of **52** (115 mg, 0.35 mmol) in ethanol (5 mL) at 0 °C was added solid potassium hydroxide (21 mg, 0.38 mmol). The reaction was stirred cold for 15 min then warmed to room temperature and stirred for an additional 3 h. The reaction was evaporated to dryness and the solid was purified by silica flash column chromatography (gradient of 0-2.5% EtOAc in DCM) to yield **92** as a white solid (88 mg, 86%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.36-7.22 (complex, 10H), 5.05 (s, 2H), 4.82 (d, J = 13.5 Hz, 1H), 4.18 (dd, J_1 = 6.5 Hz, J_2 = 14.5 Hz, 1H), 3.02-2.96 (m, 2H), 2.90 (dd, J_1 = 8 Hz, J_2 = 13.5 Hz, 1H), 2.67 (t, J = 4.5 Hz, 1H), 2.55 (br s, 1H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 155.91, 136.97, 136.29, 129.31, 128.52, 128.46, 128.07, 127.89, 126.67, 66.69, 52.38, 51.00, 44.45, 39.60. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{20}NO_3$, 298.1443; observed, 298.1437.

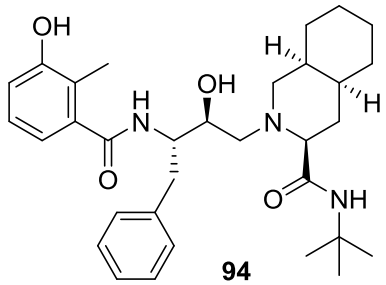


93

Benzyl ((2S,3S)-4-((3S,4aS,8aS)-3-(tert-butylcarbamoyl)octahydroisoquinolin-2(1H)-yl)-3-hydroxy-1-phenylbutan-2-yl)carbamate (93)

To a solution of **92** (70 mg, 0.235 mmol) in isopropyl alcohol (7 mL) was added (3S,4aS,8aS)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (**6**, 67 mg, 0.283 mmol). The solution was heated to reflux and stirred. After 5 h, the reaction was cooled and dried to a solid. The product was purified by silica flash column chromatography (gradient of 0-4% MeOH in DCM) to yield **93** as a white solid (110 mg, 87%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.34-7.19 (complex, 10H), 5.80 (s, 1H), 5.28 (s, 1H), 5.19

(d, J = 9.5 Hz, 1H), 5.07 (d, J = 12.5 Hz, 1H), 4.90 (d, J = 12.5 Hz, 1H), 3.70 (m, 3H), 2.94 (m, 2H), 2.69 (dd, J_1 = 2 Hz, J_2 = 11.5 Hz, 1H), 2.61 (m, 2H), 2.13 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 1.93 (m, 2H), 1.75 (m, 3H), 1.58-1.18 (complex, 16H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.01, 156.12, 138.11, 136.43, 129.36, 128.33, 127.86, 127.80, 126.28, 69.68, 66.42, 66.13, 58.49, 58.02, 54.17, 50.86, 39.36, 35.70, 33.21, 30.82, 28.55, 26.21, 25.38, 20.19. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{32}H_{46}N_3O_4$, 536.3488; observed, 536.3485.

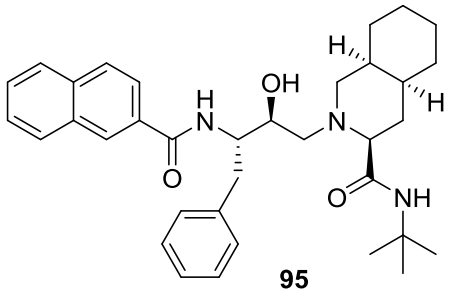


94

(3S,4aS,8aS)-N-(tert-butyl)-2-((2S,3S)-2-hydroxy-3-(3-hydroxy-2-methylbenzamido)-4-phenylbutyl)decahydroisoquinoline-3-carboxamide (94)

A solution of **93** (51 mg, 0.095 mmol) and 10% palladium on carbon (30 mg, 0.029 mmol) in methanol (10 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite and washed with excess methanol. The reaction was then dried to a solid and redissolved in THF (5 mL). 3-hydroxy-2-methylbenzoic acid (16 mg, 0.105 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (20 mg, 0.105 mmol), and hydroxybenzotriazole hydrate (17 mg, 0.105 mmol) were added to the reaction and the solution was stirred at room temperature for 18 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-4% MeOH in DCM) to yield **94** as a white solid (27 mg, 53%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.31-7.19 (complex, 8H), 6.84 (dd, J_1 = 6 Hz, J_2 = 7.5 Hz, 2H), 6.69 (dd, J_1 = 2 Hz, J_2 = 7.5 Hz, 1H), 6.28 (d, J = 9 Hz, 1H), 5.82 (s, 1H), 4.21 (q, J = 8 Hz, 1H), 3.82 (d, J = 10.5

Hz, 1H), 3.04 (m, 2H), 2.75 (d, J = 10 Hz, 1H), 2.62 (dd, J_1 = 3 Hz, J_2 = 11 Hz, 1H), 2.55 (dd, J_1 = 11 Hz, J_2 = 12 Hz, 1H), 2.15 (dd, J_1 = 3 Hz, J_2 = 11 Hz, 1H), 2.09 (dd, J_1 = 2.5 Hz, J_2 = 12.5 Hz, 1H), 2.00 (s, 3H), 1.97 (m, 1H), 1.85 (m, 1H), 1.73 (m, 2H), 1.60-1.19 (complex, 16H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.78, 170.22, 154.90, 138.14, 137.80, 129.41, 128.41, 126.41, 126.23, 122.34, 118.41, 116.59, 69.33, 67.47, 59.01, 58.14, 52.69, 51.19, 38.76, 35.80, 33.42, 31.17, 30.92, 28.62, 26.23, 25.32, 20.23, 12.26. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{32}H_{46}N_3O_4$, 536.3488; observed, 536.3476.

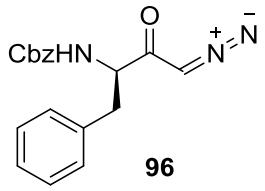


95

(3S,4aS,8aS)-2-((2S,3S)-3-(2-naphthamido)-2-hydroxy-4-phenylbutyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (95)

A solution of **93** (46 mg, 0.086 mmol) and 10% palladium on carbon (27 mg, 0.026 mmol) in methanol (15 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite and washed with excess methanol. The reaction was then dried to a solid and redissolved in THF (5 mL). 2-naphthoic acid (16 mg, 0.095 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (18 mg, 0.095 mmol), and hydroxybenzotriazole hydrate (15 mg, 0.095 mmol) were added to the reaction and the solution was stirred at room temperature for 18 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-3% MeOH in DCM) to yield **95** as a white solid (37 mg, 77%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.23 (s, 1H), 7.89 (d, 7.5 Hz, 1H), 7.83 (m, 3H), 7.53 (m, 2H), 7.34 (d, J = 7 Hz, 2H), 7.28 (t, J = 7.5 Hz, 2H), 7.19 (t, J = 7.5 Hz, 1H), 6.91 (d, J = 8.5 Hz, 1H), 5.56 (s, 1H), 4.34 (br

s, 1H), 4.24 (q, J = 8 Hz, 1H), 3.84 (d, J = 9 Hz, 1H), 3.14 (dd, J_1 = 7 Hz, J_2 = 13.5 Hz, 1H), 3.08 (dd, J_1 = 8 Hz, J_2 = 14 Hz, 1H), 2.71 (d, J = 10.5 Hz, 1H), 2.58 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 2.53 (dd, J_1 = 11 Hz, J_2 = 12.5 Hz, 1H), 2.15 (dd, J_1 = 3 Hz, J_2 = 11 Hz, 1H), 2.06 (dd, J_1 = 2.5 Hz, J_2 = 13 Hz, 1H), 1.98 (q, J = 12 Hz, 1H), 1.85 (qd, J_1 = 3 Hz, J_2 = 13 Hz, 1H), 1.73 (m, 2H), 1.59-0.94 (complex, 17H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.12, 166.96, 138.33, 134.69, 132.54, 131.43, 129.43, 128.93, 128.37, 128.21, 127.59, 127.45, 127.40, 126.51, 126.33, 123.69, 69.31, 66.67, 58.74, 58.11, 52.91, 50.90, 39.08, 35.82, 33.38, 31.04, 30.93, 28.38, 26.23, 25.35, 20.19. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{35}H_{46}N_3O_3$, 556.3539; observed, 556.3538.

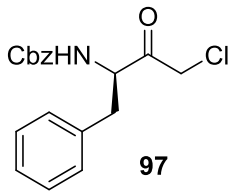


96

(R)-benzyl (4-diazo-3-oxo-1-phenylbutan-2-yl)carbamate (96)

To a stirred solution of Z-D-phenylalanine (1.11 g, 3.70 mmol) in anhydrous tetrahydrofuran (12 mL) under nitrogen at -78 °C was added anhydrous triethylamine (768 μL, 5.54 mmol) followed by isobutyl chloroformate (721 μL, 5.54 mmol). The reaction was stirred cold for 1 h and then warmed to room temperature for 1.5 h. The reaction was cooled to -78 °C before an ethereal solution of diazomethane (25 mL, 8.1 mmol) was added in one portion and the reaction was then warmed to room temperature with stirring for 1 h. The reaction was evaporated to dryness, resulting in a yellow solid. The solid was taken up in ethyl acetate, washed twice with water, twice with a saturated solution of sodium bicarbonate, and once with brine. The organic phase was dried over sodium sulfate and concentrated by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 0-6% EtOAc in $CHCl_3$) to yield **96** as a yellow solid (826 mg, 77%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.32-7.18 (complex, 8H), 7.14 (d, J = 7 Hz, 2H), 5.70 (d, J = 7.5 Hz, 1H), 5.23 (s, 1H), 5.03 (dd, J_1 = 12.5 Hz, J_2 = 19 Hz, 2H), 4.47 (d, J = 5.5 Hz, 1H), 3.04 (dd, J_1 = 7 Hz, J_2 = 14 Hz, 1H), 2.95 (dd, J_1 = 7 Hz, J_2 =

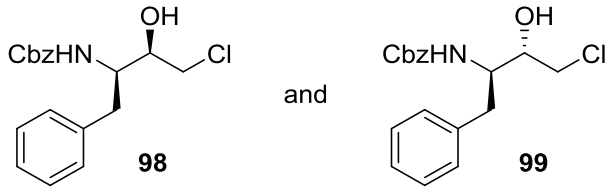
14 Hz, 1H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 192.79, 155.61, 136.03, 129.09, 128.41, 128.29, 127.94, 127.76, 126.78, 66.70, 58.74, 54.27, 38.03. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{18}N_3O_3$, 324.1348; observed, 324.1354.



97

(R)-benzyl (4-chloro-3-oxo-1-phenylbutan-2-yl)carbamate (97)

To a solution of **96** (826 mg, 2.86 mmol) in anhydrous tetrahydrofuran (8 mL) under nitrogen at 0 °C was added 1 M HCl in diethyl ether (4.28 mL, 4.28 mmol) dropwise. The reaction was stirred at 0 °C for 30 min after complete addition. The solution was dried under reduced pressure to yield **97** as an off-white solid (843 mg). No further purification was made. ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.34-7.21 (complex, 8H), 7.11 (d, J = 7 Hz, 2H), 5.53 (d, J = 7.5 Hz, 1H), 5.03 (dd, J_1 = 12 Hz, J_2 = 17.5 Hz, 2H), 4.72 (q, J = 7.5 Hz, 1H), 4.14 (d, J = 16.5 Hz, 1H), 3.97 (d, J = 16 Hz, 1H), 3.07 (dd, J_1 = 6.5 Hz, J_2 = 14 Hz, 1H), 2.95 (dd, J_1 = 7.5 Hz, J_2 = 14 Hz, 1H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 200.90, 155.70, 135.81, 135.26, 128.97, 128.76, 128.41, 128.15, 127.91, 127.22, 67.04, 58.65, 47.33, 37.31. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{19}NO_3Cl$, 332.1053; observed, 332.1060.

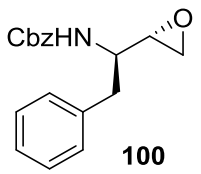


98 **99**

Benzyl ((2R,3R)-4-chloro-3-hydroxy-1-phenylbutan-2-yl)carbamate (98) and benzyl ((2R,3S)-4-chloro-3-hydroxy-1-phenylbutan-2-yl)carbamate (99)

To a solution of **97** (843 mg, 2.54 mmol) in dichloromethane (7 mL) and methanol (4 mL) under nitrogen at 0 °C was added sodium borohydride (67 mg, 1.78 mmol). The reaction was stirred at 0 °C for

1 h, and then quenched with slow addition of 2 M HCl (2 mL). The reaction was evaporated to dryness and the solid was taken up in ethyl acetate. The solution was washed twice with water followed by brine, dried over sodium sulfate, and concentrated by rotary evaporation. TLC analysis showed two diastereomers with the lower Rf compound being the (R,R)-isomer. The (R,R)-isomer was recrystallized in hexanes/ethyl acetate. The crystals were collected and washed with cold hexanes. The filtrate was collected and dried to a solid. Both diastereomers were then separated further by silica flash column chromatography (gradient of 0-6% EtOAc in DCM) to yield (R,R)-isomer (**98**) as a white solid (592 mg, 70%) and (S,R)-isomer (**99**) as a white solid (124 mg, 15%) over two steps. Characterization of the (R,R)-isomer (**98**): ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.37-7.21 (complex, 8H), 7.19 (d, J = 7 Hz, 2H), 5.04 (s, 2H), 4.83 (d, J = 8.5 Hz, 1H), 3.99 (m, 1H), 3.85 (br s, 1H), 3.66 (dd, J_1 = 4 Hz, J_2 = 11.5 Hz, 1H), 3.58 (dd, J_1 = 7.5 Hz, J_2 = 11.5 Hz, 1H), 2.96 (m, 3H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 156.24, 136.98, 136.15, 129.42, 128.63, 128.53, 128.20, 127.98, 126.75, 73.19, 66.93, 54.65, 47.61, 35.62. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{21}NO_3Cl$, 334.1210; observed, 334.1217. Characterization of the (R,S)-isomer (**99**): ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.36-7.19 (complex, 10H), 5.26 (d, J = 9 Hz, 1H), 5.06 (dd, J_1 = 12.5 Hz, J_2 = 14.5 Hz, 2H), 3.95 (q, J = 8 Hz, 1H), 3.74 (m, 1H), 3.50 (d, J = 6.5 Hz, 2H), 3.11 (d, J = 3.5 Hz, 1H), 2.96 (dd, J_1 = 7 Hz, J_2 = 13.5 Hz, 1H), 2.90 (dd, J_1 = 8 Hz, J_2 = 13.5 Hz, 1H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 156.51, 137.46, 136.18, 129.23, 128.59, 128.51, 128.16, 127.97, 126.65, 71.20, 66.94, 54.29, 47.64, 38.47. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{21}NO_3Cl$, 334.1210; observed, 334.1208.

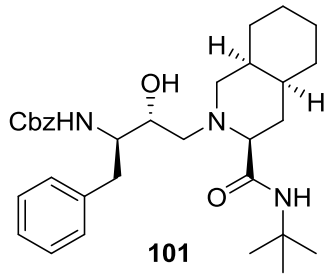


100

Benzyl ((R)-1-((S)-oxiran-2-yl)-2-phenylethyl)carbamate (100)

To a solution of **99** (124 mg, 0.371 mmol) in ethanol (10 mL) at 0 °C was added solid potassium hydroxide (23 mg, 0.409 mmol). The reaction was stirred cold for 15 min then warmed to room

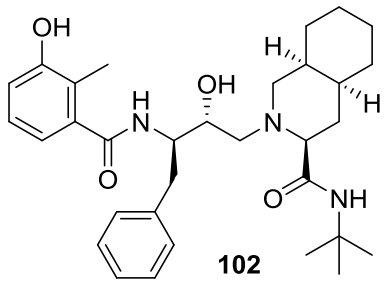
temperature and stirred for an additional 3 h. The reaction was evaporated to dryness and the solid was purified by silica flash column chromatography (gradient of 0-3% EtOAc in DCM) to yield **100** as a white solid (93 mg, 84%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.35-7.21 (complex, 10H), 5.05 (s, 2H), 4.84 (d, J = 9 Hz, 1H), 4.17 (q, J = 7.5 Hz, 1H), 2.98 (m, 2H), 2.89 (dd, J_1 = 7.5 Hz, J_2 = 13 Hz, 1H), 2.66 (t, J = 4.5 Hz, 1H), 2.54 (s, 1H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 155.90, 136.97, 136.28, 129.28, 128.49, 128.43, 128.04, 127.86, 126.65, 66.66, 52.38, 51.01, 44.43, 39.56. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{20}NO_3$, 298.1443; observed, 298.1447.



101

Benzyl ((2R,3R)-4-((3S,4aS,8aS)-3-(tert-butylcarbamoyl)octahydroisoquinolin-2(1H)-yl)-3-hydroxy-1-phenylbutan-2-yl)carbamate (101)

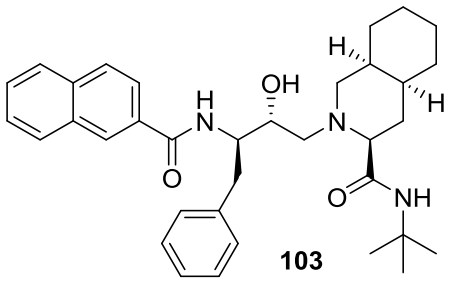
To a solution of **100** (93 mg, 0.313 mmol) in isopropyl alcohol (7 mL) was added (3S,4aS,8aS)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (**6**, 90 mg, 0.375 mmol). The solution was heated to reflux and stirred. After 5 h, the reaction was cooled and dried to a solid. The product was purified by silica flash column chromatography (gradient of 0-4% MeOH in DCM) to yield **101** as a white solid (140 mg, 83%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.35-7.15 (complex, 10H), 5.93 (s, 1H), 5.69 (d, J = 9 Hz, 1H), 5.08 (d, J = 12.5 Hz, 1H), 5.01 (d, J = 12 Hz, 1H), 3.90 (s, 1H), 3.75 (m, 2H), 3.01 (dd, J_1 = 6 Hz, J_2 = 13 Hz, 1H), 2.92 (dd, J_1 = 9 Hz, J_2 = 13 Hz, 1H), 2.53 (m, 3H), 2.17 (dd, J_1 = 7 Hz, J_2 = 13 Hz, 1H), 2.10 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 1.82 (q, J = 11.5 Hz, 1H), 1.68 (m, 3H), 1.50-1.16 (complex, 17H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.59, 156.69, 138.38, 136.41, 129.36, 128.35, 127.91, 127.88, 126.21, 70.22, 67.81, 66.48, 59.30, 58.87, 54.57, 50.67, 38.00, 35.58, 33.01, 30.61, 30.50, 28.42, 25.94, 25.52, 25.24, 20.51. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{32}H_{46}N_3O_4$, 536.3488; observed, 536.3489.



102

(3S,4aS,8aS)-N-(tert-butyl)-2-((2R,3R)-2-hydroxy-3-(3-hydroxy-2-methylbenzamido)-4-phenylbutyl)decahydroisoquinoline-3-carboxamide (102)

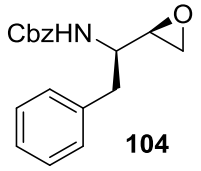
A solution of **101** (45 mg, 0.084 mmol) and 10% palladium on carbon (27 mg, 0.025 mmol) in methanol (8 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite and washed with excess methanol. The reaction was then dried to a solid and redissolved in THF (3 mL). 3-hydroxy-2-methylbenzoic acid (14 mg, 0.092 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (18 mg, 0.092 mmol), and hydroxybenzotriazole hydrate (15 mg, 0.092 mmol) were added to the reaction and the solution was stirred at room temperature for 18 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-4% MeOH in DCM) to yield **102** as a white solid (30 mg, 67%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.31-7.18 (complex, 5H), 6.94 (t, J = 7.5 Hz, 1H), 6.79 (m, 2H), 6.71 (d, J = 7 Hz, 1H), 5.85 (s, 1H), 4.18 (q, J = 8 Hz, 1H), 3.82 (t, J = 7 Hz, 1H), 3.20 (dd, J_1 = 8 Hz, J_2 = 14 Hz, 1H), 3.09 (dd, J_1 = 8 Hz, J_2 = 14 Hz, 1H), 2.70 (dd, J_1 = 2 Hz, J_2 = 11.5 Hz, 1H), 2.57 (dd, J_1 = 3 Hz, J_2 = 11 Hz, 1H), 2.45 (dd, J_1 = 6.5 Hz, J_2 = 13.5 Hz, 1H), 2.28 (m, 2H), 2.09 (s, 3H), 1.92 (q, J = 11.5 Hz, 1H), 1.80 (qd, J_1 = 3 Hz, J_2 = 13 Hz, 1H), 1.68 (m, 2H), 1.57-1.14 (complex, 19H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.66, 171.38, 154.71, 138.41, 137.87, 129.38, 128.52, 126.43, 126.37, 122.27, 118.85, 116.61, 70.21, 69.43, 60.40, 59.35, 54.25, 51.11, 37.48, 35.91, 33.20, 30.73, 30.65, 28.47, 26.08, 25.57, 20.52, 12.41. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{32}H_{46}N_3O_4$, 536.3488; observed, 536.3486.



103

(3S,4aS,8aS)-2-((2R,3R)-3-(2-naphthamido)-2-hydroxy-4-phenylbutyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (103)

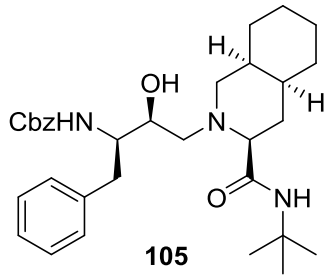
A solution of **102** (50 mg, 0.093 mmol) and 10% palladium on carbon (30 mg, 0.028 mmol) in methanol (8 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite and washed with excess methanol. The reaction was then dried to a solid and redissolved in THF (3 mL). 2-naphthoic acid (18 mg, 0.102 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (20 mg, 0.102 mmol), and hydroxybenzotriazole hydrate (17 mg, 0.102 mmol) were added to the reaction and the solution was stirred at room temperature for 18 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-3% MeOH in DCM) to yield **103** as a white solid (41 mg, 79%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.36 (s, 1H), 7.88 (dd, J_1 = 1.5 Hz, J_2 = 8.5 Hz, 2H), 7.83 (m, 3H), 7.51 (m, 2H), 7.34 (d, J = 7.5 Hz, 2H), 7.24 (t, J = 7.5 Hz, 2H), 7.15 (t, J = 7.5 Hz, 1H), 5.90 (br s, 1H), 4.31 (q, J = 8 Hz, 1H), 3.95 (t, J = 6 Hz, 1H), 3.28 (dd, J_1 = 8 Hz, J_2 = 13.5 Hz, 1H), 3.22 (dd, J_1 = 7.5 Hz, J_2 = 13.5 Hz, 1H), 2.76 (d, J = 9.5 Hz, 1H), 2.61 (m, 2H), 2.27 (dd, J_1 = 6 Hz, J_2 = 12.5 Hz, 1H), 2.19 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 1.97 (q, J = 12 Hz, 1H), 1.69-1.00 (complex, 20H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 172.97, 168.67, 138.99, 134.76, 132.57, 131.62, 129.43, 128.99, 128.41, 128.10, 128.08, 127.57, 127.47, 126.44, 126.24, 124.18, 70.71, 68.87, 59.99, 58.87, 55.45, 51.09, 36.27, 35.77, 33.40, 30.72, 28.34, 26.04, 25.38, 20.22. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{35}H_{46}N_3O_3$, 556.3539; observed, 556.3546.



104

Benzyl ((R)-1-((R)-oxiran-2-yl)-2-phenylethyl)carbamate (104)

To a solution of **98** (592 mg, 1.77 mmol) in ethanol (20 mL) and dichloromethane (5 mL) at 0 °C was added solid potassium hydroxide (109 mg, 1.95 mmol). The reaction was stirred cold for 15 min then warmed to room temperature and stirred for an additional 3 h. The reaction was evaporated to dryness and the solid was purified by silica flash column chromatography (gradient of 0-5% EtOAc in DCM) to yield **104** as a white solid (483 mg, 92%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.34-7.7.21 (complex, 8H), 7.18 (d, J = 7 Hz, 2H), 5.02 (s, 2H), 4.82 (d, J = 7 Hz, 1H), 3.76 (br s, 1H), 2.97 (dd, J_1 = 4.5 Hz, J_2 = 14 Hz, 1H), 2.92-2.83 (complex, 2H), 2.76 (s, 2H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 155.76, 136.35, 129.30, 128.55, 128.43, 128.08, 127.94, 126.73, 66.69, 53.00, 46.70, 37.39. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{20}NO_3$, 298.1443; observed, 298.1448.

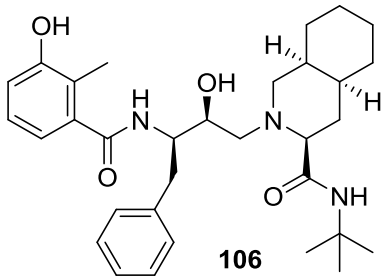


105

Benzyl ((2R,3S)-4-((3S,4aS,8aS)-3-(tert-butylcarbamoyl)octahydroisoquinolin-2(1H)-yl)-3-hydroxy-1-phenylbutan-2-yl)carbamate (105)

To a solution of **104** (483 mg, 1.62 mmol) in isopropyl alcohol (15 mL) was added (3S,4aS,8aS)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (**6**, 465 mg, 1.95 mmol). The solution was heated to reflux and stirred. After 5 h, the reaction was cooled and dried to a solid. The product was purified by silica flash column chromatography (gradient of 0-4% MeOH in DCM) to yield **105** as a white solid (799

mg, 92%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.29-7.08 (complex, 10H), 6.90 (s, 1H), 5.88 (d, *J* = 8 Hz, 1H), 4.99 (d, *J* = 12.4 Hz, 1H), 4.86 (d, *J* = 12.4 Hz, 1H), 4.39 (s, 1H), 3.87 (d, *J* = 6.4 Hz, 2H), 2.70 (m, 5H), 2.05 (d, *J* = 11.6 Hz, 2H), 1.90-1.15 (complex, 21H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.83, 156.10, 137.90, 136.26, 128.90, 127.97, 127.48, 127.40, 125.91, 69.50, 69.13, 65.95, 58.69, 57.91, 55.53, 50.39, 35.38, 34.06, 32.89, 30.72, 30.55, 28.26, 28.17, 25.97, 25.02, 19.89. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{32}H_{46}N_3O_4$, 536.3488; observed, 536.3480.

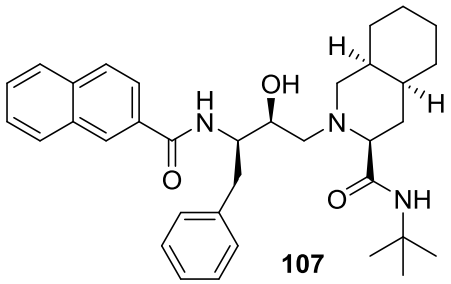


106

(3S,4aS,8aS)-N-(tert-butyl)-2-((2S,3R)-2-hydroxy-3-(3-hydroxy-2-methylbenzamido)-4-phenylbutyl)decahydroisoquinoline-3-carboxamide (106)

A solution of **105** (127 mg, 0.24 mmol) and 10% palladium on carbon (76 mg, 0.071 mmol) in methanol (15 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite and washed with excess methanol. The reaction was then dried to a solid and redissolved in THF (5 mL). 3-hydroxy-2-methylbenzoic acid (40 mg, 0.264 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (51 mg, 0.264 mmol), and hydroxybenzotriazole hydrate (43 mg, 0.264 mmol) were added to the reaction and the solution was stirred at room temperature for 18 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-4% MeOH in DCM) to yield **106** as a white solid (83 mg, 65%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.26 (t, *J* = 8.5 Hz, 2H), 7.20 (d, *J* = 7 Hz, 2H), 6.82 (t, *J* = 8 Hz, 1H), 6.75 (t, *J* = 7.5 Hz, 1H), 6.46 (s, 1H), 6.43 (d, *J* = 7.5 Hz, 1H), 6.19 (d, *J* = 9 Hz, 1H), 4.31 (m, 1H), 3.89

(d, J = 10 Hz, 1H), 2.93 (dd, J_1 = 4.5 Hz, J_2 = 14.5 Hz, 1H), 2.82 (m, 2H), 2.60 (dd, J_1 = 2.5 Hz, J_2 = 11 Hz, 1H), 2.55 (t, J = 11 Hz, 1H), 2.13 (d, J = 12 Hz, 2H), 1.95-1.19 (complex, 26H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 174.10, 170.69, 155.07, 137.73, 137.60, 129.17, 128.49, 126.55, 126.17, 122.34, 117.96, 116.46, 70.05, 69.42, 58.64, 58.21, 53.94, 51.12, 35.78, 34.86, 33.19, 30.91, 30.85, 28.60, 26.26, 25.34, 20.20, 12.19. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{32}H_{46}N_3O_4$, 536.3488; observed, 536.3491.

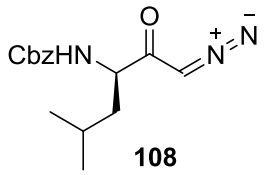


107

(3S,4aS,8aS)-2-((2S,3R)-3-(2-naphthamido)-2-hydroxy-4-phenylbutyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (107)

A solution of **105** (135 mg, 0.25 mmol) and 10% palladium on carbon (80 mg, 0.076 mmol) in methanol (15 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite and washed with excess methanol. The reaction was then dried to a solid and redissolved in THF (5 mL). 2-naphthoic acid (47 mg, 0.275 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (53 mg, 0.275 mmol), and hydroxybenzotriazole hydrate (45 mg, 0.275 mmol) were added to the reaction and the solution was stirred at room temperature for 18 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-3% MeOH in DCM) to yield **107** as a white solid (102 mg, 73%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.05 (s, 1H), 7.78 (m, 3H), 7.64 (dd, J_1 = 1 Hz, J_2 = 8 Hz, 1H), 7.50 (m, 2H), 7.27 (m, 4H), 7.20 (m, 1H), 6.79 (d, J = 8 Hz, 1H), 6.49 (s, 1H), 4.40 (m, 2H), 3.96 (d, J = 10 Hz, 1H), 3.02 (dd, J_1 = 9.5 Hz, J_2 = 14 Hz, 1H), 2.96 (dd, J_1 = 5 Hz, J_2 = 14 Hz, 1H), 2.82 (d, J = 10.5 Hz,

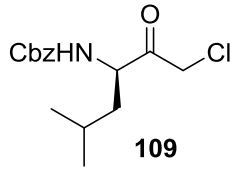
1H), 2.66 (dd, J_1 = 10.5 Hz, J_2 = 12.5 Hz, 1H), 2.61 (dd, J_1 = 3 Hz, J_2 = 11 Hz, 1H), 2.16 (d, J = 11.5 Hz, 1H), 2.10 (dd, J_1 = 3 Hz, J_2 = 12 Hz, 1H), 1.89 (q, J = 12 Hz, 1H), 1.78-1.12 (complex, 20H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.73, 168.14, 137.85, 134.63, 132.40, 131.39, 129.15, 128.82, 128.56, 128.26, 127.61, 127.57, 127.52, 126.60, 123.53, 69.82, 58.63, 58.48, 54.98, 50.79, 35.82, 35.34, 33.24, 31.06, 30.83, 28.59, 26.20, 25.38, 20.13. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{35}H_{46}N_3O_3$, 556.3539; observed, 556.3547.



108

(R)-benzyl (1-diazo-5-methyl-2-oxohexan-3-yl)carbamate (108)

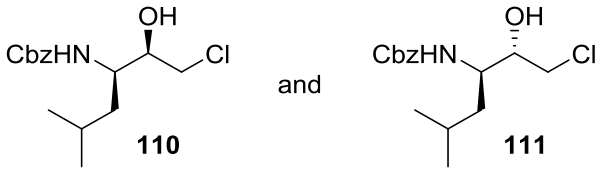
To a stirred solution of Z-D-leucine (923 mg, 3.48 mmol) in anhydrous tetrahydrofuran (12 mL) under nitrogen at -78 °C was added anhydrous triethylamine (723 μL, 5.22 mmol) followed by isobutyl chloroformate (679 μL, 5.22 mmol). The reaction was stirred cold for 1 h and then warmed to room temperature for 1.5 h. The reaction was cooled to -78 °C before an ethereal solution of diazomethane (25 mL, 8.1 mmol) was added in one portion and the reaction was then warmed to room temperature with stirring for 1 h. The reaction was evaporated to dryness, resulting in a yellow solid. The solid was taken up in ethyl acetate, washed twice with water, twice with a saturated solution of sodium bicarbonate, and once with brine. The organic phase was dried over sodium sulfate and concentrated by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 0-30% EtOAc in hexanes) to yield **108** as a yellow oil (898 mg, 89%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.28 (m, 5H), 6.01 (d, J = 8 Hz, 1H), 5.52 (s, 1H), 5.07 (dd, J_1 = 12 Hz, J_2 = 20 Hz, 2H), 4.35 (m, 1H), 1.68 (m, 1H), 1.50 (m, 2H), 0.91 (d, J = 6.5 Hz, 6H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 194.55, 155.85, 135.99, 128.09, 127.72, 127.57, 66.45, 56.31, 53.28, 40.63, 24.30, 22.68, 21.29. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{15}H_{20}N_3O_3$, 290.1505; observed, 290.1509.



109

(R)-benzyl (1-chloro-5-methyl-2-oxohexan-3-yl)carbamate (109)

To a solution of **108** (898 mg, 3.10 mmol) in anhydrous tetrahydrofuran (8 mL) under nitrogen at 0 °C was added 1 M HCl in diethyl ether (4.66 mL, 4.66 mmol) dropwise. The reaction was stirred at 0 °C for 30 min after complete addition. The solution was dried under reduced pressure to yield **109** as a yellow oil (945 mg). No further purification was made. ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.30 (m, 5H), 5.72 (d, J = 8 Hz, 1H), 5.06 (dd, J_1 = 12.5 Hz, J_2 = 17 Hz, 2H), 4.55 (m, 1H), 4.26 (dd, J_1 = 16 Hz, J_2 = 34 Hz, 2H), 1.67 (m, 1H), 1.56 (m, 1H), 1.42 (m, 1H), 0.92 (dd, J_1 = 7 Hz, J_2 = 12 Hz, 6H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 201.78, 156.01, 135.83, 128.26, 127.94, 127.71, 66.84, 56.14, 46.48, 39.69, 24.48, 22.89, 21.13. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{15}H_{21}NO_3Cl$, 298.1210; observed, 298.1210.

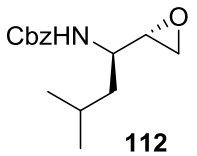


110 **111**

Benzyl ((2R,3R)-1-chloro-2-hydroxy-5-methylhexan-3-yl)carbamate (110) and benzyl ((2S,3R)-1-chloro-2-hydroxy-5-methylhexan-3-yl)carbamate (111)

To a solution of **109** (945 mg, 3.10 mmol) in dichloromethane (7 mL) and methanol (4 mL) under nitrogen at 0 °C was added sodium borohydride (82 mg, 2.17 mmol). The reaction was stirred at 0 °C for 1 h, and then quenched with slow addition of 2 M HCl (2 mL). The reaction was evaporated to dryness and the solid was taken up in ethyl acetate. The solution was washed twice with water followed by brine, dried over sodium sulfate, and concentrated by rotary evaporation. TLC analysis showed two diastereomers with the lower Rf compound being the (R,R)-isomer. Both diastereomers were purified by

silica flash column chromatography (gradient of 0-4% EtOAc in DCM) to yield the (R,R)-isomer (**110**) as a white solid (226 mg, 24%) and the (S,R)-isomer (**111**) as a white solid (37 mg, 4%) over two steps. Characterization of the (R,R)-isomer (**110**): ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.32 (m, 5H), 5.08 (s, 2H), 5.03 (d, J = 8.5 Hz, 1H), 3.81 (m, 2H), 3.58 (dd, J_1 = 3.5 Hz, J_2 = 11 Hz, 1H), 3.51 (dd, J_1 = 7.5 Hz, J_2 = 11 Hz, 1H), 3.19 (d, J = 4 Hz, 1H), 1.64 (m, 1H), 1.42 (m, 1H), 1.31 (m, 1H), 0.91 (t, J = 12 Hz, 6H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 156.52, 136.15, 128.48, 128.15, 127.99, 74.48, 66.94, 52.05, 46.84, 38.43, 24.52, 23.59, 21.35. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{15}H_{23}NO_3Cl$, 300.1366; observed, 300.1368. Characterization of the (S,R)-isomer (**111**): ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.34 (m, 5H), 5.09 (s, 2H), 4.99 (d, J = 9.5 Hz, 1H), 3.83 (m, 1H), 3.77 (m, 1H), 3.63 (dd, J_1 = 4 Hz, J_2 = 11 Hz, 1H), 3.50 (dd, J_1 = 8.5 Hz, J_2 = 11 Hz, 1H), 2.80 (d, J = 4 Hz, 1H), 1.61 (m, 2H), 1.36 (m, 1H), 0.93 (dd, J_1 = 3 Hz, J_2 = 6.5 Hz, 6H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 156.64, 136.28, 128.54, 128.19, 128.02, 73.09, 66.96, 51.05, 47.88, 41.59, 24.64, 23.00, 22.05. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{15}H_{23}NO_3Cl$, 300.1366; observed, 300.1369.

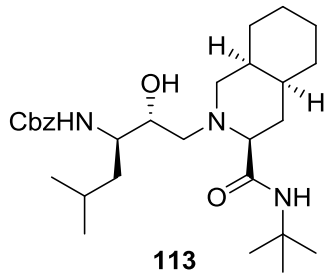


112

Benzyl ((R)-3-methyl-1-((S)-oxiran-2-yl)butyl)carbamate (112)

To a solution of **111** (37 mg, 0.123 mmol) in ethanol (5 mL) at 0 °C was added solid potassium hydroxide (8 mg, 0.136 mmol). The reaction was stirred cold for 15 min then warmed to room temperature and stirred for an additional 2 h. The reaction was evaporated to dryness and the solid was purified by silica flash column chromatography (gradient of 1-4% EtOAc in DCM) to yield **112** as a colorless oil (30 mg, 93%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.35 (m, 5H), 5.09 (dd, J_1 = 12 Hz, J_2 = 20.5 Hz, 2H), 4.58 (d, J = 8 Hz, 1H), 4.04 (m, 1H), 3.00 (s, 1H), 2.72 (m, 1H), 2.57 (s, 1H), 1.73 (m, 1H), 1.46 (m, 2H), 0.96 (s, 6H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 156.14, 136.38, 128.50, 128.12, 127.97,

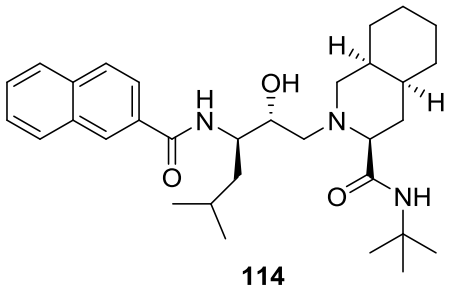
66.76, 53.80, 47.91, 44.36, 42.29, 24.62, 23.05, 22.03. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{15}H_{22}NO_3$, 264.1600; observed, 264.1598.



113

Benzyl ((2R,3R)-1-((3S,4aS,8aS)-3-(tert-butylcarbamoyl)octahydroisoquinolin-2(1H)-yl)-2-hydroxy-5-methylhexan-3-yl)carbamate (113)

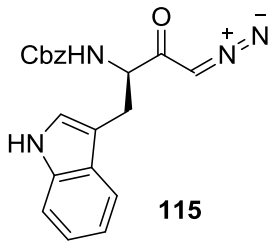
To a solution of **112** (30 mg, 0.114 mmol) in isopropyl alcohol (10 mL) was added (3S,4aS,8aS)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (**6**, 30 mg, 0.125 mmol). The solution was heated to reflux and stirred. After 5 h, the reaction was cooled and dried to a solid. The product was purified by silica flash column chromatography (gradient of 0-4% MeOH in DCM) to yield **113** as a white solid (51 mg, 89%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.32 (m, 5H), 5.89 (s, 1H), 5.35 (d, J = 9 Hz, 1H), 5.07 (dd, J_1 = 12.5 Hz, J_2 = 22 Hz, 2H), 3.72 (t, J = 6 Hz, 1H), 3.63 (m, 1H), 3.36 (br s, 1H), 2.80 (dd, J_1 = 2 Hz, J_2 = 12 Hz, 1H), 2.63 (dd, J_1 = 2.5 Hz, J_2 = 10 Hz, 1H), 2.56 (dd, J_1 = 5.5 Hz, J_2 = 13.5 Hz, 1H), 2.30 (dd, J_1 = 3 Hz, J_2 = 12 Hz, 1H), 2.23 (dd, J_1 = 8 Hz, J_2 = 13.5 Hz, 1H), 1.93 (q, J = 12 Hz, 1H), 1.82-1.20 (complex, 22H), 0.92 (t, J = 5.5 Hz, 6H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.29, 156.90, 136.62, 128.41, 127.95, 127.91, 70.60, 69.79, 66.55, 59.83, 59.31, 51.52, 50.92, 41.25, 35.60, 33.22, 30.51, 30.45, 28.63, 25.85, 24.76, 23.04, 22.22, 20.87. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{29}H_{48}N_3O_4$, 502.3645; observed, 502.3636.



114

(3S,4aS,8aS)-2-((2R,3R)-3-(2-naphthamido)-2-hydroxy-5-methylhexyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (114)

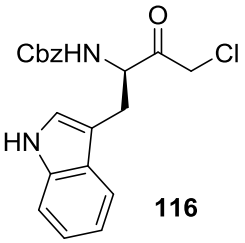
A solution of **113** (26 mg, 0.052 mmol) and 10% palladium on carbon (17 mg, 0.016 mmol) in methanol (5 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite and washed with excess methanol. The reaction was then dried to a solid and redissolved in THF (3 mL). 2-naphthoic acid (10 mg, 0.057 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (11 mg, 0.057 mmol), and hydroxybenzotriazole hydrate (9 mg, 0.057 mmol) were added to the reaction and the solution was stirred at room temperature for 20 h. The reaction was taken up in ethyl acetate and washed once with water, once with saturated sodium bicarbonate, and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-3% MeOH in DCM) to yield **114** as an off-white solid (13 mg, 48%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.47 (s, 1H), 7.98 (dd, J_1 = 1.5 Hz, J_2 = 8.5 Hz, 1H), 7.93 (d, J = 8 Hz, 1H), 7.86 (t, J = 9 Hz, 2H), 7.77 (d, J = 7.5 Hz, 1H), 7.53 (m, 2H), 5.60 (s, 1H), 4.23 (m, 1H), 3.88 (m, 1H), 2.76 (dd, J_1 = 2 Hz, J_2 = 11.5 Hz, 1H), 2.62 (dd, J_1 = 8 Hz, J_2 = 13 Hz, 1H), 2.54 (dd, J_1 = 5 Hz, J_2 = 11 Hz, 1H), 2.20 (m, 2H), 1.96 (m, 2H), 1.78-0.92 (complex, 29H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.61, 168.32, 134.79, 132.67, 131.82, 129.02, 128.05, 127.59, 127.41, 126.43, 124.32, 70.63, 70.38, 59.74, 58.91, 51.61, 51.18, 39.15, 35.82, 33.48, 30.89, 30.76, 28.55, 25.98, 25.58, 25.24, 23.25, 22.29, 20.30. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{32}H_{48}N_3O_3$, 522.3696; observed, 522.3694.



115

Benzyl (R)-(4-diazo-1-(1H-indol-3-yl)-3-oxobutan-2-yl)carbamate (115)

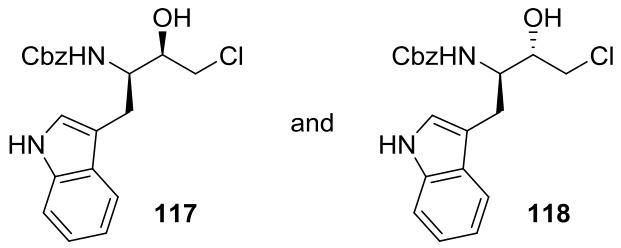
To a stirred solution of Z-D-tryptophan (995 mg, 2.94 mmol) in anhydrous tetrahydrofuran (12 mL) under nitrogen at -78 °C was added anhydrous triethylamine (611 μL, 4.41 mmol) followed by isobutyl chloroformate (574 μL, 4.41 mmol). The reaction was stirred cold for 1 h and then warmed to room temperature for 1.5 h. The reaction was cooled to -78 °C before an ethereal solution of diazomethane (25 mL, 8.1 mmol) was added in one portion and the reaction was then warmed to room temperature with stirring for 1 h. The reaction was evaporated to dryness, resulting in a yellow solid. The solid was taken up in ethyl acetate, washed twice with water, twice with a saturated solution of sodium bicarbonate, and once with brine. The organic phase was dried over sodium sulfate and concentrated by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 0-40% EtOAc in hexanes) to yield **115** as a yellowish solid (950 mg, 89%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.69 (s, 1H), 7.53 (d, *J* = 7.5 Hz, 1H), 7.24 (m, 6H), 7.12 (t, *J* = 7.5 Hz, 1H), 7.03 (t, *J* = 7.5 Hz, 1H), 6.81 (s, 1H), 5.76 (d, *J* = 7.5 Hz, 1H), 5.02 (m, 3H), 4.51 (d, *J* = 5.5 Hz, 1H), 3.12 (d, *J* = 6.5 Hz, 2H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 193.93, 155.79, 135.99, 135.92, 128.24, 127.91, 127.73, 127.08, 123.13, 121.80, 119.27, 118.26, 111.22, 109.33, 66.68, 58.16, 54.18, 27.94. Ethyl acetate peaks are visible in the NMR. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{20}H_{19}N_4O_3$, 363.1457; observed, 363.1455.



116

(R)-benzyl (4-chloro-1-(1H-indol-3-yl)-3-oxobutan-2-yl)carbamate (116)

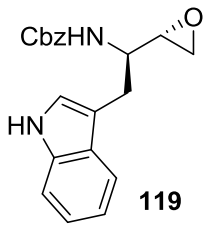
To a solution of **115** (950 mg, 2.62 mmol) in anhydrous tetrahydrofuran (8 mL) under nitrogen at 0 °C was added 1 M HCl in diethyl ether (3.93 mL, 3.93 mmol) dropwise. The reaction was stirred at 0 °C for 30 min after complete addition. The solution was dried under reduced pressure to yield **116** as an orange solid (946 mg). No further purification was made. ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.14 (s, 1H), 7.59 (d, *J* = 8 Hz, 1H), 7.33 (m, 6H), 7.22 (t, *J* = 7.5 Hz, 1H), 7.14 (t, *J* = 7.5 Hz, 1H), 6.95 (s, 1H), 5.43 (d, *J* = 7 Hz, 1H), 5.09 (s, 2H), 4.85 (q, *J* = 7 Hz, 1H), 4.09 (d, *J* = 16 Hz, 1H), 3.95 (d, *J* = 16 Hz, 1H), 3.24 (m, 2H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 201.62, 155.87, 136.19, 128.57, 128.31, 128.15, 122.95, 122.62, 120.08, 118.51, 111.42, 109.31, 67.22, 58.08, 47.67, 27.80. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{20}H_{20}N_2O_3Cl$, 371.1162; observed, 371.1169.



Benzyl ((2R,3R)-4-chloro-3-hydroxy-1-(1H-indol-3-yl)butan-2-yl)carbamate (117) and benzyl ((2R,3S)-4-chloro-3-hydroxy-1-(1H-indol-3-yl)butan-2-yl)carbamate (118)

To a solution of **116** (946 mg, 2.62 mmol) in dichloromethane (7 mL) and methanol (4 mL) under nitrogen at 0 °C was added sodium borohydride (69 mg, 1.83 mmol). The reaction was stirred at 0 °C for 1 h, and then quenched with slow addition of 2 M HCl (2 mL). The reaction was evaporated to dryness and the solid was taken up in ethyl acetate. The solution was washed twice with water followed by brine, dried over sodium sulfate, and concentrated by rotary evaporation. TLC analysis showed two diastereomers with the lower Rf compound being the (R,R)-isomer. Both diastereomers were purified by consecutive silica flash column chromatography (a gradient of 0-50% EtOAc in hexanes followed by a gradient of 1-8% EtOAc in DCM) to yield (R,R)-isomer (**117**) as a white solid (605 mg, 62%) and (R,S)-

isomer (**118**) as a white solid (80 mg, 8%) over two steps. Characterization of the (R,R)-isomer (**117**): ^{1}H NMR (500 MHz, $(CD_3)_2SO$) δ ppm 10.75 (s, 1H), 7.51 (d, *J* = 8 Hz, 1H), 7.30 (m, 3H), 7.21 (d, *J* = 7 Hz, 2H), 7.14 (d, *J* = 8.5 Hz, 1H), 7.07 (d, *J* = 1.5 Hz, 1H), 7.04 (t, *J* = 7.5 Hz, 1H), 6.93 (t, *J* = 7.5 Hz, 1H), 5.44 (d, *J* = 5.5 Hz, 1H), 4.92 (dd, J_1 = 12.5 Hz, J_2 = 19 Hz, 2H), 3.69 (m, 3H), 3.52 (dd, J_1 = 7 Hz, J_2 = 11 Hz, 1H), 3.01 (dd, J_1 = 3 Hz, J_2 = 15 Hz, 1H), 2.76 (dd, J_1 = 9 Hz, J_2 = 15 Hz, 1H). ^{13}C NMR (500 MHz, $(CD_3)_2SO$) δ ppm 155.76, 137.25, 136.09, 128.29, 127.69, 127.60, 127.36, 123.29, 120.70, 118.37, 118.11, 111.22, 110.98, 72.41, 64.95, 54.31, 47.99, 25.18. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{20}H_{22}N_2O_3Cl$, 373.1319; observed, 373.1313. Characterization of the (R,S)-isomer (**118**): ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.06 (s, 1H), 7.69 (d, *J* = 8 Hz, 1H), 7.32 (m, 5H), 7.19 (t, *J* = 7.5 Hz, 1H), 7.11 (t, *J* = 7.5 Hz, 1H), 6.96 (d, *J* = 1 Hz, 1H), 5.32 (d, *J* = 9 Hz, 1H), 5.09 (s, 2H), 4.05 (q, *J* = 8 Hz, 1H), 3.80 (m, 1H), 3.48 (m, 2H), 3.08 (m, 2H), 2.99 (d, *J* = 4 Hz, 1H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 156.66, 136.22, 136.16, 128.54, 128.19, 128.04, 127.42, 122.77, 122.15, 119.61, 118.74, 111.55, 111.18, 71.15, 66.94, 53.42, 47.91, 28.17. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{20}H_{22}N_2O_3Cl$, 373.1319; observed, 373.1335.

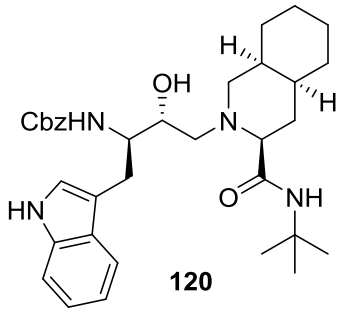


119

Benzyl ((R)-2-(1H-indol-3-yl)-1-((S)-oxiran-2-yl)ethyl)carbamate (119)

To a solution of **118** (80 mg, 0.215 mmol) in ethanol (7 mL) at 0 °C was added solid potassium hydroxide (13 mg, 0.236 mmol). The reaction was stirred cold for 15 min then warmed to room temperature and stirred for an additional 2 h. The reaction was evaporated to dryness and the solid was purified by silica flash column chromatography (gradient of 2-4% EtOAc in DCM) to yield **119** as a white solid (60 mg, 83%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.09 (s, 1H), 7.69 (d, *J* = 7 Hz, 1H), 7.34 (m, 6H), 7.20 (t, *J* = 7 Hz, 1H), 7.13 (m, 1H), 7.05 (s, 1H), 5.08 (s, 2H), 4.86 (d, *J* = 8 Hz, 1H), 4.29 (m,

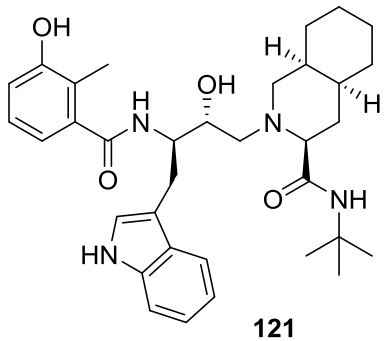
1H), 3.09 (m, 4H), 2.66 (t, J = 4 Hz, 1H), 2.58 (s, 1H). Residual ethyl acetate is visible. ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 156.12, 136.18, 128.51, 128.12, 127.99, 122.76, 122.15, 119.62, 118.83, 111.25, 111.14, 66.76, 52.51, 50.17, 44.46, 29.35. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{20}H_{21}N_2O_3$, 337.1552; observed, 337.1547.



120

benzyl ((2R,3R)-4-((3S,4aS,8aS)-3-(tert-butylcarbamoyl)octahydroisoquinolin-2(1H)-yl)-3-hydroxy-1-(1H-indol-3-yl)butan-2-yl)carbamate (120)

To a solution of **119** (60 mg, 0.178 mmol) in isopropyl alcohol (10 mL) was added (3S,4aS,8aS)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (**6**, 47 mg, 0.196 mmol). The solution was heated to reflux and stirred. After 6 h, the reaction was cooled and dried to a solid. The product was purified by silica flash column chromatography (gradient of 0-4% MeOH in DCM) to yield **120** as a white solid (88 mg, 86%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.29 (s, 1H), 7.76 (d, J = 8 Hz, 1H), 7.33 (m, 6H), 7.14 (t, J = 7.5 Hz, 1H), 7.08 (t, J = 7.5 Hz, 1H), 6.99 (s, 1H), 5.86 (s, 1H), 5.65 (d, J = 9 Hz, 1H), 5.10 (dd, J_1 = 12 Hz, J_2 = 21 Hz, 2H), 3.89 (m, 1H), 3.82 (t, J = 7 Hz, 1H), 3.63 (br s, 1H), 3.16 (dd, J_1 = 5.5 Hz, J_2 = 14.5 Hz, 1H), 3.07 (dd, J_1 = 9 Hz, J_2 = 14 Hz, 1H), 2.50 (m, 3H), 2.15 (dd, J_1 = 7.5 Hz, J_2 = 13 Hz, 1H), 2.07 (dd, J_1 = 3 Hz, J_2 = 11 Hz, 1H), 1.79 (q, J = 12 Hz, 1H), 1.60 (m, 3H), 1.44-0.97 (complex, 19H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.60, 156.84, 136.49, 136.27, 128.43, 127.99, 127.59, 122.68, 121.85, 119.36, 119.05, 112.23, 111.03, 70.20, 68.15, 66.59, 59.55, 59.11, 53.40, 50.65, 35.55, 33.01, 30.59, 30.48, 28.31, 27.90, 25.92, 25.43, 20.62. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{34}H_{47}N_4O_4$, 575.3597; observed, 575.3589.

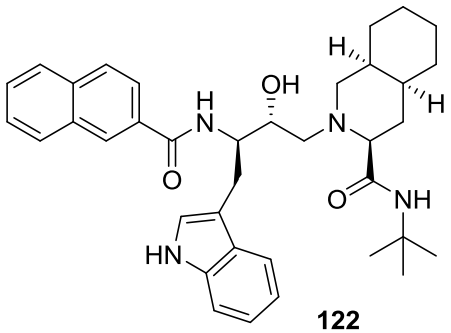


121

(3S,4aS,8aS)-N-(tert-butyl)-2-((2R,3R)-2-hydroxy-3-(3-hydroxy-2-methylbenzamido)-4-(1H-indol-3-yl)butyl)decahydroisoquinoline-3-carboxamide (121)

A solution of **120** (34 mg, 0.059 mmol) and 10% palladium on carbon (19 mg, 0.018 mmol) in methanol (5 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite and washed with excess methanol. The reaction was then dried to a solid and redissolved in THF (3 mL). 3-hydroxy-2-methylbenzoic acid (10 mg, 0.065 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (12 mg, 0.065 mmol), and hydroxybenzotriazole hydrate (11 mg, 0.065 mmol) were added to the reaction and the solution was stirred at room temperature for 20 h. The reaction was taken up in ethyl acetate and washed once with water, once with saturated sodium bicarbonate, and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-7% MeOH in DCM) to yield **121** as an off-white solid (21 mg, 62%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.19 (s, 1H), 7.73 (d, J = 8 Hz, 1H), 7.34 (d, J = 8 Hz, 1H), 7.17 (t, J = 8 Hz, 1H), 7.10 (t, J = 7.5 Hz, 1H), 7.06 (d, J = 1.5 Hz, 1H), 6.94 (t, J = 8 Hz, 1H), 6.77 (m, 3H), 6.41 (br s, 1H), 5.75 (s, 1H), 4.28 (q, J = 7.5 Hz, 1H), 3.89 (t, J = 6.5 Hz, 1H), 3.38 (dd, J_1 = 7.5 Hz, J_2 = 14.5 Hz, 1H), 3.19 (dd, J_1 = 8 Hz, J_2 = 14.5 Hz, 1H), 2.72 (dd, J_1 = 2 Hz, J_2 = 11.5 Hz, 1H), 2.55 (dd, J_1 = 3 Hz, J_2 = 11 Hz, 1H), 2.46 (dd, J_1 = 5.5 Hz, J_2 = 13 Hz, 1H), 2.27 (m, 2H), 2.13 (s, 3H), 1.88 (q, J = 11.5 Hz, 1H), 1.76-1.05 (complex, 21H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.66, 171.25, 154.55, 138.07, 136.32, 127.62, 126.46, 122.73, 122.22, 121.96, 119.42, 119.15, 118.97, 116.60, 112.42,

111.08, 70.15, 69.81, 60.29, 59.51, 53.28, 50.96, 35.82, 33.17, 30.73, 30.59, 28.33, 27.26, 26.03, 25.68, 20.60, 12.39. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{34}H_{47}N_4O_4$, 575.3597; observed, 575.3591.

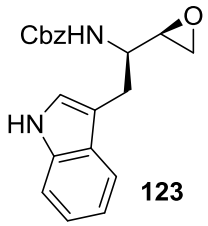


122

(3S,4aS,8aS)-2-((2R,3R)-3-(2-naphthamido)-2-hydroxy-4-(1H-indol-3-yl)butyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (122)

A solution of **120** (39 mg, 0.068 mmol) and 10% palladium on carbon (22 mg, 0.020 mmol) in methanol (5 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite and washed with excess methanol. The reaction was then dried to a solid and redissolved in THF (3 mL). 2-naphthoic acid (13 mg, 0.075 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (14 mg, 0.075 mmol), and hydroxybenzotriazole hydrate (12 mg, 0.075 mmol) were added to the reaction and the solution was stirred at room temperature for 20 h. The reaction was taken up in ethyl acetate and washed once with water, once with saturated sodium bicarbonate, and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-5% MeOH in DCM) to yield **122** as an off-white solid (29 mg, 72%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.35 (s, 1H), 8.18 (s, 1H), 7.90 (dd, J_1 = 1.5 Hz, J_2 = 8.5 Hz, 1H), 7.87 (d, J = 8 Hz, 1H), 7.82 (m, 4H), 7.51 (m, 2H), 7.33 (d, J = 8 Hz, 1H), 7.16 (t, J = 7.5 Hz, 1H), 7.10 (m, 2H), 5.64 (s, 1H), 4.46 (q, J = 7.5 Hz, 1H), 3.99 (t, J = 6 Hz, 1H), 3.45 (dd, J_1 = 7.5 Hz, J_2 = 14 Hz, 1H), 3.32 (dd, J_1 = 7.5 Hz, J_2 = 14 Hz, 1H), 2.60 (m, 2H), 2.48 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 2.21 (dd, J_1 = 6 Hz, J_2 = 13 Hz, 1H), 2.08 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 1.85 (q, J = 11.5 Hz, 1H), 1.62-

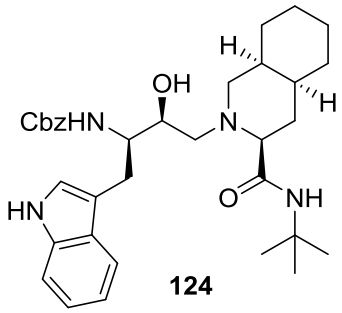
0.94 (complex, 21H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.63, 168.64, 136.29, 134.76, 132.59, 131.75, 129.01, 128.09, 127.67, 127.57, 127.46, 126.46, 124.20, 122.72, 121.83, 119.31, 119.06, 122.89, 111.06, 70.37, 69.49, 59.53, 58.86, 54.10, 50.95, 35.71, 33.30, 30.83, 30.70, 28.31, 26.12, 26.03, 25.51, 20.28. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{37}H_{47}N_4O_3$, 595.3648; observed, 595.3655.



123

Benzyl ((R)-2-(1H-indol-3-yl)-1-((R)-oxiran-2-yl)ethyl)carbamate (123)

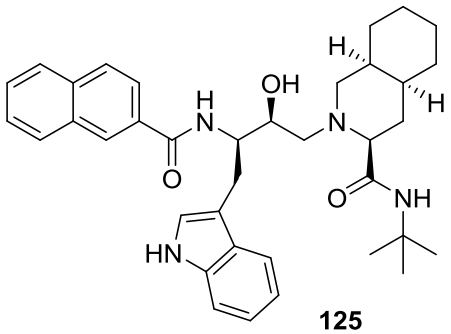
To a solution of **117** (589 mg, 1.58 mmol) in ethanol (15 mL) and dichloromethane (5 mL) at 0 °C was added solid potassium hydroxide (97 mg, 1.74 mmol). The reaction was stirred cold for 15 min then warmed to room temperature and stirred for an additional 3 h. The reaction was evaporated to dryness and the solid was purified by silica flash column chromatography (gradient of 2-5% EtOAc in DCM) to yield **123** as a white solid (470 mg, 88%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.25 (s, 1H), 7.61 (d, *J* = 8 Hz, 1H), 7.29 (m, 4H), 7.24 (m, 2H), 7.17 (t, *J* = 7.5 Hz, 1H), 7.09 (t, *J* = 7.5 Hz, 1H), 6.91 (s, 1H), 5.02 (m, 2H), 4.93 (d, *J* = 6 Hz, 1H), 3.80 (m, 1H), 3.06 (d, *J* = 3.5 Hz, 2H), 2.90 (s, 1H), 2.74 (d, *J* = 12.5 Hz, 2H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 155.92, 136.17, 128.42, 128.06, 127.94, 127.63, 122.98, 122.01, 119.58, 118.66, 111.17, 110.06, 66.68, 53.11, 52.59, 46.89, 26.94. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{20}H_{21}N_2O_3$, 337.1552; observed, 337.1547.



124

Benzyl ((2R,3S)-4-((3S,4aS,8aS)-3-(tert-butylcarbamoyl)octahydroisoquinolin-2(1H)-yl)-3-hydroxy-1-(1H-indol-3-yl)butan-2-yl)carbamate (124)

To a solution of **123** (456 mg, 1.36 mmol) in isopropyl alcohol (10 mL) was added (3S,4aS,8aS)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (**6**, 355 mg, 1.49 mmol). The solution was heated to reflux and stirred. After 6 h, the reaction was cooled and dried to a solid. The product was purified by silica flash column chromatography (gradient of 0-4% MeOH in DCM) to yield **124** as a white solid (691 mg, 88%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.95 (s, 1H), 7.56 (d, *J* = 8 Hz, 1H), 7.19 (m, 6H), 7.04 (t, *J* = 7.5 Hz, 1H), 6.90 (s, 1H), 6.73 (s, 1H), 5.45 (d, *J* = 8 Hz, 1H), 4.95 (m, 2H), 4.12 (s, 1H), 3.99 (m, 1H), 3.83 (m, 1H), 2.95 (m, 2H), 2.63 (m, 3H), 2.03 (d, *J* = 12 Hz, 1H), 1.97 (d, *J* = 9 Hz, 1H), 1.84 (q, *J* = 12 Hz, 1H), 1.70 (m, 3H), 1.55-1.13 (complex, 18H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.99, 156.60, 136.15, 128.18, 127.73, 127.61, 127.45, 122.68, 121.46, 118.87, 118.32, 111.20, 110.78, 69.42, 69.23, 66.37, 58.49, 57.88, 54.92, 50.59, 35.51, 32.96, 30.83, 30.61, 28.36, 26.05, 25.30, 25.14, 20.00. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{34}H_{47}N_4O_4$, 575.3597; observed, 575.3588.

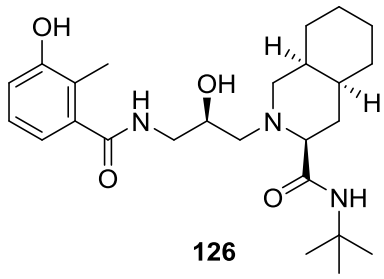


125

(3S,4aS,8aS)-2-((2S,3R)-3-(2-naphthamido)-2-hydroxy-4-(1H-indol-3-yl)butyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (125)

A solution of **124** (126 mg, 0.219 mmol) and 10% palladium on carbon (70 mg, 0.066 mmol) in methanol (5 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite and washed with excess methanol. The reaction was then dried to a solid and redissolved in THF (3 mL).

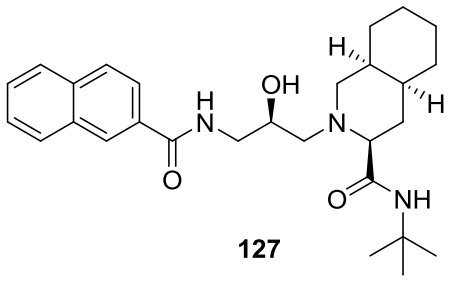
2-naphthoic acid (41 mg, 0.241 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (46 mg, 0.241 mmol), and hydroxybenzotriazole hydrate (39 mg, 0.241 mmol) were added to the reaction and the solution was stirred at room temperature for 20 h. The reaction was taken up in ethyl acetate and washed once with water, once with saturated sodium bicarbonate, and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-5% MeOH in DCM) to yield **125** as an off-white solid (68 mg, 52%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.67 (s, 1H), 7.79 (s, 1H), 7.76 (d, J = 8.5 Hz, 1H), 7.72 (d, J = 8.5 Hz, 1H), 7.66 (d, J = 8 Hz, 2H), 7.59 (dd, J_1 = 1.5 Hz, J_2 = 8.5 Hz, 1H), 7.50 (t, J = 7 Hz, 1H), 7.44 (t, J = 6.5 Hz, 1H), 7.38 (d, J = 8.5 Hz, 1H), 7.21 (t, J = 7.5 Hz, 1H), 7.14 (t, J = 7.5 Hz, 1H), 7.06 (d, J = 2 Hz, 1H), 6.70 (d, J = 7.5 Hz, 1H), 6.55 (s, 1H), 4.46 (m, 1H), 3.95 (d, J = 10 Hz, 1H), 3.22 (dd, J_1 = 9.5 Hz, J_2 = 15 Hz, 1H), 3.06 (dd, J_1 = 5 Hz, J_2 = 15 Hz, 1H), 2.76 (d, J = 10.5 Hz, 1H), 2.68 (dd, J_1 = 10.5 Hz, J_2 = 13 Hz, 1H), 2.56 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 2.11 (d, J = 12 Hz, 1H), 1.93 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 1.87 (q, J = 12 Hz, 1H), 1.76 (dq, J_1 = 3 Hz, J_2 = 13 Hz, 1H), 1.65 (m, 2H), 1.53-1.14 (complex, 17H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.94, 168.50, 136.29, 134.61, 132.37, 131.23, 128.89, 128.30, 127.83, 127.60, 127.56, 127.43, 126.60, 123.61, 122.68, 122.10, 119.59, 118.37, 111.49, 111.25, 70.02, 69.85, 58.87, 58.26, 54.75, 50.83, 35.81, 33.19, 31.07, 30.83, 28.60, 26.26, 25.42, 25.19, 20.17. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{37}H_{47}N_4O_3$, 595.3648; observed, 595.3636.



126

(3S,4aS,8aS)-N-(tert-butyl)-2-(2-hydroxy-3-(3-hydroxy-2-methylbenzamido)propyl)decahydroisoquinoline-3-carboxamide (126)

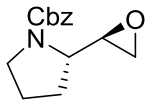
A solution of **35** (32 mg, 0.072 mmol) and 10% palladium on carbon (23 mg, 0.022 mmol) in methanol (10 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite and washed with excess methanol. The reaction was then dried to a solid and redissolved in THF (3 mL). 3-hydroxy-2-methylbenzoic acid (12 mg, 0.079 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (15 mg, 0.079 mmol), and hydroxybenzotriazole hydrate (12 mg, 0.079 mmol) were added to the reaction and the solution was stirred at room temperature for 18 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-4% MeOH in DCM) to yield **126** as a white solid (17 mg, 53%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.53 (br s, 1H), 6.95 (t, J = 8 Hz, 1H), 6.82 (d, J = 8 Hz, 1H), 6.47 (t, J = 6 Hz, 1H), 6.12 (s, 1H), 4.28 (br s, 1H), 3.91 (m, 1H), 3.63 (dq, J_1 = 3 Hz, J_2 = 6 Hz, 1H), 3.15 (m, 1H), 2.85 (d, J = 10 Hz, 1H), 2.64 (dd, J_1 = 3 Hz, J_2 = 11 Hz, 1H), 2.43 (dd, J_1 = 11 Hz, J_2 = 12.5 Hz, 1H), 2.20 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 2.11 (m, 4H), 1.95 (q, J = 12 Hz, 1H), 1.88-1.19 (complex, 21H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.82, 170.96, 154.89, 137.81, 126.41, 122.31, 118.50, 116.54, 69.58, 66.83, 59.01, 58.27, 51.15, 43.80, 35.82, 33.31, 30.96, 30.89, 28.62, 26.29, 25.44, 20.24, 12.41. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{25}H_{40}N_3O_4$, 446.3019; observed, 446.3023.



127

(3S,4aS,8aS)-2-(3-(2-naphthamido)-2-hydroxypropyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (127)

A solution of **35** (30 mg, 0.067 mmol) and 10% palladium on carbon (21 mg, 0.020 mmol) in methanol (10 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite and washed with excess methanol. The reaction was then dried to a solid and redissolved in THF (3 mL). 2-naphthoic acid (13 mg, 0.074 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (14 mg, 0.074 mmol), and hydroxybenzotriazole hydrate (12 mg, 0.074 mmol) were added to the reaction and the solution was stirred at room temperature for 18 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-3% MeOH in DCM) to yield **127** as a white solid (28 mg, 90%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.33 (s, 1H), 7.87 (m, 4H), 7.54 (m, 2H), 7.09 (m, 1H), 5.99 (s, 1H), 3.98 (m, 1H), 3.72 (ddd, J_1 = 2.5 Hz, J_2 = 6 Hz, J_3, = 14Hz, 1H), 3.37 (dt, J_1 = 6 Hz, J_2 = 13.5 Hz, 1H), 2.85 (dd, J_1 = 2 Hz, J_2 = 11.5 Hz, 1H), 2.63 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 2.51 (dd, J_1 = 10.5 Hz, J_2 = 13 Hz, 1H), 2.20 (dd, J_1 = 3 Hz, J_2 = 12 Hz, 1H), 2.13 (dd, J_1 = 2.5 Hz, J_2 = 13 Hz, 1H), 1.95 (q, J = 11.5 Hz, 1H), 1.87-1.18 (complex, 21H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.42, 167.96, 134.74, 132.58, 131.36, 128.93, 128.31, 127.66, 127.57, 126.63, 123.70, 69.71, 66.63, 58.86, 58.31, 50.93, 43.67, 35.86, 33.32, 31.00, 30.91, 28.56, 26.27, 25.47, 25.34, 20.22. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{28}H_{40}N_3O_3$, 466.3070; observed, 466.3063.

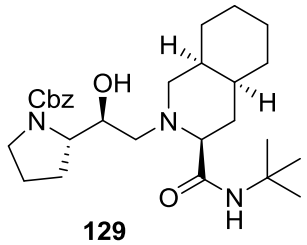


128

(S)-benzyl 2-((R)-oxiran-2-yl)pyrrolidine-1-carboxylate (128)

To a solution of **39** (429 mg, 1.51 mmol) in ethanol (15 mL) at 0 °C was added solid potassium hydroxide (93 mg, 1.66 mmol). The reaction was stirred cold for 15 min then warmed to room temperature and stirred for an additional 3 h. The reaction was evaporated to dryness and the solid was

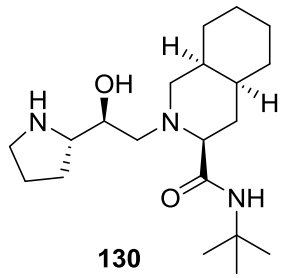
purified by silica flash column chromatography (gradient of 0.5-5.5% EtOAc in DCM) to yield **128** as a colorless oil (342 mg, 91%). ^{1}H NMR (500 MHz, $CDCl_3$, mixture of cis and trans isomers) δ ppm 7.38-7.26 (complex, 10H), 5.19-5.08 (complex, 4H), 4.21 (br s, 1H), 4.10 (br s, 1H), 3.48-3.36 (complex, 4H), 3.04 (br s, 1H), 2.99 (br s, 1H), 2.65 (br s, 1H), 2.59 (br s, 1H), 2.48 (br s, 1H), 2.43 (br s, 1H), 2.00-1.79 (complex, 8H). ^{13}C NMR (500 MHz, $CDCl_3$, mixture of cis and trans isomers) δ ppm 155.03, 154.67, 136.53, 136.47, 128.11, 127.95, 127.60, 127.45, 66.44, 56.04, 55.90, 53.56, 53.45, 46.89, 46.59, 44.00, 43.92, 28.55, 28.04, 23.97, 23.19. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{14}H_{18}NO_3$, 248.1287; observed, 248.1295.



129

(S)-benzyl 2-((S)-2-((3S,4aS,8aS)-3-(tert-butylcarbamoyl)octahydroisoquinolin-2(1H)-yl)-1-hydroxyethyl)pyrrolidine-1-carboxylate (129)

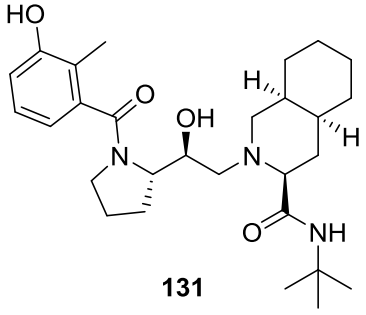
To a solution of **128** (342 mg, 1.38 mmol) in isopropyl alcohol (15 mL) was added (3S,4aS,8aS)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (**6**, 396 mg, 1.66 mmol). The solution was heated to reflux and stirred. After 5 h, the reaction was cooled and dried to a solid. The product was purified by silica flash column chromatography (gradient of 0-8% MeOH in DCM) to yield **129** as a white solid (622 mg, 93%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.35-7.30 (complex, 6H), 7.00 (br s, 1H), 5.12 (m, 2H), 4.53 (br s, 1H), 3.82-3.78 (m, 2H), 3.57 (m, 1H), 3.34 (m, 1H), 2.85 (d, *J* = 10.5 Hz, 1H), 2.62-2.50 (complex, 3H), 2.12-1.16 (complex, 25H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.85, 157.18, 136.23, 128.31, 127.88, 127.67, 71.02, 69.61, 67.12, 60.99, 59.66, 58.06, 50.32, 46.96, 35.74, 33.11, 30.90, 30.73, 28.39, 27.85, 26.17, 25.30, 24.11, 20.09. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{28}H_{44}N_3O_4$, 486.3332; observed, 486.3343.



130

(3S,4aS,8aS)-N-(tert-butyl)-2-((S)-2-hydroxy-2-((S)-pyrrolidin-2-yl)ethyl)decahydroisoquinoline-3-carboxamide (130)

A solution of **129** (622 mg, 1.28 mmol) and 10% palladium on carbon (405 mg, 0.38 mmol) in methanol (30 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite, washed with excess methanol, and dried to a solid. The solid was taken up in ethyl acetate and extracted with 1 M HCl three times. The aqueous fractions were combined, adjusted to pH 12 with 10 M NaOH, and extracted with ethyl acetate three times. The combined ethyl acetate fractions were washed with brine, dried over sodium sulfate, and concentrated by rotary evaporation to yield **130** as a white solid (435 mg). The product was used without further purification.

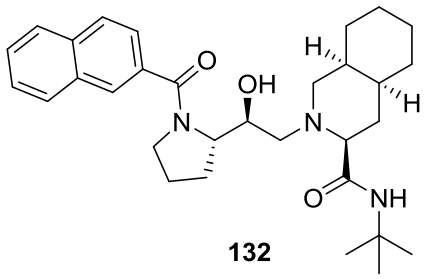


131

(3S,4aS,8aS)-N-(tert-butyl)-2-((S)-2-hydroxy-2-((S)-1-(3-hydroxy-2-methylbenzoyl)pyrrolidin-2-yl)ethyl)decahydroisoquinoline-3-carboxamide (131)

To a solution of **130** (31 mg, 0.088 mmol) in THF (3 mL) was added 3-hydroxy-2-methylbenzoic acid (15 mg, 0.097 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (19 mg, 0.097

mmol), and hydroxybenzotriazole hydrate (16 mg, 0.097 mmol). The solution was stirred at room temperature for 16 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-7% MeOH in DCM) to yield **131** as a white solid (33 mg, 74% over two steps). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.96 (br s, 1H), 7.35 (br s, 1H), 7.03 (t, J = 8 Hz, 1H), 6.84 (d, J = 8 Hz, 1H), 6.70 (d, J = 7 Hz, 1H), 5.43 (br s, 1H), 4.27 (dd, J_1 = 8 Hz, J_2 = 14 Hz, 1H), 3.88 (t, J = 9 Hz, 1H), 3.28 (m, 1H), 3.20 (m, 1H), 2.91 (d, J = 10.5 Hz, 1H), 2.62 (m, 2H), 2.20-2.05 (complex, 6H), 1.90-1.17 (complex, 24H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 174.69, 173.67, 155.37, 137.96, 126.97, 115.82, 72.32, 69.66, 61.10, 60.95, 58.54, 50.77, 36.00, 33.25, 31.14, 30.91, 28.64, 28.44, 26.32, 25.60, 24.78, 20.26, 12.66. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{28}H_{44}N_3O_4$, 486.3332; observed, 486.3335.

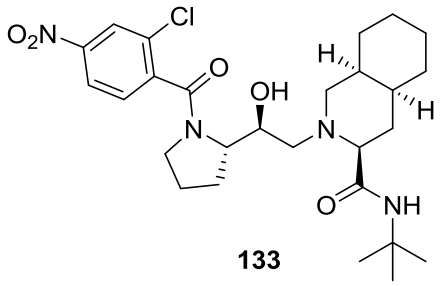


132

(3S,4aS,8aS)-2-((S)-2-((S)-1-(2-naphthoyl)pyrrolidin-2-yl)-2-hydroxyethyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (132)

To a solution of **130** (33 mg, 0.094 mmol) in THF (3 mL) was added 2-naphthoic acid (18 mg, 0.103 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (20 mg, 0.103 mmol), and hydroxybenzotriazole hydrate (17 mg, 0.103 mmol). The solution was stirred at room temperature for 16 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-5% MeOH in DCM) to yield **132** as a white solid (44 mg, 90% over two steps). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.01 (s, 1H),

7.86 (m, 3H), 7.55 (m, 3H), 7.03 (br s, 1H), 5.00 (br s, 1H), 4.41 (q, *J* = 7.5 Hz, 1H), 3.97 (t, *J* = 8.5 Hz, 1H), 3.58 (m, 1H), 2.92 (d, *J* = 11 Hz, 1H), 2.63 (m, 2H), 2.15 (m, 3H), 1.94-1.17 (complex, 24H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 174.06, 173.00, 133.93, 133.60, 132.45, 128.50, 128.15, 127.74, 127.29, 126.70, 124.18, 72.23, 69.82, 60.96, 60.35, 58.46, 51.22, 50.55, 36.01, 33.32, 31.16, 30.93, 28.64, 28.12, 26.33, 25.54, 25.26, 20.24. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{31}H_{44}N_3O_3$, 506.3383; observed, 506.3388.

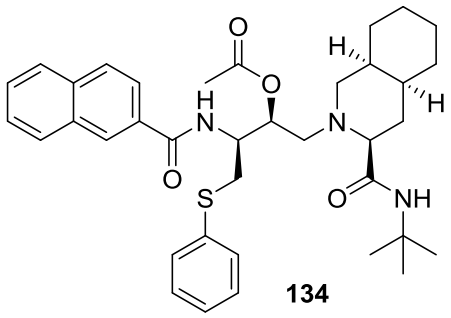


133

(3S,4aS,8aS)-N-(tert-butyl)-2-((S)-2-((S)-1-(2-chloro-4-nitrobenzoyl)pyrrolidin-2-yl)-2-hydroxyethyl)decahydroisoquinoline-3-carboxamide (133)

To a solution of **130** (218 mg, 0.620 mmol) in THF (6 mL) was added 2-chloro-4-nitrobenzoic acid (138 mg, 0.682 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (131 mg, 0.682 mmol), and hydroxybenzotriazole hydrate (111 mg, 0.682 mmol). The solution was stirred at room temperature for 18 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-5% MeOH in DCM) to yield **133** as a yellow solid (248 mg, 72% over two steps). Crystals were grown for x-ray crystallography from the layering of pentane on top of a dichloromethane solution. ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.31 (d, *J* = 2 Hz, 1H), 8.20 (dd, J_1 = 2 Hz, J_2 = 8.5 Hz, 1H), 7.54 (d, *J* = 8.5 Hz, 1H), 6.87 (s, 1H), 5.31 (s, 1H), 4.36 (m, 1H), 3.99 (m, 1H), 3.24 (m, 2H), 2.91 (d, *J* = 11 Hz, 1H), 2.63 (m, 2H), 2.18-1.13 (complex, 26H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.83, 172.95, 148.31, 142.45, 128.53, 124.96, 122.35, 121.69, 69.81, 69.73, 60.88, 58.37, 57.98, 50.55, 50.51, 49.05, 35.84, 33.23, 31.07, 30.80,

28.56, 28.20, 26.20, 25.46, 24.53, 20.16. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{27}H_{40}N_4O_5Cl$, 535.2687; observed, 535.2692.

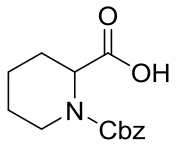


134

(2S,3S)-3-(2-naphthamido)-1-((3S,4aS,8aS)-3-(tert-butylcarbamoyl)octahydroisoquinolin-2(1H)-yl)-4-(phenylthio)butan-2-yl acetate (134)

To a stirred solution of **27** (44 mg, 0.075 mmol) in anhydrous tetrahydrofuran (3 mL) under nitrogen at room temperature was added anhydrous triethylamine (21 μL, 0.150 mmol) followed by acetic anhydride (14 μL, 0.150 mmol). The reaction was stirred for 3 h, then additional acetic anhydride was added (14 μL, 0.150 mmol) and the reaction was stirred at room temperature for 16 h. Ethyl acetate was added to the reaction. The crude material was washed once with water, once with saturated sodium bicarbonate, and once with brine. The organic phase was dried over sodium sulfate and concentrated by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 0-3% MeOH in DCM) to yield **134** as a white solid (21 mg, 44%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.24 (s, 1H), 7.88 (d, J = 8 Hz, 1H), 7.85 (d, J = 8 Hz, 2H), 7.78 (dd, J_1 = 1.5 Hz, J_2 = 8.5 Hz, 1H), 7.53 (m, 2H), 7.42 (d, J = 8 Hz, 2H), 7.26 (t, J = 8 Hz, 2H), 7.16 (t, J = 7.5 Hz, 1H), 6.23 (s, 1H), 5.52 (m, 1H), 4.58 (m, 1H), 3.28 (dd, J_1 = 6.5 Hz, J_2 = 14 Hz, 1H), 3.23 (dd, J_1 = 6 Hz, J_2 = 14 Hz, 1H), 3.00 (dd, J_1 = 2 Hz, J_2 = 11 Hz, 1H), 2.91 (dd, J_1 = 9 Hz, J_2 = 13.5 Hz, 1H), 2.53 (dd, J_1 = 3 Hz, J_2 = 11 Hz, 1H), 2.25 (dd, J_1 = 3 Hz, J_2 = 13 Hz, 1H), 2.13 (s, 3H), 2.01 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 1.81 (q, J = 11.5 Hz, 1H), 1.69 (m, 1H), 1.55-1.04 (complex, 20H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.76, 170.33, 167.33, 135.29, 134.74, 132.52, 131.59, 130.13, 129.17, 128.94, 128.25, 127.65, 127.55, 126.78, 126.58, 123.77,

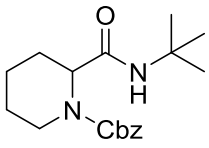
70.76, 70.11, 58.12, 55.96, 51.24, 50.75, 35.55, 34.94, 33.16, 30.88, 30.63, 28.73, 26.08, 25.12, 21.28, 20.15. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{37}H_{48}N_3O_4S$, 630.3366; observed, 630.3365.



135

1-((benzyloxy)carbonyl)piperidine-2-carboxylic acid (135)

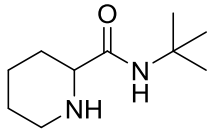
To a solution of piperidine-2-carboxylic acid (2.20 g, 17.0 mmol) in aqueous saturated sodium bicarbonate (55 mL) was added benzyl chloroformate (2.66 mL, 18.7 mmol) over 30 min. The solution was stirred at room temperature for 3 h before the pH was adjusted to 2 with sodium bisulfate. The solution was extracted with ethyl acetate three times. The combined organic extracts were washed once with brine, dried over sodium sulfate, and concentrated to a solid by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 0-4% MeOH in DCM) to yield **135** as a colorless oil (2.58 g, 58%). ^{1}H NMR (500 MHz, $CDCl_3$, mixture of rotamers) δ ppm 7.32 (m, 5H), 5.30 (s, 0.4H), 5.17 (m, 2H), 5.01 (d, *J* = 5 Hz, 0.6H), 4.90 (d, *J* = 5 Hz, 0.4H), 4.71 (s, 0.6H), 4.13 (d, *J* = 13 Hz, 0.4H), 4.05 (d, *J* = 13 Hz, 0.6H), 3.08 (t, *J* = 12.5 Hz, 0.6H), 2.99 (t, *J* = 12.5 Hz, 0.4H), 2.28 (d, *J* = 13.5 Hz, 0.6H), 2.23 (d, *J* = 13.5 Hz, 0.4H), 1.71 (m, 3H), 1.45 (m, 1H), 1.34 (m, 1H). ^{13}C NMR (500 MHz, $CDCl_3$, mixture of rotamers) δ ppm 177.26, 177.19, 156.70, 155.94, 140.71, 136.45, 128.55, 128.46, 128.00, 127.80, 127.67, 127.01, 67.50, 67.39, 65.36, 54.33, 54.13, 41.88, 41.73, 26.67, 26.52, 24.61, 24.46, 20.67, 20.62. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{14}H_{18}NO_4$, 264.1236; observed, 264.1233.



136

Benzyl 2-(tert-butylcarbamoyl)piperidine-1-carboxylate (136)

135 (551 mg, 2.09 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (441 mg, 2.30 mmol), and hydroxybenzotriazole hydrate (373 mg, 2.30 mmol) were dried and placed under nitrogen. The solids were dissolved in dry THF (25 mL) and tert-butylamine (242 μL, 2.30 mmol) was added in one portion. The solution was stirred at room temperature for 20 h. The reaction was taken up in ethyl acetate and washed once with water, once with saturated sodium bicarbonate, and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-5% ethyl acetate in DCM) to yield **136** as a white solid (379 mg, 57%). ^{1}H NMR (500 MHz, $(CD_3)_2SO$) δ ppm 7.32 (m, 5H), 5.80 (m, 1H), 5.18 (m, 2H), 4.70 (s, 1H), 4.11 (m, 1H), 2.88 (br s, 1H), 2.28 (s, 1H), 1.50 (m, 5H), 1.29 (s, 9H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 169.57, 156.28, 136.27, 128.30, 127.89, 127.56, 67.20, 54.76, 50.81, 42.91, 28.41, 25.21, 24.66, 20.09. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{18}H_{27}N_2O_3$, 319.2022; observed, 319.2028.

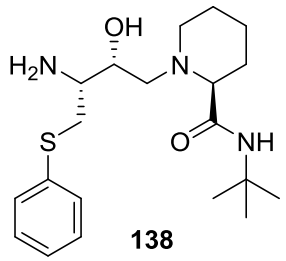


137

N-(tert-butyl)piperidine-2-carboxamide (137)

A solution of **136** (379 mg, 1.19 mmol) and 10% palladium on carbon (380 mg, 0.36 mmol) in methanol (8 mL) and ethyl acetate (8 mL) was degassed with nitrogen for 30 min. The reaction was placed under 1 atm of hydrogen and was stirred vigorously for 2 h at room temperature. The reaction was filtered through celite, washed with excess methanol, and dried to a solid. The product was purified by silica flash column chromatography (gradient of 3-10% MeOH in DCM) to yield **137** as an off white solid (207 mg, 94%). ^{1}H NMR (500 MHz, $(CD_3)_2SO$) δ ppm 7.16 (s, 1H), 3.42 (s, 2H), 2.97 (dd, J_1 = 2 Hz, J_2 = 10 Hz, 1H), 2.89 (d, J = 13 Hz, 1H), 2.49 (m, 1H), 1.69 (m, 2H), 1.45 (m, 1H), 1.26 (m, 11H). ^{13}C NMR

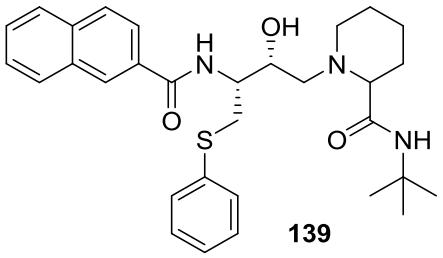
(500 MHz, $(CD_3)_2SO$) δ ppm 172.06, 59.38, 49.60, 44.78, 29.34, 28.38, 25.29, 23.62. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{10}H_{21}N_2O$, 185.1654; observed, 185.1658.



138

(S)-1-((2R,3R)-3-amino-2-hydroxy-4-(phenylthio)butyl)-N-(tert-butyl)piperidine-2-carboxamide (138)

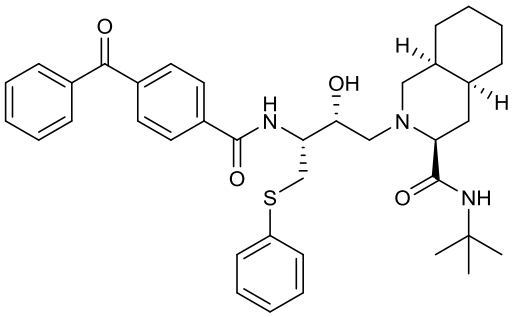
To a solution of **5** (193 mg, 0.587 mmol) in isopropyl alcohol (10 mL) was added **137** (119 mg, 0.646 mmol). The solution was heated to reflux and stirred. After 6 h, 2 M NaOH (1.47 mL, 2.93 mmol) was added and the reaction was stirred at reflux for an additional 18 h. The reaction was diluted with toluene (30 mL) and extracted with 1 M HCl (10 mL) three times. The aqueous fractions were combined and washed with toluene (15 mL) twice. The aqueous phase was adjusted to pH 11 with 10 M NaOH and was extracted with toluene (20 mL) four times. The combined toluene fractions were washed with brine (15 mL), dried over sodium sulfate, and concentrated to a solid by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 1-8% MeOH in DCM) to yield **138** as a white solid (70 mg, 31%). No further purification was made. Which diastereomer was isolated was not determined. ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.36 (d, J = 7.5 Hz, 2H), 7.24 (t, J = 8 Hz, 2H), 7.19 (t, J = 7.5 Hz, 1H), 6.47 (s, 1H), 3.81 (m, 1H), 3.23 (dd, J_1 = 3 Hz, J_2 = 13 Hz, 1H), 3.12 (d, J = 11.5 Hz, 1H), 2.88 (m, 2H), 2.80 (dd, J_1 = 10 Hz, J_2 = 13 Hz, 2H), 2.65 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 2.59 (dd, J_1 = 10.5 Hz, J_2 = 13 Hz, 1H), 2.19 (dd, J_1 = 2 Hz, J_2 = 13 Hz, 1H), 1.97 (m, 2H), 1.73 (dt, J_1 = 3 Hz, J_2 = 13 Hz, 1H), 1.63 (d, J = 11.5 Hz, 1H), 1.54-1.14 (complex, 13H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.19, 135.23, 129.70, 128.99, 126.40, 69.43, 69.05, 58.67, 53.36, 52.33, 50.69, 37.16, 31.24, 28.52, 25.15, 23.48. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{20}H_{34}N_3O_2S$, 380.2372; observed, 380.2370.



139

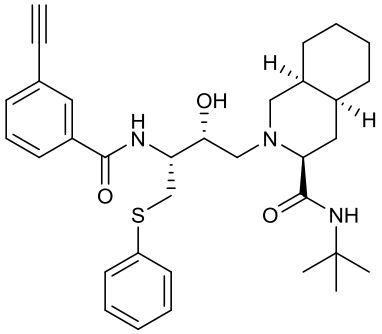
1-((2R,3R)-3-(2-naphthamido)-2-hydroxy-4-(phenylthio)butyl)-N-(tert-butyl)piperidine-2-carboxamide (139)

To a solution of **138** (55 mg, 0.145 mmol) in THF (5 mL) was added 2-naphthoic acid (27 mg, 0.159 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (30 mg, 0.159 mmol), and hydroxybenzotriazole hydrate (26 mg, 0.159 mmol). The solution was stirred at room temperature for 18 h. The reaction was taken up in ethyl acetate and washed once with water, once with saturated sodium bicarbonate, and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 1-4% MeOH in DCM) to yield **139** as a white solid (62 mg, 80% over two steps). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.15 (s, 1H), 7.82 (m, 3H), 7.74 (d, J = 8.5 Hz, 1H), 7.52 (m, 2H), 7.39 (dd, J_1 = 1 Hz, J_2 = 8.5 Hz, 2H), 7.24 (t, J = 7.5 Hz, 2H), 7.16 (t, J = 7.5 Hz, 1H), 7.04 (d, J = 7.5 Hz, 1H), 6.51 (s, 1H), 4.29 (m, 1H), 4.07 (m, 1H), 3.36 (dd, J_1 = 9 Hz, J_2 = 14 Hz, 1H), 3.24 (dd, J_1 = 4.5 Hz, J_2 = 14 Hz, 1H), 3.05 (d, J = 11.5 Hz, 1H), 2.62 (dd, J_1 = 10.5 Hz, J_2 = 13 Hz, 1H), 2.56 (dd, J_1 = 3 Hz, J_2 = 6 Hz, 1H), 2.17 (d, J = 14 Hz, 1H), 1.89 (m, 2H), 1.68 (m, 1H), 1.59-1.13 (complex, 14H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.43, 168.00, 135.47, 134.77, 132.45, 131.07, 129.90, 129.14, 128.92, 128.33, 127.68, 127.65, 126.67, 126.62, 123.62, 69.30, 69.19, 59.27, 53.50, 52.41, 50.79, 34.38, 31.26, 28.56, 25.14, 23.51. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{31}H_{40}N_3O_3S$, 534.2790; observed, 534.2784.



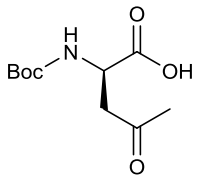
(3S,4aS,8aS)-2-((2R,3R)-3-(4-benzoylbenzamido)-2-hydroxy-4-(phenylthio)butyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (140)

This reaction was performed according to the general EDC coupling procedure (0.050 mmol 4-benzoylbenzoic acid for 40 hours). Purification by flash column chromatography (gradient of 1-3% MeOH in DCM) yielded **140** as a white solid (28.6 mg, 93%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.96 (d, J = 8 Hz, 2H), 7.93 (d, J = 8.5 Hz, 1H), 7.80-7.78 (m, 4H), 7.61 (t, J = 7.5 Hz, 1H), 7.49 (t, J = 8 Hz, 2H), 7.43 (dd, J_1 = 1.5 Hz, J_2 = 8 Hz, 2H), 7.28 (t, J = 7.5 Hz, 2H), 7.19 (t, J = 7.5 Hz, 1H), 5.49 (s, 1H), 4.53 (m, 1H), 4.16 (m, 1H), 3.90 (dd, J_1 = 9.5 Hz, J_2 = 14 Hz, 1H), 3.44 (dd, J_1 = 5 Hz, J_2 = 13.5 Hz, 1H), 2.94 (dd, J_1 = 1.5 Hz, J_2 = 11.5 Hz, 1H), 2.57 (dd, J_1 = 10.5 Hz, J_2 = 13 Hz, 1H), 2.42 (dd, J_1 = 3 Hz, J_2 = 12 Hz, 1H), 2.19 (dd, J_1 = 4 Hz, J_2 = 12.5 Hz, 1H), 2.15 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 2.03 (q, J = 12 Hz, 1H), 1.82-1.14 (complex, 11H), 1.07 (s, 9H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 195.99, 174.05, 168.97, 140.10, 137.17, 137.05, 135.42, 132.82, 130.14, 130.06, 129.78, 129.05, 128.37, 128.06, 126.46, 70.83, 70.03, 59.55, 58.87, 55.95, 51.29, 36.10, 35.57, 33.92, 31.15, 30.97, 28.47, 26.36, 26.07, 20.43. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{38}H_{48}N_3O_4S$, 642.3366; observed, 642.3362.



(3S,4aS,8aS)-N-(tert-butyl)-2-((2R,3R)-3-(3-ethynylbenzamido)-2-hydroxy-4-(phenylthio)butyl)decahydroisoquinoline-3-carboxamide (141)

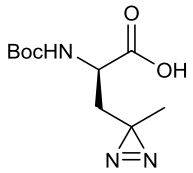
This reaction was performed according to the general EDC coupling procedure (0.065 mmol 3-ethynylbenzoic acid for 40 hours). Purification by flash column chromatography (gradient of 1-3% MeOH in DCM) yielded **141** as a slightly yellow solid (21.8 mg, 66%). ^{1}H NMR (400 MHz, $CDCl_3$) δ ppm 7.97 (s, 1H), 7.84 (d, J = 7.5 Hz, 1H), 7.80 (d, J = 8.5 Hz, 1H), 7.58 (d, J = 8 Hz, 1H), 7.42 (dd, J_1 = 1 Hz, J_2 = 8.5 Hz, 2H), 7.36 (t, J = 7.5 Hz, 1H), 7.27 (t, J = 7.5 Hz, 2H), 7.18 (t, J = 7.5 Hz, 1H), 5.46 (s, 1H), 4.50 (m, 1H), 4.14 (m, 1H), 3.87 (dd, J_1 = 9 Hz, J_2 = 13.5 Hz, 1H), 3.45 (dd, J_1 = 5 Hz, J_2 = 13.5 Hz, 1H), 2.96-2.88 (m, 2H), 2.55 (dd, J_1 = 10 Hz, J_2 = 13 Hz, 1H), 2.42 (dd, J_1 = 2.5 Hz, J_2 = 11.5 Hz, 1H), 2.21-2.14 (m, 2H), 2.04 (q, J = 11.5 Hz, 1H), 1.83-1.15 (complex, 11H), 1.08 (s, 9H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.96, 168.93, 135.40, 135.01, 134.08, 131.66, 130.17, 129.05, 128.49, 128.35, 126.51, 122.44, 82.80, 77.97, 70.85, 70.14, 59.57, 58.84, 55.88, 51.31, 36.11, 35.51, 33.94, 31.15, 30.99, 28.45, 26.37, 26.06, 20.44. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{33}H_{44}N_3O_3S$, 562.3103; observed, 562.3102.



(R)-boc-4-oxo-norvaline (142)

(R)-Boc-4,5-dehydro-leucine dicyclohexylammonium salt (ChemImpex, 1.005 g, 2.45 mmol) was dissolved in DCM and cooled to -78°C. Ozone was bubbled through the reaction until a purplish-blue color persisted. Oxygen was flushed through the reaction until the solution was colorless. The reaction was warmed to room temperature, dimethyl sulfide (449 uL, 6.12 mmol) was added, and the reaction was stirred overnight (18 h). The solvent and excess dimethyl sulfide were removed by evaporation and the residue was purified by silica flash column chromatography (gradient of 0-3% MeOH in DCM with 0.1% acetic acid) to yield **142** as an amber oil (433 mg, 76%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 5.56 (d, J = 8

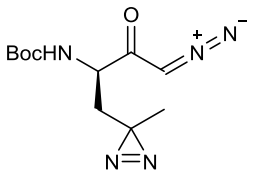
Hz, 1H), 4.46 (m, 1H), 3.13 (dd, J_1 = 4 Hz, J_2 = 18 Hz, 1H), 2.94 (dd, J_1 = 3.5 Hz, J_2 = 17.5 Hz, 1H), 2.67 (d, J = 2 Hz, 3H), 1.40 (s, 9H). Residual acetic acid from the purification is visible at 2.14 ppm. ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 206.76, 174.47, 155.60, 80.00, 49.26, 45.19, 39.72, 28.17. HRMS (*m/z*): [M + Na]$^+$ calc. for $C_{10}H_{17}NO_5Na$, 254.1004; observed, 254.1008.



(R)-boc-photo-leucine (143)

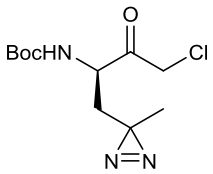
A three-necked flask containing **142** (449 mg, 1.94 mmol) was fitted with a dry ice condenser and cooled to -78°C. The flask was purged with dry nitrogen before gaseous ammonia (15 mL) was slowly condensed into the flask. The reaction was placed in an ice water bath and refluxed with stirring for 5 h. The reaction was then cooled to -78°C and a solution of hydroxylamine-*O*-sulfonic acid (253 mg, 2.23 mmol) in anhydrous methanol (1.5 mL) was added over 15 min. The reaction was warmed to reflux for another 1 h. Anhydrous methanol (3 mL) was added and the reaction was warmed to room temperature. The ammonia was purged from the system with nitrogen and the reaction was stirred at room temperature overnight, which caused a white slurry to form. The slurry was filtered, the filter cake was washed with methanol, and triethylamine (270 uL, 1.94 mmol) was added to the combined filtrate and wash. The filtrate was concentrated to 8 mL and more triethylamine (270 uL, 1.94 mmol) was added. The solution was cooled to 0°C and titrated with a solution of iodine in methanol (0.1 M) until an orange color persisted. Solvent was removed by rotary evaporation and the residue was brought up in water. The pH was adjusted to 2 with 1 M HCl and solution was extracted three times with ethyl acetate. The combined organic fractions were washed with brine, dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 0-3% MeOH in DCM with 0.1% acetic acid). The product was taken up in ethyl acetate and washed twice with 0.5M HCl and once with brine to remove residual acetic acid. The organic layer was dried over sodium sulfate and concentrated by rotary

evaporation to yield **143** as an amber oil (240 mg, 51% yield). The product was determined to exist as a mixture of rotamers by 1D selective chemical-exchange NMR[60]. ^{1}H NMR (500 MHz, $CDCl_3$, mixture of rotamers) δ ppm 11.40 (br s, 1H), 6.76 (d, J = 6.5 Hz, 0.4H), 5.18 (d, J = 8 Hz, 0.6H), 4.39 (d, J = 5 Hz, 0.6H), 4.12 (d, J = 4.5 Hz, 0.4H), 2.06 (dd, J_1 = 4.5 Hz, J_2 = 15 Hz, 0.6H), 1.88 (dd, J_1 = 4 Hz, J_2 = 15 Hz, 0.4H), 1.71 (dd, J_1 = 8.5 Hz, J_2 = 14 Hz, 0.4H), 1.61 (dd, J_1 = 9 Hz, J_2 = 14.5 Hz, 0.6H), 1.48 (s, 9H), 1.10 (s, 3H). ^{13}C NMR (500 MHz, $CDCl_3$, mixture of rotamers) δ ppm 175.95, 175.62, 156.71, 155.38, 82.37, 80.56, 50.71, 50.10, 37.73, 37.55, 28.21, 23.69, 23.48, 19.91, 19.60. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{10}H_{18}N_3O_4$, 244.1297; observed, 244.1303.



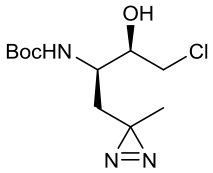
Tert-butyl (R)-(4-diazo-1-(3-methyl-3H-diazirin-3-yl)-3-oxobutan-2-yl)carbamate (144)

To a stirred solution of **143** (415 mg, 1.71 mmol) in anhydrous tetrahydrofuran (10 mL) under nitrogen at -78 °C was added anhydrous triethylamine (355 μL, 2.56 mmol) followed by isobutyl chloroformate (333 μL, 2.56 mmol). The reaction was stirred cold for 1 hour and then warmed to room temperature for 1.5 h. The reaction was cooled to -78 °C before an ethereal solution of diazomethane (25 mL, 8.1 mmol) was added in one portion and the reaction was then warmed to room temperature with stirring for 1 hour. The reaction was evaporated to dryness, resulting in a yellow solid. The solid was taken up in ethyl acetate, washed twice with water, twice with a saturated solution of sodium bicarbonate, and once with brine. The organic phase was dried over sodium sulfate and concentrated by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 0-30% EtOAc in hexanes) to yield **144** as a yellowish solid (273 mg, 60%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 5.61 (s, 1H), 5.23 (d, J = 8 Hz, 1H), 4.17 (m, 1H), 2.03 (dd, J_1 = 5 Hz, J_2 = 15 Hz, 1H), 1.47 (s, 9H), 1.44 (m, 1H), 1.08 (s, 3H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 192.86, 155.04, 80.33, 54.08, 53.69, 37.00, 28.20, 23.65, 19.76. HRMS (*m/z*): $[M + Na]^+$ calc. for $C_{11}H_{17}N_5O_3Na$, 290.1229; observed, 290.1223.



Tert-butyl (R)-(4-chloro-1-(3-methyl-3H-diazirin-3-yl)-3-oxobutan-2-yl)carbamate (145)

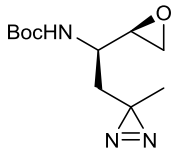
To a solution of **144** (273 mg, 1.02 mmol) in anhydrous tetrahydrofuran under nitrogen at 0 °C was added 1 M HCl in diethyl ether (1.53 mL, 1.53 mmol) dropwise. The reaction was stirred at 0 °C for 30 min after complete addition. The solution was dried under reduced pressure to yield **145** as a white solid (278 mg). No further purification was made. ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 5.13 (d, J = 7.5 Hz, 1H), 4.39 (m, 1H), 4.30 (s, 2H), 2.13 (dd, J_1 = 5 Hz, J_2 = 15 Hz, 1H), 1.49 (m, 1H), 1.47 (s, 9H), 1.09 (3H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 200.58, 155.17, 80.90, 53.67, 46.42, 36.01, 28.22, 23.55, 19.81. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{11}H_{19}N_3O_3Cl$, 276.1115; observed, 276.1117.



Tert-butyl ((2R,3R)-4-chloro-3-hydroxy-1-(3-methyl-3H-diazirin-3-yl)butan-2-yl)carbamate (146)

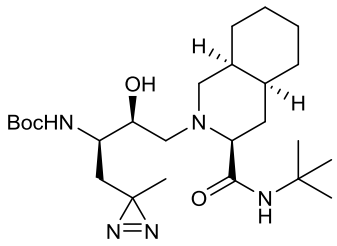
To a solution of **145** (278 mg, 1.01 mmol) in dichloromethane (5 mL) and methanol (3 mL) under nitrogen at 0 °C was added sodium borohydride (27 mg, 0.714 mmol). The reaction was stirred at 0 °C for 1 h, and then quenched with slow addition of 2 M HCl (2 mL). The reaction was evaporated to dryness and the solid was taken up in ethyl acetate. The solution was washed twice with water followed by brine, dried over sodium sulfate, and concentrated by rotary evaporation. TLC analysis showed two diastereomers with the lower Rf compound being the (R,R) product. Both diastereomers were purified by silica flash column chromatography (gradient of 0-6% EtOAc in DCM) to yield **146** as a white solid (183 mg, 65%) over two steps. ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 4.78 (d, J = 9 Hz, 1H), 3.80 (br s, 1H), 3.67

(m, 1H), 3.60 (dd, J_1 = 4 Hz, J_2 = 11.5 Hz, 1H), 3.51 (dd, J_1 = 8 Hz, J_2 = 11 Hz, 1H), 3.15 (d, J = 4 Hz, 1H), 1.92 (dd, J_1 = 3 Hz, J_2 = 15 Hz, 1H), 1.48 (s, 9H), 1.33 (dd, J_1 = 11 Hz, J_2 = 15 Hz, 1H), 1.07 (s, 3H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 155.76, 80.33, 73.89, 49.97, 46.89, 34.81, 28.32, 24.10, 19.77. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{11}H_{21}N_3O_3Cl$, 278.1271; observed, 278.1278.



Tert-butyl ((R)-2-(3-methyl-3H-diazirin-3-yl)-1-((R)-oxiran-2-yl)ethyl)carbamate (147)

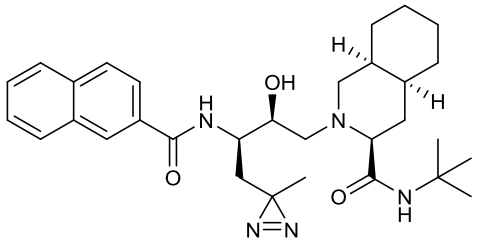
To a solution of **146** (183 mg, 0.659 mmol) in ethanol (10 mL) at 0 °C was added solid potassium hydroxide (41 mg, 0.725 mmol). The reaction was stirred cold for 15 min then warmed to room temperature and stirred for an additional 2 h. The reaction was evaporated to dryness and the solid was purified by silica flash column chromatography (gradient of 1-4% EtOAc in DCM) to yield **147** as a white solid (147 mg, 92%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 4.72 (br s, 1H), 3.51 (br s, 1H), 2.89 (s, 1H), 2.79 (t, J = 4 Hz, 1H), 2.72 (s, 1H), 1.78 (dd, J_1 = 2.5 Hz, J_2 = 14.5 Hz, 1H), 1.47 (s, 9H), 1.35 (m, 1H), 1.07 (s, 3H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 155.07, 79.79, 53.40, 48.89, 46.30, 36.58, 28.25, 23.75, 19.60. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{11}H_{20}N_3O_3$, 242.1505; observed, 242.1508.



Tert-butyl ((2R,3S)-4-((3S,4aS,8aS)-3-(tert-butylcarbamoyl)octahydroisoquinolin-2(1H)-yl)-3-hydroxy-1-(3-methyl-3H-diazirin-3-yl)butan-2-yl)carbamate (148)

To a solution of **147** (62 mg, 0.257 mmol) in isopropyl alcohol (7 mL) was added (3S,4aS,8aS)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (**6**, 67 mg, 0.283 mmol). The solution was heated to

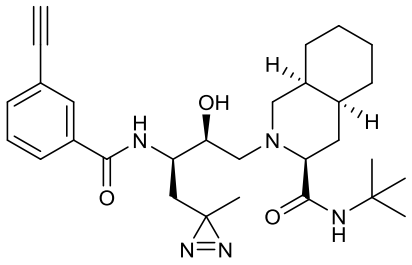
reflux and stirred. After 6 h, the reaction was cooled and dried to a solid. The product was purified by silica flash column chromatography (gradient of 0-6% MeOH in DCM) to yield **148** as a white solid (91 mg, 74%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 6.21 (s, 1H), 4.87 (d, J = 9 Hz, 1H), 3.77 (d, J = 10 Hz, 1H), 3.59 (br, 2H), 2.83 (d, J = 11 Hz, 1H), 2.62 (dd, J_1 = 3 Hz, J_2 = 11 Hz, 1H), 2.38 (dd, J_1 = 11 Hz, J_2 = 12.5 Hz, 1H), 2.15 (dd, J_1 = 3.5 Hz, J_2 = 11.5 Hz, 1H), 1.99 (d, J = 13 Hz, 1H), 1.87 (q, J = 12 Hz, 1H), 1.82-1.20 (complex, 31H), 1.05 (s, 3H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.40, 155.68, 79.68, 69.85, 69.61, 58.39, 58.22, 50.80, 50.34, 35.79, 34.79, 33.25, 30.96, 30.82, 28.60, 28.37, 26.24, 25.46, 24.16, 20.19, 19.68. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{25}H_{46}N_5O_4$, 480.3550; observed, 480.3551.



(3S,4aS,8aS)-2-((2S,3R)-3-(2-naphthamido)-2-hydroxy-4-(3-methyl-3H-diazirin-3-yl)butyl)-N-(tert-butyl)decahydroisoquinoline-3-carboxamide (149)

A solution of trifluoroacetic acid and dichloromethane (1:1, 6 mL) was used to dissolve **148** (46 mg, 0.096 mmol). The reaction was stirred at room temperature for 1 h then dried to a solid. The residue was taken up in water and the pH adjusted to 10 with 10 M NaOH. The aqueous phase was extracted three times with ethyl acetate. The ethyl acetate fractions were combined and washed once with 10 M NaOH and once with brine. The ethyl acetate was dried over sodium sulfate and dried to a solid. The solid was taken up in tetrahydrofuran (5 mL), 2-naphthoic acid (18 mg, 0.103 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (20 mg, 0.103 mmol), and hydroxybenzotriazole hydrate (17 mg, 0.103 mmol) were added, and the solution was stirred at room temperature for 18 h. The reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and

purified by silica flash column chromatography (gradient of 10-60% EtOAc in hexanes) to yield **149** as a white solid (22 mg, 44%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 8.40 (s, 1H), 7.95-7.87 (complex, 4H), 7.56 (m, 2H), 6.84 (d, J = 9 Hz, 1H), 6.06 (s, 1H), 4.27 (m, 1H), 3.96 (br, 1H), 3.90 (dt, J_1 = 2.5 Hz, J_2 = 10.5 Hz, 1H), 2.82 (dd, J_1 = 1.5 Hz, J_2 = 11.5 Hz, 1H), 2.62 (dd, J_1 = 3.5 Hz, J_2 = 11.5 Hz, 1H), 2.43 (dd, J_1 = 10.5 Hz, J_2 = 13 Hz, 1H), 2.16 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 2.10 (dd, J_1 = 2.5 Hz, J_2 = 13 Hz, 1H), 1.90 (q, J = 11.5 Hz, 1H), 1.83 (dd, J_1 = 3.5 Hz, J_2 = 10.5 Hz, 1H), 1.78-1.12 (complex, 21H), 1.09 (s, 3H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.50, 167.73, 134.82, 132.62, 131.50, 129.01, 128.49, 127.76, 127.72, 127.70, 126.73, 123.69, 69.91, 69.80, 58.49, 58.25, 51.00, 49.90, 35.82, 34.37, 33.27, 31.02, 30.86, 28.64, 26.21, 25.39, 24.46, 20.14, 19.55. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{31}H_{44}N_5O_3$, 534.3444; observed, 534.3450.



(3S,4aS,8aS)-N-(tert-butyl)-2-((2S,3R)-3-(3-ethynylbenzamido)-2-hydroxy-4-(3-methyl-3H-diazirin-3-yl)butyl)decahydroisoquinoline-3-carboxamide (150)

A solution of 47.5% trifluoroacetic acid, 47.5% dichloromethane, and 5% thioanisole (5 mL) was cooled to 0°C and used to dissolve **148** (47 mg, 0.098 mmol). The reaction was stirred cold for 30 min then dried to a solid. The residue was taken up in water and the pH adjusted to 10 with 10 M NaOH. The aqueous phase was extracted three times with ethyl acetate. The ethyl acetate fractions were combined, washed once with brine, dried over sodium sulfate, and dried to a solid. The solid was taken up in tetrahydrofuran (5 mL), 3-ethynylbenzoic acid (16 mg, 0.108 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (21 mg, 0.108 mmol), and hydroxybenzotriazole hydrate (18 mg, 0.108 mmol) were added, and the solution was stirred at room temperature for 18 h. The

reaction was taken up in ethyl acetate and washed once with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate, concentrated by rotary evaporation, and purified by silica flash column chromatography (gradient of 10-80% EtOAc in hexanes) to yield **150** as a white solid (24 mg, 48%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 7.98 (t, J = 1.5 Hz, 1H), 7.86 (dt, J_1 = 1.5 Hz, J_2 = 8 Hz, 1H), 7.62 (dt, J_1 = 1.5 Hz, J_2 = 8 Hz, 1H), 7.42 (t, J = 8 Hz, 1H), 6.71 (d, J = 9 Hz, 1H), 5.97 (s, 1H), 4.21 (m, 1H), 3.85 (dt, J_1 = 3 Hz, J_2 = 10 Hz, 1H), 3.13 (s, 1H), 2.84 (dd, J_1 = 2 Hz, J_2 = 11.5 Hz, 1H), 2.63 (dd, J_1 = 3.5 Hz, J_2 = 11.5 Hz, 1H), 2.40 (dd, J_1 = 10.5 Hz, J_2 = 13 Hz, 1H), 2.18 (dd, J_1 = 3 Hz, J_2 = 11.5 Hz, 1H), 2.10 (dd, J_1 = 2 Hz, J_2 = 13 Hz, 1H), 1.91 (q, J = 11.5 Hz, 1H), 1.82-1.14 (complex, 23H), 1.07 (s, 3H). ^{13}C NMR (500 MHz, $CDCl_3$) δ ppm 173.43, 166.66, 134.99, 134.72, 130.80, 128.70, 127.55, 122.73, 82.67, 78.24, 69.95, 69.66, 58.56, 58.10, 51.05, 49.76, 35.84, 34.29, 33.31, 31.03, 30.88, 28.66, 26.23, 25.42, 24.36, 20.19, 19.54. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{29}H_{42}N_5O_3$, 508.3288; observed, 508.3304.

2.11 References

1. Maxson, T., Deane, C. D., Molloy, E. M., Cox, C. L., Markley, A. L., Lee, S. W., and Mitchell, D. A. (2015) HIV protease inhibitors block streptolysin S production, *ACS Chem. Biol. 10*, 1217-1226.

2. Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P. C., Entian, K. D., Fischbach, M. A., Garavelli, J. S., Goransson, U., Gruber, C. W., Haft, D. H., Hemscheidt, T. K., Hertweck, C., Hill, C., Horswill, A. R., Jaspars, M., Kelly, W. L., Klinman, J. P., Kuipers, O. P., Link, A. J., Liu, W., Marahiel, M. A., Mitchell, D. A., Moll, G. N., Moore, B. S., Muller, R., Nair, S. K., Nes, I. F., Norris, G. E., Olivera, B. M., Onaka, H., Patchett, M. L., Piel, J., Reaney, M. J., Rebuffat, S., Ross, R. P., Sahl, H. G., Schmidt, E. W., Selsted, M. E., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., Sussmuth, R. D., Tagg, J. R., Tang, G. L., Truman, A. W., Vederas, J. C., Walsh, C. T., Walton, J. D., Wenzel, S. C., Willey, J. M., and van der Donk, W. A. (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature, *Nat. Prod. Rep. 30*, 108-160.

3. Bagley, M. C., Dale, J. W., Merritt, E. A., and Xiong, X. (2005) Thiopeptide antibiotics, *Chem. Rev. 105*, 685-714.

4. Chatterjee, C., Paul, M., Xie, L., and van der Donk, W. A. (2005) Biosynthesis and mode of action of lantibiotics, *Chem. Rev. 105*, 633-684.

5. Maksimov, M. O., Pan, S. J., and James Link, A. (2012) Lasso peptides: structure, function, biosynthesis, and engineering, *Nat. Prod. Rep. 29*, 996-1006.

6. Sivonen, K., Leikoski, N., Fewer, D. P., and Jokela, J. (2010) Cyanobactins-ribosomal cyclic peptides produced by cyanobacteria, *Appl. Microbiol. Biot. 86*, 1213-1225.

7. Melby, J. O., Nard, N. J., and Mitchell, D. A. (2011) Thiazole/oxazole-modified microcins: complex natural products from ribosomal templates, *Curr. Opin. Chem. Biol. 15*, 369-378.

8. Oman, T. J., and van der Donk, W. A. (2010) Follow the leader: the use of leader peptides to guide natural product biosynthesis, *Nat. Chem. Biol. 6*, 9-18.

9. Dunbar, K. L., Melby, J. O., and Mitchell, D. A. (2012) YcaO domains use ATP to activate amide backbones during peptide cyclodehydrations, *Nat. Chem. Biol. 8*, 569-575.

10. Dunbar, K. L., and Mitchell, D. A. (2013) Insights into the mechanism of peptide cyclodehydrations achieved through the chemoenzymatic generation of amide derivatives, *J. Am. Chem. Soc. 135*, 8692-8701.

11. Melby, J. O., Li, X., and Mitchell, D. A. (2014) Orchestration of enzymatic processing by thiazole/oxazole-modified microcin dehydrogenases, *Biochemistry 53*, 413-422.

12. Molloy, E. M., Cotter, P. D., Hill, C., Mitchell, D. A., and Ross, R. P. (2011) Streptolysin S-like virulence factors: the continuing sagA, *Nat. Rev. Microbiol. 9*, 670-681.

13. Cunningham, M. W. (2000) Pathogenesis of group A streptococcal infections, *Clin. Microbiol. Rev. 13*, 470-511.

14. Carapetis, J. R., Steer, A. C., Mulholland, E. K., and Weber, M. (2005) The global burden of group A streptococcal diseases, *Lancet Infect. Dis. 5*, 685-694.

15. Nizet, V., Beall, B., Bast, D. J., Datta, V., Kilburn, L., Low, D. E., and De Azavedo, J. C. (2000) Genetic locus for streptolysin S production by group A streptococcus, *Infect. Immun. 68*, 4245-4254.

16. Datta, V., Myskowski, S. M., Kwinn, L. A., Chiem, D. N., Varki, N., Kansal, R. G., Kotb, M., and Nizet, V. (2005) Mutational analysis of the group A streptococcal operon encoding streptolysin S and its virulence role in invasive infection, *Mol. Microbiol. 56*, 681-695.

17. Betschel, S. D., Borgia, S. M., Barg, N. L., Low, D. E., and De Azavedo, J. C. (1998) Reduced virulence of group A streptococcal Tn916 mutants that do not produce streptolysin S, *Infect. Immun. 66*, 1671-1679.

18. Fontaine, M. C., Lee, J. J., and Kehoe, M. A. (2003) Combined contributions of streptolysin O and streptolysin S to virulence of serotype M5 Streptococcus pyogenes strain Manfredo, *Infect. Immun. 71*, 3857-3865.

19. James, L., and McFarland, R. B. (1971) An epidemic of pharyngitis due to a nonhemolytic group A streptococcus at lowry air force base, *N. Engl. J. Med. 284*, 750-752.

20. Yoshino, M., Murayama, S. Y., Sunaoshi, K., Wajima, T., Takahashi, M., Masaki, J., Kurokawa, I., and Ubukata, K. (2010) Nonhemolytic Streptococcus pyogenes isolates that lack large regions of the sag operon mediating streptolysin S production, *J. Clin. Microbiol. 48*, 635-638.

21. Lee, S. W., Mitchell, D. A., Markley, A. L., Hensler, M. E., Gonzalez, D., Wohlrab, A., Dorrestein, P. C., Nizet, V., and Dixon, J. E. (2008) Discovery of a widely distributed toxin biosynthetic gene cluster, *Proc. Natl. Acad. Sci. U.S.A. 105*, 5879-5884.

22. Gonzalez, D. J., Lee, S. W., Hensler, M. E., Markley, A. L., Dahesh, S., Mitchell, D. A., Bandeira, N., Nizet, V., Dixon, J. E., and Dorrestein, P. C. (2010) Clostridiolysin S, a post-translationally modified biotoxin from Clostridium botulinum, *J. Biol. Chem. 285*, 28220-28228.

23. Bernheimer, A. W. (1967) Physical Behavior of Streptolysin S, *J. Bacteriol. 93*, 2024-2025.

24. Jack, R. W., Tagg, J. R., and Ray, B. (1995) Bacteriocins of gram-positive bacteria, *Microbiol. Rev. 59*, 171-200.

25. Todd, E. W. (1938) The differentiation of two distinct serological varieties of streptolysin, streptolysin O and streptolysin S, *J. Pathol. Bacteriol. 47*, 423-445.

26. Marmorek, A. (1895) Le streptocoque et le sérum antistreptococcique, *Ann. Inst. Pasteur 9*, 593-620.

27. Rasko, D. A., and Sperandio, V. (2010) Anti-virulence strategies to combat bacteria-mediated disease, *Nat. Rev. Drug Discov. 9*, 117-128.

28. Cegelski, L., Marshall, G. R., Eldridge, G. R., and Hultgren, S. J. (2008) The biology and future prospects of antivirulence therapies, *Nat. Rev. Microbiol. 6*, 17-27.

29. Baruch, M., Belotserkovsky, I., Hertzog, B. B., Ravins, M., Dov, E., McIver, K. S., Le Breton, Y. S., Zhou, Y., Cheng, C. Y., and Hanski, E. (2014) An extracellular bacterial pathogen modulates host metabolism to regulate its own sensing and proliferation, *Cell 156*, 97-108.

30. Pei, J., and Grishin, N. V. (2001) Type II CAAX prenyl endopeptidases belong to a novel superfamily of putative membrane-bound metalloproteases, *Trends Biochem. Sci. 26*, 275-277.

31. Pei, J., Mitchell, D. A., Dixon, J. E., and Grishin, N. V. (2011) Expansion of type II CAAX proteases reveals evolutionary origin of gamma-secretase subunit APH-1, *J. Mol. Biol. 410*, 18-26.

32. Bergo, M. O., Ambroziak, P., Gregory, C., George, A., Otto, J. C., Kim, E., Nagase, H., Casey, P. J., Balmain, A., and Young, S. G. (2002) Absence of the CAAX endoprotease Rce1: effects on cell growth and transformation, *Mol. Cell. Biol. 22*, 171-181.

33. Bergo, M. O., Wahlstrom, A. M., Fong, L. G., and Young, S. G. (2008) Genetic analyses of the role of RCE1 in RAS membrane association and transformation, *Methods Enzymol. 438*, 367-389.

34. Plummer, L. J., Hildebrandt, E. R., Porter, S. B., Rogers, V. A., McCracken, J., and Schmidt, W. K. (2006) Mutational analysis of the ras converting enzyme reveals a requirement for glutamate and histidine residues, *J. Biol. Chem. 281*, 4596-4605.

35. Dolence, J. M., Steward, L. E., Dolence, E. K., Wong, D. H., and Poulter, C. D. (2000) Studies with recombinant Saccharomyces cerevisiae CaaX prenyl protease Rce1p, *Biochemistry 39*, 4096-4104.

36. Kjos, M., Snipen, L., Salehian, Z., Nes, I. F., and Diep, D. B. (2010) The Abi Proteins and Their Involvement in Bacteriocin Self-Immunity, *J. Bacteriol. 192*, 2068-2076.

37. Clayton, E. M., Hill, C., Cotter, P. D., and Ross, R. P. (2011) Real-time PCR assay to differentiate Listeriolysin S-positive and -negative strains of Listeria monocytogenes, *Appl. Environ. Microb. 77*, 163-171.

38. Kaldor, S. W., Kalish, V. J., Davies, J. F., 2nd, Shetty, B. V., Fritz, J. E., Appelt, K., Burgess, J. A., Campanale, K. M., Chirgadze, N. Y., Clawson, D. K., Dressman, B. A., Hatch, S. D., Khalil, D. A., Kosa, M. B., Lubbehusen, P. P., Muesing, M. A., Patick, A. K., Reich, S. H., Su, K. S., and Tatlock, J. H. (1997) Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease, *J. Med. Chem. 40*, 3979-3985.

39. Albizati, K. F., Babu, S., Birchler, A., Busse, J. K., Fugett, M., Grubbs, A., Haddach, A., Pagan, M., Potts, B., Remarchuk, T., Rieger, D., Rodriguez, R., Shanley, J., Szendroi, R., Tibbetts, T., Whitten, K., and Borer, B. C. (2001) A synthesis of the HIV-protease inhibitor nelfinavir from D-tartaric acid, *Tetrahedron Lett. 42*, 6481-6485.

40. Ma, D., Zou, B., Zhu, W., and Xu, H. D. (2002) A short synthesis of the HIV-protease inhibitor nelfinavir via a diastereoselective addition of ammonia to the alpha,beta-unsaturated sulfoxide derived from (R)-glyceraldehyde acetonide, *Tetrahedron Lett. 43*, 8511-8513.

41. Viklund, H., Bernsel, A., Skwark, M., and Elofsson, A. (2008) SPOCTOPUS: a combined predictor of signal peptides and membrane protein topology, *Bioinformatics 24*, 2928-2929.

42. Manolaridis, I., Kulkarni, K., Dodd, R. B., Ogasawara, S., Zhang, Z., Bineva, G., O'Reilly, N., Hanrahan, S. J., Thompson, A. J., Cronin, N., Iwata, S., and Barford, D. (2013) Mechanism of farnesylated CAAX protein processing by the intramembrane protease Rce1, *Nature 504*, 301-305.

43. Markley, A. L., Jensen, E. R., and Lee, S. W. (2012) An Escherichia coli-based bioengineering strategy to study streptolysin S biosynthesis, *Anal. Biochem. 420*, 191-193.

44. Alouf, J. E. (1980) Streptococcal toxins (streptolysin O, streptolysin S, erythrogenic toxin), *Pharmacol. Ther. 11*, 661-717.

45. Caron, M., Auclair, M., Sterlingot, H., Kornprobst, M., and Capeau, J. (2003) Some HIV protease inhibitors alter lamin A/C maturation and stability, SREBP-1 nuclear localization and adipocyte differentiation, *Aids 17*, 2437-2444.

46. Coffinier, C., Hudon, S. E., Farber, E. A., Chang, S. Y., Hrycyna, C. A., Young, S. G., and Fong, L. G. (2007) HIV protease inhibitors block the zinc metalloproteinase ZMPSTE24 and lead to an accumulation of prelamin A in cells, *Proc. Natl. Acad. Sci. U.S.A. 104*, 13432-13437.

47. Coffinier, C., Hudon, S. E., Lee, R., Farber, E. A., Nobumori, C., Miner, J. H., Andres, D. A., Spielmann, H. P., Hrycyna, C. A., Fong, L. G., and Young, S. G. (2008) A potent HIV protease inhibitor, darunavir, does not inhibit ZMPSTE24 or lead to an accumulation of farnesyl-prelamin A in cells, *J. Biol. Chem. 283*, 9797-9804.

48. Quigley, A., Dong, Y. Y., Pike, A. C., Dong, L., Shrestha, L., Berridge, G., Stansfeld, P. J., Sansom, M. S., Edwards, A. M., Bountra, C., von Delft, F., Bullock, A. N., Burgess-Brown, N. A., and Carpenter, E. P. (2013) The structural basis of ZMPSTE24-dependent laminopathies, *Science 339*, 1604-1607.

49. Bernheimer, A. W. (1949) Formation of a bacterial toxin (streptolysin S) by resting cells, *J. Exp. Med. 90*, 373-392.

50. Schmittgen, T. D., and Livak, K. J. (2008) Analyzing real-time PCR data by the comparative C(T) method, *Nat. Protoc. 3*, 1101-1108.

51. Cotter, P. D., Draper, L. A., Lawton, E. M., Daly, K. M., Groeger, D. S., Casey, P. G., Ross, R. P., and Hill, C. (2008) Listeriolysin S, a novel peptide haemolysin associated with a subset of lineage I Listeria monocytogenes, *PLoS Pathog. 4*, e1000144.

52. Molohon, K. J., Melby, J. O., Lee, J., Evans, B. S., Dunbar, K. L., Bumpus, S. B., Kelleher, N. L., and Mitchell, D. A. (2011) Structure determination and interception of biosynthetic intermediates for the plantazolicin class of highly discriminating antibiotics, *ACS Chem. Biol. 6*, 1307-1313.

53. Ellermeier, C. D., and Losick, R. (2006) Evidence for a novel protease governing regulated intramembrane proteolysis and resistance to antimicrobial peptides in Bacillus subtilis, *Genes Dev. 20*, 1911-1922.

54. Ritchie, T. K., Grinkova, Y. V., Bayburt, T. H., Denisov, I. G., Zolnerciks, J. K., Atkins, W. M., and Sligar, S. G. (2009) Reconstitution of Membrane Proteins in Phospholipid Bilayer Nanodiscs, *Methods Enzymol. 464*, 211-231.

55. Biswas, I., Jha, J. K., and Fromm, N. (2008) Shuttle expression plasmids for genetic studies in Streptococcus mutans, *Microbiology 154*, 2275-2282.

56. Gantt, S., Casper, C., and Ambinder, R. F. (2013) Insights into the broad cellular effects of nelfinavir and the HIV protease inhibitors supporting their role in cancer treatment and prevention, *Curr. Opin. Oncol. 25*, 495-502.

57. Mitchell, D. A., Lee, S. W., Pence, M. A., Markley, A. L., Limm, J. D., Nizet, V., and Dixon, J. E. (2009) Structural and functional dissection of the heterocyclic peptide cytotoxin streptolysin S, *J. Biol. Chem. 284*, 13004-13012.

58. Chen, X. H., Scholz, R., Borriss, M., Junge, H., Mogel, G., Kunz, S., and Borriss, R. (2009) Difficidin and bacilysin produced by plant-associated Bacillus amyloliquefaciens are efficient in controlling fire blight disease, *J. Biotechnol. 140*, 38-44.

59. Deane, C. D., Melby, J. O., Molohon, K. J., Susarrey, A. R., and Mitchell, D. A. (2013) Engineering unnatural variants of plantazolicin through codon reprogramming, *ACS Chem. Biol. 8*, 1998-2008.

60. Hu, D. X., Grice, P., and Ley, S. V. (2012) Rotamers or diastereomers? An overlooked NMR solution, *J. Org. Chem. 77*, 5198-5202.

Chapter 3: Targeting aldehydes and ketones for reactivity based natural product discovery

I am grateful to Jonathan Tietz for helping to set up NMR experiments and advice on analyzing the data.

3.1 Introduction

Natural products have historically been a valuable source of medically important drugs and drug leads or have served as inspirations for synthetic approaches.[1-3] While novel natural products are still being discovered at a rapid pace, quickly filtering out known structures has become increasingly challenging.[4, 5] This is especially true for traditional activity-based screening which selects for abundantly produced products and often results in rediscovery of compounds commonly produced by many species, such as streptomycin in antibiotic screening.[6] Many natural products are easily missed by traditional screening due to production at very low levels (if at all) under laboratory conditions and but can be extremely interesting as tools to understand biological systems or as starting points for drug development.[3, 7, 8] A number of inventive methods to facilitate natural product discovery that circumvent the issues surrounding activity-based screening have been advanced, including new cultivation techniques, bioinformatics guided discovery, activation of silent gene clusters, and chemoselective enrichment.[5, 9-12] Similar to chemoselective enrichment, another strategy for the rapid identification of select natural products is to target a specific, reactive chemical handle that is generally not present on other cellular components. A recent example of this approach successfully targeted the α,β-unsaturated carbonyls present in dehydrated amino acid-containing natural products with a thiol nucleophile, resulting in the discovery of a novel thiopeptide.[13] We aimed to expand the scope of this approach to additional reactive handles as well as introduce a method for the facile identification of labeled products based on a unique isotopic signature.

While not exceedingly common, aldehydes and ketones can be found on various natural products, especially those produced by polyketide synthases (PKS).[14] For example, the ketone in erythromycin is formed due to the absence of a ketoreductase domain in the third module of the PKS and a similar strategy is commonly utilized in other PKS systems.[14, 15] One method for the installation of aldehydes is through the oxidation of hydroxyl groups by cytochrome P450 type enzymes, as with the PKS products

tylosin and rosamicin.[16, 17] Due to their rarity in most biological systems, aldehydes have previously been used for bioconjugation[18, 19] with numerous examples of labeling on oligonucleotides[20, 21] and glycoproteins.[22, 23] These bioconjugations are typically carried out with aminooxy or hydrazine groups to afford oxime or hydrazone linkages respectively.[18] Although oxime formation is sluggish for many substrates at neutral pH, sufficient reaction progress for detection of labeling can be achieved for aldehydes and more electrophilic ketones under mild conditions within a few hours.[24, 25] We therefore reasoned that the aminooxy group could likewise be utilized in a probe for the discovery of aldehyde, and possibly ketone, containing natural products.

A major hurdle of natural product discovery that we sought to overcome with this strategy is the often extremely low quantities produced by most biosynthetic gene clusters under laboratory conditions.[3] Coupled with a complicated metabolic background, labeling events could be easily overlooked for low abundance compounds. We believed that the introduction of a unique isotopic signature to our probe would help ameliorate this issue by providing a distinct peak distribution for detection with mass spectrometry (MS). This concept was recently expounded by Bertozzi and coworkers, who found that a dibromide motif provides a unique and readily detectable pattern by MS.[26] The naturally occurring 1:1 ratio of Br^{79} to Br^{81} results in a symmetrical triplet with major peaks at M, M+2, and M+4. Based on this distribution, Bertozzi and coworkers developed a mass pattern prediction program, IsoStamp,[26] which was successfully utilized for glycoproteome profiling.[27, 28] By including a dibromide motif on an aldehyde reactive probe, we hoped to facilitate the discovery of low abundance natural products through reactivity based screening tied to facilitated peak detection, especially in the context of a complicated metabolic background (Figure 3.1).

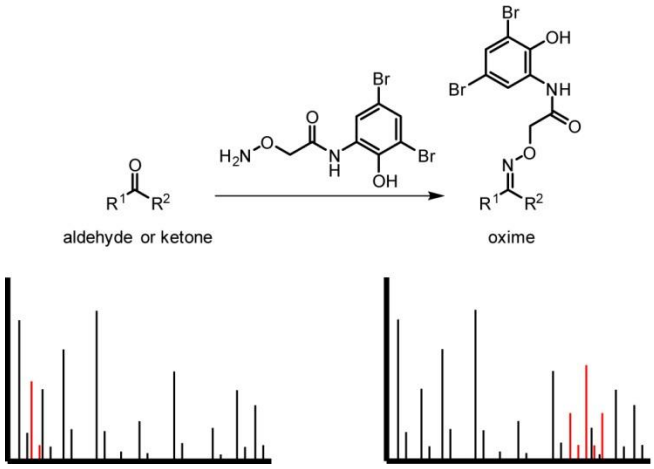
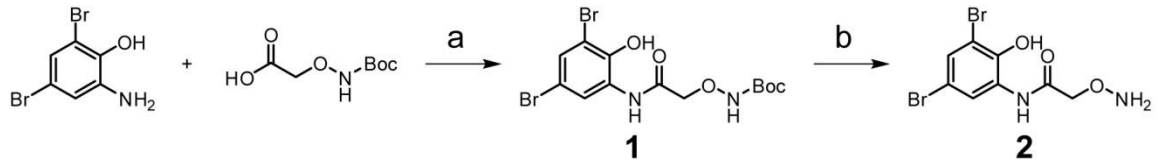


Figure 3.1. Reaction scheme for labeling aldehydes and ketones with a dibrominated aminooxy probe. Labeled compounds will show a mass shift accompanied by a distinctive isotopic distribution by mass spectrometry, allowing facile identification of peaks of interest. Diagrams representing the mass shift and isotopic distribution upon labeling are shown beneath the reaction.

3.2 Probe design and validation

Probe **2** was designed with an aminooxy warhead for the specific labeling of aldehyde and ketone chemical handles. While oxime formation can require hours or even days to reach completion at room temperature and neutral pH, complete labeling of the natural product is unnecessary for detection. We thus believed that **2** would be sufficient for labeling target natural products in 1 to 2 hours under mild conditions,[24] while providing a highly stable oxime linkage.[29] Probe **2** is also easily synthesized from readily available starting materials, requiring only a standard amide bond formation by EDC coupling followed by an acid deprotection (Scheme 3.1). A dibromide tag was included on **2** to provide a unique isotopic signature to aid in detection of labeled peaks by MS analysis (Figure 3.1). We believed that this signature would prove useful in situations where numerous extraneous signals from background metabolites resulted in overlapping peaks. The signature was also expected to be beneficial for low abundance compounds and for compounds that do not ionize well by MS. Finally, the distinctive isotopic distribution simplifies spectral analysis by providing an easily recognizable pattern that can be detected without comparison to a spectrum of the unreacted sample. Detection of labeled peaks could also be automated with software such as the IsoStamp program developed by Bertozzi and coworkers,[26] but we found visual analysis of the resultant spectra to be extremely rapid and sufficient for our purposes.



Scheme 3.1. Synthesis of dibromo aminooxy probe 2[a]

[a](a) 3-hydroxy-2-methylbenzoic acid, EDC, HOBt, THF (63%). (b) 4 M HCl in dioxane (55%).

With probe **2** in hand, we sought to establish a simple labeling and detection procedure and to validate the reactivity of the probe towards known aldehyde and ketone containing natural products. We selected a small panel of commercially available compounds with aldehydes and ketones in different chemical environments for this purpose. As oxime formation on aldehydes is much more rapid than on many ketones,[25] we first tested labeling on streptomycin. Streptomycin is an aminoglycoside antibiotic produced by a number of Actinomycetes, including *Streptomyces griseus*, that contains an aldehyde group installed enzymatically in the penultimate step of its biosynthesis (Figure 3.2A).[30] Reactions were performed with streptomycin and probe **2** in water for 2 hours at room temperature. The reaction mixture was then subjected to matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) analysis along with an unreacted standard. Near complete conversion of the peak corresponding to streptomycin (*m/z* 582.3 Da) to a peak with the unique isotopic distribution of a dibromide tag corresponding to labeled streptomycin (*m/z* 902.2 Da) was observed (Figure 3.2B). To confirm that probe **2** would also effectively label aldehydes in the context of a bacterial extract, the streptomycin producer *Steptomyces griseus* WC-3480 was grown on solid media and several colonies were used for a cell-surface extraction with water. As with commercial streptomycin, efficient labeling was readily observed (Figure 3.3). The labeling reaction was also successful with a range of organic solvents, including methanol, butanol, acetonitrile, ethyl acetate, and chloroform, indicating that the labeling reaction could be performed on extracts directly without the need to remove or dilute the extraction solvent. However, acetone and other aldehyde/ketone containing solvents cannot be used as the presence of the carbonyl rapidly quenches the probe.

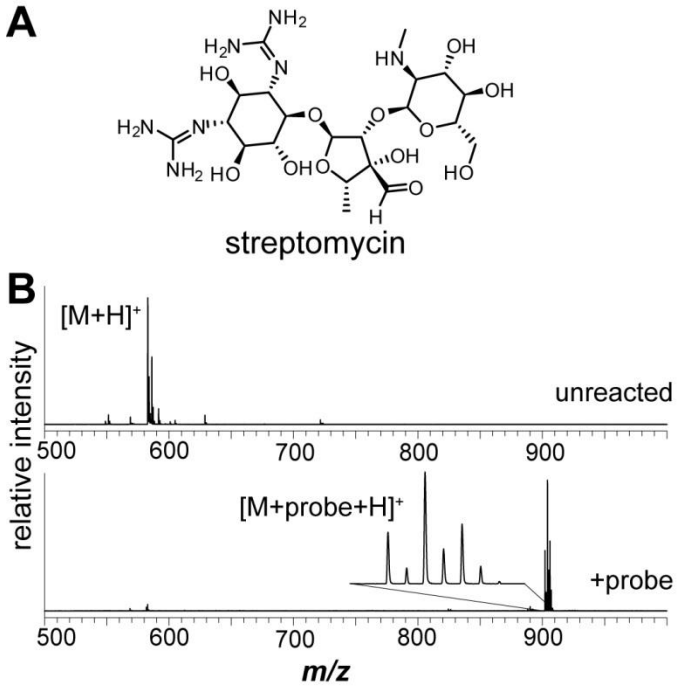


Figure 3.2. Labeling of commercial streptomycin. (A) Structure of streptomycin. (B) MALDI-TOF MS spectra of streptomycin either unreacted (top) or labeled with aminooxy probe **2** (bottom). The labeled streptomycin peak is shown magnified to display the unique dibromide isotopic pattern.

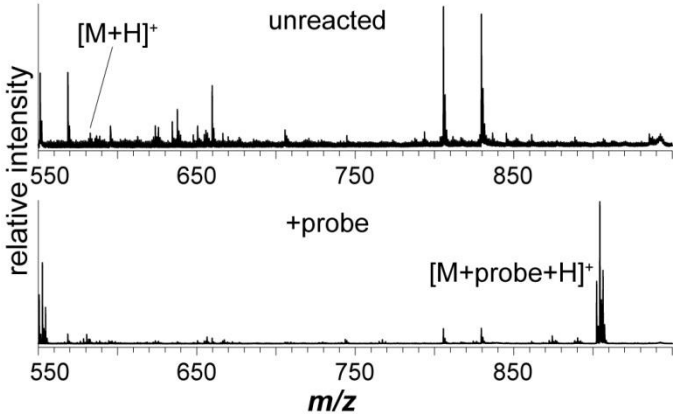


Figure 3.3. Labeling of streptomycin from *S. griseus*. MALDI-TOF MS spectra of extract from *S. griseus* WC3480 either unreacted (top) or labeled with aminooxy probe **2** (bottom). The labeled streptomycin peak is shown magnified to display the unique dibromide isotopic pattern.

As aryl aldehydes are known to form oximes more slowly than alkyl aldehydes,[25] the labeling reaction was also performed on anisaldehyde (Figure 3.4A). Although the unreacted compound was too small for detection with the MALDI-TOF-MS instrument employed in this study, a peak with a dibromine isotopic distribution was observed at the expected mass of the labeled product (*m/z* 456.9 Da)

(Figure 3.4B). In addition to confirming that probe **2** is effective at labeling aryl aldehydes, this result also supports the utility of this labeling strategy for the detection of very small molecules that are below the mass range of detection. Without the presence of the dibromide tag, the appearance of peaks at low *m/z* ratios in labeling reactions would be difficult to interpret due to the lack of detection of the unreacted natural product.

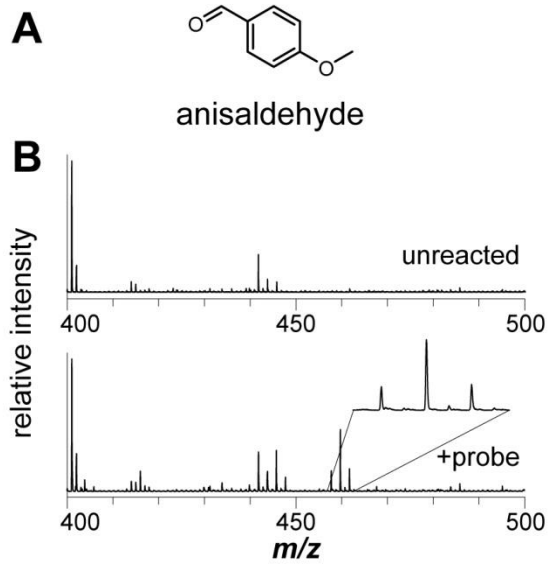


Figure 3.4. Labeling of anisaldehyde. (A) Structure of anisaldehyde. (B) MALDI-TOF MS spectra of anisaldehyde either unreacted (top) or labeled with aminooxy probe **2** (bottom). Unlabeled anisaldehyde is below the mass to charge ratio cut-off of the MALDI-TOF MS instrument used and is thus not shown. The labeled anisaldehyde peak is shown magnified to display the unique dibromide isotopic pattern.

The labeling reaction was next tested on ketone-containing natural products to determine if the reaction would proceed rapidly enough to be of use in labeling this reactive handle. The macrolide antibiotic erythromycin, which contains a ketone in the macrocyclic ring (Figure 3.5A), was tested first in the labeling reaction. Surprisingly, no detectable labeling was observed under the mild reaction conditions employed, even when the reaction was allowed to proceed for multiple days (Figure 3.5B). This result was unexpected given that alkyl ketones with similar neighboring groups are known to react efficiently with hydrazine reagents.[25] It is possible that erythromycin adopts a structure in solution that occludes the ketone and prevents oxime formation. If this were the case, other alkyl ketones would potentially be more susceptible to nucleophilic attack. The ketone containing natural products daunomycin and FK506 were

thus also tested and were found to efficiently label (Figure 3.6 and Figure 3.7). These data indicate that the exact chemical context of the ketone functionality is important in determining whether detectable labeling will be observed.

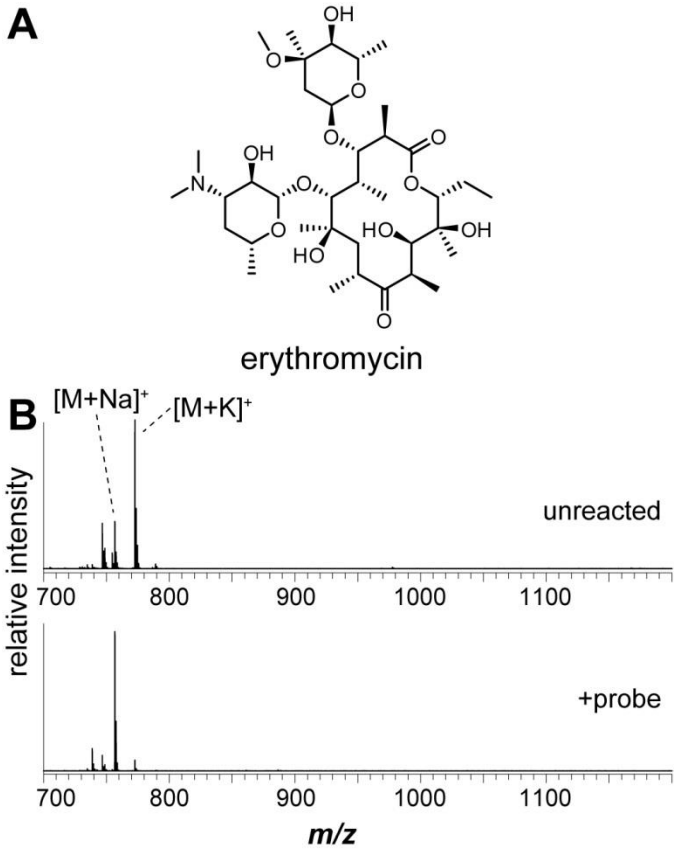


Figure 3.5. Labeling of erythromycin. (A) Structure of erythromycin. (B) MALDI-TOF MS spectra of erythromycin either unreacted (top) or reacted with aminooxy probe **2** (bottom). No labeling of erythromycin was observed.

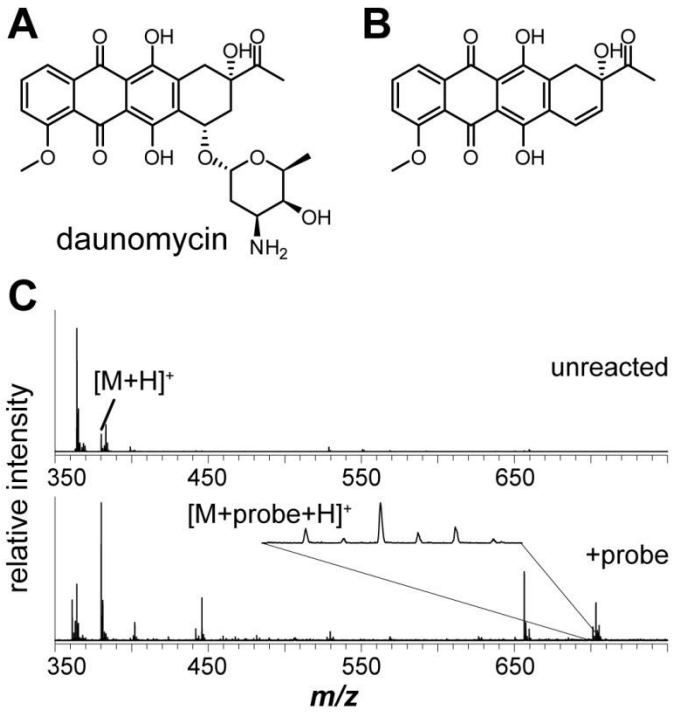


Figure 3.6. Labeling of commercial daunomycin (daunorubicin). (A) Structure of daunomycin. (B) Possible structure of an ionization-induced degradation product of daunomycin generated during MALDI ionization. Fully intact daunomycin produces a very weak or no signal by MALDI-TOF MS or ESI-MS. (C) MALDI-TOF MS spectra of daunomycin either unreacted (top) or labeled with aminooxy probe **2** (bottom). The labeled peak of the ionization-induced degradation product is shown magnified to display the unique dibromide isotopic pattern.

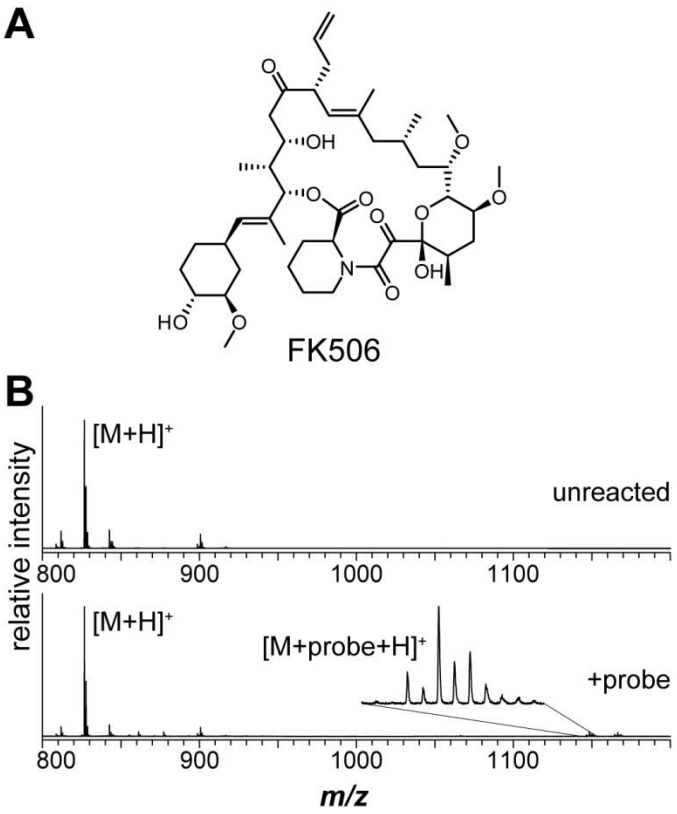


Figure 3.7. Labeling of commercial FK506 (tacrolimus). (A) Structure of FK506. (B) MALDI-TOF MS spectra of FK506 either unreacted (top) or labeled with aminooxy probe **2** (bottom). The labeled FK506 peak is close to the baseline and is shown magnified to display the unique dibromide isotopic pattern.

3.3 Screening of the NCI NP collection

Due in part to the results during probe validation on ketones, the National Cancer Institute's Natural Product (NCI NP) collection of 117 compounds was screening with probe **2**. The collection contains a large number of diverse ketones in the context of natural products, providing an ideal set of test cases to further elucidate the specificity of this method. Further, many of the compounds in the collection contain additional functionality, including other potential electrophiles. This set of compounds thus provides an indirect screen of functional groups that may display off-target reactivity with the probe.

The reactions were run under standardized conditions in methanol at room temperature and were analyzed by MALDI-TOF-MS in positive mode. A peak corresponding to the natural product was observed for 83 of 117 compounds; the remainder either did not ionize effectively or appeared to be degraded (either due to instability of the compounds in storage or due to ionization induced

fragmentation). Of the compounds that did not effectively ionize, most were small, hydrophobic, and devoid of functionality that might easily ionize. These compounds also tended to lack any functionality that could reasonably be expected to react with a strong nucleophile such as an aminooxy group and thus their exclusion from the screening data is likely not significant.

The collection contained two compounds with aldehyde groups. One was severely degraded but labeling was observed for the other. Both of the two methyl ketones and both of the two α-keto-carbonyl containing compounds labeled as well, although one of the methyl ketones was daunomycin which had been previously tested independently of the NCI NP collection (Figure 3.8). The remaining ketones in the collection can be separated into three groups, each of which had one or two compounds label: saturated aliphatic (2/5), unsaturated aliphatic (2/13), and aryl (1/7) ketones (Figure 3.8 and Figure 3.9). These compounds did not display robust labeling, with the labeled peak having an ion intensity less than 5% of the unlabeled ion. The lone exception to this trend was virginiamycin S1, which interestingly showed near complete labeling (Figure 3.10). The differences between compounds that labeled and those that did not were often minute, indicating that subtle differences in the chemical context of the natural product can sufficiently adjust the reactivity of the ketone and dictate labeling.

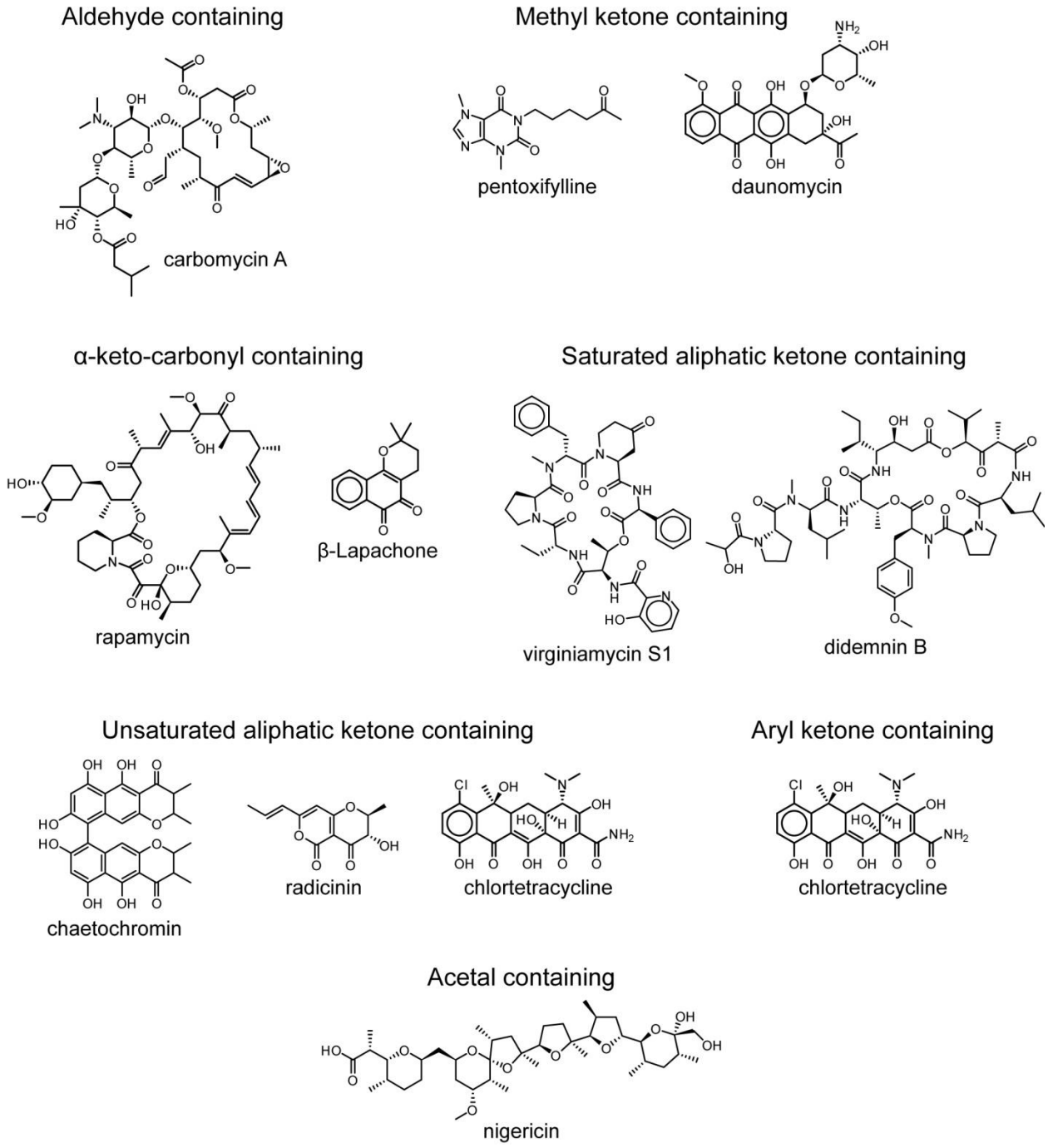


Figure 3.8. Structures of NCI NP collection compounds that displayed observable labeling with probe **2**.

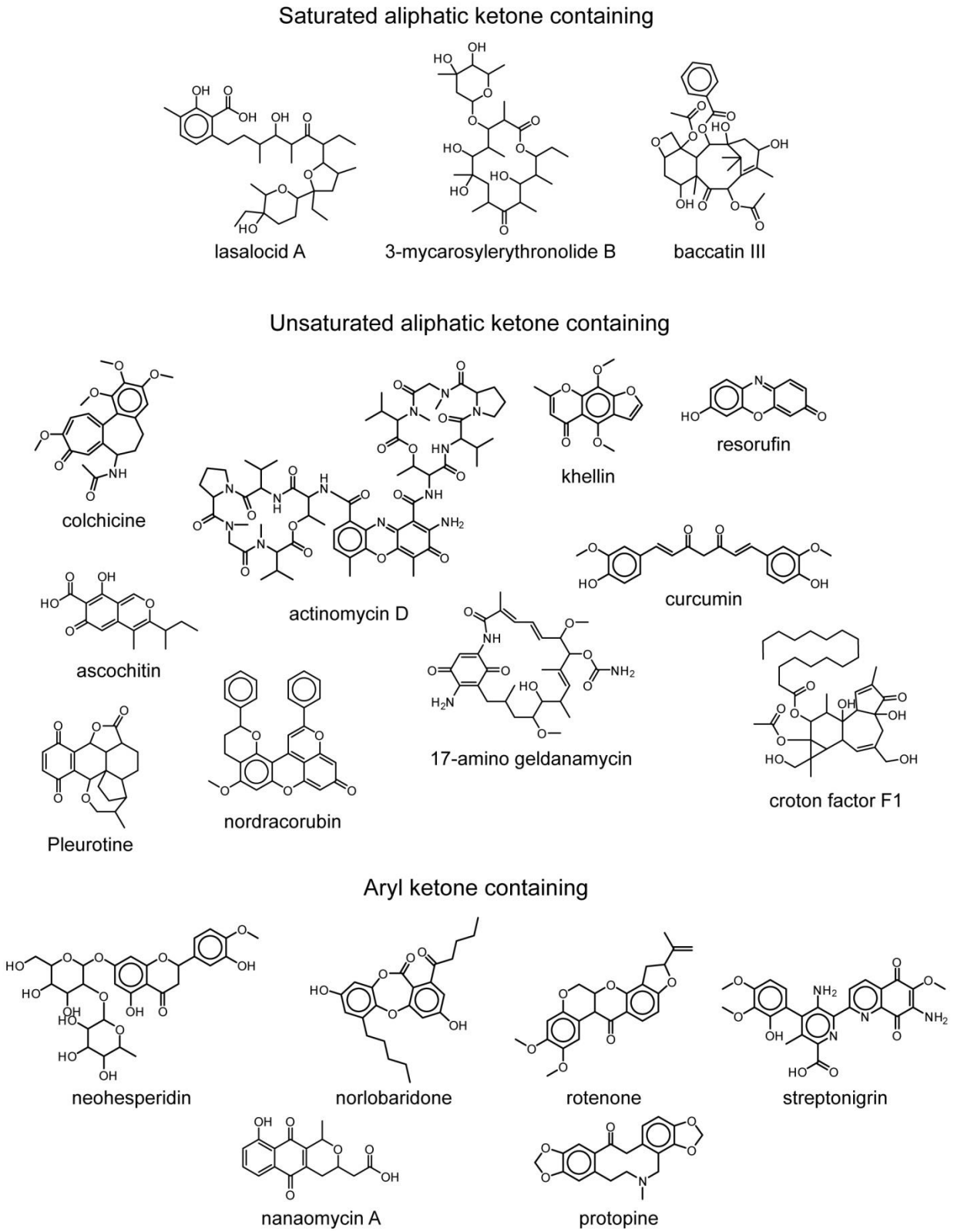


Figure 3.9. Structures of NCI NP collection compounds that did not labeling with probe **2**.

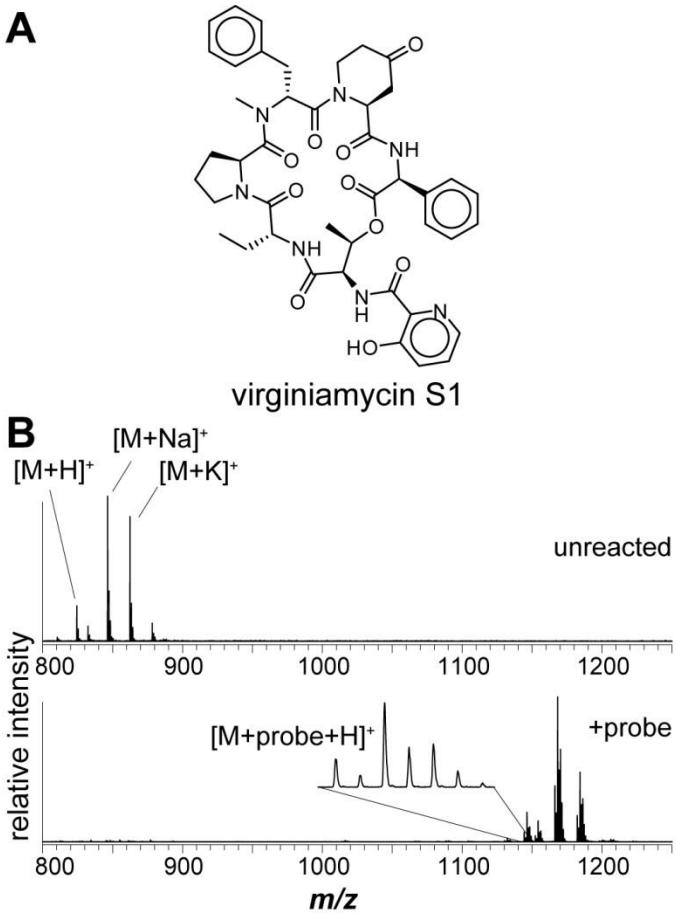


Figure 3.10. Labeling of commercial virginiamycin S1. (A) Structure of virginiamycin S1. (B) MALDI-TOF MS spectra of virginiamycin S1 either unreacted (top) or labeled with aminooxy probe **2** (bottom). The labeled virginiamycin S1 $[M+H]^+$ peak is shown magnified to display the unique dibromide isotopic pattern.

Compounds containing other potential electrophiles, including other carbonyls, epoxides, chloroalkanes, and nitriles, displayed no reactivity towards the probe. None of the glycoside-containing natural products (6 compounds) labeled, although the sugars in all of these cases were non-reducing. Given that the utility of oxime bioconjugation in oligonucleotide research stems from the ability of aminooxy warheads to label the open form of deoxyribose at abasic sites,[21] natural products containing reducing sugars may label. Interesting, 1 of 8 non-glycosidic acetals labeled, although the reason for this is unclear (Figure 3.8). A caveat to the analysis of any unlabeled compounds in this screen is that the NCI collection used may contain degraded compounds. Thus, a lack of labeling may not be due to a lack of reactivity, but rather from the degradation of the functional group of interest.

3.4 Screening of bacterial extracts

A previously generated extract collection that was created for Doroghazi et al.[10] was obtained for screening by oxime labeling. The collection contains extracts from 334 strains of Actinomycetes grown on solid media. Extracts from an additional 14 strains of Actinomycetes that were collected prior to obtaining the larger collection were also screened. From these 348 extracts, 36 hits were initially identified. Eleven of the strains containing hits were selected for verification based on their availability, of which six were confirmed to contain compounds that were labeled with probe **2**. The original extracts from the strains in which hits were not replicated had been in long term storage and may not have been stable over that time, perhaps forming ketones or aldehydes from the degradation of extracted glycans.

One of the hit strains contained two compounds that underwent complete labeling (*m/z* 582.3 and 744.3). The species was identified as *Streptomyces bikiniensis* subsp. *bikiniensis* ISP-5582 and an analysis of the strain quickly revealed it to be a known producer of streptomycin and mannosidostreptomycin (Figure 3.11). High resolution masses matched the hit masses to these known antibiotics and thus these hits were considered as further validation for the screening method. Another hit found in *Streptomyces albulus* B-3066 (*m/*z 606.3 Da) was then chosen for follow up as a high resolution mass did not correspond to any known structures in the natural product database Dictionary of Natural Products and due to the ease of growing the organism on a large scale (Figure 3.12A). The hit was found to be an analog of the natural product antipain, which we have named antipain-cit (Figure 3.12B).

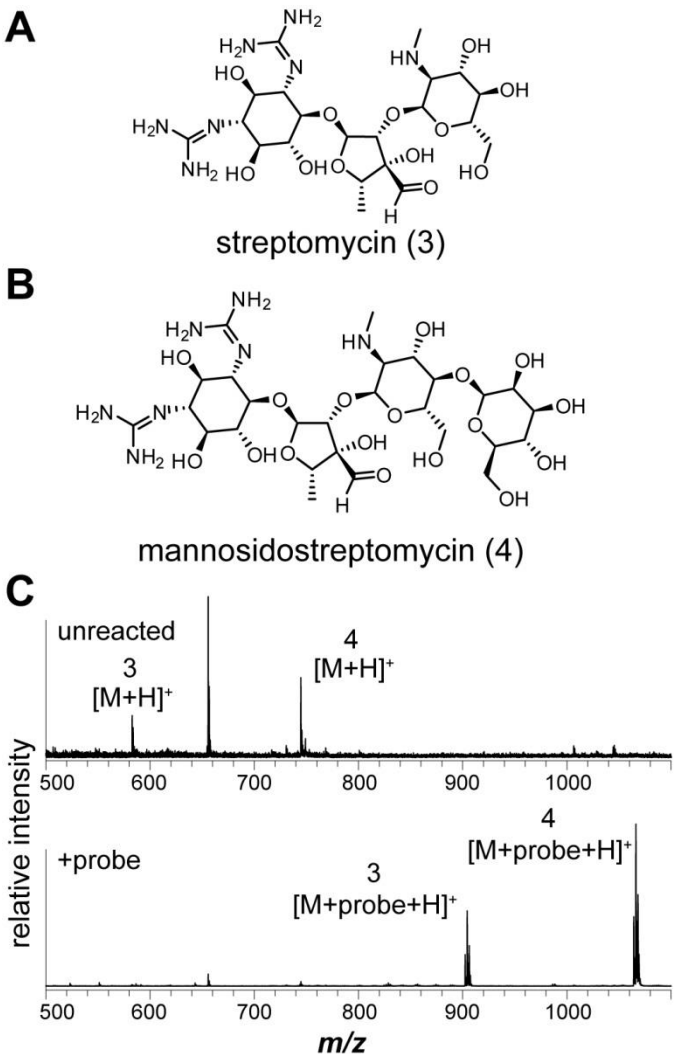


Figure 3.11. Labeling of streptomycin and mannosidostreptomycin (streptomycin B) in an extract from *S. bikiniensis* subsp. *bikiniensis* ISP-5582. (A) Structure of streptomycin (**3**). (B) Structure of mannosidostreptomycin (**4)**. (C) MALDI-TOF MS spectra of *S. bikiniensis* extract either unreacted (top) or labeled with aminooxy probe **2** (bottom).

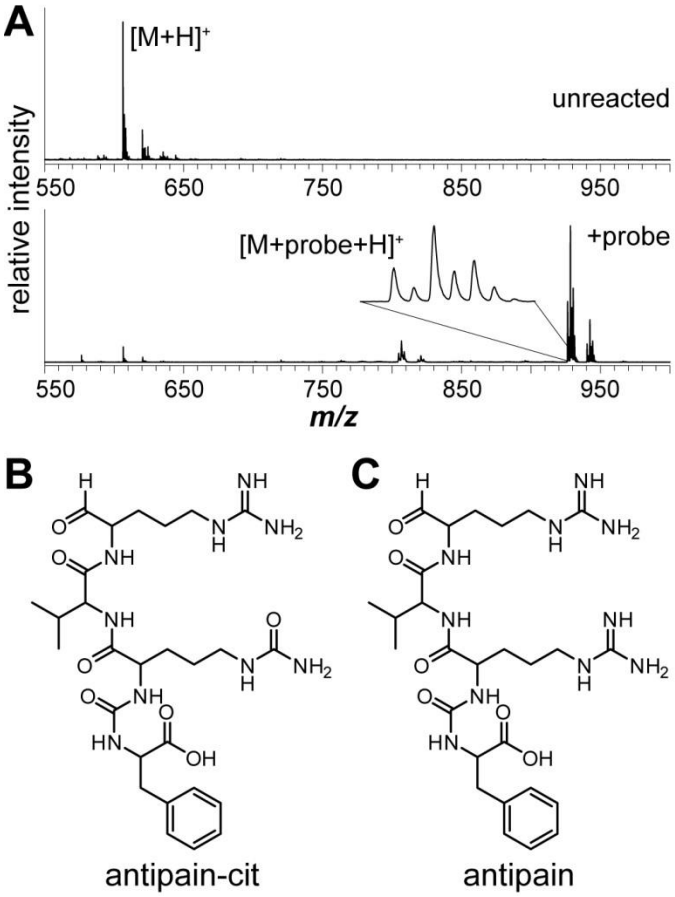


Figure 3.12. Labeling antipain-cit in *S. albulus* B3066. (A) MALDI-TOF MS spectra of partially purified bacterial extract either unreacted (top) or labeled with aminooxy probe **2** (bottom). The labeled peak is shown magnified to display the unique dibromide isotopic pattern. (B) Proposed structure of antipain-cit. (C) Structure of antipain.

3.5 Isolation and structural determination of antipain-cit

Despite numerous efforts to purify antipain-cit by MPLC, HPLC, and ion exchange chromatography, conditions that did not cause excessive smearing of the compound were not identified. This was exacerbated by a weak chromophore, resulting in no defined peak in the UV trace during HPLC purification. Therefore, extract from *S. albulus* B3066 was labeled after desalting with a solid phase extraction column. The chromophore present on probe **2** provided a distinct signature in the UV trace and it quickly became apparent that antipain-cit was actually a collection of two major and two minor isomers (Figure 3.13A). One of the major isomers (antipain-cit-2) was purified in sufficient quantities to allow structural characterization (Figure 3.13B).

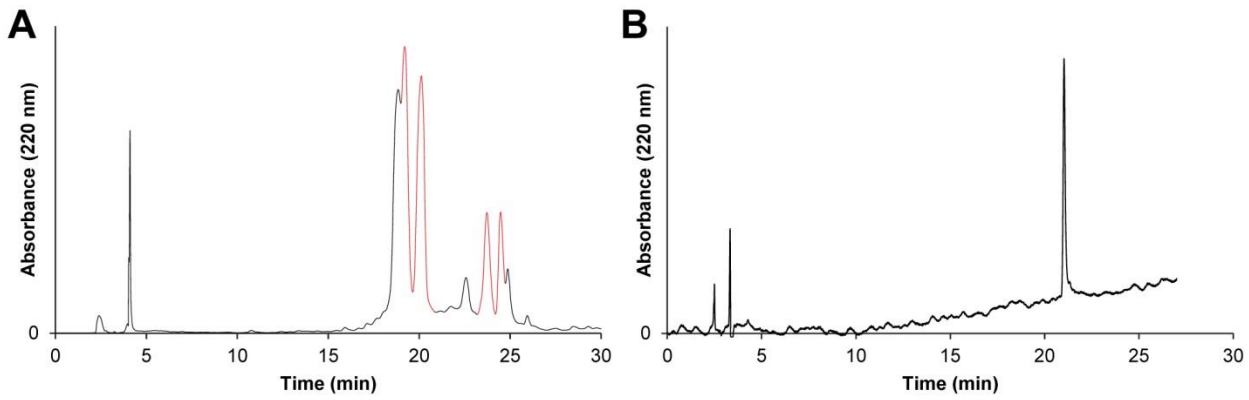


Figure 3.13. HPLC traces of labeled antipain-cit. (A) HPLC trace showing the 220 nm absorbance of partially purified, labeled antipain-cit. The peaks corresponding to the isomers of labeled antipain-cit are highlighted in red. (B) HPLC trace showing the 220 nm absorbance of purified antipain-cit for quality control. The two small peaks eluting before 5 min are part of the void and appear in blank runs.

The mass spectrum of the labeled product displayed an $[M + H]^+$ ion of *m/z* 926.2156 while an analysis of the collision induced dissociation (CID) spectrum only suggested the presence of a phenylalanine residue (Figure 3.14). NMR spectroscopy was then used to establish that the structure was highly similar to the known natural product antipain, with the high resolution mass suggesting the substitution of one of the arginine residues for a citrulline (Figure 3.12B, C). Connectivity was determined by ^{1}H-^{1}H COSY, ^{1}H-^{1}H TOCSY, ^{1}H-^{1}H NOESY, ^{1}H-^{13}C HSQC, and ^{1}H-^{13}C HMBC (Table 3.1, Figure 3.15, and Figure 3.16). However, the location of the citrulline residue could not be conclusively determined due to the absence of cross peaks for the protons on the guanidine and urea groups of the arginine and citrulline respectively. Additional experiments will be required to conclusively determine the location of each residue.

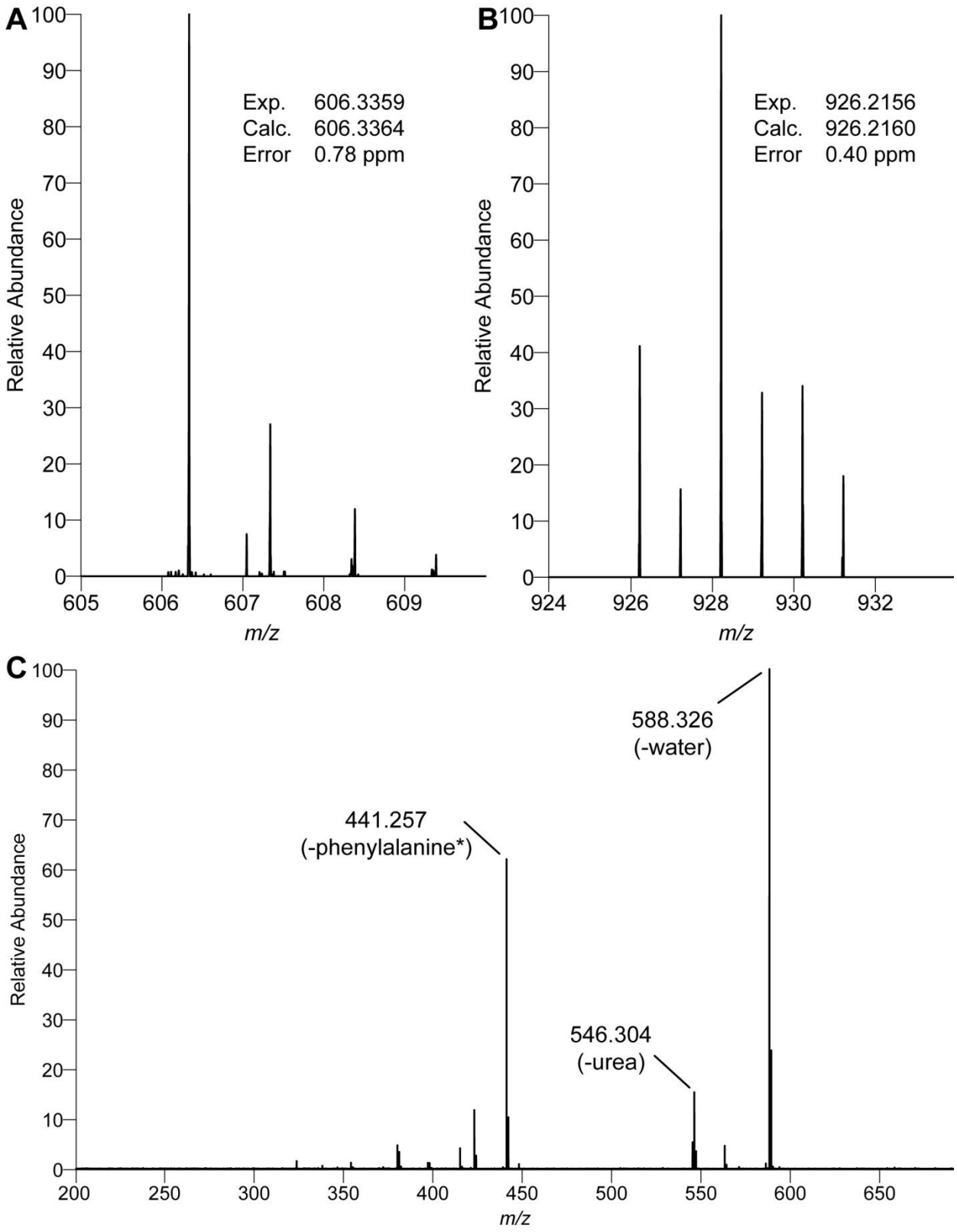


Figure 3.14. High resolution Fourier transform mass spectrometry (FT-MS) analysis of antipain-cit (A) and labeled antipain-cit (B). CID spectrum of *m/z* 606 Da with monoisotopic masses given. *The loss of

Figure 3.14. (cont.)
phenylalanine corresponds to the complete amino acid rather than the loss of 147 Da usually seen in peptide fragmentation. This is due to the urea connection between the phenylalanine and the remainder of the natural product.

Position	^{1}H δ, ppm (*J*)	^{13}C δ, ppm	COSY	TOCSY	HMBC	NOESY
A	8.18 d (3.0 Hz)	119.42	B	B	B	-
B	7.16 d (3.0 Hz)	128.70	A	A, C	A	-
C	8.16 d (3.0 Hz)	-	-	B	-	-
D	-	168.30	-	-	-	-
E	4.56 s	72.71	-	-	D	-
F	7.57 d (5.25 Hz)	152.50	G	G, H, I, J_a, J_b, L	G, H	G
G	4.47 q (6.75 Hz)	48.68	F, H	F, H, I, J_a, J_b, L	F, M	F, H
H	1.79 m	28.38	G, I	F, G, I, J_a, J_b, L	F, G, J	G
I	1.59 m	24.68	H, J_a, J_b	F, G, H, J_a, J_b, L	J	J
J	a: 3.12 m	40.40	I, J_b	F, G, H, I, J_b, L	H, I, K	I
	b: 3.17 m		I, J_a	F, G, H, I, J_a, L	H, I, K	I
K	-	157.24	-	-	-	-
L	8.12 m	-	-	F, G, H, I, J_a, J_b	-	-
M	-	172.06	-	-	-	-
N	4.06 d (6.0 Hz)	59.27	O	O, P, Q, R	M, O, P, Q, S	L, O, R
O	2.11 dq (6.75, 13.7 Hz)	29.84	N, P, Q	N, P, Q, R	N, P, Q	N, P, Q
P	0.87 d (6.75 Hz)	17.11	O	N, O, Q, R	N, O, Q	O
Q	0.91 d (6.75 Hz)	18.44	O	N, O, P, R	N, O, P	O
R	7.58 m	-	-	N, O, P, Q	-	N
S	-	174.40	-	-	-	-
T	4.03 dd (5.25, 8.25 Hz)	54.55	U_a, U_b	U_a, U_b, V, W, X, Z	S	U_a, Z
U	a: 1.76 m	28.99	T, U_b, V	T, U_b, V, W, X, Z	-	T
	b: 1.52 m		T, U_a, V	T, U_b, V, W, X, Z	-	-
V	1.51 m	26.07	U_a, U_b, W	T, U_a, U_b, W, X, Z	W	W
W	3.10 m	38.97	V	T, U_a, U_b, V, X, Z	V, Y	V
X	6.21 m	-	-	T, U_a, U_b, V, W, Z	-	-
Y	-	160.95	-	-	-	-
Z	6.57 d (6.75 Hz)	-	-	T, U_a, U_b, V, W, X	-	T, A'
A'	6.00 d (5.25 Hz)	-	-	B', D'_a, D'_b	-	Z
B'	4.30 m	56.87	D'_a, D'_b	A', D'_a, D'_b	C', D'	D'_a, D'_b
C'	-	177.96	-	-	-	-
D'	a: 3.13 m	38.30	B', D'_b	A', B', D'_b	B', E', F'	B', D'_b, F'
	b: 2.94 dd (7.5, 13.5 Hz)		B', D'_a	A', B', D'_a	B', E', F'	B', D'_a, F'

Table 3.1. NMR assignments for labeled antipain-cit.

E’	-	138.36	-	-	-	-
F'	7.23 m	129.15	G'	G', H'	D', E', G', H'	D'_a, D'_b
G'	7.22 m	127.82	F', H'	F', H'	H'	-
H'	7.14 m	125.86	G'	G', F'	-	-

Table 3.1. (cont.)
Abbreviations: s, singlet; d, doublet; q, quartet; m, multiplet

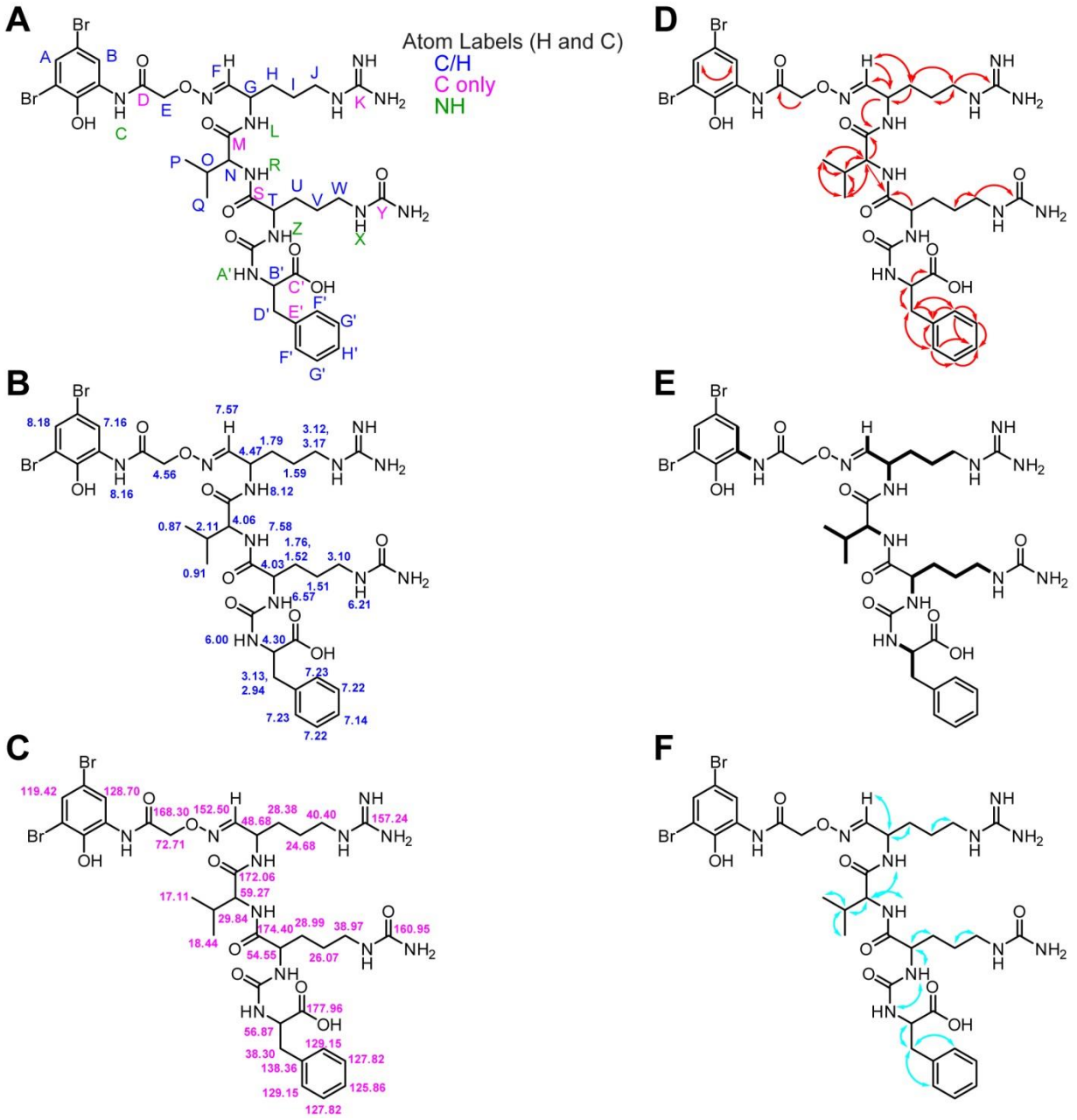


Figure 3.15. NMR assignments for labeled antipain-cit. (A) Labeling scheme for NMR peak assignments. (B) ^{1}H peak assignments (δ ppm). (C) ^{13}C peak assignments (δ ppm). (D) Though-molecular correlations from ^{1}H-^{13}C HMBC connectivity. (E) Though-molecular correlations from ^{1}H-^{1}H COSY/TOCSY connectivity. (F) Though-space correlations from ^{1}H-^{1}H NOESY connectivity.

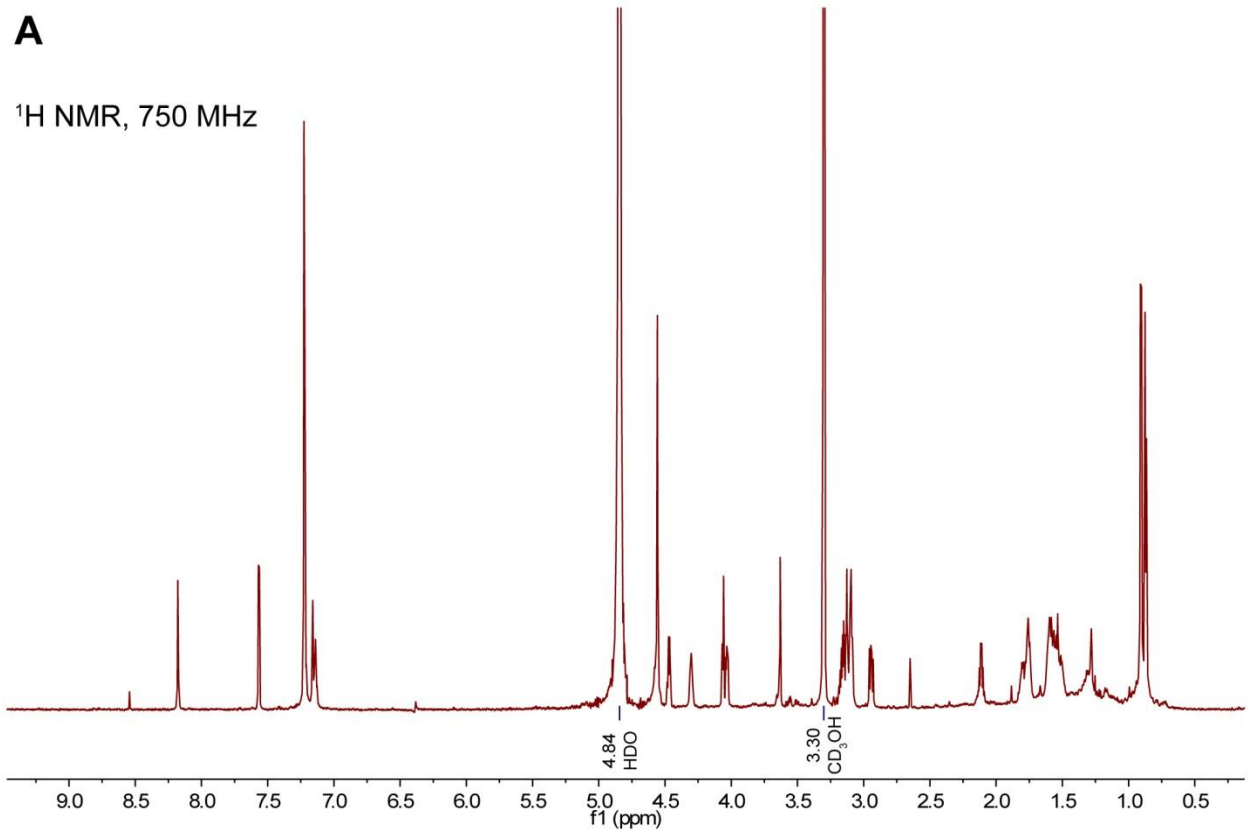


Figure 3.16. NMR spectra for labeled antipain-cit. (A) ^{1}H NMR in CD_3OD. (B) ^{1}H-^{1}H COSY in CD_3OD. (C) ^{1}H-^{1}H TOCSY in CD_3OD. (D) ^{1}H-^{1}H TOCSY in CD_3OH. (E) ^{1}H-^{13}C HSQC in CD_3OD. (F) ^{1}H-^{13}C HMBC in CD_3OD. (G) ^{1}H-^{1}H NOESY in CD_3OH.

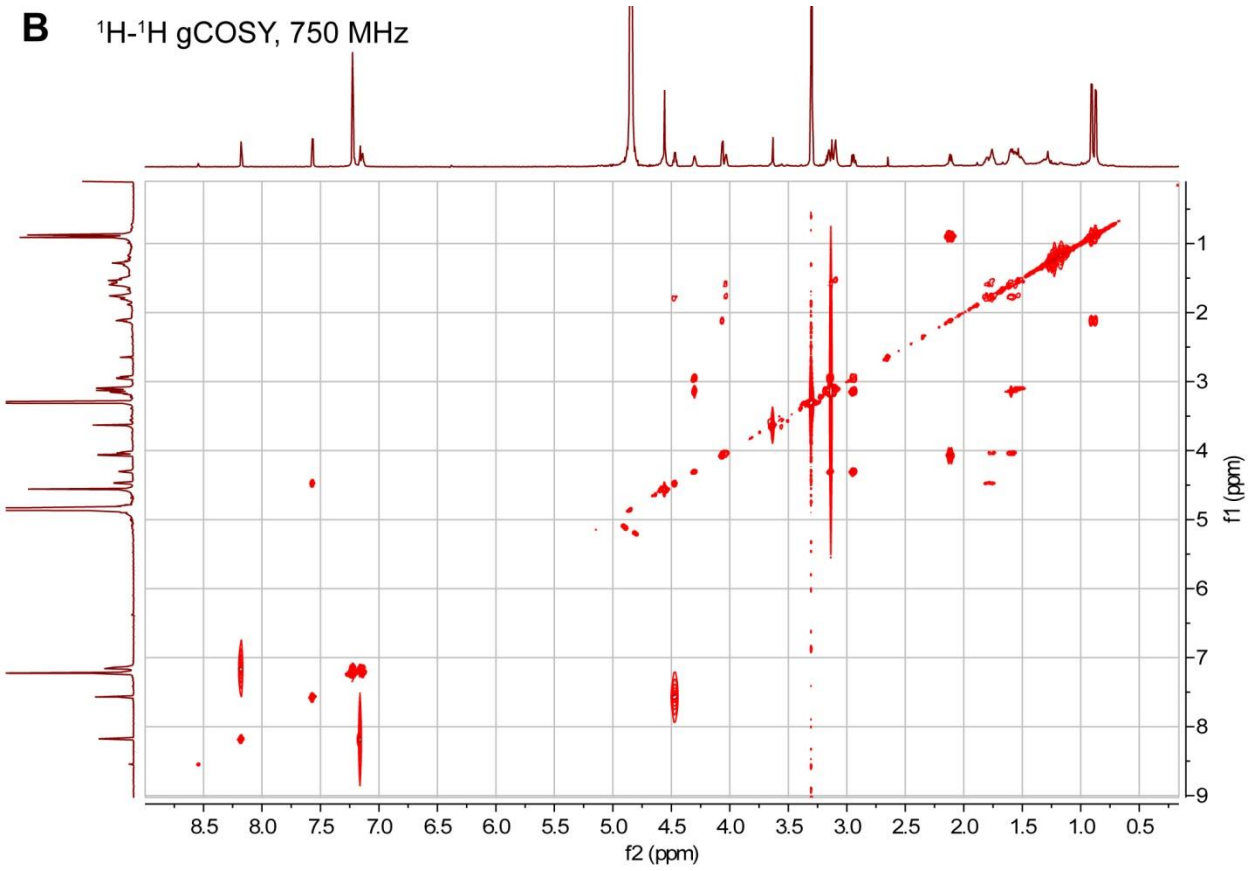


Figure 3.16. (cont.)

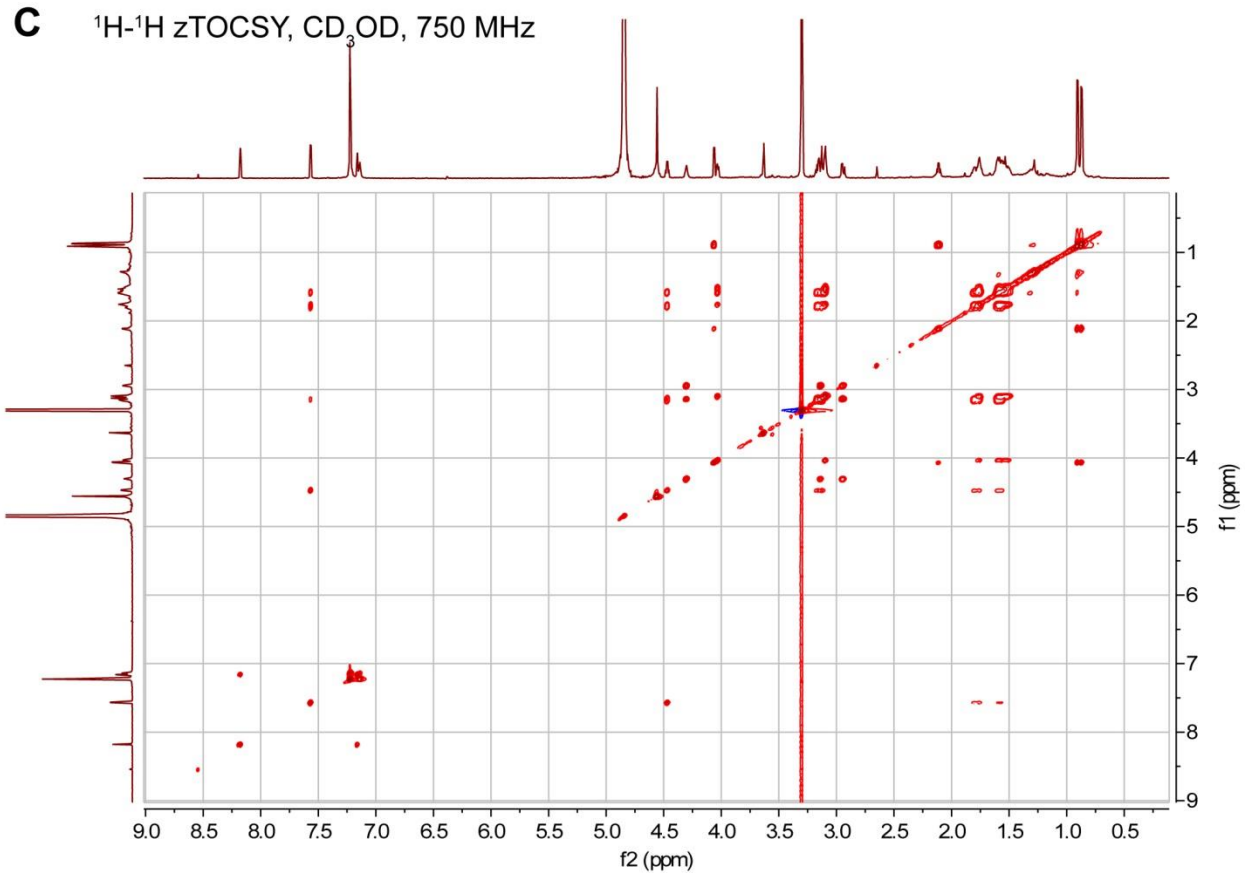


Figure 3.16. (cont.)

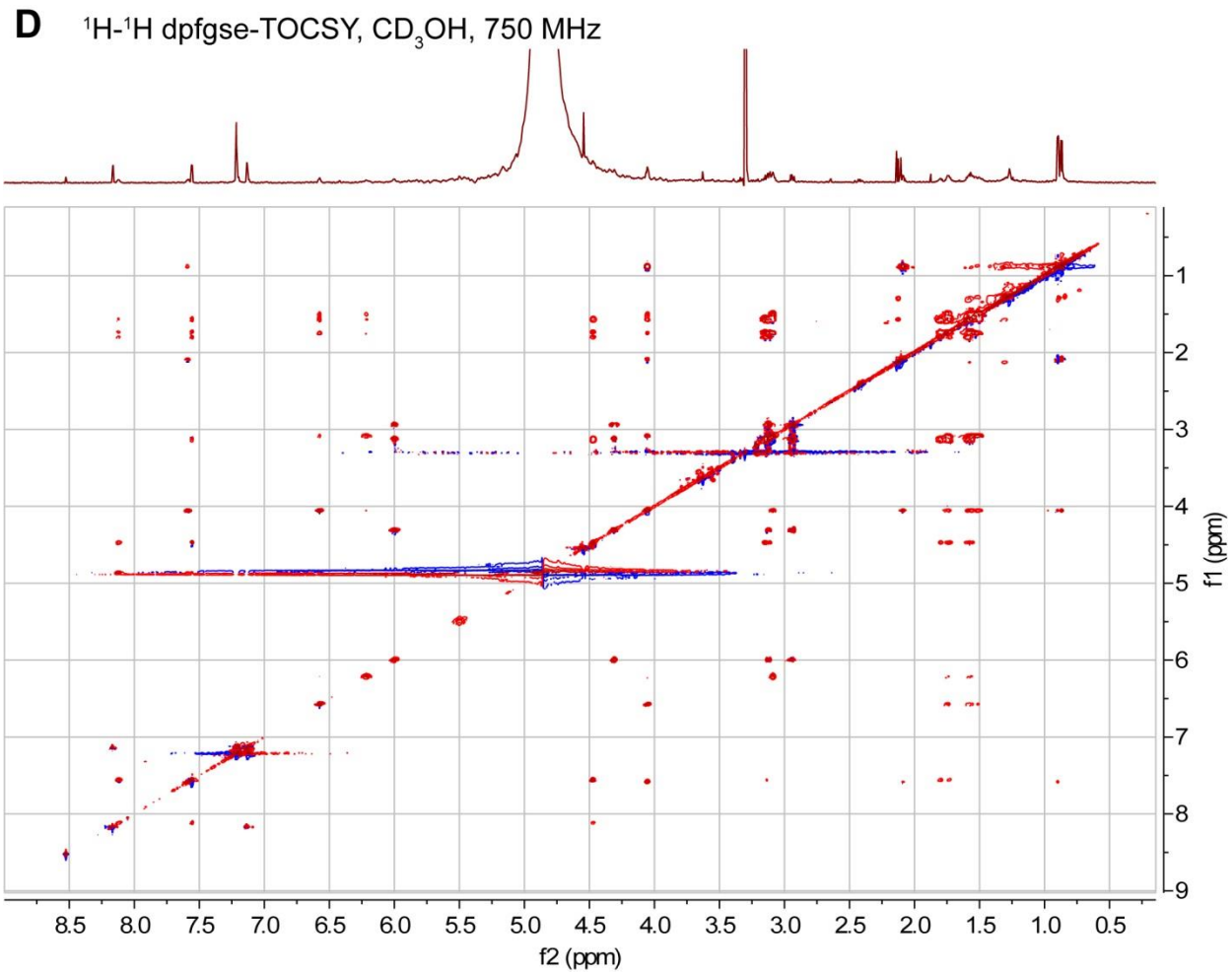


Figure 3.16. (cont.)

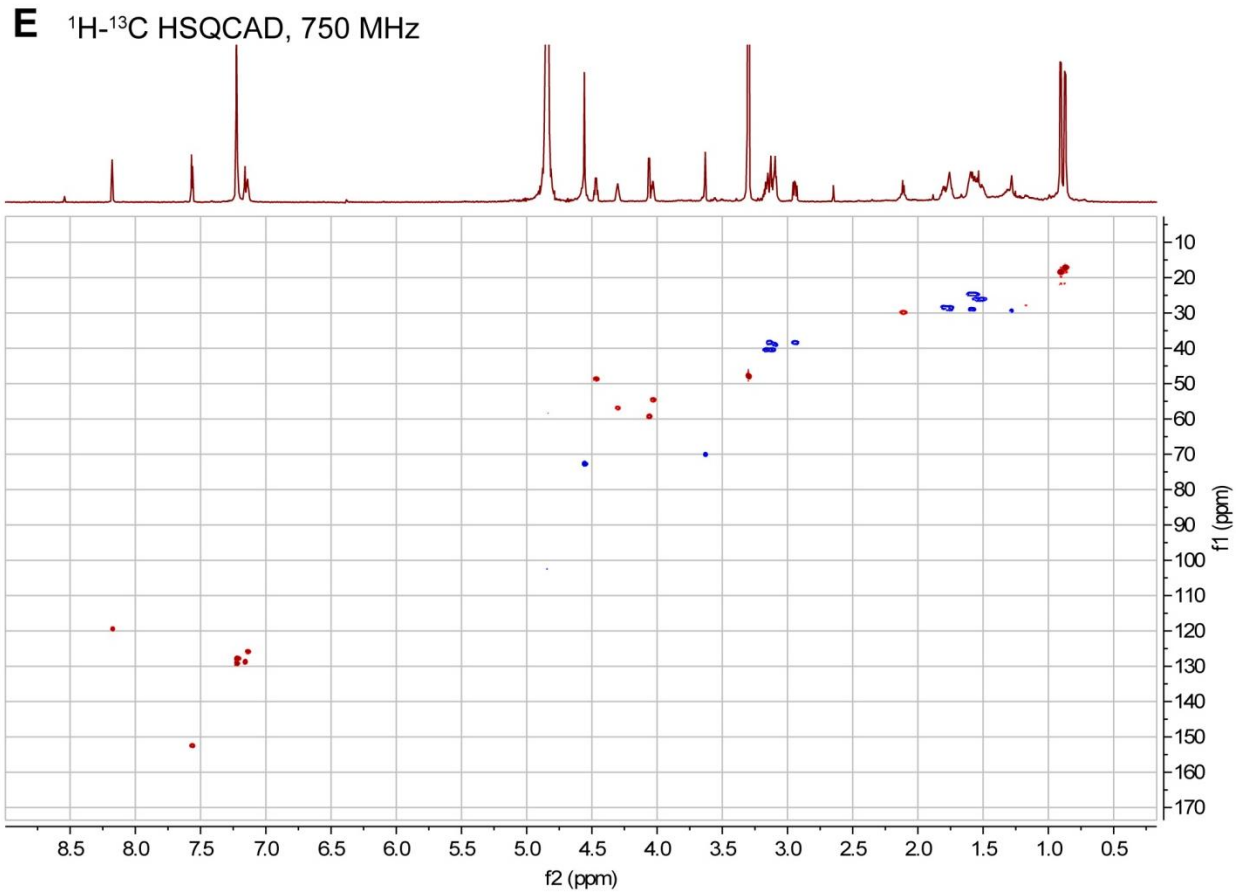


Figure 3.16. (cont.)

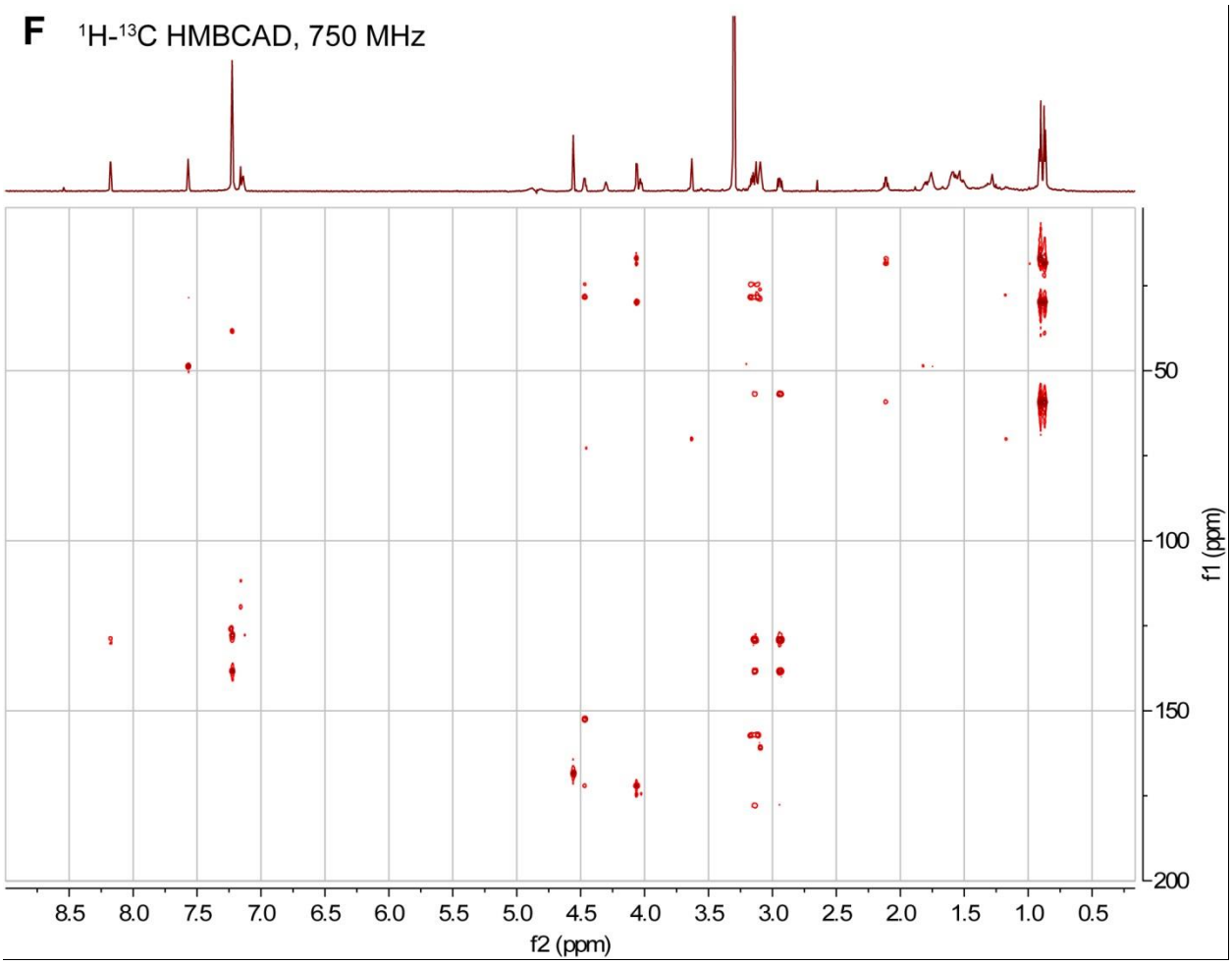


Figure 3.16. (cont.)

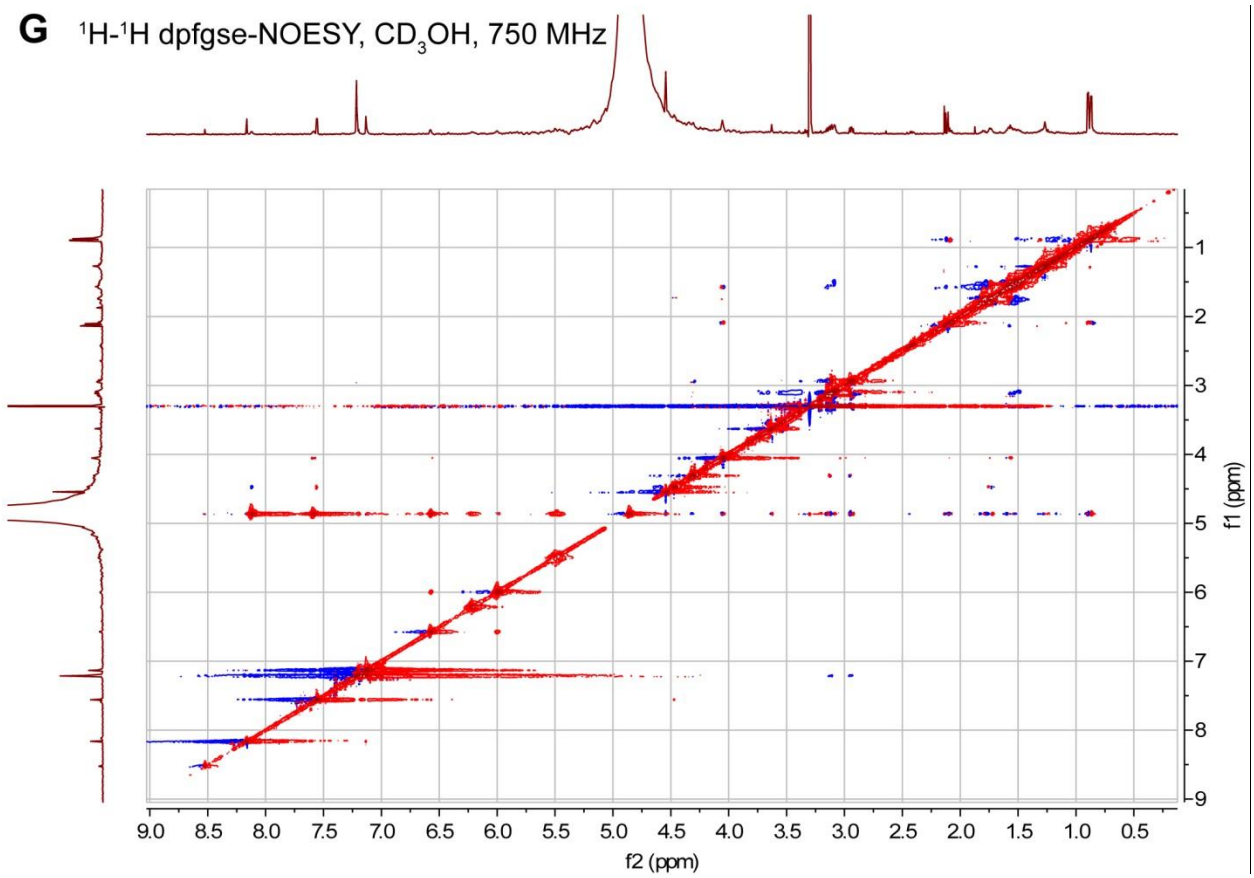


Figure 3.16. (cont.)

Antipain is a natural product produced by several species of Actinomycetes including *Streptomyces yokosukanensis* and *Streptomyces michiganensis*.[31] The compound was found to possess inhibitory activity against serine and cysteine proteases and has been used commercially.[31] One other analog of antipain (antipain Y) was recently described that contains a tyrosine in place of the phenylalanine residue.[32] Antipain is known to exist as a mixture of the D- and L-arginal, with the aldehyde primarily tied up in either a cyclic hemiaminal forming from the ε-nitrogen or in a hydrate.[33] As two stereoisomers result from formation of the cyclic hemiaminal, antipain thus exists as a complex mixture of eight isoforms: four cyclic hemiaminals, two hydrates, and two unhydrated aldehydes.[33] This complexity as well as continuing conversion between the various isoforms is likely the cause of the difficulties encountered during purification with reverse phase liquid chromatography. The presence of both D- and L-arginal also likely accounts for two of the detected isomers in the labeled product, while

the other two isomers may arise from racemization of a second amino acid. Purification of unlabeled antipain-cit is ongoing and will allow for biological assays, although it is expected that it will also function as a protease inhibitor. The biosynthetic origin of antipain is currently unknown but we hope analysis of the genome after sequencing is completed will illuminate a gene cluster with the capacity to produce this modified tetrapeptide.

3.6 Summary and outlook

In this study, we have developed a new tool for reactivity based screening for natural product discovery based on oxime formation. Probe **2** selectively labels aldehydes and certain ketones under extremely mild conditions, and the method utilizes commonly available laboratory equipment. Although many ketones do not label to an observable extent under the reaction conditions used here, more aggressive reaction conditions or previously reported catalysts[34, 35] could be employed to widen the scope of the reaction. However, utilizing conditions with significantly highly reactivity may cause problems with labeling other cellular components, such as sugars. Aldehydes are not generally found on primary metabolites so hits from screening with mild conditions have a high probability of being natural products of interest. The lack of off-target reactivity also allows large excesses of probe to be used without the need for accurate stoichiometric calculations. The dibromide tag on probe **2** allows for rapid analysis of mass spectra for labeled products. The unique isotopic distribution aids in determining true labeling events in noisy spectra and for products that are produced in small quantities or that do not ionize well.

The use of oxime ligations in natural product discovery was validated by labeling several aldehyde or ketone containing natural products, including streptomycin, daunomycin, and virginiamycin S1. Screening of a collection of 348 bacterial extracts led to the discovery of a new natural product, antipain-cit, as well as the dereplication of streptomycin and mannosidostreptomycin. Labeling of antipain-cit with probe **2** facilitated the isolation of a single isomer for structural elucidation. The structure of antipain-cit was determined through high resolution MS and NMR spectroscopy. Biological activity testing will be performed once unlabeled material is isolated, although it is likely that antipain-cit also acts as a serine and cysteine protease inhibitor. Sequencing of the genome of *S. albulus* B3066 is

ongoing and identification of the gene cluster responsible for the production of antipain-cit will hopefully provide insight into how the aldehyde and urea linkage moieties are installed, as these are both fairly uncommon and understudied groups in natural products.

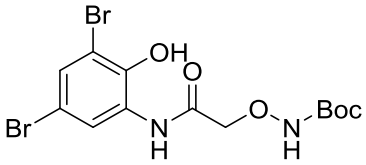
3.7 Experimental

3.7.1 Materials

All chemicals were purchased from Sigma-Aldrich, VWR, or Fisher Scientific and used without further purification unless otherwise specified.

3.7.2 Synthesis

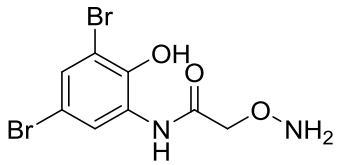
^{1}H and ^{13}C NMR spectra were collected on Varian Inova 500 MHz spectrometers. All ^{1}H and ^{13}C spectra were referenced to the solvent peaks. High-resolution mass spectrometry (HRMS) data were obtained on a Micromass Q-TOF Ultima tandem quadrupole mass-spectrometer at the University of Illinois at Urbana-Champaign Mass Spectrometry Laboratory.



Tert-butyl (2-((3,5-dibromo-2-hydroxyphenyl)amino)-2-oxoethoxy)carbamate (1)

To a solution of (Boc-aminooxy)acetic acid (200 mg, 1.05 mmol) in dry tetrahydrofuran (10 mL) was added 2-amino-4,6-dibromophenol (294 mg, 1.10 mmol), 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (220 mg, 1.15 mmol), and hydroxybenzo-triazole hydrate (186 mg, 1.15 mmol). The solution was stirred at room temperature for 18 h. The reaction was then taken up in ethyl acetate and washed twice with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate and concentrated by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 0-25% ethyl acetate in hexanes) to yield **1** as an orange solid (292 mg, 63%). ^{1}H NMR (500 MHz, $CDCl_3$) δ ppm 10.66 (br, 1H), 9.64 (br, 1H), 7.92 (s, 1H), 7.52 (d, J = 2.5 Hz, 1H), 7.43 (d, J = 2 Hz, 1H), 4.52 (s, 2H), 1.52 (s, 9H). ^{13}C NMR

(500 MHz, $CDCl_3$) δ ppm 169.49, 158.67, 145.26, 132.43, 127.16, 124.61, 114.28, 111.52, 84.51, 76.28, 28.08. HRMS (*m/z*): $[M + Na]^+$ calc. for $C_{13}H_{16}N_2O_5Br_2Na$, 460.9324; observed, 460.9320.



2-(aminooxy)-N-(3,5-dibromo-2-hydroxyphenyl)acetamide (2)

The boc protected probe **1** (122 mg, 0.277 mmol) was dissolved in 4 M HCl in dioxane (3 mL) and stirred at room temperature for 3 h. The reaction was taken up in ethyl acetate and washed twice with saturated sodium bicarbonate and once with brine. The ethyl acetate fraction was dried over sodium sulfate and concentrated by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 0-5% methanol in dichloromethane) to yield **2** as a white solid (52 mg, 55%). ^{1}H NMR (500 MHz, $(CD_3)_2SO$) δ ppm 8.02 (d, *J* = 2 Hz, 1H), 7.51 (d, *J* = 2.5 Hz, 1H), 4.18 (s, 2H). ^{13}C NMR (500 MHz, $(CD_3)_2SO$) δ ppm 169.83, 144.49, 129.60, 129.35, 123.62, 112.33, 110.91, 74.31. HRMS (*m/z*): $[M - H]^-$ calc. for $C_8H_7N_2O_3Br_2$, 336.8823; observed, 336.8816.

3.7.3 Test labeling reactions

Labeling reactions for streptomycin, daunomycin, 4-anisaldehyde, and FK-506 were prepared in water, methanol, or ethanol with a final concentration of 1 μM aldehyde or ketone and 1 mM probe **2** from 10x stocks (in water or ethanol). The choice of reaction solvent did not significantly affect labeling. The reactions were run at room temperature for 3 h with occasional manual shaking. The reactions were analyzed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) in reflector positive mode using α-cyano-4-hydroxycinnamic acid (CHCA) as a matrix.

3.7.4 Screening of the NCI NP collection

The Natural Products Set III, consisting of a select set of 117 natural products, was obtained from the National Cancer Institute's Developmental Therapeutics Program. The compound set was reconstituted by the addition of 18 μL of DMSO to give a final stock of 20 μL at 10 mM for each compound. The compounds were further diluted to 1 mM with DMSO before use in screening. The screen

was performed by the addition of 3 µL of the 1 mM natural product stock solutions to 24 µL methanol followed by the addition of 3 µL of a 10 mM stock of **2** in methanol. The reactions were sealed and incubated at room temperature without stirring for 24 h. The reactions were analyzed by MALDI-TOF MS in reflector positive mode using α-cyano-4-hydroxycinnamic acid (CHCA) as a matrix. Matrix suppression was disabled and the detection range was set from the minimum allowed (200 Da) to 2000 Da.

3.7.5 Verification of streptomycin labeling in *S. griseus* and *S. bikiniensis* extract

Frozen stocks of *S. griseus* WC-3480 and *S. bikiniensis* ISP5582 were streaked out on agar plates of ATCC172 media (10 g/L glucose, 20 g/L soluble starch, 5 g/L yeast extract, 5 g/L N-Z amine type A (Sigma C0626), 1 g/L reagent grade calcium carbonate, pH 7.3) and grown at 30 °C for 10 d. Several colonies were then scrapped from the surface of the agar and suspended in 60% acetonitrile in water. The suspension was vortexed vigorously before use in labeling reactions. Reactions were set up with 9 µL of extract and 1 µL of a 10 mM solution of **2** in methanol or just methanol as a control. The reactions were incubated at room temperature overnight. The reactions were analyzed by MALDI-TOF MS in reflector positive mode using α-cyano-4-hydroxycinnamic acid (CHCA) as a matrix.

3.7.6 Growth and extraction for extracts collected in this study

Seed cultures of actinobacteria (5 mL) were grown in ATCC172 media at 30 °C on a tube roller for 4-7 d. A 1 mL portion of the seed cultures were used to inoculate 15 cm agar plates (60 mL media per plate) of ATCC172 (with 15 g/L agar), ISP4 (10 g/L soluble starch, 1 g/L potassium phosphate dibasic, 1 g/L magnesium sulfate heptahydrate, 1 g/L sodium chloride, 2 g/L ammonium sulfate, 2 g/L calcium carbonate, 1 mg/L iron sulfate heptahydrate, 1 mg/L zinc sulfate heptahydrate, 1 mg/L manganese chloride tetrahydrate, 15 g/L agar, pH 7.2), or MS (10 g/L mannitol, 10 g/L soy flour (Kinako, Wel-Pac), 10 g/L malt extract, 15 g/L agar). Agar plates were grown at 30 °C for 10 days. Bacteria and the top layer of agar were scrapped off the plate and extracted with methanol overnight. Solid material was removed by centrifugation at 20,000 x g for 30 min followed by careful removal of the liquid extract and concentration under reduced pressure.

3.7.7 Bacterial extract screening

A previously described collection of extracts from actinobacteria[10] combined with the extracts from above was screened using the aminooxy probe **2**. The previously collected extracts had been partially purified on Oasis HLB extraction columns (Waters) and were dissolved in 50% acetonitrile in water. Labeling reactions were set up with 9 µL of extract solution and 1 µL of **2** from a 10 mM stock in ethanol in 0.2 mL tubes. The reactions were run for at least 3 h at room temperature with occasional manual shaking. Each reaction was submitted to MALDI-TOF MS in reflector positive mode using CHCA as a matrix. Spectra were analyzed manually for peaks displaying an isotope pattern consistent with the presence of two bromine atoms. Initial hits were verified by a follow up screen using the same conditions as above but with the commercially available 1-[(aminooxy)methyl]-4-chlorobenzene hydrochloride (Santa Cruz Biotechnology).

3.7.8 Purification of antipain-cit from *S. albulus* B3066

A frozen stock of *S. albulus* B3066 was streaked out on agar plates of ATCC172 media (10 g/L glucose, 20 g/L soluble starch, 5 g/L yeast extract, 5 g/L N-Z amine type A (Sigma C0626), 1 g/L reagent grade calcium carbonate, pH 7.3) and grown at 30 °C for 4 d. The plates were used to inoculate liquid cultures of ATCC172 media. The liquid cultures were grown at 30 °C for 7 days before they were used to inoculate 600 mL x 6 half sheet cake pans of solid ATCC172 media. The cake pans were incubated at 30 °C for 10 d before harvesting. To harvest, the agar was frozen at -20 °C, warmed to room temperature, and squeezed to collect the liquid. The agar was then washed with 50% methanol in water and this wash was combined with the agar squeeze. The methanol was removed by evaporation and the remaining aqueous material was loaded onto a HyperSep C18 SPE 10 g column (Thermo Scientific). The column was washed with 200 mL water before product was eluted with 10-100% acetonitrile in water. Fractions containing antipain-cit were combined and dried to a solid. The material was taken up in 8 mL water and 2 mL methanol and 10 mg of probe **2** was added. The reaction was stirred for 2 d at room temperature and was then adsorbed onto celite 545 and solid loaded onto an 86 g C18 reversed-phase column with a CombiFlash Rf 200 (Teledyne Isco). Chromatography was performed with a flow rate of 60 mL/min

using a gradient of 10-100% acetonitrile in water. Fractions containing antipain-cit were pooled and dried to a solid. The solid was resuspended in water and loaded onto a 250 mm x 10 mm inner diameter, 5 µm particle size Betasil C18 reversed-phase LC column (Thermo Scientific). Chromatography was performed with a flow rate of 4 mL/min using water with 10 mM ammonium bicarbonate (solvent A) and acetonitrile (solvent B) with a gradient of: 5 min of 17% B, 20 min ramp from 17% to 27% B, 5 min ramp from 27% to 95% B. Fractions containing the second isomer to elute were pooled and concentrated. Purity of the isolated labeled antipain-cit was evaluated by HPLC with a 250 mm x 4.6 mm inner diameter, 5 µm particle size Betasil C18 reversed-phase LC column (Thermo Scientific) using the same conditions as above but with a 1 mL/min flow rate. Isolated yield was approximately 2 mg for a single major isomer.

3.7.9 FT-MS/MS analysis of antipain-cit

Partially purified antipain-cit and labeled antipain-cit were dissolved in water. Samples were directly infused using an Advion Nanomate 100 into a ThermoFisher Scientific LTQ-FT hybrid linear ion trap, operating at 11T (calibrated weekly). The FT-MS was operated using the following parameters: minimum target signal counts, 5,000; resolution, 100,000; isolation width (MS/MS), 5 m/z; normalized collision energy (MS/MS), 35; activation q value (MS/MS), 0.4; activation time (MS/MS), 30 ms. Data analysis was conducted using the Qualbrowser application of Xcalibur software (Thermo-Fisher Scientific).

3.7.10 NMR of labeled antipain-cit

NMR spectra were collected on a Varian NMR System 750 MHz narrow bore magnet spectrometer (VNS750NB employing a 5 mm Varian $^{1}H[^{13}C/^{15}N]$ PFG X, Y, Z probe) The spectrometer was operated at 750 MHz for 1 H detection and 188 MHz for indirect 13C detection. Carbon resonances were assigned only via indirect detection (HSQC and HMBC experiments). Resonances were referenced internally to the CD_3OH solvent peak (3.30 ppm). Default Varian pulse sequences were employed for ^{1}H, gCOSY, zTOCSY, dpfgse-TOCSY, HSQCAD, HMBCAD, and dpfgse-NOESY experiments. Samples were prepared by dissolving approximately 2 mg of labeled antipain-cit (HPLC-purified and lyophilized)

in CD_3OD or CD_3OH. CD_3OD or CD_3OH were obtained from Sigma. Samples were held at 25 °C during acquisition.

3.8 References

1. Newman, D. J., and Cragg, G. M. (2012) Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010, *J. Nat. Prod. 75*, 311-335.

2. Dias, D. A., Urban, S., and Roessner, U. (2012) A historical overview of natural products in drug discovery, *Metabolites 2*, 303-336.

3. Harvey, A. L., Edrada-Ebel, R., and Quinn, R. J. (2015) The re-emergence of natural products for drug discovery in the genomics era, *Nat. Rev. Drug Discov. 14*, 111-129.

4. Sashidhara, K. V., and Rosaiah, J. N. (2007) Various dereplication strategies using LC-MS for rapid natural product lead identification and drug discovery, *Nat. Prod. Commun. 2*, 193-202.

5. Lewis, K. (2013) Platforms for antibiotic discovery, *Nat. Rev. Drug Discov. 12*, 371-387.

6. Baltz, R. H. (2006) Marcel Faber Roundtable: Is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration?, *J. Ind. Microbiol. Biotechnol. 33*, 507-513.

7. Crews, C. M., and Splittgerber, U. (1999) Chemical genetics: exploring and controlling cellular processes with chemical probes, *Trends Biochem. Sci. 24*, 317-320.

8. Carlson, E. E. (2010) Natural Products as Chemical Probes, *ACS Chem. Biol. 5*, 639-653.

9. Luo, Y., Cobb, R. E., and Zhao, H. (2014) Recent advances in natural product discovery, *Curr. Opin. Biotechnol. 30*, 230-237.

10. Doroghazi, J. R., Albright, J. C., Goering, A. W., Ju, K. S., Haines, R. R., Tchalukov, K. A., Labeda, D. P., Kelleher, N. L., and Metcalf, W. W. (2014) A roadmap for natural product discovery based on large-scale genomics and metabolomics, *Nat. Chem. Biol. 10*, 963-968.

11. Scherlach, K., and Hertweck, C. (2009) Triggering cryptic natural product biosynthesis in microorganisms, *Org. Biomol. Chem. 7*, 1753-1760.

12. Odendaal, A. Y., Trader, D. J., and Carlson, E. E. (2011) Chemoselective enrichment for natural products discovery, *Chem. Sci. 2*, 760-764.

13. Cox, C. L., Tietz, J. I., Sokolowski, K., Melby, J. O., Doroghazi, J. R., and Mitchell, D. A. (2014) Nucleophilic 1,4-additions for natural product discovery, *ACS Chem. Biol. 9*, 2014-2022.

14. Staunton, J., and Weissman, K. J. (2001) Polyketide biosynthesis: a millennium review, *Nat. Prod. Rep. 18*, 380-416.

15. Staunton, J., and Wilkinson, B. (1997) Biosynthesis of Erythromycin and Rapamycin, *Chem. Rev. 97*, 2611-2630.

16. Chiou, K. C. (2013) Exploration of the Diverse Functions of Cytochrome P450 Monooxygenases Towards the Development of Biocatalysts, In *Chemistry*, p 98, University of Michigan, Ann Arbor.

17. Iizaka, Y., Higashi, N., Ishida, M., Oiwa, R., Ichikawa, Y., Takeda, M., Anzai, Y., and Kato, F. (2013) Function of cytochrome P450 enzymes RosC and RosD in the biosynthesis of rosamicin macrolide antibiotic produced by Micromonospora rosaria, *Antimicrob. Agents Chemother. 57*, 1529-1531.

18. Kalia, J., and Raines, R. T. (2010) Advances in Bioconjugation, *Curr. Org. Chem. 14*, 138-147.

19. Shieh, P., and Bertozzi, C. R. (2014) Design strategies for bioorthogonal smart probes, *Org. Biomol. Chem. 12*, 9307-9320.

20. Zatsepin, T. S., Stetsenko, D. A., Gait, M. J., and Oretskaya, T. S. (2005) Use of carbonyl group addition-elimination reactions for synthesis of nucleic acid conjugates, *Bioconjugate Chem. 16*, 471-489.

21. Kojima, N., Takebayashi, T., Mikami, A., Ohtsuka, E., and Komatsu, Y. (2009) Development of novel chemical probes to detect abasic sites in DNA, *Nucleic Acids Symp. Ser.*, 45-46.

22. Zeng, Y., Ramya, T. N., Dirksen, A., Dawson, P. E., and Paulson, J. C. (2009) High-efficiency labeling of sialylated glycoproteins on living cells, *Nat. Methods 6*, 207-209.

23. Bayer, E. A., Benhur, H., and Wilchek, M. (1988) Biocytin Hydrazide - a Selective Label for Sialic Acids, Galactose, and Other Sugars in Glycoconjugates Using Avidin Biotin Technology, *Anal. Biochem. 170*, 271-281.

24. Kool, E. T., Crisalli, P., and Chan, K. M. (2014) Fast Alpha Nucleophiles: Structures that Undergo Rapid Hydrazone/Oxime Formation at Neutral pH, *Org. Lett. 16*, 1454-1457.

25. Kool, E. T., Park, D. H., and Crisalli, P. (2013) Fast Hydrazone Reactants: Electronic and Acid/Base Effects Strongly Influence Rate at Biological pH, *J. Am. Chem. Soc. 135*, 17663-17666.

26. Palaniappan, K. K., Pitcher, A. A., Smart, B. P., Spiciarich, D. R., Iavarone, A. T., and Bertozzi, C. R. (2011) Isotopic signature transfer and mass pattern prediction (IsoStamp): an enabling technique for chemically-directed proteomics, *ACS Chem. Biol. 6*, 829-836.

27. Breidenbach, M. A., Palaniappan, K. K., Pitcher, A. A., and Bertozzi, C. R. (2012) Mapping yeast N-glycosites with isotopically recoded glycans, *Mol. Cell. Proteomics 11*, M111 015339.

28. Woo, C. M., Iavarone, A. T., Spiciarich, D. R., Palaniappan, K. K., and Bertozzi, C. R. (2015) Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact N- and O-glycopeptide discovery and analysis, *Nat. Methods 12*, 561-567.

29. Kalia, J., and Raines, R. T. (2008) Hydrolytic stability of hydrazones and oximes, *Angew. Chem. Int. Ed. 47*, 7523-7526.

30. Maier, S., and Grisebach, H. (1979) Biosynthesis of Streptomycin - Enzymic Oxidation of Dihydrostreptomycin (6-Phosphate) to Streptomycin (6-Phosphate) with a Particulate Fraction of Streptomyces-Griseus, *Biochim. Biophys. Acta 586*, 231-241.

31. Suda, H., Takeuchi, T., Umezawa, H., Aoyagi, T., and Hamada, M. (1972) Antipain, a New Protease Inhibitor Isolated from Actinomycetes, *J. Antibiot. 25*, 263-&.

32. Nakae, K., Kojima, F., Sawa, R., Kubota, Y., Igarashi, M., Kinoshita, N., Adachi, H., Nishimura, Y., and Akamatsu, Y. (2010) Antipain Y, a new antipain analog that inhibits neurotransmitter release from rat dorsal root ganglion neurons, *J. Antibiot. 63*, 41-44.

33. Hoebeke, J., Busattosamsoen, C., Davoust, D., and Lebrun, E. (1994) H-1-Nmr Study of the Diastereomeric Forms of the Protease Inhibitor Antipain, *Magn. Reson. Chem. 32*, 220-224.

34. Rashidian, M., Mahmoodi, M. M., Shah, R., Dozier, J. K., Wagner, C. R., and Distefano, M. D. (2013) A Highly Efficient Catalyst for Oxime Ligation and Hydrazone-Oxime Exchange Suitable for Bioconjugation, *Bioconjugate Chem. 24*, 333-342.

35. Larsen, D., Pittelkow, M., Karmakar, S., and Kool, E. T. (2015) New Organocatalyst Scaffolds with High Activity in Promoting Hydrazone and Oxime Formation at Neutral pH, *Org. Lett. 17*, 274-277.

Appendix A: Review of molecular diagnostics for bacterial infections

This chapter is taken in part from Maxson and Mitchell[1] and is reproduced with permission from Tetrahedron.

A.1 The need for advanced diagnostics

Rapid diagnostic tests to determine an infecting organism and its resistance profile are crucial as treatment with inappropriate antibiotics not only promotes resistance but also destroys a patient's microbiome. This is especially true for narrow-spectrum antibiotics for which it is imperative to know which organism is causing the infection or treatment will be a complete failure. In some cases, the manifestation of disease is indicative of a specific pathogen, such as an erythema migrans rash with Lyme disease, and any future narrow-spectrum antibiotic could likely be prescribed before further testing. Similarly, around 80% of UTI cases are caused by *E. coli*[2] and, since they are generally non-life threatening, could be initially treated with an *E. coli*-specific antibiotic until diagnostic testing indicates a different pathogen is present. Narrow-spectrum antibiotics could also be used for prophylaxis, targeting the pathogens most often responsible for complications following surgeries or implants. However, most diseases can be caused by a multitude of different pathogens and thus the need for rapid and accurate diagnosis inherently goes hand in hand with the use of narrow-spectrum antibiotics.

Traditional culture-based diagnostic methods combined with Gram staining can take 1-2 days to identify a pathogen, or even longer for slow growing organisms.[3-5] Subsequent determination of resistance by phenotypic methods (Kirby-Bauer testing, microbroth dilutions, Etests) then requires up to an additional day before results are available.[6, 7] During this time frame, patients are typically treated empirically with broad-spectrum agents that may or may not be de-escalated when appropriate after testing is complete. After this delay, the advantage of treating with a narrow-spectrum antibiotic may be lost, with the microbiome already perturbed and potentially problematic organisms on the rise. The sensitivity and specificity of some gold standards leaves something to be desired as well, with blood culture failing to detect pathogens up to 50% of the time after empiric treatment has started[8, 9] and sputum samples often resulting in the growth of multiple species in culture.[3]

To sufficiently improve upon traditional methods such that narrow-spectrum antibiotics can be readily employed, several important requirements must be met (Table A.1). Specifically focusing on diagnostics for narrow-spectrum antibiotic use, the most important criteria are the time to result and the accuracy and sensitivity of the test, especially with regards to the ability to rule out polymicrobial infections. Ideally, a point-of-care diagnostic test would be rapid enough to provide a result in time for the initial treatment decision. For outpatient care, this would ideally occur during the visit. For inpatients, especially those in critical condition, getting results as soon as possible is essential.[10, 11] Reducing delays in appropriate antibiotic treatment for critical conditions such as sepsis has been repeatedly shown to have a major impact on patient survival.[11-13] Fulfilling this speed requirement will likely mean development of devices that are portable or can be set up in hospital clinical microbiology labs, as sending samples to a third party would cause a significant delay.

The other central requirements of high accuracy and sensitivity for diagnostic tests are particularly important for pathogen-specific antibiotics. Misidentification or the failure to detect a pathogen would lead to an incorrect treatment strategy with potentially dire consequences.[11] Identification of polymicrobial infections would also be crucial to prevent treatment failure due to a second, untreated pathogen.[14, 15] Many newer diagnostic methods are designed around technologies that simply do not work well when multiple species are present in the sample.[16, 17] False negatives with rarer pathogens resulting from molecular diagnostics that only test for the most common causes of infection is an additional concern, although this certainly applies to treatment with any antibiotic regardless of spectrum. Thus, tests that can incorporate the detection of the largest number of potential pathogens while maintaining a high degree of specificity and sensitivity are warranted.

Beyond speed and accuracy, several other factors should be considered when comparing new diagnostic technologies.[5] Tests that can incorporate rapid resistance determination have a significant advantage in utility, especially if high levels of resistance are known to exist for the narrow-spectrum antibiotic that could be employed. From a financial standpoint, expensive instrumentation may be

prohibitive for smaller institutions while a high cost per sample may discourage routine use, particularly if the test is not readily accepted for reimbursement by insurers. Another consideration is the amount of hands on time required to perform the test. A test that only takes an hour to complete but requires the attention of a technician the entire time for a single sample may not be feasible. More sample manipulation also leads to more opportunities for contamination, leading to inaccurate results. Finally, tests that provide quantification may be useful for the differentiation of colonizing and infectious organisms although the benefit of this has been called into question.[18]

While improved diagnostics are discussed here in terms of their ability to aid in narrow-spectrum antibiotic use, it is important to keep in mind that there are other advantages to improved diagnostics such as tracking outbreaks and epidemics, monitoring resistance, and discontinuing isolation of patients falsely suspected of having highly infectious diseases. We will first review the two main areas of advancement in diagnostics, nucleic acid- and MS-based methods. Many of the techniques from these two methods have been extensively tested and have already begun to enter the clinic. We will then briefly touch on emerging technologies that have the potential to significantly improve diagnostic methods in the future.

Table A.1. Summary of molecular diagnostic techniques.

Technique	Turnaround time	Number of species possible per assay	Resistance detection	Approximate Instrumentation cost
PNA-FISH	< 1 h after blood culture	1	No	< $1,000
NanoString	8 h	Undeveloped, > 800 targets possible	Yes	> $100,000
Microring resonators (RNA hybridization)	< 1 h	Undeveloped, multiplexing possible	Yes	~ $100,000
PCR from culture	1 - 3 h	> 50 for commercial assay	Yes	$20,000 - 50,000 (for real time)
PCR from blood	6 - 12 h	> 300 for commercial assay	Currently no	$20,000 - 50,000
PCR HRMA	3 h after culture	All	No	$20,000 - 50,000
WGS	> 24 h	All	Yes	> $100,000
MALDI-MS	< 15 min after culture	All	In development	> $100,000
PCR ESI-MS	5 h after culture	All	Yes	> $100,000
Rapid antigen testing	< 30 min	1	No	N/A
Microring resonators (Direct pathogen detection)	< 1 h	Undeveloped, multiplexing possible	No	~ $100,000
T2MR	3 h	Undeveloped, multiplexing possible	Yes	Undeveloped
IC 3D	4 h	Undeveloped, multiplexing possible	No	Undeveloped

A.2 Nucleic acid-based technologies

New molecular biology techniques developed over the last few decades were quickly seized upon by the diagnostics field for novel strategies for pathogen identification. Nucleic acid-based technologies (NATs) utilize bacterial DNA or RNA to rapidly pinpoint the infecting pathogen either through hybridization leading directly to a signal or through the amplification of DNA. Both strategies have been extensively developed with commercial tests available that offer significant improvements over traditional culture-based methods.

A.2.1 Non-amplification NATs

The field of non-amplification based NATs is primarily focused on peptide nucleic acid (PNA) fluorescence in situ hybridization (FISH) techniques. PNA-FISH based tests were among the first new molecular diagnostics to receive FDA clearance and have been available for clinical use since the early 2000s.[19] The tests report the presence of the target bacteria via a fluorescent signal generated upon hybridization of a short PNA probe (<25 bases) to bacterial RNA or DNA.[20] Due to its high cellular abundance and sequence differences between organisms in the variable regions, ribosomal RNA (rRNA) is often used as the target for PNA probes.[21] The use of PNA probes instead of DNA minimizes electrostatic repulsion, allowing for rapid and tight binding.[22] Short PNA probes are also able to cross the cell membrane, removing any need for a lysis step and reducing hands on time.[23] Extremely rapid commercial kits are available that can provide results in under an hour (e.g. *Quick*FISH, AdvanDx)[24] and it has been demonstrated that PNA-FISH can provide excellent accuracy in a clinical setting.[25, 26] However, PNA-FISH methods require a high bacterial count for detection and have thus far only been designed for use on positive blood cultures, resulting in a 1-2 day delay from when a patient sample is collected to when the test can be performed.[19] While this is a significant improvement on identification of pathogens by subculturing from a positive blood culture, the sensitivity of the method would need to be dramatically improved to allow the direct testing of patient samples required for a truly rapid result. Additionally, since rRNA is targeted, the presence of resistance markers is not analyzed at all with current methods and an amplification step would likely be required to detect the relatively low abundance of

RNA from resistance genes. Like most NAT methods, PNA-FISH requires a specific probe or set of probes for each target organism. This results in an increasing cost per sample as additional probes for other pathogens are included, although the initial instrumentation costs are minimal.[19] As PNA-FISH does not multiplex (simultaneously measure multiple analytes in a single assay) effectively and will miss the presence of any pathogens not specifically targeted, it is not ideally suited for diagnosis prior to narrow-spectrum antibiotic use.

Several other promising hybridization techniques have also been proposed for bacterial detection and identification, although these have not been nearly as extensively validated as PNA-FISH. One such technology is NanoString, an RNA detection method that is commercially available with the nCounter system.[27] NanoString allows for bacterial detection through the use of capture and reporter probes specific to targeted RNA sequences, with the reporter probe containing sets of colored indicators that are read through microscopy.[27, 28] The technique has been shown to be effective for bacterial, viral, and fungal pathogen detection with the use of several probe pairs per species.[28, 29] Up to 800 probes can be multiplexed in a single reaction, allowing a large number of species to be targeted at once.[27] However, the upfront cost of the instrumentation is in the hundreds of thousands of dollars and the technique requires a long hybridization time (>12 h),[27] although a recent report shortened this significantly, allowing the entire technique to be performed in around 8 h.[30]

An emerging technique for RNA detection via hybridization is the use of microring resonators. Microring resonators detect changes in the local refractive index upon binding of target molecules.[31] By functionalizing the rings with DNA capture probes designed for species specificity, bacterial detection based on target RNA molecules can be achieved in <1 h after RNA isolation.[32-34] Microring resonators can be inexpensively manufactured with many rings on a single, small chip, allowing for highly multiplexed detection of a large number of targets.[32] Both microring resonators and NanoString allow for a degree of quantification,[27, 33] which, unlike PNA-FISH, provides the advantage of possible resistance determination by examining transcriptome changes during antibiotic treatment in susceptible organisms.[28] Despite the ease of multiplexing these two technologies, untargeted organisms will still be missed,

resulting in rarer pathogens evading detection. Additionally, both methods have only been tested with relatively high bacterial counts so the limits of detection may not be sufficient for direct analysis of certain patient samples like blood.

A.2.2 Amplification NATs

In contrast to the hybridization-based techniques, many other NATs rely on DNA amplification to generate a readout, commonly through gel electrophoresis or an increasing fluorescent signal. While some tests target rRNA (directly as complementary DNA (cDNA) or as the gene) similar to the PNA-FISH based tests, amplification allows lower abundance, species-specific genes to be used.[35] This opens the possibility of determining the antibiotic susceptibility of pathogens through the detection of common resistance-conferring genes.[36] The earliest amplification based tests were introduced in the mid-1990s for the specific detection of *M. tuberculosis*,[37] and a number of tests capable of identifying a wide range of pathogens and resistance markers have since been released.[35-38] The advent of real-time PCR allowed for further improvements to amplification based technologies, eliminating the need for post-amplification analysis through techniques like gel electrophoresis and allowing for quantitative results.[39] An early example of a commercial real-time PCR device is the Cepheid Xpert MRSA/SA test, which can identify the presence of *S. aureus* and the *mecA* gene for methicillin resistance in an hour from a positive blood culture.[40] Real-time PCR systems with multiplexing to allow the detection of multiple pathogens have also been developed, such as the SeptiFast system for the identification of the 25 most common pathogens responsible for blood stream infections.[41]

There are several important considerations when discussing amplification based techniques, which have been discussed in several comprehensive reviews.[35, 36, 42] One issue is the amount of sample handling required to prepare DNA and set up reactions. The possibility of contamination is relatively high during this time so significant effort has gone into the development of automated, cartridge based systems for DNA extraction and preparation.[42] Generally, free DNA is not removed prior to lysis of bacterial cells so many amplification-based tests will report on the presence of dead cells and DNA from recent infections.[36] This can be advantageous in some cases, especially if a patient sample is collected after

antibiotic treatment has already begun, but can also lead to false positives as the presence of small amounts of bacterial DNA in samples like blood is not always indicative of infection.[43] Another consideration concerning resistance profiling is that only known resistance genes can be targeted, so the reliability of susceptibility determination through these methods may be questionable for some pathogens. Financially, real-time PCR instruments require a moderate initial investment and highly multiplexed tests for the detection of numerous pathogens can carry a significant cost per sample. Finally, as with direct hybridization techniques, only specifically targeted pathogens can be detected.

A method to circumvent the need to target specific primers to each species is to use broad range PCR. This approach relies on amplification of common genes such as the 16S RNA with primers that bind in conserved regions, similar to 16S typing for bacterial identification. Rapid species determination can then be performed with high-resolution melt analysis (HRMA) by comparing the melting curves of a set of amplification products to an existing database.[44, 45] Sequence differences between species lead to slight changes in the melting temperature for each amplification product and have been used to differentiate clinically relevant pathogens.[16] The accuracy of the technique is inherently limited by the quality of the reference database and a large number of known clinical isolates will need to analyzed and recorded to ensure sufficient species coverage for comparison. Strain variants with polymorphisms in the amplified gene regions could pose a problem however, and polymicrobial samples are impossible to analyze with HRMA, severely limiting its applicability.[46] Other detection methods for broad range PCR exist, including sequencing and electrospray ionization (ESI)-MS (discussed in section 3.3.2).

Moving a step beyond sequencing with broad range PCR, whole genome sequencing (WGS) with next-generation sequencing (NGS) methods has the potential to supplant traditional clinical microbiology as the gold standard in identifying the causative agent(s) of infectious disease.[5, 47] After sequencing, the infective species can be identified with the highest degree of confidence by alignment to a reference genome (if available) or by multilocus sequence typing (MLST) if a member of the same species has not yet been sequenced.[5, 48] In addition to species identification, WGS provides information on the presence of resistance markers and virulence factors and is an unparalleled tool for differentiating strains for

epidemiological purposes.[5, 49] A number of platforms for NGS have been developed utilizing different chemistry, including 454, Illumina, Ion Torrent, and PacBio, among others.[50, 51] Currently, WGS is too slow and expensive for routine pathogen identification, but may be a viable option in the near future if the rapid advancement in sequencing technology over the past decades continues.[5] Pathogen identification directly from patient samples using a metagenomics approach removes the >24 h delay required for DNA extraction from culture,[52, 53] while improvements in NGS have already drastically reduced the time and price required for sequencing (a single bacterial genome with sufficient depth coverage can now be obtained in less than a day for $200 by academic core facilities, with cost-savings truly becoming significant when many bacteria are sequenced).[5, 47, 54] With continued advances in NGS and automated data analysis software,[52] it appears that WGS is posed to replace all other NAT diagnostics.

While amplification techniques have proven to be accurate enough to be used with narrow-spectrum antibiotics, the time currently required for amplification is of concern. Many commercial tests are designed to work with the high bacterial counts available after initial culture and can give a result in as little as an hour,[40, 55] but the delay from culture time is significant. Testing directly from patient samples suffers from low starting bacterial counts and the presence of substances inhibitory to amplification, such as heme in blood.[39] The presence of vast excesses of host DNA can also be problematic in PCR assays.[39, 56] Several tests have been developed that overcome these issues and can be used directly on blood samples but they require much longer amplification times of 6-12 h.[35] This is still an improvement over waiting for cultures to grow however and can likely be further improved through methods of isolating bacteria or high quality bacterial DNA from patient samples, as discussed in section 3.2.3. No resistance determination is available in the tests currently marketed for use on blood but could presumably be included for known resistance genes in the future.

A.2.3 Bacterial isolation from patient samples

While some clinical sample types such as urine from UTIs can contain high bacterial counts (>10^5 colony forming units (CFU)/mL),[57] many samples have far fewer bacteria present. Blood in particular can have as little as 1-10 CFU/mL of an invading pathogen in adult sepsis patients.[58] Therefore,

methods for isolating high quality bacterial DNA from samples that often contain large amounts of human DNA can improve the reliability of NAT methods. The simplest way to achieve this is through the coupling of standard DNA purification systems with the selective removal of human DNA. One approach is to gently lyse human cells while leaving bacterial cells intact, followed by enzymatic or chemical degradation of human DNA. After washing, bacterial cells are lysed through a more vigorous method and the DNA is isolated.[59, 60] Another approach involves exploiting the differences between bacterial and human DNA, specifically in methylation of cytidylate-phosphate-deoxyguanylate (CpG) motifs by human CpG-binding protein. Unmethylated prokaryotic CpG motifs are selectively bound with an immobilized protein, allowing human DNA to be washed away to give an affinity purification of the bacterial DNA.[59, 61]

An alternative strategy for the isolation of bacteria or bacterial DNA from patient samples involves the use of microfluidics technologies. Patient samples can be passed through devices that contain bacterial capture materials, including antibodies,[62-64] synthetic ligands,[65] or even human opsonin.[66] After capture, human cells and other sample components are readily washed away before release of bacterial DNA via cell lysis.[63] A different microfluidics based approach that does not rely on specific bacterial capture but rather on the intrinsic physical properties of the bacterial cells was recently reported.[30] The method utilizes inertial lift forces to separate bacteria from blood cells based on size differences. This type of method also has the advantage of collecting intact cells, which could then potentially be used in non-DNA based detection methods such as MALDI.

A.3 Mass spectrometry-based technologies

The advent of so called “soft” ionization methods that are capable of ionizing large biomolecules without significant fragmentation has made the development of MS-based methods appealing for pathogen identification. These methods are based on the detection of sets of genus- or species-specific biomolecules, typically proteins or nucleic acids.[17, 67] The collected spectra are compared to a database assembled from known isolates to identify the species present, similar to PCR HRMA. Although MS methods require a large initial investment with the purchase of instrumentation, they have the advantage

of extremely low per-sample costs and very fast turnaround times. In contrast to NATs, MS-based methods do not require specific probes for each species of interest and can thus theoretically identify any species in a sample without any prior knowledge if robust databases are available. Two different general approaches to MS-based identification have been developed thus far, matrix-assisted laser desorption/ionization (MALDI)-MS and PCR ESI-MS, although MALDI-based techniques have been more extensively investigated.

A.3.1 MALDI-MS

MALDI-MS functions through desorption and ionization of sample molecules after laser irradiation with the assistance of matrix compounds followed by analyte detection, typically in a time of flight (TOF) analyzer. The application of MALDI-TOF-MS to clinical microbiology has been extensively reviewed.[17, 68, 69] Differences in bacterial proteomes are commonly used for species identification with MALDI-TOF-MS, although other biomolecules can also be detected. Generally, the focus is on ribosomal proteins as they are abundantly expressed throughout all stages of growth.[17] Since no prior knowledge of the organism in a sample is needed, fungi can also be identified without additional testing assuming a reference database including the organisms of interest is available.[68] While samples can be analyzed directly using whole cells, a protein extraction step improves signal intensity and reproducibility and is required for some organisms.[68] Sample preparation and analysis with MALDI-TOF-MS is extremely rapid, generally taking less than 15 minutes. With a sufficiently extensive database, the presence of any pathogen in a sample can be accurately detected.[70] Several MALDI-TOF-MS instruments designed for microorganism detection are available, including the Bruker BioTyper and the Shimadzu Axima Assurance system.[71]

However, there are several important limitations in the application of MALDI-TOF-MS for diagnostics that still need to be overcome. Currently, MALDI-TOF-MS for bacterial identification generally requires a high CFU count and thus is often not feasible directly on patient samples.[17, 69] Samples are usually analyzed after culture, which adds a significant amount of time to obtain results. However, several studies have reported success when analyzing urine[72-75] and cerebral spinal fluid[76]

samples directly. Bacteria could potentially be concentrated from patient samples as well using microfluidics technologies as described in section 3.2.3 to help circumvent the lack of sensitivity. Another issue facing the widespread implementation of MALDI-TOF-MS as a clinical tool is the often poor reproducibility between labs, as different sample preparation techniques and instrument types can cause substantial variation in analyte detection,[77, 78] although standardization between labs (perhaps through the Clinical and Laboratory Standards Institute or a similar organization) will improve this. The detection of antibiotic resistance is difficult with MALDI-TOF-MS but strategies to accomplish this have been advanced,[79] including the detection of methylation on rRNA[80] and reaction monitoring of β-lactam hydrolysis.[81] Differentiation between highly similar species and deconvolution of polymicrobial samples also pose significant challenges. Despite these problems, MALDI-TOF-MS has been successfully validated as a highly accurate method for pathogen identification in numerous clinical samples and several commercial platforms are currently available.[17, 70, 71] MALDI-TOF-MS systems have been adopted in many advanced clinical labs and reductions in size and cost of the instruments will enable smaller labs to utilize this technology as well. If improvements in analyzing patient samples directly can be made, the ability to identify any pathogen in the available databases makes MALDI-TOF-MS a promising possibility for diagnosis prior to narrow-spectrum antibiotic use.

A.3.2 PCR ESI-MS

ESI is another MS ionization method whereby a high voltage is used to disperse a sample into an aerosol containing charged droplets.[82] It has been used as a general detection method for bacterial identification after PCR amplification with broad range primers, similar to PCR HRMA.[46] The amplified DNA is analyzed by ESI-MS and species identification is achieved by comparing the experimental mass to charge (*m/z*) ratio of the PCR products to the expected *m/z* ratio based on the gene sequence.[36, 83, 84] Like with MALDI-MS based methods, a robust database is required, but databases for ESI-MS can be built from existing sequencing information in addition to clinical isolates.[36] This can ease the identification of strain variants and results that do not have an exact match in the database can have the

nearest neighbor identified by a BLAST-like search algorithm.[36] As with PCR HRMA and MALDI-MS, no prior knowledge of the sample species is required before testing.

However, the requirement of an amplification step introduces many of the issues described in section 3.2.2 for amplification based NAT techniques, such as sample processing and the time needed for amplification. The ESI-MS step itself is very rapid however and can be set up to run in an automated fashion. Several sets of primers for different gene regions are used to mitigate the risk posed by genotypic strain variants, although it is still possible that a divergent strain could be misidentified.[84, 85] Resistance determination can be performed in a manner similar to other amplification based methods by including primers for specific resistance-conferring genes at the cost of increased complexity and number of reactions that must be run on each sample.[85] PCR ESI-MS also has the advantage of working well with mixtures of organisms, with the identification of both species and the relative ratios of abundance usually determined.[85, 86] As with MALDI-MS instruments, ESI-MS instruments tend to be large and expensive, but improvements in miniaturized ESI-MS units[87] and in amplification technologies offer the potential for PCR ESI-MS to be a rapid point-of-care diagnostic. The Iridica system from Abbott Molecular is currently on the market and is specifically designed for pathogen identification directly from patient samples although it is not yet available in the United States.[88]

A.4 Alternative detection and identification technologies

Outside of NAT and MS-based methods, there are a number of other promising technologies that could be utilized for bacterial detection and identification prior to narrow-spectrum antibiotic use. One established method for diagnosis is rapid antigen testing, which relies on a visible readout for antibody-antigen binding. Rapid antigen testing has been used for years for *S. pyogenes* and provides results in ~15 min at the point-of-care, compared to 1-2 d for traditional culture on blood agar plates.[89-91] However, antigen testing in its current form only exists for a few select pathogens, likely due in part to the limited scope of such tests. It is thus generally used where clinical signs point to a certain type of infection, as with pharyngitis. Rapid antigen testing could be extended to additional pathogens (e.g. *E. coli* for UTIs) or potentially even multiplexed to allow broader identification.

The remaining techniques discussed in this section are relatively new and have not undergone the extensive testing with patient samples that most of the previous methods have, but they offer some promising possibilities for future development. Microring resonators were discussed in section 3.2.1 as an emerging technology for bacterial identification by hybridization of RNA but they could also be used for the direct detection of pathogens with species-specific antibody functionalized rings. This concept has been demonstrated for detection of intact bacterial cells[32] and virus particles.[92] Microring resonators designed around this concept rather than RNA detection would eliminate the need for an RNA isolation step and could be extremely rapid, but would require highly specific antibody generation for each pathogen of interest.

Another method that is being applied to pathogen detection and identification is T2 magnetic resonance (T2MR)-based biosensing. T2MR measures the transverse relaxation time of the nuclear magnetic resonance signal of water and is sensitive to changes in the microenvironments in aqueous solutions.[93] Recently, DNA functionalized nanoparticles that cluster upon binding of target nucleic acids were reported, resulting in easily detected T2 relaxation rate changes.[94] In combination with a specially designed, portable T2MR instrument that is compatible with standard PCR tubes, detection of *Candida* species at concentrations <5 CFU/mL in blood directly from patient samples was reported.[94, 95] Although an amplification step is required, results could be obtained in <3 h, which is less than half the time required for real-time PCR based methods. The method is also currently in development for bacterial detection by T2 Biosystems. Although the ability of T2MR to determine resistance profiles has not been discussed in the literature, it would theoretically be possible by functionalizing additional nanoparticles with DNA oligomers targeting known resistance genes.

The final technology we will discuss is integrated comprehensive droplet digital detection (IC 3D).[96] IC 3D provides a readout for highly sensitive bacterial detection through single fluorescent particles suspended in picoliter droplets. The fluorescent droplets are produced by mixing a DNAzyme sensor solution and the sample solution with lysis buffer in a microfluidic channel that encapsulates them into the droplets. The DNAzymes are activated by specific target bacterial molecules, which results in the

cleavage of a nucleic acid substrate to release a fluorophore from its associated quencher. The fluorophore is then detected by a recently developed 3D particle counter.[97] The entire process takes around 4 h for bacterial identification and is sensitive enough to detect <10 CFU/mL while also providing quantitative results directly from blood samples. While this method has many promising advantages, it will need to be developed into a multiplexed system for multiple species detection and suffers from the requirement for a specific DNAzyme for each species.

A.5 References

1. Maxson, T., and Mitchell, D. A. (2015) Targeted treatment for bacterial infections: prospects for pathogen-specific antibiotics coupled with rapid diagnostics, *Tetrahedron*, in press doi:10.1016/j.tet.2015.09.069

2. Zalewska-Piatek, B. M. (2011) Urinary tract infections of Escherichia coli strains of chaperone-usher system, *Pol. J. Microbiol. 60*, 279-285.

3. Tenover, F. C. (2011) Developing Molecular Amplification Methods for Rapid Diagnosis of Respiratory Tract Infections Caused by Bacterial Pathogens, *Clin. Infect. Dis. 52*, S338-S345.

4. Paolucci, M., Landini, M. P., and Sambri, V. (2010) Conventional and molecular techniques for the early diagnosis of bacteraemia, *Int. J. Antimicrob. Agents 36*, S6-S16.

5. Didelot, X., Bowden, R., Wilson, D. J., Peto, T. E., and Crook, D. W. (2012) Transforming clinical microbiology with bacterial genome sequencing, *Nat. Rev. Genet. 13*, 601-612.

6. Bou, G. (2007) Minimum inhibitory concentration (MIC) analysis and susceptibility testing of MRSA, *Methods Mol. Biol. 391*, 29-49.

7. Gherardi, G., Angeletti, S., Panitti, M., Pompilio, A., Di Bonaventura, G., Crea, F., Avola, A., Fico, L., Palazzo, C., Sapia, G. F., Visaggio, D., and Dicuonzo, G. (2012) Comparative evaluation of the Vitek-2 Compact and Phoenix systems for rapid identification and antibiotic susceptibility testing directly from blood cultures of Gram-negative and Gram-positive isolates, *Diagn. Microbiol. Infect. Dis. 72*, 20-31.

8. Dellinger, R. P., Levy, M. M., Rhodes, A., Annane, D., Gerlach, H., Opal, S. M., Sevransky, J. E., Sprung, C. L., Douglas, I. S., Jaeschke, R., Osborn, T. M., Nunnally, M. E., Townsend, S. R., Reinhart, K., Kleinpell, R. M., Angus, D. C., Deutschman, C. S., Machado, F. R., Rubenfeld, G. D., Webb, S. A., Beale, R. J., Vincent, J. L., Moreno, R., and Surviving Sepsis Campaign Guidelines Committee including the Pediatric, S. (2013) Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012, *Crit. Care Med. 41*, 580-637.

9. Fenollar, F., and Raoult, D. (2007) Molecular diagnosis of bloodstream infections caused by non-cultivable bacteria, *Int. J. Antimicrob. Agents 30 Suppl 1*, S7-15.

10. Kumar, A., Roberts, D., Wood, K. E., Light, B., Parrillo, J. E., Sharma, S., Suppes, R., Feinstein, D., Zanotti, S., Taiberg, L., Gurka, D., Kumar, A., and Cheang, M. (2006) Duration of

hypotension before initiation of effective antimicrobial therapy is the critical determinant of survival in human septic shock, *Crit. Care Med. 34*, 1589-1596.

11. Ibrahim, E. H., Sherman, G., Ward, S., Fraser, V. J., and Kollef, M. H. (2000) The influence of inadequate antimicrobial treatment of bloodstream infections on patient outcomes in the ICU setting, *Chest 118*, 146-155.

12. Leibovici, L., Shraga, I., Drucker, M., Konigsberger, H., Samra, Z., and Pitlik, S. D. (1998) The benefit of appropriate empirical antibiotic treatment in patients with bloodstream infection, *J. Intern. Med. 244*, 379-386.

13. Kollef, M. H., Sherman, G., Ward, S., and Fraser, V. J. (1999) Inadequate antimicrobial treatment of infections - A risk factor for hospital mortality among critically ill patients, *Chest 115*, 462-474.

14. Pavlaki, M., Poulakou, G., Drimousis, P., Adamis, G., Apostolidou, E., Gatselis, N. K., Kritselis, I., Mega, A., Mylona, V., Papatsoris, A., Pappas, A., Prekates, A., Raftogiannis, M., Rigaki, K., Sereti, K., Sinapidis, D., Tsangaris, I., Tzanetakou, V., Veldekis, D., Mandragos, K., Giamarello, H., and Dimopoulos, G. (2013) Polymicrobial bloodstream infections: Epidemiology and impact on mortality, *J. Glob. Antimicrob. Resist. 1*, 207-212.

15. Jindal, A., Moreker, M. R., Pathengay, A., Khera, M., Jalali, S., Majji, A., Mathai, A., Sharma, S., Das, T., and Flynn, H. W., Jr. (2013) Polymicrobial endophthalmitis: prevalence, causative organisms, and visual outcomes, *J. Ophthalmic Inflamm. Infect. 3*, 6.

16. Won, H., Rothman, R., Ramachandran, P., Hsieh, Y. H., Kecojevic, A., Carroll, K. C., Aird, D., Gaydos, C., and Yang, S. (2010) Rapid Identification of Bacterial Pathogens in Positive Blood Culture Bottles by Use of a Broad-Based PCR Assay Coupled with High-Resolution Melt Analysis, *J. Clin. Microbiol. 48*, 3410-3413.

17. Clark, A. E., Kaleta, E. J., Arora, A., and Wolk, D. M. (2013) Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry: a Fundamental Shift in the Routine Practice of Clinical Microbiology, *Clin. Microbiol. Rev. 26*, 547-603.

18. Berton, D. C., Kalil, A. C., Cavalcanti, M., and Teixeira, P. J. Z. (2008) Quantitative versus qualitative cultures of respiratory secretions for clinical outcomes in patients with ventilator-associated pneumonia, *Cochrane Database Syst. Rev. 4*.

19. Forrest, G. N. (2007) PNA FISH: present and future impact on patient management, *Expert Rev. Mol. Diagn. 7*, 231-236.

20. Stender, H. (2003) PNA FISH: an intelligent stain for rapid diagnosis of infectious diseases, *Expert Rev. Mol. Diagn. 3*, 649-655.

21. Delong, E. F., Wickham, G. S., and Pace, N. R. (1989) Phylogenetic Stains - Ribosomal Rna-Based Probes for the Identification of Single Cells, *Science 243*, 1360-1363.

22. Oliveira, K., Procop, G. W., Wilson, D., Coull, J., and Stender, H. (2002) Rapid identification of Staphylococcus aureus directly from blood cultures by fluorescence in situ hybridization with peptide nucleic acid probes, *J. Clin. Microbiol. 40*, 247-251.

23. Stender, H., Mollerup, T. A., Lund, K., Petersen, K. H., Hongmanee, P., and Godtfredsen, S. E. (1999) Direct detection and identification of Mycobacterium tuberculosis in smear-positive sputum samples by fluorescence in situ hybridization (FISH) using peptide nucleic acid (PNA) probes, *Int. J. Tuberc. Lung Dis. 3*, 830-837.

24. Deck, M. K., Anderson, E. S., Buckner, R. J., Colasante, G., Coull, J. M., Crystal, B., Della Latta, P., Fuchs, M., Fuller, D., Harris, W., Hazen, K., Klimas, L. L., Lindao, D., Meltzer, M. C., Morgan, M., Shepard, J., Stevens, S., Wu, F., and Fiandaca, M. J. (2012) Multicenter Evaluation of the Staphylococcus QuickFISH Method for Simultaneous Identification of Staphylococcus aureus and Coagulase-Negative Staphylococci Directly from Blood Culture Bottles in Less than 30 Minutes, *J. Clin. Microbiol. 50*, 1994-1998.

25. Harris, D. M., and Hata, D. J. (2013) Rapid identification of bacteria and candida using pna-fish from blood and peritoneal fluid cultures: a retrospective clinical study, *Ann. Clin. Microbiol. Antimicrob. 12*.

26. Parcell, B. J., and Orange, G. V. (2013) PNA-FISH assays for early targeted bacteraemia treatment, *J. Microbiol. Methods 95*, 253-255.

27. Kulkarni, M. M. (2011) Digital multiplexed gene expression analysis using the NanoString nCounter system, *Curr. Protoc. Mol. Biol. 94:B:25B.10:25B.10.1–25B.10.17*.

28. Barczak, A. K., Gomez, J. E., Kaufmann, B. B., Hinson, E. R., Cosimi, L., Borowsky, M. L., Onderdonk, A. B., Stanley, S. A., Kaur, D., Bryant, K. F., Knipe, D. M., Sloutsky, A., and Hung, D. T. (2012) RNA signatures allow rapid identification of pathogens and antibiotic susceptibilities, *Proc. Natl. Acad. Sci. U. S. A. 109*, 6217-6222.

29. Hsu, J. L., Binkley, J., Clemons, K. V., Stevens, D. A., Nicolls, M. R., and Holodniy, M. (2014) Application of a non-amplification-based technology to detect invasive fungal pathogens, *Diagn. Microbiol. Infect. Dis. 78*, 137-140.

30. Hou, H. W., Bhattacharyya, R. P., Hung, D. T., and Han, J. (2015) Direct detection and drug-resistance profiling of bacteremias using inertial microfluidics, *Lab Chip 15*, 2297-2307.

31. Bogaerts, W., De Heyn, P., Van Vaerenbergh, T., De Vos, K., Selvaraja, S. K., Claes, T., Dumon, P., Bienstman, P., Van Thourhout, D., and Baets, R. (2012) Silicon microring resonators, *Laser Photon. Rev. 6*, 47-73.

32. Ramachandran, A., Wang, S., Clarke, J., Ja, S. J., Goad, D., Wald, L., Flood, E. M., Knobbe, E., Hryniewicz, J. V., Chu, S. T., Gill, D., Chen, W., King, O., and Little, B. E. (2008) A universal biosensing platform based on optical micro-ring resonators, *Biosens. Bioelectron. 23*, 939-944.

33. Kindt, J. T., and Bailey, R. C. (2012) Chaperone Probes and Bead-Based Enhancement To Improve the Direct Detection of mRNA Using Silicon Photonic Sensor Arrays, *Anal. Chem. 84*, 8067-8074.

34. Scheler, O., Kindt, J. T., Qavi, A. J., Kaplinski, L., Glynn, B., Barry, T., Kurg, A., and Bailey, R. C. (2012) Label-free, multiplexed detection of bacterial tmRNA using silicon photonic microring resonators, *Biosens. Bioelectron. 36*, 56-61.

35. Mancini, N., Carletti, S., Ghidoli, N., Cichero, P., Burioni, R., and Clementi, M. (2010) The Era of Molecular and Other Non-Culture-Based Methods in Diagnosis of Sepsis, *Clin. Microbiol. Rev. 23*, 235-251.

36. Ecker, D. J., Sampath, R., Li, H. J., Massire, C., Matthews, H. E., Toleno, D., Hall, T. A., Blyn, L. B., Eshoo, M. W., Ranken, R., Hofstadler, S. A., and Tang, Y. W. (2010) New technology for rapid molecular diagnosis of bloodstream infections, *Expert Rev. Mol. Diagn. 10*, 399-415.

37. Ginocchio, C. C. (2011) Strengths and Weaknesses of FDA-Approved/Cleared Diagnostic Devices for the Molecular Detection of Respiratory Pathogens, *Clin. Infect. Dis. 52*, S312-S325.

38. Kothari, A., Morgan, M., and Haake, D. A. (2014) Emerging Technologies for Rapid Identification of Bloodstream Pathogens, *Clin. Infect. Dis. 59*, 272-278.

39. Ince, J., and McNally, A. (2009) Development of rapid, automated diagnostics for infectious disease: advances and challenges, *Expert Rev. Med. Devices 6*, 641-651.

40. Wolk, D. M., Struelens, M. J., Pancholi, P., Davis, T., Della-Latta, P., Fuller, D., Picton, E., Dickenson, R., Denis, O., Johnson, D., and Chapin, K. (2009) Rapid Detection of Staphylococcus aureus and Methicillin-Resistant S. aureus (MRSA) in Wound Specimens and Blood Cultures: Multicenter Preclinical Evaluation of the Cepheid Xpert MRSA/SA Skin and Soft Tissue and Blood Culture Assays, *J. Clin. Microbiol. 47*, 823-826.

41. Lehmann, L. E., Hunfeld, K. P., Emrich, T., Haberhausen, G., Wissing, H., Hoeft, A., and Stuber, F. (2008) A multiplex real-time PCR assay for rapid detection and differentiation of 25 bacterial and fungal pathogens from whole blood samples, *Med. Microbiol. Immunol. 197*, 313-324.

42. Caliendo, A. M., Gilbert, D. N., Ginocchio, C. C., Hanson, K. E., May, L., Quinn, T. C., Tenover, F. C., Alland, D., Blaschke, A. J., Bonomo, R. A., Carroll, K. C., Ferraro, M. J., Hirschhorn, L. R., Joseph, W. P., Karchmer, T., MacIntyre, A. T., Reller, L. B., Jackson, A. F., and Idsa. (2013) Better Tests, Better Care: Improved Diagnostics for Infectious Diseases, *Clin. Infect. Dis. 57*, S139-S170.

43. Nikkari, S., McLaughlin, I. J., Bi, W. L., Dodge, D. E., and Relman, D. A. (2001) Does blood of healthy subjects contain bacterial ribosomal DNA?, *J. Clin. Microbiol. 39*, 1956-1959.

44. Wittwer, C. T., Reed, G. H., Gundry, C. N., Vandersteen, J. G., and Pryor, R. J. (2003) High-resolution genotyping by amplicon melting analysis using LCGreen, *Clin. Chem. 49*, 853-860.

45. Yang, S., Ramachandran, P., Rothman, R., Hsieh, Y. H., Hardick, A., Won, H., Kecojevic, A., Jackman, J., and Gaydos, C. (2009) Rapid Identification of Biothreat and Other Clinically Relevant Bacterial Species by Use of Universal PCR Coupled with High-Resolution Melting Analysis, *J. Clin. Microbiol. 47*, 2252-2255.

46. Jeng, K., Gaydos, C. A., Blyn, L. B., Yang, S., Won, H., Matthews, H., Toleno, D., Hsieh, Y. H., Carroll, K. C., Hardick, J., Masek, B., Kecojevic, A., Sampath, R., Peterson, S., and Rothman, R. E. (2012) Comparative analysis of two broad-range PCR assays for pathogen detection in positive-blood-culture bottles: PCR-high-resolution melting analysis versus PCR-mass spectrometry, *J. Clin. Microbiol. 50*, 3287-3292.

47. Fournier, P. E., Dubourg, G., and Raoult, D. (2014) Clinical detection and characterization of bacterial pathogens in the genomics era, *Genome Med. 6*, 114.

48. Jolley, K. A., Bliss, C. M., Bennett, J. S., Bratcher, H. B., Brehony, C., Colles, F. M., Wimalarathna, H., Harrison, O. B., Sheppard, S. K., Cody, A. J., and Maiden, M. C. (2012) Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain, *Microbiology 158*, 1005-1015.

49. Reuter, S., Ellington, M. J., Cartwright, E. J., Koser, C. U., Torok, M. E., Gouliouris, T., Harris, S. R., Brown, N. M., Holden, M. T., Quail, M., Parkhill, J., Smith, G. P., Bentley, S. D., and Peacock, S. J. (2013) Rapid bacterial whole-genome sequencing to enhance diagnostic and public health microbiology, *JAMA Intern. Med. 173*, 1397-1404.

50. Hodkinson, B. P., and Grice, E. A. (2015) Next-Generation Sequencing: A Review of Technologies and Tools for Wound Microbiome Research, *Adv. Wound Care 4*, 50-58.

51. Mardis, E. R. (2013) Next-generation sequencing platforms, *Annu. Rev. Anal. Chem. 6*, 287-303.

52. Hasman, H., Saputra, D., Sicheritz-Ponten, T., Lund, O., Svendsen, C. A., Frimodt-Moller, N., and Aarestrup, F. M. (2014) Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples, *J. Clin. Microbiol. 52*, 139-146.

53. Loman, N. J., Constantinidou, C., Christner, M., Rohde, H., Chan, J. Z., Quick, J., Weir, J. C., Quince, C., Smith, G. P., Betley, J. R., Aepfelbacher, M., and Pallen, M. J. (2013) A culture-independent sequence-based metagenomics approach to the investigation of an outbreak of Shiga-toxigenic Escherichia coli O104:H4, *J. Am. Med. Assoc. 309*, 1502-1510.

54. van Dijk, E. L., Auger, H., Jaszczyszyn, Y., and Thermes, C. (2014) Ten years of next-generation sequencing technology, *Trends Genet. 30*, 418-426.

55. Altun, O., Almuhayawi, M., Ullberg, M., and Ozenci, V. (2013) Clinical Evaluation of the FilmArray Blood Culture Identification Panel in Identification of Bacteria and Yeasts from Positive Blood Culture Bottles, *J. Clin. Microbiol. 51*, 4130-4136.

56. Cogswell, F. B., Bantar, C. E., Hughes, T. G., Gu, Y., and Philipp, M. T. (1996) Host DNA can interfere with detection of Borrelia burgdorferi in skin biopsy specimens by PCR, *J. Clin. Microbiol. 34*, 980-982.

57. Kwon, J. H., Fausone, M. K., Du, H., Robicsek, A., and Peterson, L. R. (2012) Impact of laboratory-reported urine culture colony counts on the diagnosis and treatment of urinary tract infection for hospitalized patients, *Am. J. Clin. Pathol. 137*, 778-784.

58. Reimer, L. G., Wilson, M. L., and Weinstein, M. P. (1997) Update on detection of bacteremia and fungemia, *Clin. Microbiol. Rev. 10*, 444-465.

59. Horz, H. P., Scheer, S., Huenger, F., Vianna, M. E., and Conrads, G. (2008) Selective isolation of bacterial DNA from human clinical specimens, *J. Microbiol. Methods 72*, 98-102.

60. Loonen, A. J., Bos, M. P., van Meerbergen, B., Neerken, S., Catsburg, A., Dobbelaer, I., Penterman, R., Maertens, G., van de Wiel, P., Savelkoul, P., and van den Brule, A. J. (2013)

Comparison of pathogen DNA isolation methods from large volumes of whole blood to improve molecular diagnosis of bloodstream infections, *PLoS One 8*, e72349.

61. Sachse, S., Straube, E., Lehmann, M., Bauer, M., Russwurm, S., and Schmidt, K. H. (2009) Truncated Human Cytidylate-Phosphate-Deoxyguanylate-Binding Protein for Improved Nucleic Acid Amplification Technique-Based Detection of Bacterial Species in Human Samples, *J. Clin. Microbiol. 47*, 1050-1057.

62. Xia, N., Hunt, T. P., Mayers, B. T., Alsberg, E., Whitesides, G. M., Westervelt, R. M., and Ingber, D. E. (2006) Combined microfluidic-micromagnetic separation of living cells in continuous flow, *Biomed. Microdevices 8*, 299-308.

63. Cho, Y. K., Lee, J. G., Park, J. M., Lee, B. S., Lee, Y., and Ko, C. (2007) One-step pathogen specific DNA extraction from whole blood on a centrifugal microfluidic device, *Lab Chip 7*, 565-573.

64. Boehm, D. A., Gottlieb, P. A., and Hua, S. Z. (2007) On-chip microfluidic biosensor for bacterial detection and identification, *Sens. Actuators, B 126*, 508-514.

65. Lee, J. J., Jeong, K. J., Hashimoto, M., Kwon, A. H., Rwei, A., Shankarappa, S. A., Tsui, J. H., and Kohane, D. S. (2014) Synthetic ligand-coated magnetic nanoparticles for microfluidic bacterial separation from blood, *Nano Lett. 14*, 1-5.

66. Kang, J. H., Super, M., Yung, C. W., Cooper, R. M., Domansky, K., Graveline, A. R., Mammoto, T., Berthet, J. B., Tobin, H., Cartwright, M. J., Watters, A. L., Rottman, M., Waterhouse, A., Mammoto, A., Gamini, N., Rodas, M. J., Kole, A., Jiang, A., Valentin, T. M., Diaz, A., Takahashi, K., and Ingber, D. E. (2014) An extracorporeal blood-cleansing device for sepsis therapy, *Nat. Med. 20*, 1211-1216.

67. Baldwin, M. A. (2005) Mass spectrometers for the analysis of biomolecules, *Methods Enzymol. 402*, 3-48.

68. Croxatto, A., Prod'hom, G., and Greub, G. (2012) Applications of MALDI-TOF mass spectrometry in clinical diagnostic microbiology, *FEMS Microbiol. Rev. 36*, 380-407.

69. Nomura, F. (2015) Proteome-based bacterial identification using matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS): A revolutionary shift in clinical diagnostic microbiology, *BBA-Proteins Proteom. 1854*, 528-537.

70. Seng, P., Drancourt, M., Gouriet, F., La Scola, B., Fournier, P. E., Rolain, J. M., and Raoult, D. (2009) Ongoing Revolution in Bacteriology: Routine Identification of Bacteria by Matrix-Assisted Laser Desorption Ionization Time-of-Flight Mass Spectrometry, *Clin. Infect. Dis. 49*, 543-551.

71. Cherkaoui, A., Hibbs, J., Emonet, S., Tangomo, M., Girard, M., Francois, P., and Schrenzel, J. (2010) Comparison of two matrix-assisted laser desorption ionization-time of flight mass spectrometry methods with conventional phenotypic identification for routine identification of bacteria to the species level, *J. Clin. Microbiol. 48*, 1169-1175.

72. Ferreira, L., Sanchez-Juanes, F., Gonzalez-Avila, M., Cembrero-Fucinos, D., Herrero-Hernandez, A., Gonzalez-Buitrago, J. M., and Munoz-Bellido, J. L. (2010) Direct identification of urinary

tract pathogens from urine samples by matrix-assisted laser desorption ionization-time of flight mass spectrometry, *J. Clin. Microbiol. 48*, 2110-2115.

73. Ferreira, L., Sanchez-Juanes, F., Munoz-Bellido, J. L., and Gonzalez-Buitrago, J. M. (2011) Rapid method for direct identification of bacteria in urine and blood culture samples by matrix-assisted laser desorption ionization time-of-flight mass spectrometry: intact cell vs. extraction method, *Clin. Microbiol. Infect. 17*, 1007-1012.

74. Wang, X. H., Zhang, G., Fan, Y. Y., Yang, X., Sui, W. J., and Lu, X. X. (2013) Direct identification of bacteria causing urinary tract infections by combining matrix-assisted laser desorption ionization-time of flight mass spectrometry with UF-1000i urine flow cytometry, *J. Microbiol. Methods 92*, 231-235.

75. Kim, Y., Park, K. G., Lee, K., and Park, Y. J. (2015) Direct Identification of Urinary Tract Pathogens From Urine Samples Using the Vitek MS System Based on Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry, *Ann. Lab. Med. 35*, 416-422.

76. Nyvang Hartmeyer, G., Kvistholm Jensen, A., Bocher, S., Damkjaer Bartels, M., Pedersen, M., Engell Clausen, M., Abdul-Redha, R., Dargis, R., Schouenborg, P., Hojlyng, N., Kemp, M., and Christensen, J. J. (2010) Mass spectrometry: pneumococcal meningitis verified and Brucella species identified in less than half an hour, *Scand. J. Infect. Dis. 42*, 716-718.

77. Williams, T. L., Andrzejewski, D., Lay, J. O., and Musser, S. M. (2003) Experimental factors affecting the quality and reproducibility of MALDI TOF mass spectra obtained from whole bacteria cells, *J. Am. Soc. Mass Spectrom. 14*, 342-351.

78. Albrethsen, J. (2007) Reproducibility in protein profiling by MALDI-TOF mass spectrometry, *Clin. Chem. 53*, 852-858.

79. Hrabak, J., Chudackova, E., and Walkova, R. (2013) Matrix-Assisted Laser Desorption Ionization-Time of Flight (MALDI-TOF) Mass Spectrometry for Detection of Antibiotic Resistance Mechanisms: from Research to Routine Diagnosis, *Clin. Microbiol. Rev. 26*, 103-114.

80. Kirpekar, F., Douthwaite, S., and Roepstorff, P. (2000) Mapping posttranscriptional modifications in 5S ribosomal RNA by MALDI mass spectrometry, *RNA 6*, 296-306.

81. Sparbier, K., Schubert, S., Weller, U., Boogen, C., and Kostrzewa, M. (2012) Matrix-assisted laser desorption ionization-time of flight mass spectrometry-based functional assay for rapid detection of resistance against beta-lactam antibiotics, *J. Clin. Microbiol. 50*, 927-937.

82. Banerjee, S., and Mazumdar, S. (2012) Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte, *Int. J. Anal. Chem. 2012*, 282574.

83. Muddiman, D. C., Anderson, G. A., Hofstadler, S. A., and Smith, R. D. (1997) Length and base composition of PCR-amplified nucleic acids using mass measurements from electrospray ionization mass spectrometry, *Anal. Chem. 69*, 1543-1549.

84. Baldwin, C. D., Howe, G. B., Sampath, R., Blyn, L. B., Matthews, H., Harpin, V., Hall, T. A., Drader, J. J., Hofstadler, S. A., Eshoo, M. W., Rudnick, K., Studarus, K., Moore, D., Abbott, S., Janda, J. M., and Whitehouse, C. A. (2009) Usefulness of multilocus polymerase chain reaction

followed by electrospray ionization mass spectrometry to identify a diverse panel of bacterial isolates, *Diagn. Microbiol. Infect. Dis. 63*, 403-408.

85. Kaleta, E. J., Clark, A. E., Johnson, D. R., Gamage, D. C., Wysocki, V. H., Cherkaoui, A., Schrenzel, J., and Wolk, D. M. (2011) Use of PCR Coupled with Electrospray Ionization Mass Spectrometry for Rapid Identification of Bacterial and Yeast Bloodstream Pathogens from Blood Culture Bottles, *J. Clin. Microbiol. 49*, 345-353.

86. Jeng, K., Hardick, J., Rothman, R., Yang, S., Won, H., Peterson, S., Hsieh, Y. H., Masek, B. J., Carroll, K. C., and Gaydos, C. A. (2013) Reverse transcription-PCR-electrospray ionization mass spectrometry for rapid detection of biothreat and common respiratory pathogens, *J. Clin. Microbiol. 51*, 3300-3307.

87. Janfelt, C., Graesboll, R., and Lauritsen, F. R. (2012) Portable electrospray ionization mass spectrometry (ESI-MS) for analysis of contaminants in the field, *Int. J. Environ. Anal. Chem. 92*, 397-404.

88. Bacconi, A., Richmond, G. S., Baroldi, M. A., Laffler, T. G., Blyn, L. B., Carolan, H. E., Frinder, M. R., Toleno, D. M., Metzgar, D., Gutierrez, J. R., Massire, C., Rounds, M., Kennel, N. J., Rothman, R. E., Peterson, S., Carroll, K. C., Wakefield, T., Ecker, D. J., and Sampath, R. (2014) Improved sensitivity for molecular detection of bacterial and Candida infections in blood, *J. Clin. Microbiol. 52*, 3164-3174.

89. Cohen, J. F., Cohen, R., and Chalumeau, M. (2013) Rapid antigen detection test for group A streptococcus in children with pharyngitis (Protocol), *Cochrane Database Syst. Rev. 4*.

90. Leung, A. K., Newman, R., Kumar, A., and Davies, H. D. (2006) Rapid antigen detection testing in diagnosing group A beta-hemolytic streptococcal pharyngitis, *Expert Rev. Mol. Diagn. 6*, 761-766.

91. Stewart, E. H., Davis, B., Clemans-Taylor, B. L., Littenberg, B., Estrada, C. A., and Centor, R. M. (2014) Rapid antigen group A streptococcus test to diagnose pharyngitis: a systematic review and meta-analysis, *PLoS One 9*, e111727.

92. McClellan, M. S., Domier, L. L., and Bailey, R. C. (2012) Label-free virus detection using silicon photonic microring resonators, *Biosens. Bioelectron. 31*, 388-392.

93. Skewis, L. R., Lebedeva, T., Papkov, V., Thayer, E. C., Massefski, W., Cuker, A., Nagaswami, C., Litvinov, R. I., Kowalska, M. A., Rauova, L., Poncz, M., Weisel, J. W., Lowery, T. J., and Cines, D. B. (2014) T2 magnetic resonance: a diagnostic platform for studying integrated hemostasis in whole blood--proof of concept, *Clin. Chem. 60*, 1174-1182.

94. Neely, L. A., Audeh, M., Phung, N. A., Min, M., Suchocki, A., Plourde, D., Blanco, M., Demas, V., Skewis, L. R., Anagnostou, T., Coleman, J. J., Wellman, P., Mylonakis, E., and Lowery, T. J. (2013) T2 magnetic resonance enables nanoparticle-mediated rapid detection of candidemia in whole blood, *Sci. Transl. Med. 5*, 182ra154.

95. Mylonakis, E., Clancy, C. J., Ostrosky-Zeichner, L., Garey, K. W., Alangaden, G. J., Vazquez, J. A., Groeger, J. S., Judson, M. A., Vinagre, Y. M., Heard, S. O., Zervou, F. N., Zacharioudakis, I. M., Kontoyiannis, D. P., and Pappas, P. G. (2015) T2 magnetic resonance assay for the rapid diagnosis of candidemia in whole blood: a clinical trial, *Clin. Infect. Dis. 60*, 892-899.

96. Kang, D. K., Ali, M. M., Zhang, K. X., Huang, S. S., Peterson, E., Digman, M. A., Gratton, E., and Zhao, W. A. (2014) Rapid detection of single bacteria in unprocessed blood using Integrated Comprehensive Droplet Digital Detection, *Nat. Commun. 5*, 5427.

97. Skinner, J. P., Swift, K. M., Ruan, Q. Q., Perfetto, S., Gratton, E., and Tetin, S. Y. (2013) Simplified confocal microscope for counting particles at low concentrations, *Rev. Sci. Instrum. 84*, 074301.

Appendix B: Further publications with minor contributions

B.1 Identification of the minimal cytolytic unit for streptolysin S and an expansion of the toxin family

Evelyn M. Molloy, Sherwood R. Casjens, Courtney L. Cox, Tucker Maxson, Nicole A. Ethridge, Gabriele Margos, Volker Fingerle, Douglas A. Mitchell

This chapter was reproduced from Molloy et al.[1] and is reproduced with permission from BMC Microbiology.

I aided in the *in vitro* reconstitution of cytolysin production and in testing the lytic activity of SLS truncation mutants.

Abstract

Background

Streptolysin S (SLS) is a cytolytic virulence factor produced by the human pathogen *Streptococcus pyogenes* and other *Streptococcus* species. Related "SLS-like" toxins have been characterized in select strains of *Clostridium* and *Listeria*, with homologous clusters bioinformatically identified in a variety of other species. SLS is a member of the thiazole/oxazole-modified microcin (TOMM) family of natural products. The structure of SLS has yet to be deciphered and many questions remain regarding its structure-activity relationships.

Results

In this work, we assessed the hemolytic activity of a series of C-terminally truncated SLS peptides expressed in SLS-deficient *S. pyogenes*. Our data indicate that while the N-terminal poly-heterocyclizable (NPH) region of SLS substantially contributes to its bioactivity, the variable C-terminal region of the toxin is largely dispensable. Through genome mining we identified additional SLS-like clusters in diverse *Firmicutes*, *Spirochaetes* and *Actinobacteria*. Among the *Spirochaete* clusters, naturally truncated SLS-like precursors were found in the genomes of three Lyme disease-causing *Borrelia burgdorferi* sensu lato (*Bbsl*) strains. Although unable to restore hemolysis in SLS-deficient *S. pyogenes*, a *Bbsl* SLS-like precursor peptide was converted to a cytolysin using purified SLS biosynthetic

enzymes. A PCR-based screen demonstrated that SLS-like clusters are substantially more prevalent in *Bbsl* than inferred from publicly available genome sequences.

Conclusions

The mutagenesis data described herein allowed us to define the minimal cytolytic unit of SLS as the NPH region. Interestingly, this region is found in all characterized TOMM cytolysins, as well as the novel putative TOMM cytolysins we discovered. We propose that this conserved region represents the defining feature of the SLS-like TOMM family. We demonstrate the cytolytic potential of a *Bbsl* SLS-like precursor peptide, which is of similar length to the SLS minimal cytolytic unit, when modified with purified SLS biosynthetic enzymes. As such, we speculate that some *Borrelia* have the potential to produce a TOMM cytolysin, although the biological significance of this finding remains to be determined. In addition to providing new insight into the structure-activity relationships of SLS, this study greatly expands the cytolysin group of TOMMs.

Background

Streptococcus pyogenes (Group A *Streptococcus*) commonly causes mild diseases such as pharyngitis and impetigo, and in rare cases, invasive infections such as necrotizing fasciitis and streptococcal toxic shock syndrome.[2] While the characteristic β-hemolytic phenotype of *S. pyogenes* grown on blood agar has long been known,[3] more than four decades passed before the responsible hemolytic/cytolytic factor,[4] streptolysin S (SLS), was identified.[5] The eventual discovery of the SLS biosynthetic gene cluster[6] prompted investigations that subsequently designated SLS a member of the thiazole/oxazole-modified microcin (TOMM) group of natural products.[7-9] TOMMs are a class of functionally and structurally diverse ribosomal peptides that are posttranslationally modified to contain the eponymous thiazole and (methyl)oxazole heterocycles derived from select cysteine, serine, and threonine residues.[9] The SLS biosynthetic operon (Figure B.1A) encodes a precursor peptide, SagA (Figure B.1B), and three heterocycle-forming proteins, SagBCD.[6, 8] Similar to other characterized ribosomal natural products, the SagA N-terminal leader peptide contains residues recognized by the modifying enzymes, while the C-terminal core peptide undergoes heterocyclization[8, 10, 11] (Figure B.1B &

C). The suspected protease, SagE, is believed to remove the leader peptide,[12, 13] permitting the export of mature and bioactive SLS via a dedicated ABC transporter, SagGHI (Figure B.1A).

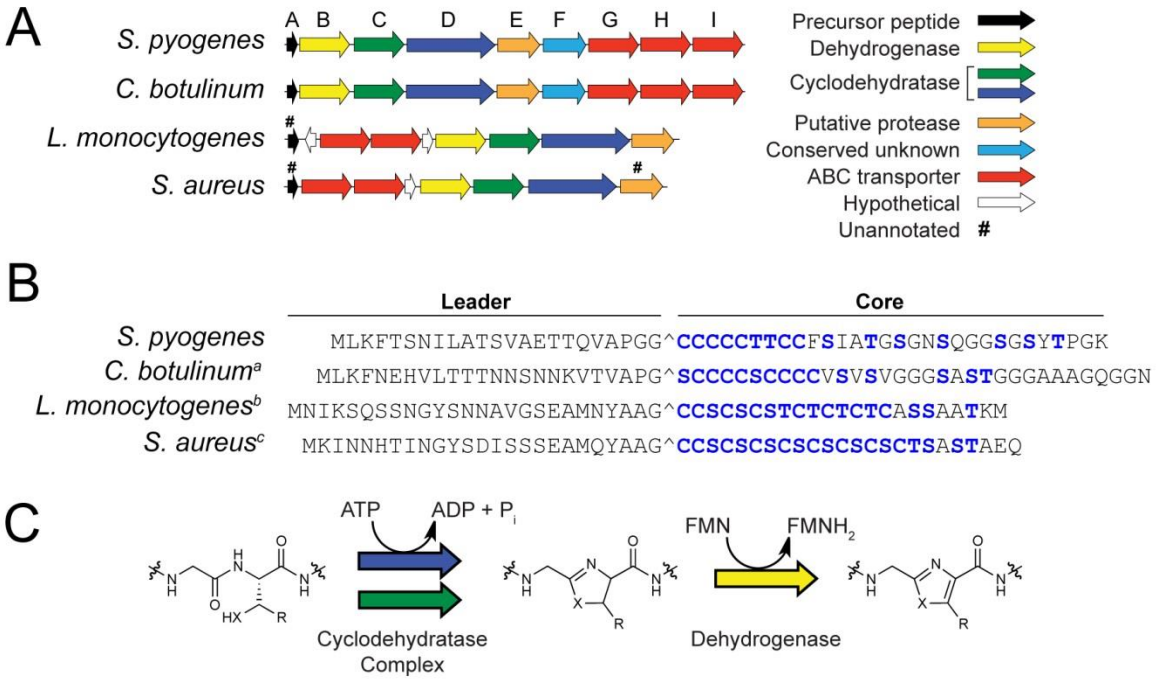


Figure B.1. Gene cluster organization and precursor peptide sequences of SLS-like cytolysins, and post-translational modification scheme. (A) Lettering corresponds to the SLS operon "*sag*" genes. The proposed function of each gene is color-coded according to the legend. A representative is shown, as in some cases, organization varies among strains and/or not all strains of a species possess the cluster. (B) Heterocyclizable residues (Cys, Ser, Thr) are shown in blue. The predicted leader cleavage sites are indicated with carets. [a]bioactivity of "Clostridiolysin S" shown from *C. sporogenes*[18] and upon heterologous expression in *S. pyogenes ΔsagA*[8] [b]bioactivity of "listeriolysin S" shown from native producers,[19, 20] but neither the precursor peptide nor core peptide fused to SagA leader were cytolytic upon heterologous expression in *S. pyogenes ΔsagA*[8]. [c]bioactivity of putative cytolysin not yet shown from a native producer, but core peptide fused to SagA leader cytolytic upon heterologous expression in *S. pyogenes ΔsagA*[8]. (C) Azole heterocycles are installed on a TOMM precursor peptide by the successive action of a cyclodehydratase (C and D proteins) and dehydrogenase (B protein). X = S or O; R = H or CH_3.

SLS is the founding member of a group of homologous TOMM cytolysins that have been characterized in other *Streptococcus* species (Figure B.2A), including invasive human isolates of the β-hemolytic Group C and Group G streptococci, which belong to *S. dysgalactiae* subsp. *equisimilis*.[14] SLS variants are also produced by the animal pathogens *S. iniae*[15] and *S. equi*.[16] More recently, *S. anginosus*, a human commensal organism and emerging opportunistic pathogen, has been shown to produce twin SLS

homologs.[17] Notably, SLS-like toxins are not limited to *Streptococcus*, being produced by other *Firmicutes* including select strains of *Clostridium*[18] and *Listeria*[19-21] (Figure B.1A & B).

A

```
                    5    10    15    20    25    30    35    40    45    50
S. pyogenes         MLKFTSNILATSVAETTQVAPGG^CCCCCTTCCFSIATGSG-NSQGGSGSYTPGK--
S. dysgalactiae     MLQFTSNILATSVAETTQVAPGG^CCCCCTTCCFSINVGGG-SAQGGSGSYTPGK--
S. iniae            MLQFTSNILATSVAETTQVAPGG^CCCCCCTCCVAVNVGSG-SAQGGSGTPAPAPK-
S. equi             MLQFASNILATSVAETTQVAPGG^CCCCCSCCCCVSASWGNTTINNNYGAAEPKA--
S. anginosus_1      MLKFSSNVLATSVADTTQVAPGG^CCCCCCTCCFSVAVGGN-A-TGGSTTGSVAPTK
S. anginosus_2      MLKLDSHIMATSVAETTQVAPGG^CCCCCCTCCFSVAVGGN-A-TGGSTNIKP----
                    **:: *.::*****:******** *****  **      .
```

B

SagA point mutant	Bioactivity	Core peptide region	Reference
C24A	–	N-terminal heterocyclizable	[7] & [22]
C25A	+	N-terminal heterocyclizable	[7]
C26A	++	N-terminal heterocyclizable	[7]
C27A	+ / –	N-terminal heterocyclizable	[7] / [22]
C28A	+++	N-terminal heterocyclizable	[7]
T29V	+++	N-terminal heterocyclizable	[7]
T30V	++	N-terminal heterocyclizable	[7]
C31A	+++	N-terminal heterocyclizable	[7]
C32A	–	N-terminal heterocyclizable	[7]
S34A	+	C-terminal variable	[7]
T37A	+++	C-terminal variable	[7]
S39A	–	C-terminal variable	[7]
S42A	+++	C-terminal variable	[7]
S46A	+	C-terminal variable	[7]
S48A	++	C-terminal variable	[7]
T50V	+++	C-terminal variable	[7]
K53A	–	C-terminal variable	[22]

Figure B.2. Alignment of experimentally validated streptococcal SLS variants and SLS structure-activity relationships. (A) Alignment using Clustal Omega[50] reveals that the SLS core region possesses a highly conserved N-terminus containing 9–10 contiguous heterocylizable residues (underlined), while the C-

terminus is variable in terms of both identity of residues and length. This phylogenetic data is deposited in the Dryad Digital Repository, doi:10.5061/dryad.d4863. The putative leader peptide cleavage sites are shown as carets. Residues of SagA from *S. pyogenes* are numbered. C-terminal residues where alanine substitutions were previously installed and that are specifically referred to in the article are marked in red (see panel B). (B) Summary of cytolytic activity of SagA core peptide mutants measured by expression in *S. pyogenes ΔsagA*. Activity on blood agar equal to complementation with wild-type SagA is designated as +++, 30–70 % activity is ++, less than 30 % activity is +, and non-detectable activity is –. C-terminal mutations specifically referred to in the article are marked in red. Mutants were expressed from pDCerm under control of a constitutive promoter[7] or from pAD under control of the native *sagA* promoter.[23]

Despite intense study, the precise mechanism of SLS toxicity towards mammalian cells remains incompletely understood.[22] SLS has also thus far been recalcitrant to structural elucidation owing to its poor physicochemical properties, although an extensive mass spectral analysis detected two oxazole moieties at positions S46 and S48 of SLS modified with purified biosynthetic enzymes[18] (Figure B.3A). Previous mutagenesis studies indicate that residues spanning the entire SLS core peptide are required for *S. pyogenes* hemolysis[8, 23] (Figure B.2B). This contrasts with the highly conserved N-terminal poly-heterocyclizable (NPH) region and variable C-terminus of SLS natural variants (Figure B.2A). Here, we sought to reconcile these discrepancies by probing the contribution of the C-terminus to SLS bioactivity. We demonstrate that severely C-terminally truncated SLS peptides expressed in a *sagA* deletion mutant of *S. pyogenes* M1 5448 (*S. pyogenes ΔsagA*) retain hemolytic activity and posit that the NPH region found in all TOMM cytolysins is the defining feature of the family (Figure B.3A, underlined portion). Using bioinformatics, we identified a set of SagA homologs containing the NPH region in TOMM biosynthetic clusters in additional *Firmicutes*, as well as the *Actinobacteria* and *Spirochaete* phyla. We demonstrate that a naturally truncated precursor peptide encoded by a member of *Borrelia burgdorferi* sensu lato (*Bbsl*) species group can be matured into a cytolysin using purified SLS biosynthetic enzymes. Notably, several *Bbsl* cause Lyme disease, the most commonly reported tick-borne illness in the Northern hemisphere,[24, 25] and our PCR-based screen revealed that genes encoding SLS-like precursors are prevalent in diverse *Bbsl* isolates. While our preliminary findings are intriguing, it remains to be seen if these putative cytolytic TOMMs are produced by *Borrelia*, and if so, whether they function to enhance virulence. Together, our results identify the minimal cytoytic determinant of SLS and indicate that there are potentially considerably more producers of related cytolysins than currently appreciated.

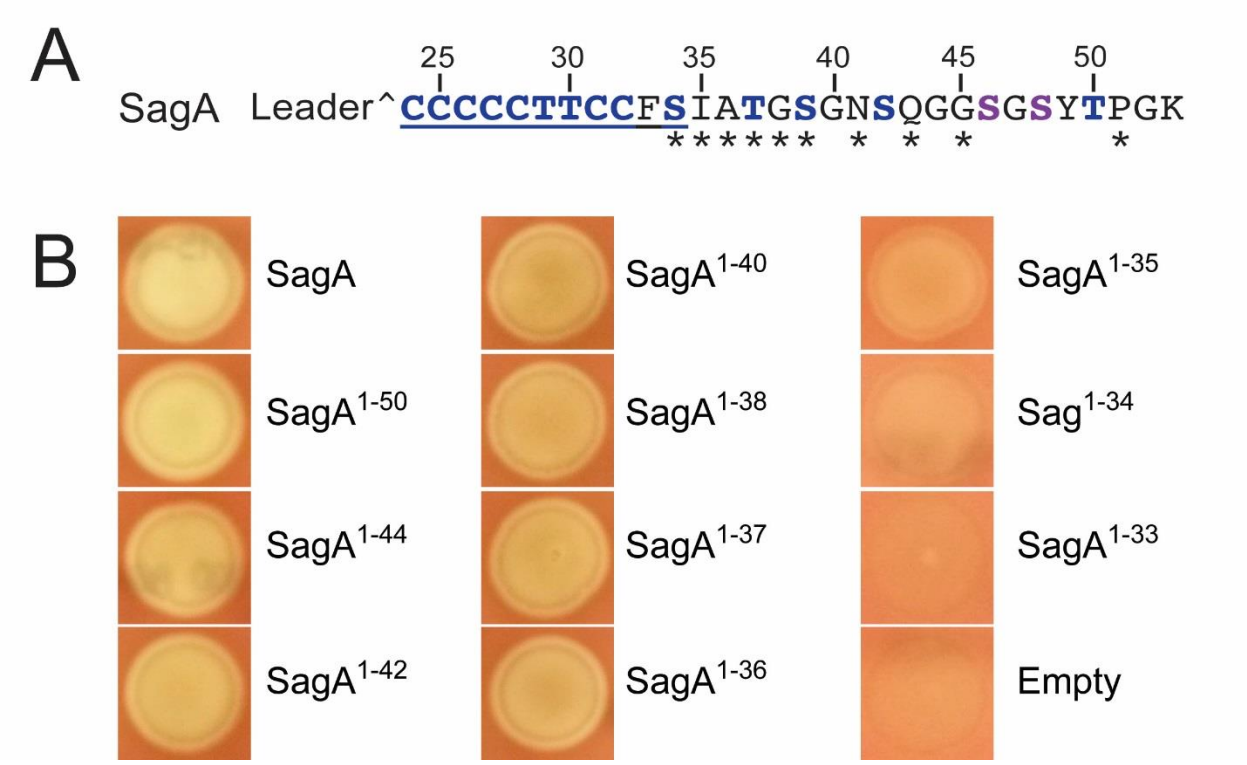


Figure B.3. Assessment of the hemolytic activity of truncated SLS precursor peptides expressed in *S. pyogenes ΔsagA*. (A) Residues of the SagA core peptide are numbered. Potentially modified residues of SLS are blue, while those known to be heterocyclized when SagA is modified with purified SagBCD are purple. The predicted leader cleavage site is indicated with a caret. The positions of stop codons introduced in SagA are marked with asterisks (yields truncation mutants listed in panel B). The minimal unit required for hemolytic activity of SLS in *S. pyogenes ΔsagA* ($SagA^{24-34}$) is underlined (see panel B). (B) The labels correspond to the SagA truncation mutants described in panel A. A representative colony on blood agar of GAS M1 *ΔsagA* expressing each precursor from pDCerm is shown. The colony labeled "Empty" was a negative control containing pDCerm vector.

Results and Discussion

Expression of truncated SLS precursors in *S. pyogenes* Δ*sagA* reveals the importance of N-terminal heterocyclizable region for hemolytic activity

An alignment of experimentally validated SLS core peptides revealed a highly conserved N-terminal region, containing 9-10 contiguous heterocyclizable residues, while the C-terminus is considerably more variable in terms of both identity of residues and length (Figure B.2A). Since only the NPH region of the core peptide is apparently under strict evolutionary pressure, we postulated that the C-terminal region might be less important for SLS function. Indeed, heterologous expression data support this, as peptides with divergent C-termini were able to restore the β-hemolytic phenotype in *S. pyogenes*

Δ*sagA*[8] (Figure B.1B). To test our hypothesis, we installed premature stop codons in *sagA* (Figure B.3A) and expressed the resultant C-terminally truncated SLS derivatives in *S. pyogenes ΔsagA*. Unfortunately, the ring-state of processed SagA cannot be readily determined, even with advanced spectroscopic methods.[7, 18] We thus relied on a simple phenotypic assay, i.e. β-hemolysis on blood agar, to simultaneously evaluate successful production and export of SLS variants.

Consistent with the divergence of the core peptides (Figure B.2A), SLS retained significant activity upon C-terminal truncation. In fact, *S. pyogenes ΔsagA* expressing SLS derivatives missing up to 17 residues ($SagA^{1-36}$) showed comparable levels of β-hemolysis on blood agar to that complimented with full-length SagA (Figure B.3B). Hemolytic activity substantially decreased with additional truncation to yield $SagA^{1-35}$ and $SagA^{1-34}$, with activity being undetectable in the case of $SagA^{1-33}$ (Figure B.3B). These findings allowed us to define $SagA^{1-34}$ as the minimal unit required for hemolytic activity of SLS expressed by *S. pyogenes*. Interestingly, this segment encompasses the entirety of the conserved NPH region (Figure B.2A), as well as just two additional residues (Figure B.3A). Based on the predicted leader peptide cleavage site,[23] this indicates that only 11 core residues are required for detectable bioactivity of SLS expressed by *S. pyogenes*.

The retention of activity by such drastically C-terminally truncated SLS derivatives expressed in *S. pyogenes ΔsagA* stands in contrast to alanine substitution studies that implicated several of the "missing" residues in SLS activity by both expression in *S. pyogenes ΔsagA* and by modification using purified enzymes (Figure B.2B).[8, 23] While a K53A mutation eliminated SLS hemolytic activity from *S. pyogenes*,[23] our truncation series demonstrated that its removal is not detrimental to activity (Figure B.3B). Our results are consistent with the existence of natural SLS variants that lack a terminal lysine (Figure B.2A). Moreover, known SLS-like toxins encoded by non-*Streptoccocus* species also lack terminal lysines (Figure B.1B). Even more striking is our observation that peptides that lack S39 retain cytolytic activity (Figure B.3B). S39 was identified as a critical residue because *S. pyogenes ΔsagA* expressing SagA-S39A was devoid of cytolytic activity and avirulent in a mouse infection model[8] (Figure B.2B). As before, this position is not conserved in the SLS family (Figure B.2A) or in SLS-like toxins

encoded by non-*Streptoccocus* species (Figure B.1B). Finally, many of our mutants that lack the unconserved residues S46 and S48 are hemolytic to levels indistinguishable from WT. Previous mutagenesis suggested that these positions contribute to the hemolytic properties of SLS both in *S. pyogenes* and when modified with purified enzymes.[8] S46 and S48 have also been shown to be cyclized when SLS is modified with its cognate purified biosynthetic enzymes,[18] although it is not known if this holds true for native SLS. Thus, many positions that were previously implicated in SLS activity by point mutation are not conserved across the SLS family and were herein found to be dispensable. In the absence of a complete structure for SLS or detailed mechanism of action, it is unclear why mutation of a residue can be more detrimental than removal, but it is possible that the "alanine-substituted" toxins adopt an unproductive topology that is incapable of lysing cells.

Bioinformatics-guided discovery of novel TOMM cytolysins

Given that the number of bioinformatically identifiable TOMM biosynthetic clusters more than quadrupled since 2011,[9, 26] we sought to determine if naturally truncated precursor peptides are present in any novel SLS-like clusters. Our genome mining effort revealed TOMM clusters encoding SagB, -C and –D homologs in the vicinity of SagA homologs in *Firmicutes* genera not previously known to possess such clusters (i.e. *Bacillus*, *Lactobacillus*, *Enterococcus*, *Exiguobacterium*, *Oenococcus* and *Virgibacillus*) (Figure B.4 & Table B.1). Interestingly, a number of *Actinobacteria* (*Mobiluncus* and *Propionibacterium*) and *Spirochaetes* (*Borrelia*, *Brachyspira*) encoded SLS-like clusters (Figure B.4A & Table B.1). The *Brachyspira* TOMM clusters were independently noted in a recent publication.[27] We hypothesized that the newly-annotated SLS-like precursors (Figure B.4B) represent novel TOMM cytolysins since the sequence of a TOMM precursor dictates the function of the natural product after posttranslational modification.[7, 9] Moreover, while the novel SagA-like precursor peptides were of variable length, all contained the aforementioned NPH region. Our identification of these putative SLS-like cytolysins greatly expands the TOMM cytolysin family (Figure B.5 & Table B.2) and suggests that the TOMM cytolysins are not confined to *Firmicutes*.

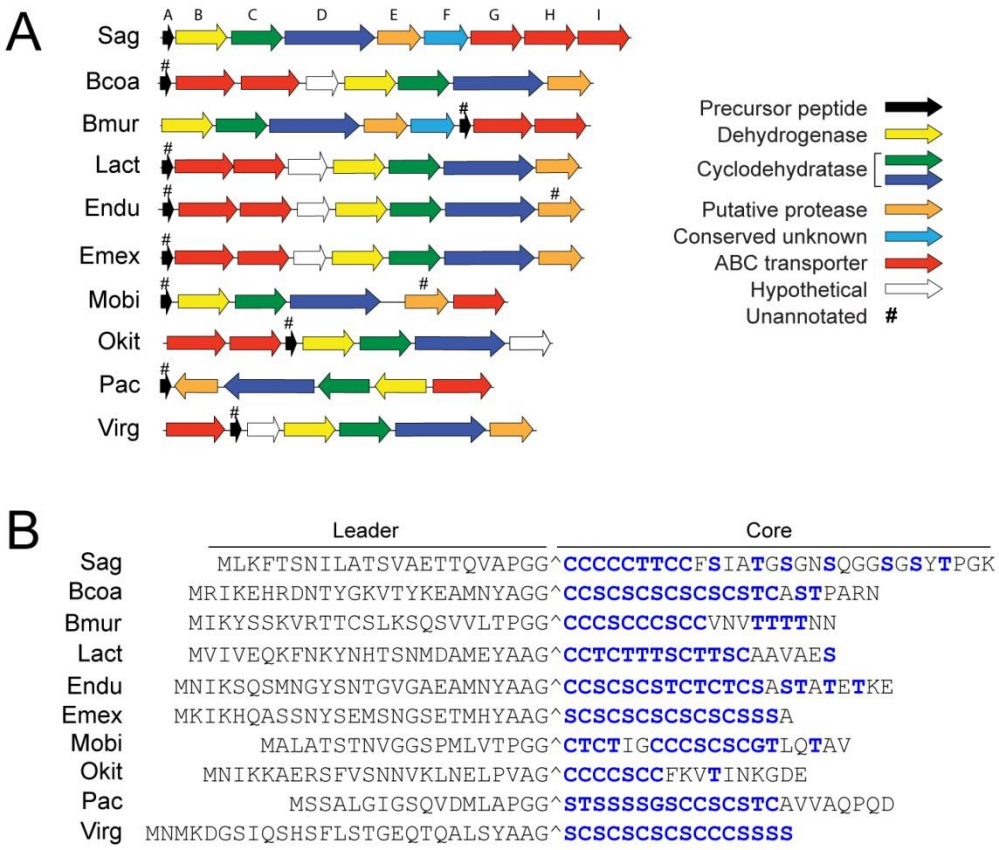


Figure B.4. Gene cluster organization and precursor peptide sequences of predicted SLS-like TOMM biosynthetic gene clusters. (A) One example is given for each genus with strain abbreviations as per Table B.1A. Lettering corresponds to the SLS operon "*sag*" genes. The proposed function of each gene is color-coded according to the legend. In the case of *Lactococcus* and *Catelliglobosispora* (included in Fig. B.1.5), the protein sequences are not associated with a genome in GenBank, making it difficult to confirm all surrounding genes; thus, a representative cluster organization was not given. Those clusters identified in *Borrelia* are depicted in Fig. B.1.6 and Fig. B.1.7. (B) Potentially modified residues are shown in blue. The predicted leader cleavage sites are indicated with a caret.

A

Abr.	Strain
Bval	*Borrelia valasiana* VS116
Sag	*Streptococcus pyogenes* M1 GAS
Bts	*Clostridium botulinum* A str. ATCC 3502
Lls	*Listeria monocytogenes* serotype 4b str. F2365
Sts	*Staphylococcus aureus* subsp. *aureus* JKD6159
Bmur	*Brachyspira murdochii* DSM 12563
Lact	*Lactocacillus crispatus* MV-1A-US
Endu	*Enterococcus durans* ATCC 6056
Emex	*Exiguobacterium mexicanum* str. HUD
Okit	*Oenococcus kitaharae* DSM 17330
Pac	*Propionibacterium acnes* HL099PA1
Virg	*Virgibacillus alimentarius* str. J18T
Mobi	*Mobiluncus mulieris* FB024-16
Bcoa	*Bacillus coagulans* XZL9

B

	B protein (Accession)	C protein (Accession)	D protein (Accession)
Bval	ACN53116.1	ACN53117.1	ACN53119.1
Sag	AAK33685.1	AAK33686.1	AAK33687.1
Bts	YP_001386444.1	YP_001386445.1	YP_001386446.1
Lls	YP_013716.1	YP_013717.1	YP_013718.1
Sts	ADL23377.1	ADL23376.1	ADL23375.1
Bmur	ADG71093.1	ADG71092.1	ADG71091.1
Lact	WP_005721909.1	WP_005729239.1	WP_005721911.1
Endu	EOT25570.1	EOT25569.1	EOT25568.1
Emex	KGI86368.1	KGI86369.1	KGI86370.1
Okit	EHN58990.1	EHN58991.1	EHN58992.1
Pac	EGF73207.1	EGF73206.1	EGF73205.1
Virg	WP_029270587.1	Wp_029270585.1	WP_02970583.1
Mobi	EFN92400.1	EFN92374.1	EFN92445.1
Bcoa	WP_017550909.1	WP_026104562.1	WP_017550911.1

C

	B protein (e-value)	C protein (e-value)	D protein (e-value)
Bval	0	0	0
Sag	2e-37	1e-29	1e-58
Bts	2e-40	4e-51	5e-80
Lls	4e-38	3e-06	2e-36
Sts	4e-39	3e-06	2e-40
Bmur	3e-33	2e-38	9e-67
Lact	1e-34	3e-04	9e-37
Endu	8e-37	3e-07	4e-36
Emex	4e-37	2e-05	2e-32
Okit	2e-33	2e-13	1e-47
Pac	2e-23	2e-04	5e-16
Virg	2e-38	3e-08	1e-44
Mobi	1e-19	7e-10	3e-22
Bcoa	2e-36	1e-08	1e-34

Table B.1. Accession numbers and expectation values of BCD proteins from representative SLS-like TOMM clusters. (A) Strains used for the similarity analysis and the corresponding abbreviations (continued in panels B and C). (B) GenBank accession numbers (ncbi.nlm.nih.gov) of the B, C and D proteins used for the analysis. (C) BLAST e-values for the designated proteins upon comparison with the *B. valaisiana* VS116 homolog. E-values were determined using BLAST-P with standard settings and limiting the searched organisms to include only the strains in panel A, except *B. valaisiana* VS116.

Figure B.5. Phylogenetic analysis of the TOMM cytolysin family. A maximum-likelihood tree based on the D protein (listed in Table B.2) for the majority of known and putative producers of cytolytic TOMMs (as of December 2014). This phylogenetic data is deposited in the Dryad Digital Repository, doi:10.5061/dryad.d4863. Clades are color-coded based on the originating genus for each SLS-like toxin, as represented in the legend. The relevant phyla are given in parentheses.

Organism	Accession
Bacillus coagulans	WP_017550911.1
Borrelia afzelii	WP_014486303.1
Borrelia afzelii PKo	YP_005592169.1
Borrelia garinii	WP_029362437.1
Borrelia spielmanii	WP_012665778.1
Borrelia spielmanii A14S	ACN53165.1
Borrelia valaisiana	WP_012665421.1
Borrelia valaisiana VS116	YP_002642260.1
Brachyspira alvinipulli	WP_028330607.1
Brachyspira hampsonii	WP_008723840.1
Brachyspira hampsonii 30446	EKV56787.1
Brachyspira hampsonii 30599	ELV05843.1
Brachyspira hyodysenteriae	WP_012671839.1
Brachyspira hyodysenteriae WA1	YP_002722513.1
Brachyspira innocens	WP_020004891.1
Brachyspira intermedia	WP_014488740.1
Brachyspira intermedia PWS/A	AEM22928.1
Brachyspira murdochii	WP_013113517.1
Brachyspira murdochii DSM 12563	YP_003633290.1
Brachyspira pilosicoli	WP_014933816.1
Brachyspira pilosicoli 95/1000	YP_003785416.1
Brachyspira pilosicoli B2904	AFR69796.1
Brachyspira pilosicoli P43/6/78	AGA67271.1
Brachyspira pilosicoli WesB	CCG57701.1
Catelliglobosispora koreensis	WP_026208092.1
Clostridium algidicarnis	WP_029453539.1
Clostridium botulinum	KGO12978.1
Clostridium botulinum A str. ATCC 19397	YP_001382880.1
Clostridium botulinum A str. ATCC 3502	YP_001253033.1
Clostridium botulinum A str. Hall	ABS36674.1
Clostridium botulinum A2 str. Kyoto	YP_002802818.1
Clostridium botulinum A2B3 87	KEI98503.1
Clostridium botulinum A3 str. Loch Maree	YP_001785837.1
Clostridium botulinum Af84	EPS53302.1
Clostridium botulinum B str. Osaka05	GAE00980.1
Clostridium botulinum B1 str. Okra	YP_001780137.1
Clostridium botulinum Ba4 str. 657	ACQ55104.1
Clostridium botulinum Bf	EDT87439.1
Clostridium botulinum CFSAN001627	EKN42007.1
Clostridium botulinum CFSAN001628	EKX80885.1
Clostridium botulinum F str. Langeland	ABS40086.1
Clostridium botulinum H04402 065	YP_005677074.1
Clostridium botulinum NCTC 2916	EDT83567.1
Clostridium papyrosolvens DSM 2782	WP_004620508.1
Clostridium sporogenes	KCZ68546.1
Clostridium sporogenes ATCC 15579	EDU38309.1
Clostridium sporogenes PA 3679	EHN15787.1
Enterococcus caccae	WP_010772870.1
Enterococcus caccae ATCC BAA-1240	EOL43456.1
Enterococcus casseliflavus	WP_010748138.1
Enterococcus durans	WP_016176860.1
Enterococcus durans ATCC 6056	EOT25568.1
Enterococcus faecalis	WP_010829081.1
Enterococcus faecalis 02-MB-BW-10	EPH79378.1
Enterococcus faecalis 20.SD.W.06	EPI05532.1
Enterococcus faecalis ATCC 10100	EOK06979.1
Enterococcus faecalis ATCC 27275	EOJ36344.1
Enterococcus faecalis ATCC 6055	EOK09817.1
Enterococcus faecalis B373	ETU56355.1
Enterococcus faecalis B375	ETU59690.1
Enterococcus faecalis B388	ETU66734.1
Enterococcus faecalis BM4654	ERT22326.1
Enterococcus faecalis CH188	EEU83936.1
Enterococcus faecalis DS5	EEU65361.1
Enterococcus faecalis EnGen0067	EOE36355.1
Enterococcus faecalis EnGen0068	EOK60899.1
Enterococcus faecalis EnGen0069	EOK61941.1
Enterococcus faecalis EnGen0070	EOE40348.1
Enterococcus faecalis EnGen0078	EOE07317.1
Enterococcus faecalis EnGen0079	EOE11069.1
Enterococcus faecalis EnGen0085	EOE69387.1
Enterococcus faecalis EnGen0087	EOF17228.1
Enterococcus faecalis EnGen0088	EOE45978.1
Enterococcus faecalis EnGen0089	EOE52128.1
Enterococcus faecalis EnGen0090	EOE54055.1
Enterococcus faecalis EnGen0091	EOE59226.1
Enterococcus faecalis EnGen0092	EOE68244.1
Enterococcus faecalis EnGen0093	EOE77632.1
Enterococcus faecalis EnGen0094	EOE81271.1
Enterococcus faecalis EnGen0095	EOE86625.1
Enterococcus faecalis EnGen0096	EOE88375.1
Enterococcus faecalis EnGen0097	EOE93276.1
Enterococcus faecalis EnGen0098	EOF02369.1
Enterococcus faecalis EnGen0099	EOF03324.1
Enterococcus faecalis EnGen0100	EOF12774.1
Enterococcus faecalis EnGen0101	EOF34985.1
Enterococcus faecalis EnGen0106	EOE37509.1
Enterococcus faecalis EnGen0107	EOF15906.1
Enterococcus faecalis EnGen0108	EOF22854.1
Enterococcus faecalis EnGen0109	EOE67233.1
Enterococcus faecalis EnGen0110	EOE60833.1
Enterococcus faecalis EnGen0111	EOE71939.1
Enterococcus faecalis EnGen0112	EOE98929.1
Enterococcus faecalis EnGen0113	EOF06702.1
Enterococcus faecalis EnGen0114	EOF06780.1
Enterococcus faecalis EnGen0119	EOE78791.1
Enterococcus faecalis EnGen0120	EOE50960.1
Enterococcus faecalis EnGen0194	EOG32052.1
Enterococcus faecalis EnGen0195	EOG41201.1
Enterococcus faecalis EnGen0196	EOG40520.1
Enterococcus faecalis EnGen0197	EOG45747.1
Enterococcus faecalis EnGen0198	EOG49159.1
Enterococcus faecalis EnGen0199	EOG52184.1
Enterococcus faecalis EnGen0200	EOG55956.1
Enterococcus faecalis EnGen0201	EOM47171.1
Enterococcus faecalis EnGen0202	EOM47547.1
Enterococcus faecalis EnGen0203	EOM54058.1
Enterococcus faecalis EnGen0204	EOG59260.1
Enterococcus faecalis EnGen0205	EOG64505.1
Enterococcus faecalis EnGen0206	EOG71661.1
Enterococcus faecalis EnGen0207	EOG59433.1
Enterococcus faecalis EnGen0208	EOG77170.1
Enterococcus faecalis EnGen0209	EOG81470.1
Enterococcus faecalis EnGen0210	EOG85030.1
Enterococcus faecalis EnGen0211	EOG83552.1
Enterococcus faecalis EnGen0212	EOG90137.1
Enterococcus faecalis EnGen0213	EOG93561.1
Enterococcus faecalis EnGen0214	EOG94599.1
Enterococcus faecalis EnGen0215	EOG99089.1
Enterococcus faecalis EnGen0216	EOH04293.1
Enterococcus faecalis EnGen0217	EOH03814.1
Enterococcus faecalis EnGen0218	EOH06555.1
Enterococcus faecalis EnGen0219	EOH14023.1
Enterococcus faecalis EnGen0220	EOH14498.1
Enterococcus faecalis EnGen0221	EOH15060.1
Enterococcus faecalis EnGen0222	EOH24906.1
Enterococcus faecalis EnGen0223	EOH23478.1
Enterococcus faecalis EnGen0224	EOH25096.1
Enterococcus faecalis EnGen0225	EOH33630.1
Enterococcus faecalis EnGen0226	EOH33308.1
Enterococcus faecalis EnGen0228	EOG68742.1
Enterococcus faecalis EnGen0231	EOI26955.1
Enterococcus faecalis EnGen0232	EOM28329.1
Enterococcus faecalis EnGen0235	EOH60041.1
Enterococcus faecalis EnGen0244	EOI09977.1
Enterococcus faecalis EnGen0245	EOL60151.1
Enterococcus faecalis EnGen0246	EOL64825.1
Enterococcus faecalis EnGen0247	EOL68252.1
Enterococcus faecalis EnGen0251	EOI27763.1
Enterococcus faecalis EnGen0252	EOI19958.1
Enterococcus faecalis EnGen0253	EOM22453.1
Enterococcus faecalis EnGen0280	EOL73649.1
Enterococcus faecalis EnGen0298	EOI96701.1
Enterococcus faecalis EnGen0302	EOJ00228.1
Enterococcus faecalis EnGen0303	EOL87177.1
Enterococcus faecalis EnGen0306	EOJ07468.1
Enterococcus faecalis EnGen0310	EOI79100.1
Enterococcus faecalis EnGen0311	EOI95446.1
Enterococcus faecalis EnGen0334	EOK27181.1
Enterococcus faecalis EnGen0336	EOJ65123.1
Enterococcus faecalis EnGen0337	EOJ74371.1
Enterococcus faecalis EnGen0338	EOK19004.1
Enterococcus faecalis EnGen0340	EOJ96539.1
Enterococcus faecalis EnGen0344	EOL96003.1
Enterococcus faecalis EnGen0348	EOL36132.1
Enterococcus faecalis EnGen0355	EOJ75184.1
Enterococcus faecalis EnGen0359	EOJ95323.1
Enterococcus faecalis EnGen0360	EOK03757.1
Enterococcus faecalis EnGen0362	EOL34134.1
Enterococcus faecalis EnGen0374	EOG72352.1
Enterococcus faecalis GA2	KAJ57765.1
Enterococcus faecalis GAN13	KAJ80272.1
Enterococcus faecalis MD6	KAJ71606.1
Enterococcus faecalis MN16	KAJ64407.1
Enterococcus faecalis MTUP9	KAJ81196.1
Enterococcus faecalis MTmid8	KAJ70704.1
Enterococcus faecalis NJ44	KAJ85908.1
Enterococcus faecalis TX0027	EFT47976.1
Enterococcus faecalis TX0309A	EFU95029.1
Enterococcus faecalis TX0470	EFQ69779.1
Enterococcus faecalis TX0645	EFU04723.1
Enterococcus faecalis TX1341	EFU13276.1
Enterococcus faecalis TX4244	EFT92251.1
Enterococcus faecalis TX4248	EFM84307.1
Enterococcus flavescens ATCC 49996	EOH83464.1
Exiguobacterium mexicanum	KGI86370.1
Lactobacillus crispatus	WP_005721911.1
Lactobacillus crispatus 125-2-CHN	EEU18905.1

Table B.2. List of accession identifiers for D proteins of putative TOMM cytolysin clusters. The genus, species, strain, and GenBank accession identifiers are given for all D proteins used in the phylogenetic analysis (Fig. B.1.5), with inherent duplications from GenBank RefSeq annotations deleted.

Organism	Accession
Lactobacillus crispatus 2029	ETW72903.1
Lactobacillus crispatus CTV-05	EFQ45481.1
Lactobacillus crispatus FB049-03	EKB60155.1
Lactobacillus crispatus FB077-07	EKB62797.1
Lactobacillus crispatus MV-1A-US	EEU28751.1
Lactobacillus crispatus MV-3A-US	EEX30204.1
Lactobacillus helveticus	WP_014918795.1
Lactobacillus helveticus R0052	YP_006655842.1
Lactococcus lactis	WP_029344607.1
Listeria innocua	AHK25022.1
Listeria innocua ATCC 33091	EHN61024.1
Listeria monocytogenes	KEX21484.1
Listeria monocytogenes 07PF0776	AFH79679.1
Listeria monocytogenes FSL N1-017	EFK41963.1
Listeria monocytogenes FSL R2-503	EEW18969.1
Listeria monocytogenes J1816	EGF37831.1
Listeria monocytogenes L312	YP_006754227.1
Listeria monocytogenes SLCC2378	CBY72882.1
Listeria monocytogenes SLCC2755	CBY48694.1
Listeria monocytogenes WSLC1042	AHJ35174.1
Listeria monocytogenes serotype 4b str. CLIP 80459	YP_002757821.1
Listeria monocytogenes serotype 4b str. F2365	AAT03895.1
Listeria monocytogenes serotype 4b str. LL195	CCO63704.1
Listeria monocytogenes serotype 4bV str. LS542	ERH87138.1
Listeria monocytogenes serotype 4bV str. LS642	ERH79260.1
Listeria monocytogenes serotype 4bV str. LS643	ERH87937.1
Listeria monocytogenes serotype 4bV str. LS644	ERH82705.1
Listeria monocytogenes serotype 4bV str. LS645	ERH81297.1
Listeria monocytogenes serotype 7 str. SLCC2482	YP_006693101.1
Listeria monocytogenes str. 4b H7858	EAL10365.1
Listeria seeligeri FSL S4-171	EFS02713.1
Mobiluncus mulieris	WP_004018421.1
Mobiluncus mulieris FB024-16	EFN92445.1
Oenococcus kitaharae	WP_007745654.1
Oenococcus kitaharae DSM 17330	EHN58992.1
Propionibacterium	WP_002519916.1
Propionibacterium acnes	WP_013069418.1
Propionibacterium acnes HL005PA1	EFT29310.1
Propionibacterium acnes HL007PA1	EFS69869.1
Propionibacterium acnes HL038PA1	EFT69160.1
Propionibacterium acnes HL043PA1	EGE95685.1
Propionibacterium acnes HL043PA2	EGE93488.1
Propionibacterium acnes HL045PA1	EFT21539.1
Propionibacterium acnes HL053PA1	EFT19487.1
Propionibacterium acnes HL056PA1	EFS70668.1
Propionibacterium acnes HL074PA1	EFS37246.1
Propionibacterium acnes HL078PA1	EFT54407.1
Propionibacterium acnes HL083PA1	EFS49402.1
Propionibacterium acnes HL086PA1	EFS77213.1
Propionibacterium acnes HL096PA1	YP_007886541.1
Propionibacterium acnes HL096PA2	EGE74424.1
Propionibacterium acnes HL099PA1	EGF73205.1
Propionibacterium acnes SK137	ADD99586.1
Propionibacterium acnes SK182	EGR94682.1
Propionibacterium sp. KPL1854	ERS36581.1
Propionibacterium sp. KPL2009	ERS25567.1
Staphylococcus aureus	WP_000627618.1
Staphylococcus aureus RF122	YP_416846.1
Staphylococcus aureus subsp. *aureus* JKD6159	ADL23375.1
Staphylococcus aureus subsp. *aureus* USA300_TCH959	EES92840.1
Staphylococcus lugdunensis	WP_002492067.1
Staphylococcus lugdunensis ACS-027-V-Sch2	EKS22906.1
Staphylococcus lugdunensis HKU09-01	YP_003470294.1
Staphylococcus lugdunensis M23590	EFU83565.1
Staphylococcus lugdunensis N920143	YP_005759171.1
Staphylococcus lugdunensis UCIM6116	EVI52859.1
Staphylococcus lugdunensis VCU139	EHS05291.1
Staphylococcus lugdunensis VCU148	KAK62221.1
Staphylococcus lugdunensis VCU150	KAK58134.1
Streptococcus anginosus	WP_003036343.1
Streptococcus anginosus C238	YP_008500116.1
Streptococcus anginosus SK52 = DSM 20563	GAD36919.1
Streptococcus anginosus T5	GAD46826.1
Streptococcus anginosus group	WP_003031070.1
Streptococcus anginosus subsp. *whileyi*	BAO73279.1
Streptococcus anginosus subsp. *whileyi* CCUG 39159	EID21406.1
Streptococcus anginosus subsp. *whileyi* MAS624	BAN62320.1
Streptococcus canis	WP_003047821.1
Streptococcus canis FSL Z3-227	EIQ81603.1
Streptococcus castoreus	WP_027969506.1
Streptococcus constellatus	WP_006267405.1
Streptococcus constellatus subsp. constellatus	BAO73259.1
Streptococcus constellatus subsp. *pharyngis*	BAO73270.1
Streptococcus constellatus subsp. *pharyngis* C1050	AGU79357.1
Streptococcus constellatus subsp. *pharyngis* C232	AGU72233.1
Streptococcus constellatus subsp. *pharyngis* C818	AGU73989.1
Streptococcus constellatus subsp. pharyngis SK1060 = CCUG 46377	EGV10767.1
Streptococcus constellatus subsp. *viborgensis* SK1359	BAO73288.1
Streptococcus didelphis	WP_018366953.1
Streptococcus dysgalactiae	WP_015016859.1
Streptococcus dysgalactiae subsp. *equisimilis* 167	BAN93159.1
Streptococcus dysgalactiae subsp. *equisimilis* AC-2713	YP_006904304.1
Streptococcus dysgalactiae subsp. *equisimilis* ATCC 12394	YP_006012853.1
Streptococcus dysgalactiae subsp. *equisimilis* GGS_124	BAH81209.1
Streptococcus dysgalactiae subsp. *equisimilis* RE378	YP_006859332.1
Streptococcus dysgalactiae subsp. *equisimilis* SK1249	EGL48335.1
Streptococcus dysgalactiae subsp. *equisimilis* SK1250	EGR89087.1
Streptococcus equi	WP_021320450.1
Streptococcus equi subsp. *equi* 4047	CAW92802.1
Streptococcus equi subsp. *ruminatorum* CECT 5772	KED03797.1
Streptococcus equi subsp. *zooepidemicus*	CAX00154.1
Streptococcus equi subsp. *zooepidemicus* ATCC 35246	AEJ24716.1
Streptococcus equi subsp. *zooepidemicus* CY	AIA67973.1
Streptococcus equi subsp. *zooepidemicus* MGCS10565	YP_002122878.1
Streptococcus equi subsp. *zooepidemicus* SzS31A1	EQB24140.1
Streptococcus ictaluri	WP_008087931.1
Streptococcus ictaluri 707-05	EHI70399.1
Streptococcus iniae	AAN16974.1
Streptococcus iniae 9117	EKB52950.1
Streptococcus iniae IUSA1	ESR10088.1
Streptococcus iniae SF1	YP_008056991.1
Streptococcus intermedius SK54	GAD40641.1
Streptococcus phocae C-4	KGR72224.1
Streptococcus porcinus	WP_003085810.1
Streptococcus porcinus str. Jelinkova 176	EGJ28166.1
Streptococcus pseudoporcinus	WP_007895544.1
Streptococcus pseudoporcinus LQ 940-04	EHI65561.1
Streptococcus pseudoporcinus SPIN 20026	EFR43668.1
Streptococcus pyogenes	AIL11847.1
Streptococcus pyogenes A20	YP_006932647.1
Streptococcus pyogenes AA216	KGE56761.1
Streptococcus pyogenes AA472	KGE55663.1
Streptococcus pyogenes ABC020004984	EZL07464.1
Streptococcus pyogenes ABC020005405	EZL04521.1
Streptococcus pyogenes ABC020005716	EZM86425.1
Streptococcus pyogenes ABC020005773	EZM98094.1
Streptococcus pyogenes ABC020005887	EZL66568.1
Streptococcus pyogenes ABC020006103	EZL69740.1
Streptococcus pyogenes ABC020006298	EZL01609.1
Streptococcus pyogenes ABC020006309	EZK99877.1
Streptococcus pyogenes ABC020006345	EZM90158.1
Streptococcus pyogenes ABC020006715	EZM91974.1
Streptococcus pyogenes ABC020007794	EZL00284.1
Streptococcus pyogenes ABC020009156	EZK97848.1
Streptococcus pyogenes ABC020010041	EZK94809.1
Streptococcus pyogenes ABC020013256	EZN18369.1
Streptococcus pyogenes ABC020013551	EZK93305.1
Streptococcus pyogenes ABC020013952	EZL68340.1
Streptococcus pyogenes ABC020014529	EZL63105.1
Streptococcus pyogenes ABC020014690	EZK92279.1
Streptococcus pyogenes ABC020014764	EZM94430.1
Streptococcus pyogenes ABC020014897	EZL62452.1
Streptococcus pyogenes ABC020014925	EZL60662.1
Streptococcus pyogenes ABC020015277	EZK89616.1
Streptococcus pyogenes ABC020015285	EZK91630.1
Streptococcus pyogenes ABC020015292	EZK83714.1
Streptococcus pyogenes ABC020015294	EZK84743.1
Streptococcus pyogenes ABC020016937	EZK86148.1
Streptococcus pyogenes ABC020017280	EZN20443.1
Streptococcus pyogenes ABC020017526	EZL59288.1
Streptococcus pyogenes ABC020017774	EZN21628.1
Streptococcus pyogenes ABC020021452	EZN23373.1
Streptococcus pyogenes ABC020025676	EZK78818.1
Streptococcus pyogenes ABC020026287	EZL54648.1
Streptococcus pyogenes ABC020026799	EZL57966.1
Streptococcus pyogenes ABC020026946	EZK80982.1
Streptococcus pyogenes ABC020029706	EZM22867.1
Streptococcus pyogenes ABC020029711	EZL53727.1
Streptococcus pyogenes ABC020029793	EZL50852.1
Streptococcus pyogenes ABC020030020	EZM96178.1
Streptococcus pyogenes ABC020030063	EZL95891.1
Streptococcus pyogenes ABC020030925	EZL49398.1
Streptococcus pyogenes ABC020031290	EZL48255.1
Streptococcus pyogenes ABC020031898	EZK78048.1
Streptococcus pyogenes ABC020032057	EZN01129.1
Streptococcus pyogenes ABC020032182	EZL46699.1
Streptococcus pyogenes ABC020032183	EZK76765.1
Streptococcus pyogenes ABC020032186	EZM21945.1
Streptococcus pyogenes ABC020032523	EZL44883.1
Streptococcus pyogenes ABC020032535	EZM99319.1
Streptococcus pyogenes ABC020033020	EZL94073.1
Streptococcus pyogenes ABC020033193	EZL43961.1
Streptococcus pyogenes ABC020033233	EZM25036.1
Streptococcus pyogenes ABC020033799	EZL40926.1
Streptococcus pyogenes ABC020033919	EZM19899.1
Streptococcus pyogenes ABC020035427	EZK76235.1
Streptococcus pyogenes ABC020035446	EZM17426.1
Streptococcus pyogenes ABC020037542	EZM16213.1

Table B.2 (cont.)

Organism	Accession
Streptococcus pyogenes ABC020038544	EZM48551.1
Streptococcus pyogenes ABC020038545	EZN02450.1
Streptococcus pyogenes ABC020038558	EZL37853.1
Streptococcus pyogenes ABC020039031	EZM15684.1
Streptococcus pyogenes ABC020039391	EZL38778.1
Streptococcus pyogenes ABC020040643	EZM43240.1
Streptococcus pyogenes ABC020041397	EZN25885.1
Streptococcus pyogenes ABC020041424	EZM10046.1
Streptococcus pyogenes ABC020043540	EZL91280.1
Streptococcus pyogenes ABC020044010	EZK71661.1
Streptococcus pyogenes ABC020044173	EZL35755.1
Streptococcus pyogenes ABC020044188	EZL34919.1
Streptococcus pyogenes ABC020044193	EZL33973.1
Streptococcus pyogenes ABC020044203	EZL84609.1
Streptococcus pyogenes ABC020044412	EZK69905.1
Streptococcus pyogenes ABC020046158	EZM40179.1
Streptococcus pyogenes ABC020046230	EZM59851.1
Streptococcus pyogenes ABC020046264	EZL31600.1
Streptococcus pyogenes ABC020046470	EZL25746.1
Streptococcus pyogenes ABC020046534	EZM34639.1
Streptococcus pyogenes ABC020046559	EZM38035.1
Streptococcus pyogenes ABC020046589	EZN29605.1
Streptococcus pyogenes ABC020046688	EZN27624.1
Streptococcus pyogenes ABC020046986	EZN30941.1
Streptococcus pyogenes ABC020047076	EZL79721.1
Streptococcus pyogenes ABC020047086	EZM60893.1
Streptococcus pyogenes ABC020047328	EZN32978.1
Streptococcus pyogenes ABC020047395	EZK70416.1
Streptococcus pyogenes ABC020047506	EZM54206.1
Streptococcus pyogenes ABC020047619	EZL74515.1
Streptococcus pyogenes ABC020047635	EZM55483.1
Streptococcus pyogenes ABC020047925	EZK65930.1
Streptococcus pyogenes ABC020047955	EZM61725.1
Streptococcus pyogenes ABC020047959	EZK65117.1
Streptococcus pyogenes ABC020047977	EZM76338.1
Streptococcus pyogenes ABC020047993	EZL27947.1
Streptococcus pyogenes ABC020048184	EZK63120.1
Streptococcus pyogenes ABC020048262	EZM34042.1
Streptococcus pyogenes ABC020048387	EZM79356.1
Streptococcus pyogenes ABC020048395	EZL28801.1
Streptococcus pyogenes ABC020048503	EZM63476.1
Streptococcus pyogenes ABC020048541	EZK62834.1
Streptococcus pyogenes ABC020048543	EZL23177.1
Streptococcus pyogenes ABC020048606	EZN06458.1
Streptococcus pyogenes ABC020048885	EZM04805.1
Streptococcus pyogenes ABC020049250	EZL20252.1
Streptococcus pyogenes ABC020049540	EZM02720.1
Streptococcus pyogenes ABC020049545	EZM65212.1
Streptococcus pyogenes ABC020050193	EZN33996.1
Streptococcus pyogenes ABC020050731	EZL21730.1
Streptococcus pyogenes ABC020050818	EZN04456.1
Streptococcus pyogenes ABC020051161	EZM08650.1
Streptococcus pyogenes ABC020051269	EZN39726.1
Streptococcus pyogenes ABC020052023	EZN36612.1
Streptococcus pyogenes ABC020052203	EZL18913.1
Streptococcus pyogenes ABC020052211	EZM58730.1
Streptococcus pyogenes ABC020052216	EZL17908.1
Streptococcus pyogenes ABC020052218	EZN07637.1
Streptococcus pyogenes ABC020052291	EZK60904.1
Streptococcus pyogenes ABC020052309	EZM66994.1
Streptococcus pyogenes ABC020052313	EZM68739.1
Streptococcus pyogenes ABC020052375	EZN09671.1
Streptococcus pyogenes ABC020052378	EZL90141.1
Streptococcus pyogenes ABC020052420	EZL15911.1
Streptococcus pyogenes ABC020052497	EZL75544.1
Streptococcus pyogenes ABC020052519	EZM03953.1
Streptococcus pyogenes ABC020052543	EZN40534.1
Streptococcus pyogenes ABC020052553	EZN11415.1
Streptococcus pyogenes ABC020052558	EZK56964.1
Streptococcus pyogenes ABC020052877	EZN13714.1
Streptococcus pyogenes ABC020052898	EZM77592.1
Streptococcus pyogenes ABC020052980	EZM70484.1
Streptococcus pyogenes ABC020053060	EZM45746.1
Streptococcus pyogenes ABC020053240	EZM01429.1
Streptococcus pyogenes ABC020053250	EZM84630.1
Streptococcus pyogenes ABC020054184	EZM44515.1
Streptococcus pyogenes ABC020054871	EZL13291.1
Streptococcus pyogenes ABC020054955	EZL12666.1
Streptococcus pyogenes ABC020054973	EZK55532.1
Streptococcus pyogenes ABC020055897	EZM00332.1
Streptococcus pyogenes ABC020056020	EZM12753.1
Streptococcus pyogenes ABC020056060	EZL83926.1
Streptococcus pyogenes ABC020056064	EZL97559.1
Streptococcus pyogenes ABC020056068	EZM82887.1
Streptococcus pyogenes ABC020056181	EZN15172.1
Streptococcus pyogenes ABC020056765	EZM72615.1
Streptococcus pyogenes ABC020056794	EZM33516.1
Streptococcus pyogenes ABC020056883	EZN17038.1
Streptococcus pyogenes ABC020056884	EZM56071.1
Streptococcus pyogenes ABC020056885	EZL09064.1
Streptococcus pyogenes ABC020056891	EZL78578.1
Streptococcus pyogenes ABC020056894	EZM50647.1
Streptococcus pyogenes ABC020056898	EZL76407.1
Streptococcus pyogenes ABC020057168	EZL91515.1
Streptococcus pyogenes ABC020057192	EZL70617.1
Streptococcus pyogenes ABC020057288	EZM10802.1
Streptococcus pyogenes ABC020057442	EZM74377.1
Streptococcus pyogenes ABC020057449	EZL85075.1
Streptococcus pyogenes ABC020058759	EZL05836.1
Streptococcus pyogenes ABC020059502	EZM81136.1
Streptococcus pyogenes ABC020060220	EZM27129.1
Streptococcus pyogenes ABC020060777	EZM49091.1
Streptococcus pyogenes ABC020060793	EZN37311.1
Streptococcus pyogenes ABC020061424	EZM88171.1
Streptococcus pyogenes ABC020062474	EZM28932.1
Streptococcus pyogenes ABC020062601	EZM29499.1
Streptococcus pyogenes ABC020064181	EZM38970.1
Streptococcus pyogenes ATCC 10782	EFM33242.1
Streptococcus pyogenes Alab49	YP_006071750.1
Streptococcus pyogenes GA03455	ESU92116.1
Streptococcus pyogenes GA03747	ESU92489.1
Streptococcus pyogenes GA03799	ESU85817.1
Streptococcus pyogenes GA03805	ESA58106.1
Streptococcus pyogenes GA06023	ERL14940.1
Streptococcus pyogenes GA16797	ESU93624.1
Streptococcus pyogenes GA19681	EQL81989.1
Streptococcus pyogenes GA19700	ESA44718.1
Streptococcus pyogenes GA19702	ESU91867.1
Streptococcus pyogenes GA40056	ESA53006.1
Streptococcus pyogenes GA40377	ESA58487.1
Streptococcus pyogenes GA40468	ESA52401.1
Streptococcus pyogenes GA40634	EPZ47781.1
Streptococcus pyogenes GA40884	ESU88551.1
Streptococcus pyogenes GA41039	ESA46931.1
Streptococcus pyogenes GA41046	ERL18782.1
Streptococcus pyogenes GA41208	ESA50455.1
Streptococcus pyogenes GA41345	EPZ46951.1
Streptococcus pyogenes GA41394	ESA54459.1
Streptococcus pyogenes HKU QMH11M0907901	EIK42294.1
Streptococcus pyogenes HSC5	AGQ27390.1
Streptococcus pyogenes M1 476	YP_007587151.1
Streptococcus pyogenes M1 GAS	NP_268966.1
Streptococcus pyogenes MGAS10270	ABF33684.1
Streptococcus pyogenes MGAS10394	YP_059899.1
Streptococcus pyogenes MGAS10750	ABF37597.1
Streptococcus pyogenes MGAS15252	YP_005388641.1
Streptococcus pyogenes MGAS1882	YP_005411340.1
Streptococcus pyogenes MGAS2096	YP_600222.1
Streptococcus pyogenes MGAS2111	KGE60575.1
Streptococcus pyogenes MGAS315	AAM79090.1
Streptococcus pyogenes MGAS5005	AAZ51183.1
Streptococcus pyogenes MGAS6180	AAX71657.1
Streptococcus pyogenes MGAS8232	NP_606967.1
Streptococcus pyogenes MGAS9429	ABF31805.1
Streptococcus pyogenes NS88.2	CCG27133.1
Streptococcus pyogenes NZ131	YP_002285592.1
Streptococcus pyogenes SS1447	KGE59378.1
Streptococcus pyogenes SSI-1	BAC64466.1
Streptococcus pyogenes STAB1102	AIG48613.1
Streptococcus pyogenes STAB901	AIG50128.1
Streptococcus pyogenes STAB902	AIG47640.1
Streptococcus pyogenes UTMEM-1	EQL80632.1
Streptococcus pyogenes UTSW-2	EQL82124.1
Streptococcus pyogenes str. Manfredo	YP_001128783.1
Virgibacillus alimentarius	WP_029270583.1

Table B.2 (cont.)

Of particular interest were the three potential SLS-like TOMM biosynthetic clusters we discovered among the 23 published *Bbsl* genomes (Figure B.6A; hereafter referred to as "Bor TOMM clusters"), namely in *B. afzelii* PKo [GenBank: CP002947],[28] *B. valaisiana* VS116 [GenBank: CP001442],[29] and *B. spielmanii* A14S [GenBank: CP001465][29] (Figure B.6A). The novel *Bbsl*-encoded precursor peptides consisted of SagA homologs that were of similar length to our "minimal cytolytic unit" for SLS (Figure B.6B). We preliminarily classified the putative *Bbsl* SLS-like products as TOMM cytolysins based on sequence similarity to SLS, with support from a maximum likelihood tree of a subset of the TOMM family based on the associated D proteins (Figure B.6C). Thus, we assigned names to the homologous genes within the clusters following the SLS nomenclature, with *borA* encoding the precursor peptide (refers collectively to *bafzA* from *B. afzelii* PKo, *bvalA* from *B. valaisiana* VS116, and *bspiA* from *B. spielmanii*; Figure B.6B), *borB* encoding the dehydrogenase and *borC/D* encoding the cyclodehydratase and so forth.

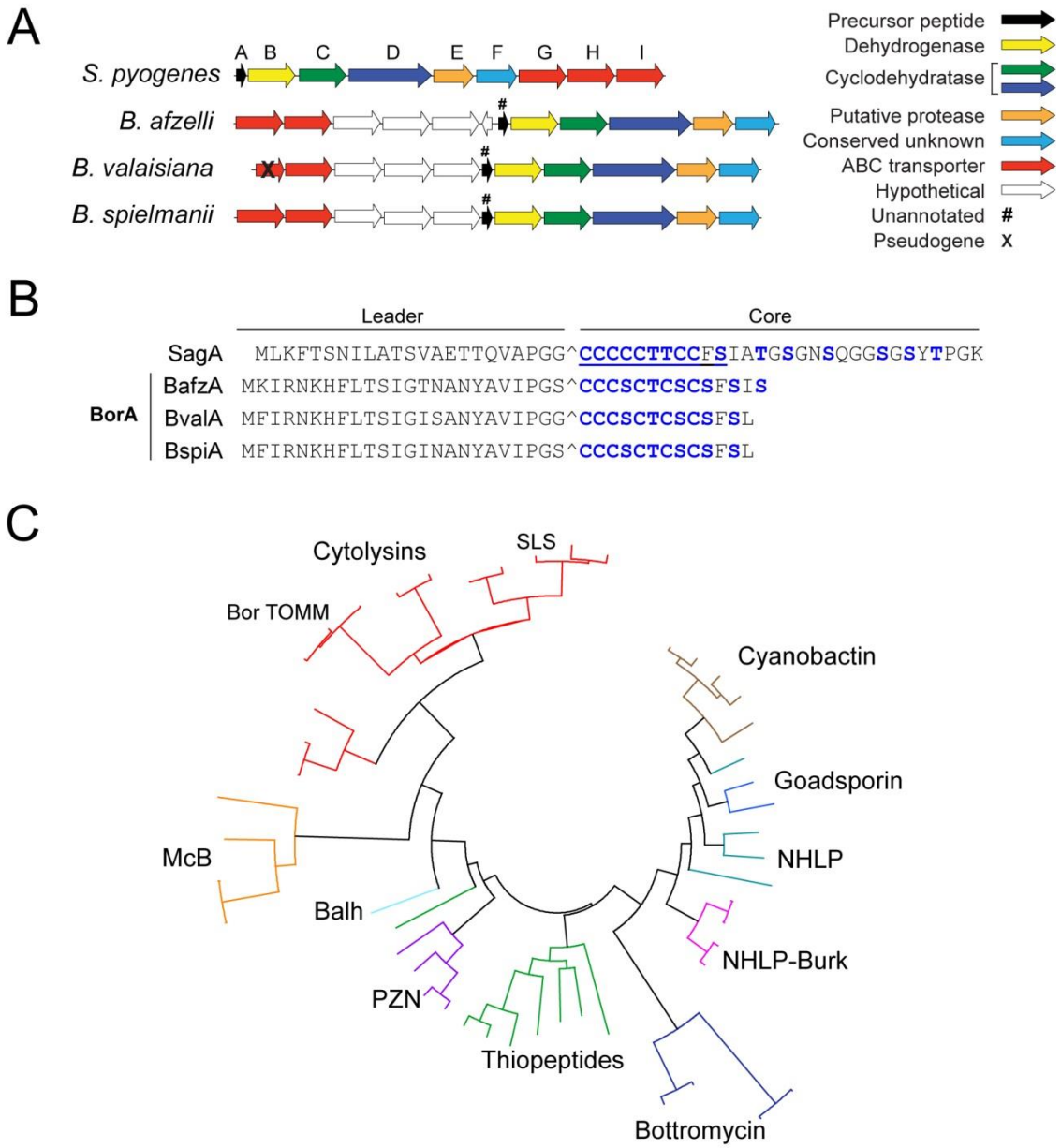


Figure B.6. Gene cluster organization and precursor peptide sequences of SLS and Bor TOMM, and phylogenetic analysis of the TOMM family. (A) Lettering corresponds to the SLS operon "*sag*" genes. The proposed function of each gene is color-coded according to the legend. (B) BorA refers collectively to BafzA from *B. afzelii* PKo, BvalA from *B. valaisiana* VS116, and BspiA from *B. spielmanii* A14S. Potentially modified residues are shown in blue. The predicted leader cleavage sites are indicated with a caret. The minimal core length required for hemolytic activity of SLS in *S. pyogenes ΔsagA* ($SagA^{24-34}$) is underlined. (C) A maximum-likelihood tree of a representative sample of TOMMs based on the D protein from each cluster. This phylogenetic data is depostied in the Dryad Digital Repository, doi:10.5061/dryad.d4863. Clades are color-coded based on the predicted class of TOMM natural product. NHLP, nitrile hydratase leader peptide; NHLP-Burk, NHLP from *Burkholderia*; PZN, plantazolicin; Balh, uncharacterized TOMM from *Bacillus* sp. Al Hakam; McB, microcin B from *Gammaproteobacteria*. SLS, streptolysin S; Bor TOMM, putative SLS-like cytolysin from *Bbsl*.

SLS-like TOMM biosynthetic genes are widely distributed among *Bbsl*

We noted that the Bor TOMM clusters identified in *B. afzelii* PKo, *B. valaisiana* VS116 and *B. spielmanii* A14S were present on lp28-8 plasmids (Figure B.7). These plasmids exhibit 7-9% nucleotide sequence divergence from each other, which is consistent with that observed on the chromosome and other linear plasmids in these species.[30] Moreover, the G+C content of the cluster is 25-26% in all three species, which is typical for *Bbsl* plasmids. It thus appears that the Bor TOMM cluster did not arrive into the genus recently by horizontal transfer, but has likely been present for a long time, diverging in parallel with the chromosome approximately when the three species diverged.

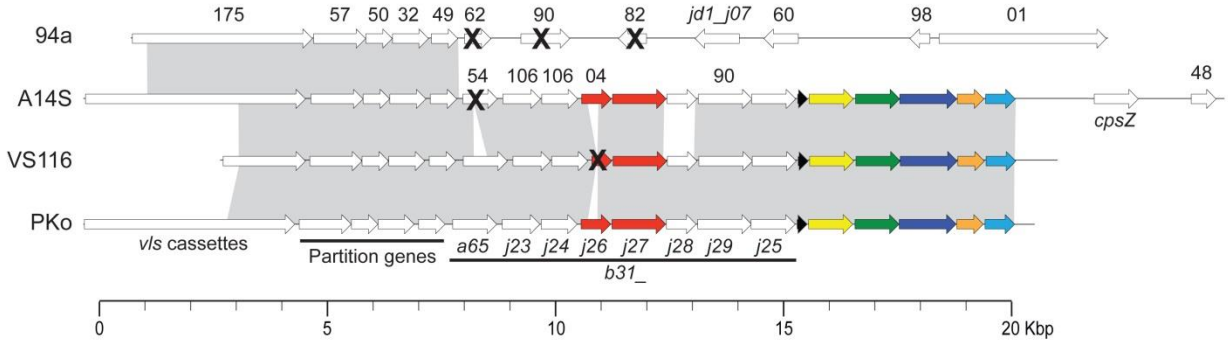


Figure B.7. Open reading frame (ORF) diagrams for *Borrelia* lp28-8 plasmids. Four published *Bbsl* genomes harbor plasmids in the lp28-8 compatibility group.[51] As is typical for *Bbsl* linear plasmids,[52] the lp28-8 plasmids are mosaically related; gray areas between maps denote regions of sequence similarity. Lp28-8 from *B. burgdorferi* 94a lacks the Bor TOMM biosynthetic cluster so is not further discussed. The ORFs of the Bor TOMM clusters of *B. spielmanii* A14S, *B. valaisiana* VS116 and *B. afzelii* PKo are color-coded as in Fig. B.1.1A; white arrows are hypothetical or have functions unrelated to the Bor TOMM biosynthetic cluster. The numbers above an ORF give the Pfam designation. Names of the ORFs known from type strain *B. burgdorferi* B31 are labeled "*b31_*" and are shown below the maps. Each lp28-8 carries a truncated *vls* cassette region, which allows alteration of the antigenic properties of VlsE outer surface proteins, a key pathogenic strategy. An "X" over an ORF indicates a pseudogene. In VS116, the ABC transporter gene has a substantial deletion and may be nonfunctional, suggesting that a different transporter performs export or that VS116 does not produce the Bor TOMM.

We hypothesized that the three Bor TOMM clusters identified in *Bbsl* genomes is an underestimation of the true prevalence given that *Bbsl* are prone to losing plasmids when cultured *in vitro*.[31, 32] In order to estimate the distribution of Bor TOMM clusters in *Bbsl*, we performed a PCR screen for the associated biosynthetic genes in 140 genomic DNA samples from tick or human patient isolates (Table B.3). Seven *Bbsl* species were represented, originating from North America and 13 countries in

Europe and Asia. Because of the similarity among the VS116, PKo, and A14S Bor TOMM cluster sequences, primers specific to *bvalB, -C* and *-D* were used to detect *borB, -C* and *-D* in the isolates (Figure B.8 & Table B.4).

					Presence of gene			
Bbsl* **species**	**Strain**	**Geographic Origin**	**Biological Source**	**P#**	***borB***	***borC***	***borD***	**Source**
B. burgdorferi **sensu stricto (pathogen, North America and Western Europe)**	B297	New York	Skin	4	-	-	-	IS
	B331	New York (Westchester County)	Skin (EM)	5	-	-	-	IS
	B356	New York (Westchester County)	Skin (EM)	3	-	-	-	IS
	B376	New York (Westchester County)	Skin (EM)	4	-	-	-	IS
	B379	New York	Skin (EM)	5	-	-	-	IS
	B418	New York (Westchester County)	Skin (EM)	2	-	-	-	IS
	B477	New York	Skin (EM)	2	-	-	-	IS
	B500	New York (Westchester County)	Skin (EM)	2	-	-	-	IS
	B515	New York (Westchester County)	Skin (EM)	4	-	-	-	IS
	BL206	New York (Westchester County)	Blood (LD)	4	-	-	-	IS
	BL268	New York	Blood	3	-	-	-	IS
	IPT2	France (Alsace)	*I. ricinus* (M)	Unknown	-	-	-	IP
	IPT19	France (Alsace)	*I. ricinus* (M)	Unknown	-	-	-	IP
	IPT23	France (Alsace)	*I. ricinus* (F)	Unknown	-	-	-	IP
	IPT39	France (Alsace)	*I. ricinus* (M)	Unknown	-	-	-	IP
	IPT58	France (Alsace)	*I. ricinus* (F)	Unknown	-	-	-	IP
	IPT69	France (Alsace)	*I. ricinus* (N)	Unknown	-	-	-	IP
	IPT135	France (Auvergne)	*I. ricinus* (F)	Unknown	-	-	-	IP
	IPT137	France (Alsace)	*I. ricinus* (F)	Unknown	-	-	-	IP
	IPT190	France (Normandie)	*I. ricinus* (M)	Unknown	-	-	-	IP
	IPT191	France (Normandie)	*I. ricinus* (F)	Unknown	-	-	-	IP
	IPT193	France (Normandie)	*I. ricinus* (M)	Unknown	-	-	-	IP
	PAli	Germany (Regensburg)	Vitreous body	1	-	-	-	VF
	PDri	Germany (Munich)	Skin (EM & arthritis)	2	-	-	-	VF
B. garnii **(pathogen, Eurasia)**	IPT28	France (Alsace)	*I. ricinus* (M)	Unknown	-	-	-	IP
	IPT114	France (Alsace)	*I. ricinus* (M)	Unknown	-	-	-	IP
	IPT130	France (Alsace)	*I. ricinus* (M)	Unknown	-	-	-	IP
	IPT139	France (Alsace)	*I. ricinus* (F)	Unknown	-	-	-	IP
	IPT140	France (Alsace)	*I. ricinus* (M)	Unknown	-	-	-	IP
	IPT156	France (Auvergne)	*I. ricinus* (M)	Unknown	-	-	-	IP
	IPT157	France (Limousin)	*I. ricinus* (F)	Unknown	-	+	-	IP
	IPT158	France (Limousin)	*I. ricinus* (M)	Unknown	-	-	-	IP

	IPT165	France (Auvergne)	*I. ricinus* (M)	Unknown	-	-	-	IP
	IPT167	France (Limousin)	*I. ricinus* (F)	Unknown	-	-	-	IP
	IPT168	France (Limousin)	*I. ricinus* (M)	Unknown	-	-	-	IP
	IPT169	France (Auvergne)	*I. ricinus* (F)	Unknown	-	+	-	IP
	IPT171	France (Auvergne)	*I. ricinus* (F)	Unknown	-	-	-	IP
	IPT172	France (Auvergne)	*I. ricinus* (F)	Unknown	-	-	-	IP
	IPT178	France (Auvergne)	*I. ricinus* (F)	Unknown	-	-	-	IP
	PBes	Germany (Minden)	CSF (NB)	5	-	+	-	VF
	PBol	Germany (Nurnberg)	CSF (NB)	3	-	-	-	VF
	PBr	Germany (Günzburg)	CSF (NB)	7	-	-	-	VF
	PBu	Germany (Ulm)	Skin (EM)	4	-	+	-	VF
	PFr	Germany (Munich)	CSF (NB)	6	-	-	-	VF
	PHei	Germany (Göttingen)	CSF (NB)	6	-	-	-	VF
	PKi	Germany (Tübingen)	CSF (NB)	11	-	-	-	VF
	PLa	Germany (Koblenz)	CSF (NB)	5	-	-	-	VF
	PRef	Germany (Günzburg)	CSF (NB)	11	-	+	-	VF
	PWudII	Germany (Munich)	Skin (EM)	5	-	-	-	VF
	20047	France	*I. ricinus*	3	-	-	-	SC from RM
***B. afzelii* (pathogen, Eurasia)**	ACA-1[a]	Sweden	Skin (ACA)	<10	+	+	+	SC from SB
	ACA-1[a]	Sweden	Skin (ACA)	3	+	+	+	SC from IS
	ACA-1[a]	Sweden	Skin (ACA)	4	-	-	-	SC from IS
	B023	Germany	Skin (EM)	<10	-	-	-	SC from UM
	DK26	Denmark	Skin (EM)	<10	-	-	-	SC from RM
	EMC1	Sweden	Skin (EM)	3	-	-	+	SC from RM
	HT10	Japan	*I. persulcatus*	<10	+	+	+	SC from RM
	IBS11	France (Alsace)	Skin (EM)	Unknown	-	-	-	IP
	IBS12	France (Alsace)	Skin (EM)	Unknown	+	+	-	IP
	IBS13	France (Alsace)	Skin (ACA)	Unknown	-	+	-	IP
	IPT109	France (Alsace)	*I. ricinus* (F)	Unknown	-	+	-	IP
	IPT110	France (Alsace)	*I. ricinus* (F)	Unknown	+	+	-	IP
	IPT118	France (Auvergne)	*I. ricinus* (F)	Unknown	-	-	-	IP
	IPT122	France (Auvergne)	*I. ricinus* (F)	Unknown	+	+	-	IP
	IPT138	France (Alsace)	*I. ricinus* (F)	Unknown	+	+	-	IP
	IPT142	France (Alsace)	*I. ricinus* (M)	Unknown	-	+	-	IP
	IPT152	France (Limousin)	*I. ricinus* (M)	Unknown	-	+	-	IP
	IPT154	France (Limousin)	*I. ricinus* (F)	Unknown	-	+	-	IP
	IPT164	France (Auvergne)	*I. ricinus* (M)	Unknown	+	+	-	IP
	IPT179	France (Auvergne)	*I. ricinus* (M)	Unknown	+	+	-	IP
	Isau1	Germany (Munich)	*I. ricinus*	4	+	+	+	VF

	J1 ("IPF")	Japan	*I. persulcatus*	<10	-	+	-	SC from RM
	PBas	Germany (Munich)	Skin (borrelial lymphocytoma)	2	+	+	+	VF
	PBec	Germany (Hamburg)	Skin (ACA)	3	+	+	-	VF
	PBil	Germany (Minden)	Skin (EM)	4	+	+	-	VF
	PBo	Germany (Munich)	CSF (NB)	3	+	+	+	VF
	PBog	Germany (Berchtesgarden)	Skin (EM)	6	+	+	+	VF
	PEbe	Germany (Altötting)	Skin (EM)	2	+	+	+	VF
	PFiII	Germany (Munich)	Skin (ACA)	6	-	+	-	VF
	PFuk	Germany (Munich)	Skin (ACA)	3	+	+	+	VF
	PHak	Germany (Munich)	Skin (EM)	3	+	+	+	VF
	PHes	Germany (Bad Homburg)	Skin (ACA)	4	+	+	-	VF
	PKap	Germany (Tübingen)	Skin (ACA)	2	+	+	+	VF
	PKo[b]	Germany	Skin, EM	7	-	-	-	SC from RM
	PKo[b]	Germany (Munich)	Skin (EM)	10	+	+	+	VF
	PKr	Germany (Munich)	CSF (NB)	2	+	+	+	VF
	PObf	Germany (Bad Mertengheim)	Skin	1	-	-	+	VF
	PWe	Germany (Munich)	Skin (ACA)	2	-	-	-	VF
	R-IP3	Russia	*I. persulcatus*	<10	+	+	+	SC from PR
	R-IP21 ("IP21")	Russia	*I. persulcatus*	<10	-	+	+	SC from RM
	TDe	Germany (unknown)	*I. ricinus*	7	+	+	-	VF
	T10	Germany (Munich)	*I. ricinus*	<10	+	+	-	VF
	UM01	Sweden	Skin (EM)	<10	-	+	+	SC from RM
	UO1	Sweden	Skin (EM)	<10	-	-	-	SC from RM
	VS461	Switzerland	*I. ricinus*	<10	-	-	-	SC from RM
	61BV3	Germany (Berlin)	*I. ricinus*	<10	+	+	-	AS
	97B3	Germany (Berlin)	*I. ricinus*	<10	-	-	+	AS
B. spielmanii **(rare pathogen, Eurasia)**	Anzic	Slovenia	Skin (EM)	Unknown	-	-	-	FS
	A14S	The Netherlands	Skin	Low	+	+	+	AvD
	Isau2	Germany (Munich)	*I. ricinus*	2	-	-	-	VF
	PHap	Germany (Munich)	Skin (EM)	2	+	+	+	VF
	PMai	Germany (Munich)	Skin (EM)	7	+	+	+	VF
	PMEW	Germany (Munich)	Skin (EM)	5	+	+	+	VF
	PSigII	Germany (Munich)	Skin (ACA)	7	+	+	+	VF
B. bavariensis **(rare pathogen, Eurasia)**	PBaeI	Germany (Munich)	Skin (EM)	Low	-	-	-	GM from VF[1]
	PBi	Germany (Ingolstadt)	CSF (NB)	5	-	-	-	VF
	PBN	Germany (Munich)	CSF	Low	-	-	-	GM from VF
	PFin	Germany (Munich)	CSF (NB)	8	-	-	-	VF
	PFlk	Germany (Munich)	CSF (EM & NB)	Low	-	-	-	GM from VF

	PHoe	Germany (Munich)	CSF (NB)	Low	-	-	-	GM from VF
	PNi	Germany (Halle)	Skin (EM)	6	-	-	-	VF
	POb	Germany (Munich)	Skin (EM)	Low	-	-	-	GM from VF
	PRab	Austria (Villach)	Hand joint synovia	Low	-	-	-	GM from VF
	PRof	Germany (Munich)	Skin (ACA)	3	-	-	-	VF
	PScf	Germany (Munich)	CSF (NB)	Low	-	-	-	GM from VF
	PTrob	Slovenia	Skin	Low	-	-	-	FS
***B. valaisiana* (probable pathogen, Eurasia)**	IPT29	France (Meuse)	*I. ricinus* (N)	Unknown	-	+	-	IP
	IPT31	France (Meuse)	*I. ricinus* (F)	Unknown	-	+	-	IP
	IPT33	France (Meuse)	*I. ricinus* (N)	Unknown	-	+	-	IP
	IPT47	France (Alsace)	*I. ricinus* (F)	Unknown	-	+	-	IP
	IPT85	France (Alsace)	*I. ricinus* (M)	Unknown	-	+	-	IP
	IPT102	France (Auvergne)	*I. ricinus* (F)	Unknown	-	+	-	IP
	IPT111	France (Alsace)	*I. ricinus* (F)	Unknown	-	+	-	IP
	IPT121	France (Alsace)	*I. ricinus* (F)	Unknown	-	+	-	IP
	IPT144	France (Limousin)	*I. ricinus* (F)	Unknown	+	+	+	IP
	IPT163	France (Auvergne)	*I. ricinus* (F)	Unknown	-	+	-	IP
	IPT174	France (Auvergne)	*I. ricinus* (F)	Unknown	-	+	-	IP
	IPT177	France (Limousin)	*I. ricinus* (F)	Unknown	+	+	+	IP
	IPT184	France (Limousin)	*I. ricinus* (M)	Unknown	-	-	-	IP
	IPT186	France (Limousin)	*I. ricinus* (F)	Unknown	-	+	-	IP
	IPT187	France (Limousin)	*I. ricinus* (F)	Unknown	-	-	-	IP
	IPT188	France (Normandie)	*I. ricinus* (M)	Unknown	-	+	-	IP
	VS116	Switzerland	*I. ricinus*	11	+	+	+	SC from RM
	10MT	Korea	*I. nipponensis*	<10	-	-	-	SC from RM
	61104BT	UK (Bath)	*I. ricinus*	0 (not cultured)	-	-	-	GM
	61214BT	UK (Bath)	*I. ricinus*	0 (not cultured)	-	-	-	GM
	61306L	UK (Bath)	*I. ricinus*	0 (not cultured)	-	-	-	GM
***B. lusitaniae* (probable pathogen, Eurasia)**	Poti B2	Portugal	*I. ricinus*	7	-	+	+	IP
	Poti B3	Portugal	*I. ricinus*	5	-	+	+	IP

Table B.3. PCR screen for Bor TOMM biosynthetic genes. These results provide more detail on those given in Table B.5. Strains marked in gray have published genomes. P# indicates number of passages where known. Male (M), female (F) or nymph (N) is indicated for tick isolates where known. In the case of human isolates, CSF indicates isolation from cerebrospinal fluid. Disease presentation is given where known: EM, erythema migrans; LD, Lyme disease; NB, neuroborreliosis; ACA, acrodermatitis chronica atrophicans. Source abbreviations: IP, Institute Pasteur; VF, Volker Fingerle; IS, Ira Schwartz; SC, Sherwood Casjens; RM, Richard Marconi; AS, Arno Schönberg; GM: Gabriele Margos; SB, Sven Bergstrom; PR, Patricia Rosa; UM, Ulrike Munderloh; FS, Franc Strle; AvD, Alje van Dam. LGL Bayern, Germany. **Bbsl* species group currently contains 21 confirmed and proposed species, including

human pathogens, suspected pathogens and non-pathogenic species.[53-55] [a]The reported *B. afzelii* ACA-1 genome sequence does not contain Bor TOMM genes;[28] however, we detected *borBCD* in two of three ACA-1 strains of differing passage numbers. Presumably lp28-8 was present in the original ACA-1 isolate but was lost prior to the sequencing of its genome. This hypothesis is supported by a report of the sequence of ACA-1 *vls* cassette region, also missing from the published ACA-1 genome, indicating the presence of the lp28-8 in their culture.[56] [b]*BorBCD* were not detected in passage 7 of a PKo strain that at passage 6 had its genome sequenced.[28] In contrast, *borBCD* were detected in another isolate of PKo at passage 10, again highlighting the well-known heterogeneous nature of *Bbsl* upon cultivation.

A

borB **(5' mismatches = 3; 3' mismatches = 3)**

```
BafzB           ATGGTAAAAAATGACAATGAAAGTTCTGAAGTTTCTAATATGAAGCTATCAGGGATGTAT 60
BspiB           ATGATAAAAAATGATAATGAAAGTTTTGAAGTTTCTAATATGAAGCTATCAGGGATGTAT 60
BvalB           ATGGTAAAAAATGATAATGAAAGTTCTGAAGTTTTTAATATGAAGCTATCAGGGATGTAT 60
                *** ********** ********** ******** *************************

BafzB           GAAATAAATAATGTTTCAAGAAGTGTTTTTAATAGCAATATCATTGCTAATCCTTCTAAT 120
BspiB           GAAATAAATAATATTTCAAGAAGTGTGTTTAATAGCAATATTATTGCTAATCCTTCTAAT 120
BvalB           GAAATAAATAATGTTTCAAGAAGTGTGTTTAATAGCAATATTATTGCTATTCCTTCTAAT 120
                ************ ************* ************** ******* **********

BafzB           ATTAGAATTCAAGCATCTTTATTTGATAAAGATATAGAAAGCTTGACTCTTAATTATTTA 180
BspiB           ATGAGAATTCAAGCATCTTTGTTTGATAAAGATACAGAAAGTTTGACTCTTAATTATTTA 180
BvalB           ATTAGAATTAAAGCATCTTTATTTGACAAAGATGCAGAAAGTTTGACTCTTAATTATTTA 180
                ** ****** ********** ***** ******  ****** ******************

BafzB           TTAAATTTTTCTGAAATTAAAAAATTTTATTTTAATGAAAATATTTCTAAATTTTTGCAA 240
BspiB           TTAAATTTTTCTGAAATTAAAAAATTTTATTTTAATGAAAATATTTCTCAATTTTTGCAA 240
BvalB           TTAAATTTTTCTGAAATTAAAAAATTTTATTTTAATGAAAATATTTCTAAATTTTTGCAA 240
                ********************************************** ***********

BafzB           AACCCCGCTGCTATTGCAAATATATCTAATAGAGAATATCGCGAATTTGACGA-GA---- 295
BspiB           AATTCCGCTGCTATTGCAAATATATCTAACAGAGAATATCGTGAATTTAATGA-GA---- 295
BvalB           GATCCCGCTGCTATTGCAAATATATCTAACAGAGAATATCGTGAATTTGATGATGATGAT 300
                 *  ************************* *********** ****** * ** **

BafzB           -ATGTAATTTATTTGCCTAAAGTTAAAAGACTTAAAATTAAACTTGAAGATGCTTTACTT 354
BspiB           -ATGTAATTTATTTGCCTAAAGTTAAAAGACTCAAAATTAAACTTGAAGATGCTTTGCTT 354
BvalB           GATGTAGTTTTTTTACCTAAAGTTAAAAGACTTAAAATTAAACTTGAAGATGCGTTGCTT 360
                 ***** *** *** ***************** ******************** ** ***

BafzB           AAAAGGAAATCTTCTCGAAAGTTTAGAAATCAATATTTAAATTTGGTGGATTTATCGACG 414
BspiB           AAAAGAAAATCTTCTAGAAAGTTTAAAAATCAATATTTAAATTTGGTAGATTTGTCTACG 414
BvalB           AAAAGAAAGTCTTATAGAAAGTTTAAAAATCAATATTTAAATTTGGTGGATTTATCGACG 420
                ***** ** **** * ********* ********************** ***** ** ***

BafzB           CTTCTTTATTATGCTGCTGGTGATGTTCGATATGATTATGTAGAATTTGGAAATTCAAAA 474
BspiB           CTTCTTTATTATTCTGCTGGTGATGTTAGATATGATTATGTAGAATTTGGAAATTCAAAA 474
BvalB           CTTCTTTATTATGCTGCTGGCGATGTTCGATATGATTATGTAGAATTTGGAAATTCAAAA 480
                ************ ******* ****** ********************************

BafzB           TTTAGGCAATCTAGAAAATCTTATCCATCAGGAGGCGGAATGTATCCGATAGATATTTAT 534
BspiB           TTTAGGCAATCTAGAAAGACTTATCCATCAGGAGGAGGAATGTATCCTATAGATATTTAT 534
BvalB           CTTAGGCAATCTAGGAAGCCTTATCCATCAGGGGGAGGAATGTATCCTATAGATATTTAT 540
                 ************* **  ************* ** *********** ************

BafzB           TTTTACGCAAATAATGTTGAAAAGCTTGATAAAGGTTTTTATTTATATCAATCTTCAAAT 594
BspiB           TTTTACGCAAATAATGTTGAAAAGCTAGATAAAGGTTTTTATTTATATCAATCTTCAAAT 594
BvalB           TTTTATGTAAATAATGTTGAAAAGCTTGATAAAGGTTTTTATTTATATCAATCTTCAAAT 600
                ***** * ****************** ********************************

BafzB           AATTCAATTGTTAAGATCAATATTAGTTCGATTAAGATTGAAAATTTACTTACTATTTCG 654
BspiB           AATTCAATTGTTAAGATTAATATAAGTTCGATTAAGATTGAAAATTTACTTAATATTTCG 654
BvalB           AATTCAATTGTTAAGATCAATATTGGTTCGATTAAGATTGAAAATTTACTTACTATTTCG 660
                ***************** *****  ************************** *******
```

```
BafzB           TTTTTTGAAGGTTCTTCTAATTTTAATATTGCTTTCTTTTTTGTTTATAGATATGAGGTT 714
BspiB           TTTTTTGAGGGTTCTTCTAATTTTAATGTTGCTATCTTTTTTGTTTATAGATATGAGGTT 714
BvalB           TTTTTTGAAGGTTCTTCTAATTTTAATATTGCTTTCTTTTTTGTTTATAGATATGAGGTT 720
                ******** ****************** ***** ************************

BafzB           AATTATTTAAAGTATTCAGAAATGAGCTTGTCTTTTGCTTTTATAGAGCTAGGTGCTATT 774
BspiB           AATTATTTGAAGTATTCAGAAATGAGTTTGTCTTTTACTTTTATAGAGTTAGGTGCTATT 774
BvalB           AATTATTTAAAGTATTCAGAGATGAGTTTGTCTTTTGCTTTTATAGAGCTAGGCGCTATT 780
                ******** *********** ***** ********* *********** **** ******

BafzB           CATCAGAATTTTTCTTTGGTTTCTGCTGCTTTGAATTTGGGGTATTGCAGTTGGGGGGGA 834
BspiB           CATCAAAATTTTTCTTTGGTTTCTACTGATTTAAATCTGGGATATTGTGTCTGGGGAGGA 834
BvalB           CATCAAAATTTTTCTTTAGTTTCTACTGCTTTGAATTTGGGATATTGTAGTTGGGGAGGG 840
                ***** *********** ****** *** *** *** **** *****    ***** **

BafzB           TTTGATAAATCAGAGTCAGAAAAGTATTTAAAGCTAGATGGAATTAGTGAGCATGTAGTT 894
BspiB           TTTGATAAGCCAGAGCTGGAAAAGTGTTTAAAGCTAGATGGAATTAGTGAACATGTAGTT 894
BvalB           TTTGATAAACCAGAGCTAGAAAAGTGCTTAAAGTTAGATGGAATTAGTGAACATGTAGTT 900
                ********  *****   *******  ****** ***************** *********

BafzB           TCAGCTGCTATTTTAGGGAATGTTTAA 921
BspiB           TCAGCTTCTATTTTAGGAAATGTTTAA 921
BvalB           TCAGCTGCTATTTTAGGGAATGTCTAA 927
                ****** ********** ***** ***
```

B

borC **(5' mismatches = 2; 3' mismatches = 1)**

```
BafzC           ATGAATAATGGTTTATATTATTTTTCAGATAATGTGAGAGTTTTAAAAGAATCCAAAGGC 60
BspiC           ATGAATAATGGTTTATATTATTTTTCAGATAATGTGAGAGTTTTAAAAGAATCCAAAGGC 60
BvalC           ATGAATAATGGTTTGTATTATTTTTCAGATAATGTAAGAGTTTTAAAAGAATCCAAAGGG 60
                ************** ******************** ***********************

BafzC           TTTTTGATTATTAGGAAGGGAGTTTGGAATTATGAGGATTTGATAATTTCTATTGAAGAT 120
BspiC           TTTTTGATTGTTAGGAAGGGAGTTTGGAATTATGAGGATCTAACAATTTCTATTGAAGAT 120
BvalC           TTTTTGATTGTTAGGAAGGGGGTTTGGAATTATGAGGACCTGATAATTTCTATTGAAGAT 120
                ********* ********** *****************  * * ****************

BafzC           TCTTATATTAATGAATTGAGCAAAGTTTTTAATTCTTTATCTTCTAATGAGGGAATTGAT 180
BspiC           TCTTATATTAATGAATTGAGCAAAGTTTTTAATTCTTTGTCTTCTGATGAAGGAATTGAT 180
BvalC           TCTTATATTAATGAATTGAGCAAAGTTTTTAATTCTTTGTCTTCTGATGAGGGAATTGAT 180
                ************************************ ****** **** *********

BafzC           TTGATTCAAATAAAAAATGATAGGGTTAAAAATATTGTTAAAGAATTAATTAATTCTGGA 240
BspiC           TTGATTCAAATAAAAAATGATAAGGTCAAAAATATTGTTAAAGAATTAATTAATTCTGGT 240
BvalC           TTAATTCAAATAAAAAACAATAAGGTTAAAAATATTGTTAAAGAATTAATTAATTCTGGG 240
                ** **************  *** *** ********************************

BafzC           TTTGTTCATTTGAATGAGTTTAGGAAAAATAATAGATTTTTAAATTATTTATATCTTGGT 300
BspiC           TTTGTTCATTTGAATGAGTTTGGAGAAAATAATAGATTTTTAAATTATTTATATCTTGGT 300
BvalC           TTTGTTCATTTGAATGAGTTCAGAATTAATAATAGATTTTTAAATTATTTATATTTTGGG 300
                ********************  *    *************************** ****

BafzC           CAATATTTAGATAGTGTTCCCCAAGAGAATCCTTATTTGCTGATTTCAGATGATAGCAGC 360
BspiC           CAATATTTAGAAGTTGCTCCCCAAGAGAATCCTTATTTATTGATTTCAGATGATAATAGT 360
BvalC           CAATATTTGGAAGTTGTTCCCCAAGAGAATCCTTATTTGCTGATTTCAGATGATAGTAGT 360
                ******** **   ** ********************  ***************  **

BafzC           TTGATTAATGTTTTAGAAACCACTGCTAAATCGTTTGATTTAAAGTATGAGTTTTTGAGT 420
BspiC           TTGATTAATATTTTAGAAACAACCGCTAAATCGTTTAATTTAAAGTATGAGCTTTTGAGT 420
BvalC           TTGATTAATATTCTAGAAACTACTGCTAAATCGTTTGATTTAAAGTATGAGCTTTTGAGT 420
                ********* ** ******* ** ************ ************** ********
```

```
BafzC           CAGGAAAACTTGGAATTTTTAAAAACACCTGTTTTTTTAAATAAATTGGATTTAATAAAT 480
BspiC           CGGGAAAACTTGGAATTTTTAGAAACACCTATTTTTTTAAATAAATTAGATTTAATAAAT 480
BvalC           CATGAAAACTTGGAATTTTTAAAAACACCTGTTTTTTTAAATAAATTGGATTTAATGCGG 480
                *  ****************** ******** *************** ********

BafzC           TATTCAAAAAGTTTAGACCATTTTAGATCCAAATTTATGCAATTTGAAGGGGTTATTCTG 540
BspiC           TATTCAAAAAGTTTAGACCATTTTAGATCGAAATTTATGCCATTTGAAGGGGTTATTCTA 540
BvalC           TATTCAAAAAGTTTAGATCATTTTAGATCCAAATTTATGCCATTTGAAGGGGTTATTATG 540
                ***************** *********** ********** *************** *

BafzC           TTGTTGGGCAATTTGAATCCATTTTTACTTAGAAATATTAACAGAATTTTAATAGGACTT 600
BspiC           TTATTGGGTAATTTGAATCCATTTTTACTTAGGAATATTAATAGAATTTTAATAGGACTT 600
BvalC           TTGTTGGGTAATTTGAACCCATTTTTACTTAGAAATATTAATAGAATTTTAATAGGACTT 600
                ** ***** ******** ************** ******** ******************

BafzC           AATAAACCAATTTTTTTTTGGCACCATAGATGGTCCTTTCATGTTGATTACATGTTTAGAG 660
BspiC           AATAAACCAATTTTTTTTTGGTACCATAGATGGTCCTTTTATGTTGATTACATGTTTAGAG 660
BvalC           AATAAACCAATTTTTTTTTGGCACCATAGATGGGCCTTTTATGCTGATTACATGTTTAGAG 660
                ******************** *********** ***** *** *****************

BafzC           CCCAAAAGAACAGCGTGTATTGAATGTTATGAGCAAAGGATAATTGCTAGGATGACGGAT 720
BspiC           CCTAAAAGAACAGCATGTATTGAATGTTATGAGCAAAGGATAATTGCTAGGATGACAGAT 720
BvalC           CCTAAAAGAACAGCATGTATTGAATGTTATGAACAAAGGATAATTGCTAGAATGACAGAT 720
                ** *********** ***************** ****************** ***** ***

BafzC           CATGTACTTTATCATAGATTTAATAAGGAGTATGAAAAACACAAGAGGGATTCAAAGTTA 780
BspiC           CATGTGCTTTATCATAGATTTAATAAGCAGTATGAAAAACATAAGAGGGATCCAAATTTA 780
BvalC           CATGTACTTTATCATAGATTTAATAAGGAGTATGAAAAACATAAGAGGAACACAGAGTTA 780
                ***** ********************** ************* ****** *  ** * ***

BafzC           GTTTCTTTAGGGGATTTTTTAAGTCCAGCGTATTATCATATTGCTTACATTTTAGTTGGG 840
BspiC           GTTTCTTCAGGGGATTTTTTAAGTCCTGCTTATTACCATATTGCTTACATTTTAGTTGGG 840
BvalC           ATTTCTTCAGGAGATTTTTTAAGTCCAGCTTATTATCATATTGCTTACATTTTAGTTGGG 840
                 ****** *** ************** ** ***** ************************

BafzC           GATGCTTTTACTTTTTTTCAGGTTAAAACATCTAAACTTATGGGGAGAGCTTTGCATATT 900
BspiC           GATGCTTTTACCTTTTTTCAGATTAAAACATCTAAGCTCATGGGAAGAGCTTTACATATT 900
BvalC           GATGCTTTTACCTTTTTTCAGGTTAAAACATCTAAACTTATGGGGAGAGCTTTGCATATT 900
                *********** ********* ************* ** ***** ******** ******

BafzC           TATGTGCCTTTTTTTTGAGTTCCAAAATCAAGATGTTTTGAGAATATCATCTTGTCCTGCT 960
BspiC           TATGTACCTTTTTTTTGAGTTTCAAAATCAAGATGTTTTGAGGATATCATCTTGTCCTGCT 960
BvalC           TATGTGCCTTTTTTTTGAATTCCAAAATCAAGATGTTTTGAGAATATCATCTTGTCCTGCT 960
                ***** *********** ** ******************** ******************

BafzC           TGTGGATATTTGTCAACTTATAAATCAAGATGTTTGAATATTGAAACACAAAAGCTTGTT 1020
BspiC           TGTGGATATTTGTCGACTTATAAATCAAGATGTTTGAATATTGAAACACAAAAACTTGTT 1020
BvalC           TGTGGATATTTGTCGACTTATAAATCAAGATGTTTGAATATTGAAACTCAAAAGCTTGTT 1020
                ************** ********************************* ***** ******

BafzC           TCAACATTAATTTCTAATTCTAATTAA 1047
BspiC           TCAACATTAATTTCTAATTCTAATTAA 1047
BvalC           TCAACATTAATTTCTAATTCTAATTAA 1047
                **************************
```

C

borD **(5' mismatches = 7; 3' mismatches = 3)**

```
BafzD           ATGTTGAATTATTATCCTTATTCAAGTAAGCTCTATAGAGATCTTGTTTTTACTAATTCT 60
BspiD           GTGTTTAATTATTATCCTTATTCAAGCAAGCTTTATAGAGATCTTGTTTTTACTAATTCT 60
BvalD           GTGATTAACTATTATCCTTATTCTAGTAAGCTTTATAGAGATCTTGTTTTTACTAATTCT 60
                 ** * ** ************** ** ***** ***************************

BafzD           CCGGCTACAGGCATTGGTCCTTCGTTAGTTACATTGTTGCCTTTTCAAAAGGGATTGCCT 120
BspiD           CCGGCAACAGGTATTGGTCCGTCTTTAGTTACATTATTGCCTTTTCAAAAGGGATTGCCT 120
BvalD           CCGGCTACAGGTATTGGTCCAGCTTTAGTTACATTGTTACCCTTTCAAAAGGGATTGCCT 120
                ***** ***** ********  * ************ ** ** ******************

BafzD           TTATTATATTCTTCTACATGTATTTTGCCTAATTATCACAAAATATTAATAGGAGAGACA 180
BspiD           TTATTATATTCTTCTACATGTGTTTTACCTAATTACCATAAAATATTAATAGGGGAGACA 180
BvalD           TTATTATATTCTTCTACATGTATTTTGCCCAATTATCATAAAATATTAATAGGAGAGGCA 180
                ********************** **** ** ***** ** ************** *** **

BafzD           TGTAATATGGAATATCATATTTCTGGTTATGGGCGTTCTTATGAGGAGGCTATTACTAGA 240
BspiD           TTTAATATGGAATATCATATTTCTGGTTATGGACGCTCTTATGAGGAGGCTATTACTAGA 240
BvalD           TACAATATGGAATACCATATTTCTGGTTATGGGCGTTCTTATGAGGAGGCTATTACTAGA 240
                *  ************ ***************** ** ************************

BafzD           TTACAAGGCGAAACTATTGAGAGATATTCTCTTTTGATGTCCAAAACTTTATTTCAAGAA 300
BspiD           TTACAAGGCGAAACTATTGAGAGATATTCCCTTTTGATGTCCAAAACTTTATTTCAAGAA 300
BvalD           TTGCAAGGTGAAACTATTGAGAGATATTCTCTTTTAATGTCTAAAACTTTATTTCAAGAA 300
                ** ***** ******************** ***** ***** ******************

BafzD           GATTTTATTCTTTCTTCTAGAAAGGCCCTTTTAAATACCTCAAAGTATAAAGTTATGCCT 360
BspiD           GATTTTATTCTTTCTTCTAGAAAGGAGCTTCTAAATACTTCAAAGTATAAGGTTATGCCT 360
BvalD           GATTTTATTCTTTCTTCTAGAAAGTCTCTTCTAAATACTTCAAAGTATAAAGTTATGCCT 360
                ************************   *** ******* *********** *********

BafzD           TTAGAATATAGAAATGTTTTTTCAGATTTAGATAGAAACTTTAATCTTCCTTATTCAAAA 420
BspiD           TTGGAATATAGAAATGTTTTTTCAGATTTAGATAGAAATTGTAATCTTCCTTATTCAAAA 420
BvalD           TTGGAGTATAGAAACGTTTTTTCAGATTTGGATAGAAATTTTAACCTTCCTTATACAAAA 420
                ** ** ******** ************** ******** * *** ********* *****

BafzD           GTTAATGAGAGCGATGAAATTTATTGGATTTTATTGCCATCTTTAATTCATTCTAATGAA 480
BspiD           GTTTATGAGAGCGATGAAATTTATTGGATTTTATTGCCATCCTTAATTCATTCCAATGAA 480
BvalD           GTTTGTGAGAGTGATGAGATTTATTGGATTTTATTGCCATCTTTAATTCATTCTAATGAA 480
                ***  ****** ***** ********************** *********** ******

BafzD           AAAACATGGGTTCCTTGTGATATGGTTTTTATGGGTATCCAGCAAGATTTTAAAATGCCA 540
BspiD           AAAATATGGGTTCCTTGCGATATGATTTTTATGGGTATTCATCAGAATTTTAAAATGCCA 540
BvalD           AAAATATGGGTCCCTTGCGATATGATTTTTATGGGTATCCATCAGAATTTTAAAATGCCA 540
                **** ****** ***** ****** ************* ** **  **************

BafzD           GCTTTTAGTACAGGAACAGCAGTTCATAGGACAGTGGAGATTGCACTTAGTAATGCAATA 600
BspiD           ACTTTTAGTACAGGAACAGCAGTTCATAGGACAGTAGAGATTGCGCTTAGTAATGCAATA 600
BvalD           ACTTTTAGTACAGGAACAGCAGTTCATAGGTCAGTAGAGATTGCACTTAGCAATGCAATA 600
                 **************************** **** ******** ***** *********

BafzD           ATAGAGGTAATACAATTACATTGTTATATTTCTAATTGGTACGTAAAATCCAAAAGGCCT 660
BspiD           ATAGAGGTAATACAATTACATTGTTATATTTCTAATTGGTACGTAAAATCCAAAAGGCCC 660
BvalD           ATAGAGGTAATACAATTACATTGCTATATTTCTAATTGGTACATGAAATCTAAAAGGCCT 660
                ********************** ****************** * ***** ********
```

```
BafzD           GTGATTGACTGGAAAAGTAATAAATTTTTGAGAAAATCAATTAGTGATTTAGAACTTGAG 720
BspiD           GTGATTGACTGGAAAAGTAATAAATTTTTGAGAAAATCAATTAGTGATTTAGAACTTGAA 720
BvalD           GTGATTGACTGGAAAAGTAATAAATTTTTGAGAAAATCAATTGGTGATTTAGAACTTGAA 720
                **************************************** ***************

BafzD           TCCAATTTTGATTTAATAATTTTAGATTACTCTGATAGAGATTTGGGAATGCCAGTTTAT 780
BspiD           TCCAATTTTGATTTAATAGTTTTAGATTATTCTGATAAAGATTTAGGAATGCCCGTTTAT 780
BvalD           TCTTATTTTGATTTAGTTATTTTGGATTATTCTGATAGGGATTTGGGAATGCCAGTTTAT 780
                **  *********** *  **** ***** *******  ***** ******** ******

BafzD           GCAGCATTTTTAAGAAATAAGAAAAAATCGATTCCGTATTTAGTTTGTGGAATTCAAGGG 840
BspiD           GCAGCATTCCTAAGAAATAAGAAAAAATCGATTCCATATTTAGTTTGTGGAATTCAAGGT 840
BvalD           GCAGCATTCTTAAGAAATAAGAAAAAATCGATTCCATATTTAGTTTGTGGAATTCAAGGA 840
                ********  ************************ ***********************

BafzD           GGCTTTAATAAAGAGTATGTTTTGTTAAGAGCCGTAGAAGAAGCCGCTGTGATTGCTCAA 900
BspiD           GGCTTTAATAAAGAGTATGTTTTTTTAAGAGCTGTAGAAGAAGCTGCTGTAGTTGCTCAA 900
BvalD           GGTTTTAATAAAGAGTATGTTTTGTTAAGAGCTGTAGAAGAAGCTGCTGTAGTTGCTCAG 900
                ** ********************* ******** *********** *****  *******

BafzD           TCTTTACCTTTAATTTATTTTTTAAAGGCTAAAGAGATAAGCAAGTTAACATTAAATAGT 960
BspiD           TCTTTGCCCCTAATTTATTTTTTTAAGGCTAAAGAAATAAGTAAGTTAACATTAAATAGT 960
BvalD           TCTTTACCTTTAATTTATTTTTTTAAGGCTAGAGAGATAAGTAATTTAACATTAAATAGT 960
                ***** **  ************* ******* *** ***** ** ***************

BafzD           CTAAGACATAGTTTCAATTTGGATGATAATTTTTTATATTACTCTAACTTAAATGAAATT 1020
BspiD           CTAAGACACAGTTTTAATTTGGATGATAATTTTTTCTATTACTCCAATTTAAATGAAATT 1020
BvalD           CTGAGGCACAGTTTTAATTTGGATGATAATTTTTTATATTATTCCAACCTGAATGAAATT 1020
                ** ** ** ***** ******************** ***** ** **  * *********

BafzD           TCTTTAAAAGATTCTTTATTAAATTCTATTATAGATGATGAAACTAGATTAGAGATATTT 1080
BspiD           TCTTTAAAAGATTCTATGTTAAATTCTATTATAGATAATGAAACTAGATTAGAGATATCT 1080
BvalD           TCTTTAAAAGATTCTTTGTTGAATTCTATTATGGATAATGAAACTAGATTAGAGATATCT 1080
                *************** * ** *********** *** ********************** *

BafzD           TCTGAAGAGAACTTAGTTGAAAATGAATTGGATATGAGTTTAAATATTTTGAAGTCTGTA 1140
BspiD           TCTGAAGAAAGCTTAATTAAAAATGAATTGGATATGAGCTTAAATATTTTGCGATCTGTA 1140
BvalD           TCTGAAGAAAACTTAGTTGAAAATGAATTGAATATGAGTTTAAGTATTTTACGATCTGTA 1140
                ******** * **** ** *********** ******* **** ******    ******

BafzD           AGTAAATATGCTGTTTACTGTGATATAACTCCCCCCGAATTAGAAAGTACATCTTTTAGG 1200
BspiD           AGTAAATATGCTGTTTATTGCGATATAACTCCTCCCGAATTAGAAGGTAGATCTTTTAGA 1200
BvalD           AGTAAATATGCTGTTTATTGTGATATAACTCCTCCTGAATTAGAAAGCACATCTTTTAAA 1200
                ***************** ** *********** ** ********* * * ********

BafzD           GCTATAAGGATTCTTGTTCCTGAGTTGCTTAAAATGTGTTTTCCATTTCATCCCTTTGAT 1260
BspiD           TCTATAAGGATTCTTGTTCCTGAATTGCTTAAGATGTGTTTTCCATTTCATCCCTTTGAT 1260
BvalD           TCTATAAGGATTCTTGTTCCAGAGTTGCTTAAAATGTGTTTTCCATTTCACCCCTTTGAT 1260
                 ******************* ** ******** ***************** *********

BafzD           AAGCATCCATATTTTAAAAAGAAAGGAGAAATTTTAGATGGATATTTCCCACACCCAATA 1320
BspiD           GAGCATCCATATTTTAAAAAGAAAGGAGAAATTTTAGATGAATATTTCCCACATCCAATA 1320
BvalD           GAGCATCCATATTTTAAAAAGAAAGGAGAAATTTTAGATGAATATTACCCACACCCAATA 1320
                 ************************************ ***** ****** ******

BafzD           CCTTAG 1326
BspiD           CCTTAG 1326
BvalD           CCTTAG 1326
                ******
```

Figure B.8. Nucleotide alignments of Bor TOMM biosynthetic genes. *BorB* (A), *borC* (B) and *borD* (C) from *Borrelia afzelii* PKo (Bafz), *B. spielmanii* A14S (Bspi), and *B. valaisiana* VS116 (Bval) were aligned using ClustalW. This phylogenetic data is deposited in the Dryad Digital Repository,

doi:10.5061/dryad.d4863. Sites where PCR screening primers annealed are underlined. The primers used to detect *borB/C/D* (Table B.5 and Table B.3) were based on *bvalB/C/D* from *B. valaisiana* VS116 (Table B.4). The number of mismatches per primer when compared to the two other sequenced examples of the gene is indicated in parentheses for each.

Primer Name	Sequence (5' → 3')	Description
PCR Screening		
BvalA-fwd	ATGTTTATTAGAAATAAACATTTTTTGACTAGC	Forward primer to amplify *borA* based on *bvalA*
BvalA-rev	TTATAGACTAAAGCTGCAACTACAAGTGC	Reverse primer to amplify *borA* based on *bvalA*
BvalB-fwd	ATGGTAAAAAATGATAATGAAAGTTCTGAAGT	Forward primer to amplify *borB* based on *bvalB*
BvalB-rev	TTAGACATTCCCTAAAATAGCAGCTG	Reverse primer to amplify *borB* based on *bvalB*
BvalC-fwd	ATGAATAATGGTTTGTATTATTTTTCAGATAATGTAAG	Forward primer to amplify *borC* based on *bvalC*
BvalC-rev	TTAATTAGAATTAGAAATTAATGTTGAAACAAGCTTT	Reverse primer to amplify *borC* based on *bvalC*
BvalD-fwd	GTGATTAACTATTATCCTTATTCTAGTAAGCTTTATAGA	Forward primer to amplify *borD* based on *bvalD*
BvalD-rev	CTAAGGTATTGGGTGTGGGTAATATTC	Reverse primer to amplify *borD* based on *bvalD*
pDCerm Constructs		
BvalA-5'XbaI	AAATCTAGAATGTTTATTAGAAATAAACATTTTTTGAC	Forward primer to subclone *bvalA* from pIDTSMART construct
BvalA-3'BamHI	AAAGGATCCTTATAGACTAAAGCTGCAACTACAAGTG	Reverse primer to subclone *bvalA* from pIDTSMART construct
SagA-BvalA-5'XbaI	AAATCTAGAATGTTAAAATTTACTTCAAATATTTTAGCTAC	Forward primer to subclone *sagA-bvalA* from pIDTSMART construct
SagA-BvalA-3'BamHI	AAAGGATCCTTATAGACTAAAGCTGCAACTACAAGTGCAAC	Reverse primer to subclone *sagA-bvalA* from pIDTSMART construct
sagA$^{1\text{-}50}$For	GAAGTTATACGtaAGGTAAATAAGGATCCTGATCTTCAGAA	Forward primer for incorporation of stop codon at position corresponding to P51 in *sagA*
sagA$^{1\text{-}50}$Rev	CCTTATTTACCTtaCGTATAACTTCCGCTACCACCTTGAGA	Reverse primer for incorporation of stop codon at position corresponding to P51 in *sagA*
sagA$^{1\text{-}44}$For	ATTCTCAAGGTtaaAGCGGAAGTTATACGCCAGGTAAATAA	Forward primer for incorporation of stop codon at position corresponding to G45 in *sagA*
sagA$^{1\text{-}44}$Rev	TAACTTCCGCTttaACCTTGAGAATTACCACTTCCAGTAGC	Reverse primer for incorporation of stop codon at position corresponding to G45 in *sagA*
sagA$^{1\text{-}42}$For	GTGGTAATTCTtAAGGTGGTAGCGGAAGTTATACGCCAGGT	Forward primer for incorporation of stop codon at position corresponding to Q43 in *sagA*
sagA$^{1\text{-}42}$Rev	CCGCTACCACCTTaAGAATTACCACTTCCAGTAGCAATTGA	Reverse primer for incorporation of stop codon at position corresponding to Q43 in *sagA*
sagA$^{1\text{-}40}$For	CTGGAAGTGGTtAaTCTCAAGGTGGTAGCGGAAGTTATACG	Forward primer for incorporation of stop codon at position corresponding to N41 in *sagA*
sagA$^{1\text{-}40}$Rev	CCACCTTGAGAtTaACCACTTCCAGTAGCAATTGAGAAGCA	Reverse primer for incorporation of stop codon at position corresponding to N41 in *sagA*
sagA$^{1\text{-}38}$For	TTGCTACTGGAtaaGGTAATTCTCAAGGTGGTAGCGGAAGT	Forward primer for incorporation of stop codon at position corresponding to S39 in *sagA*
sagA$^{1\text{-}38}$Rev	TGAGAATTACCttaTCCAGTAGCAATTGAGAAGCAACAAGT	Reverse primer for incorporation of stop codon at position corresponding to S39 in *sagA*

sagA$^{1\text{-}37}$For	CAATTGCTACTtaAAGTGGTAATTC TCAAGGTGGTAGCGGA	Forward primer for incorporation of stop codon at position corresponding to G38 in *sagA*
sagA$^{1\text{-}37}$Rev	GAATTACCACTTtaAGTAGCAATTG AGAAGCAACAAGTAGT	Reverse primer for incorporation of stop codon at position corresponding to G38 in *sagA*
sagA$^{1\text{-}36}$For	TCTCAATTGCTtaaGGAAGTGGTAA TTCTCAAGGTGGTAGC	Forward primer for incorporation of stop codon at position corresponding to T37 in *sagA*
sagA$^{1\text{-}36}$Rev	TTACCACTTCCttaAGCAATTGAGA AGCAACAAGTAGTACA	Reverse primer for incorporation of stop codon at position corresponding to T37 in *sagA*
sagA$^{1\text{-}35}$For	GCTTCTCAATTtaaACTGGAAGTGG TAATTCTCAAGGTGGT	Forward primer for incorporation of stop codon at position corresponding to A36 in *sagA*
sagA$^{1\text{-}35}$Rev	CCACTTCCAGTttaAATTGAGAAGC AACAAGTAGTACAGCA	Reverse primer for incorporation of stop codon at position corresponding to A36 in *sagA*
sagA$^{1\text{-}34}$For	GTTGCTTCTCAtaaGCTACTGGAAG TGGTAATTCTCAAGGT	Forward primer for incorporation of stop codon at position corresponding to I35 in *sagA*
sagA$^{1\text{-}34}$Rev	CTTCCAGTAGCttaTGAGAAGCAAC AAGTAGTACAGCAGCA	Reverse primer for incorporation of stop codon at position corresponding to I35 in *sagA*
sagA$^{1\text{-}33}$For	CTTGTTGCTTCtaaATTGCTACTGGA AGTGGTAATTCTCAA	Forward primer for incorporation of stop codon at position corresponding to S34 in *sagA*
sagA$^{1\text{-}33}$Rev	CCAGTAGCAATttaGAAGCAACAAG TAGTACAGCAGCAACA	Reverse primer for incorporation of stop codon at position corresponding to S34 in *sagA*
sagA-bvalA+A_For	GCTTTAGTCTAgctTAAGGATCCTG ATCTTCAGAACTTAAG	Forward primer for appending alanine codon after position corresponding to L36 in *sagA-bvalA*
sagA-bvalA+A_Rev	CAGGATCCTTAagcTAGACTAAAGC TGCAACTACAAGTGCA	Reverse primer for appending alanine codon after position corresponding to L36 in *sagA-bvalA*
sagA-bvalA-S27CFor	GCTGCTGTTGTtgcTGCACTTGTAG TTGCAGCTTTAGTCTA	Forward primer for substitution with cysteine codon at position corresponding to S27 in *sagA-bvalA*
sagA-bvalA-S27CRev	CTACAAGTGCAgcaACAACAGCAG CCTCCAGGAGCAACTTG	Reverse primer for substitution with cysteine codon at position corresponding to S27 in *sagA-bvalA*
pDC_SeqMCS_F	GGGAGGAAATAAAGCGCTAGG	Forward primer for verification of site-directed mutants of *sagA* by DNA sequencing
S. pyogenes M1 _ΔsagA_ Transformant Confirmation		
pDC_SeqMCS_F	GGGAGGAAATAAAGCGCTAGG	Forward primer to detect pDCerm containing gene of interest
pDCerm_PCR_R	ACAAGTAACCAGTGACTGCCG	Reverse primer to detect pDCerm containing gene of interest
SagB_F	ATGTCATTTTTTACAAAGGAACAA	Forward primer to detect *sagB* gene
SagB_R	ATTGACGATGACTTCTTCG	Reverse primer to detect *sagB* gene
spy1258F	AAAGACCGCCTTAACCACCT	Forward primer to detect gene unique to *S. pyogenes* (Liu et al., 2005)
spy1258R	TGGCAAGGTAAACTTCTAAAGCA	Reverse primer to detect gene unique to *S. pyogenes* (Liu et al., 2005)
pET28b-MBP Constructs		
BvalA-5'BamHI	AAAAGGATCCATGTTTATTAGAAA TAAAC	Forward primer to subclone *bvalA* from pIDTSMART construct
BvalA-3'NotI	AAAAGCGGCCGCTTATAGACTAA AGCTGC	Reverse primer to subclone *bvalA* from pIDTSMART construct
SagA-BvalA-5'BamHI	AAAAGGATCCATGTTAAAATTTAC TTCAAA	Forward primer to subclone *sagA-bvalA* from pIDTSMART construct

SagA-BvalA-3'NotI	AAAAGCGGCCGCTTATAGACTAAAGCTGCAAC	Reverse primer to subclone *sagA-bvalA* from pIDTSMART construct
sagA$^{1\text{-}35}$For	GCTTCTCAATTtaaACTGGAAGTGGTAATTCTCAAGGTGGT	Forward primer for incorporation of stop codon at position corresponding to A36 in *sagA*
sagA$^{1\text{-}35}$Rev	CCACTTCCAGTttaAATTGAGAAGCAACAAGTAGTACAGCA	Reverse primer for incorporation of stop codon at position corresponding to A36 in *sagA*

Table B.4. Primers used in this study. In all cases, *bvalA/B/C/D* or *BvalA/B/C/D* refers to the relevant TOMM genes or proteins from *B. valaisiana* VS116. Restriction endonuclease recognition sites are underlined. Lower-case letters indicate bases targeted for mutagenesis.

Our results reveal that Bor TOMM biosynthetic genes are present in select *B. afzelii, B. spielmanii, B. valaisiana*, *B. lusitaniae*, and *B. garinii* isolates (Table B.3, summarized in Table B.5). In line with the well-known plasmid instability of *Borrelia* plasmids, we observed cases of passage-dependent amplification of Bor TOMM genes (Table B.3). While all three biosynthetic genes could be amplified from some of the aformentioned strains, there were many instances where only one or two biosynthetic genes were detected (Table B.3 & Table B.5). Overall, *borC* was successfully amplified more frequently than *borB* or *borD*, likely due to higher conservation at the extreme 5' and 3' ends where the primers were designed to anneal (Table B.4). Since mutational inactivation of nonessential genes is quite common on *Bbsl* linear plasmids, another possibility is that some target genes are too divergent to detect with the primers employed.[33]

		PCR Positive			
***Bbsl* species**	**No. of strains screened**	***borB***	***borC***	***borD***	**All**
Bbss	24	0	0	0	0
B. garinii	26	0	5	0	0
B. afzelii	47	26	34	19	14
B. spielmanii	7	5	5	5	5
B. bavariensis	13	0	0	0	0
B. valaisiana	21	3	15	3	3
B. lusitaniae	2	0	2	2	0

Table B.5. The Bor TOMM biosynthetic cluster is present among diverse *Bbsl* isolates. Positive PCR hits for Bor TOMM biosynthetic genes in *Bbsl* isolates. Strains originated in Europe and Asia, except 11 *Borrelia burgdorferi* sensu stricto (*Bbss*) from North America. "All" indicates the number of isolates testing positive for genes encoding all three putative heterocycle forming enzymes, BorBCD. More details are provided in Table B.3.

In contrast, Bor TOMM genes were absent from all *B. burgdorferi* sensu stricto (*Bbss*) and *B. bavariensis* isolates screened (Table B.3 & Table B.5), suggesting that they do not possess a Bor TOMM cluster, or again, that the target genes are too divergent to detect. The fact that the Bor TOMM is not present in the 13 *Bbss* published genomes[34] suggests that the former is likely true for this species. Notably, *Bbss* has been reported to produce a number of hemolytic factors.[35, 36]

Mapping the distribution of a subset of the Bor TOMM cluster data onto a maximum likelihood tree, which was generated from a multilocus sequence typing (MLST) analysis, further supports our hypothesis that the Bor TOMM cluster has long been present in *Borrelia* (Figure B.9). The combined effects of a non-exhaustive PCR screen with non-degenerate primers and the unavoidable plasmid loss upon cultivation means that data reported here represent the minimum distribution and prevalence of the Bor TOMM cluster in *Bbsl*. However, these data demonstrate that the newly-identified Bor TOMM cluster is underrepresented in available genome sequences.

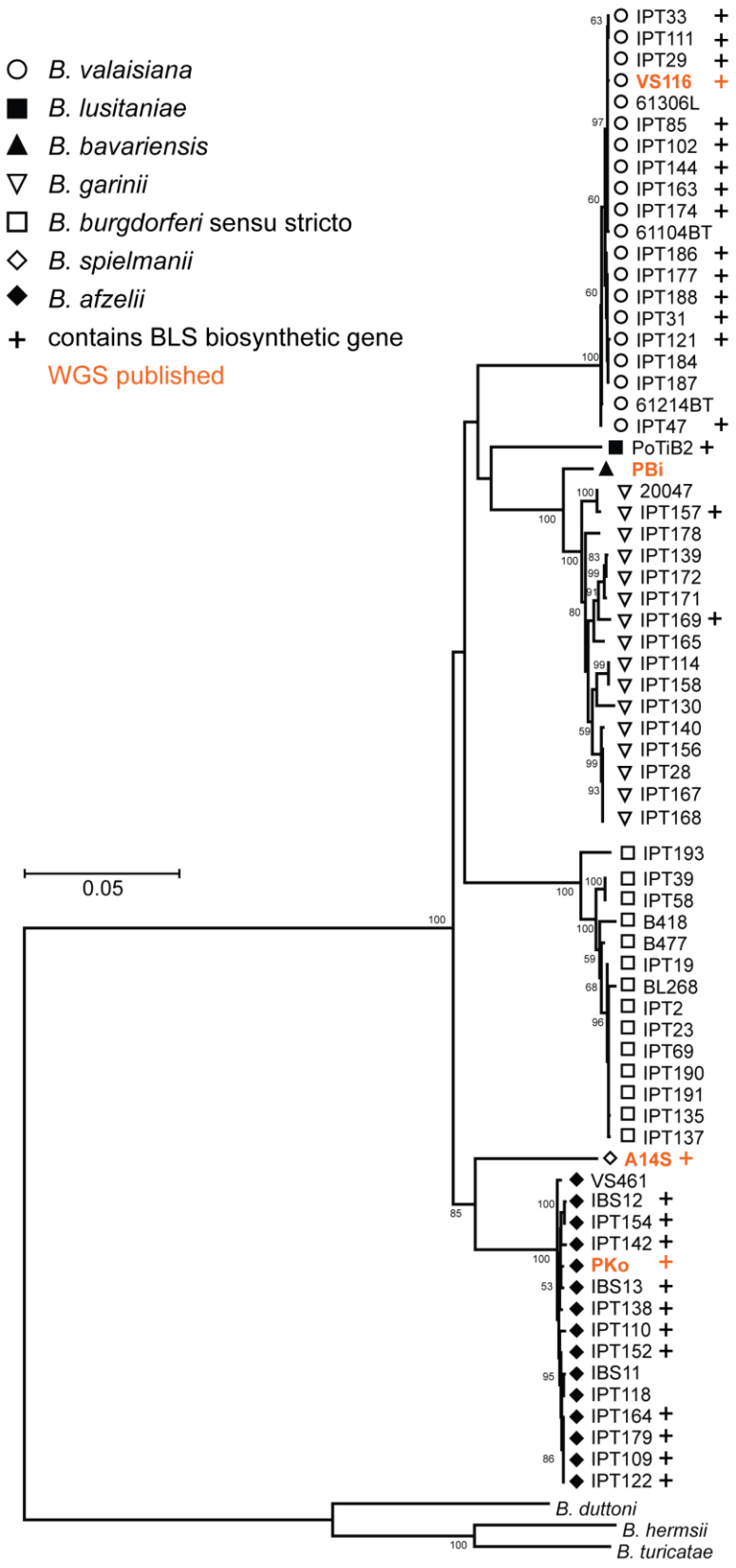

B. valaisiana
B. lusitaniae
B. bavariensis
B. garinii
B. burgdorferi sensu stricto
B. spielmanii
B. afzelii
+ contains BLS biosynthetic gene
WGS published
0.05
IPT33 +
IPT111 +
IPT29 +
VS116 +
61306L
IPT85 +
IPT102 +
IPT144 +
IPT163 +
IPT174 +
61104BT
IPT186 +
IPT177 +
IPT188 +
IPT31 +
IPT121 +
IPT184
IPT187
61214BT
IPT47 +
PoTiB2 +
PBi
20047
IPT157 +
IPT178
IPT139
IPT172
IPT171
IPT169 +
IPT165
IPT114
IPT158
IPT130
IPT140
IPT156
IPT28
IPT167
IPT168
IPT193
IPT39
IPT58
B418
B477
IPT19
BL268
IPT2
IPT23
IPT69
IPT190
IPT191
IPT135
IPT137
A14S +
VS461
IBS12 +
IPT154 +
IPT142 +
PKo +
IBS13 +
IPT138 +
IPT110 +
IPT152 +
IBS11
IPT118
IPT164 +
IPT179 +
IPT109 +
IPT122 +
B. duttoni
B. hermsii
B. turicatae

Figure B.9. Maximum likelihood tree of *Bbsl* strains used in the Bor TOMM PCR screen. The tree shows the relationships of a number of *Borrelia* strains, including a subset of the *Bbsl* strains included in our PCR screen. This phylogenetic data is deposited in the Dryad Digital Repository, doi:10.5061/dryad.d4863. A plus sign (+) next to the strain designation indicates that gene(s) from the Bor TOMM biosynthetic cluster (*borB/borC/borD*) was/were detected (Table B.3). WGS: whole genome sequence.

The SLS-like peptides from *Borrelia* are potential cytolysins

Given that cytolytic TOMM production has been implicated in the virulence of some pathogens,[8, 19, 23] coupled with the prevelance of SLS-like TOMM clusters in *Bbsl*, we decided to study the cytolytic potential of one of the naturally C-terminally truncated SLS-like peptides from *Bbsl*. In light of the complexity of *Borrelia* cultivation requirements, the aformentioned tendency to lose plasmids when cultured and inefficient genetic tools,[37, 38] we opted to take a reductionist approach and assess the cytolytic potential of BvalA by heterologous expression in *S. pyogenes ΔsagA.* Importantly, there is precedent for this approach, e.g. the SLS-like precursor from *C. sporogenes* (Figure B.1B) successfully complemented *S. pyogenes ΔsagA*[8] and was later confirmed to function as a cytolysin in the context of its native producer [18]. BvalA bears ~70% amino acid similarity to SagA (Figure B.10A) and the SagBCD proteins are ~50% similar to their BvalBCD equivalents. However, no hemolytic activity was observed upon heterologous expression of the non-cognate substrate BvalA in *S. pyogenes ΔsagA* (Figure B.10B).

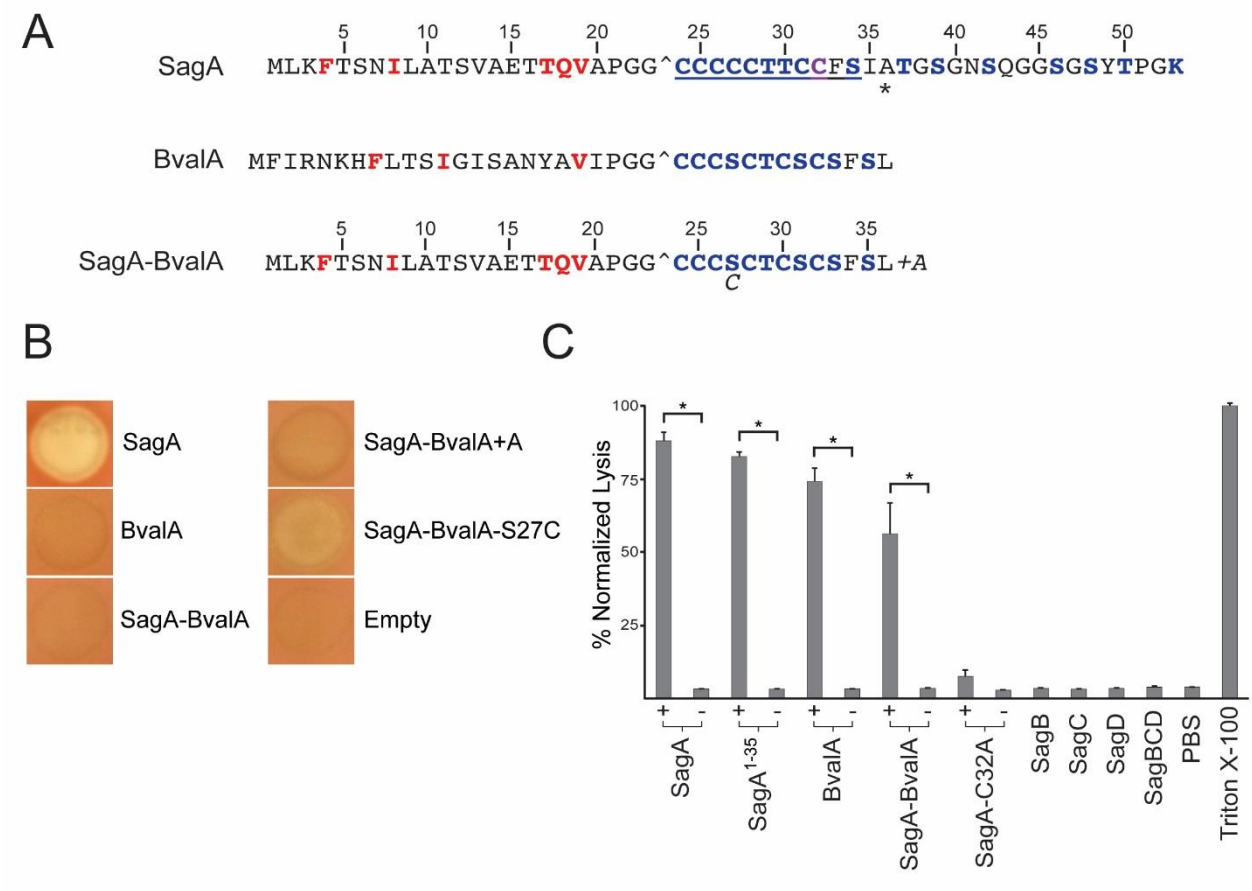


Figure B.10. Hemolytic activity of SLS-like peptides expressed in *S. pyogenes ΔsagA* and modified by purified SagBCD enzymes. (A) Residues of the SagA and SagA-BvalA precursors are numbered. The motifs known to be important for binding SagC[8] are shown in red for the SagA leader. Partial binding motifs also shown in red for the BvalA leader. Potentially modified residues are blue except C32, the position of a previously generated alanine substitution that abolished cytolytic activity of SagA, which is marked in purple. The predicted leader cleavage sites are indicated with a caret. The position of a stop codon introduced in SagA to yield the $SagA^{1-35}$ (BvalA mimic) is marked with an asterisk. SagA-BvalA-S27C and SagA-BvalA+A point mutants are represented by italicized letters adjacent to the SagA-BvalA precursor sequence. The precursor peptide labels continue in panel B & C. (B) A representative colony on blood agar of *S. pyogenes ΔsagA* expressing each precursor peptide from pDCerm is shown. The colony labeled "Empty" was a negative control containing pDCerm vector. (C) Where appropriate, bars are labeled with + or – indicating presence or absence, respectively, of MBP-SagBCD in the reaction using the indicated MBP-tagged precursor peptide. Triton X-100, positive control for 100% lysis; PBS, negative control for background hemolysis. Error is reported as the standard deviation of the mean (n = 3). * indicates $P < 0.01$ for each synthetase reaction versus the relevant unreacted precursor. $P < 0.01$ for all lytic samples versus all negative controls, i.e. SagA-C32A reaction, enzymes alone, PBS.

Substrate binding during TOMM biosynthesis occurs via recognition of key motifs within N-terminal leader region of the precursor peptide by the C-protein component of the synthetase.[8, 26] Although partial binding motifs appear to be present in the BorA leader peptides (Figure B.10A), we

cannot rule out the possibility that BvalA was not bound by the *S. pyogenes* biosynthetic machinery during heterologous expression. In light of this, we generated a chimeric construct encoding the cognate SagA leader peptide fused to the non-cognate BvalA core peptide (SagA-BvalA; Figure B.10A). Such a strategy has been exploited to replace highly divergent leader peptides with the SagA leader to allow native interaction with the heterocycle-forming synthetase.[8] In this way, hemolytic activity was demonstrated for the SLS-like core from *S. aureus* (Figure B.1B) upon expression in *S. pyogenes ΔsagA*.[8] However, SagA-BvalA was not hemolytic under identical conditions (Figure B.10B). Co-expression of rare streptococcal tRNAs[39] also failed to yield hemolysis from BvalA or SagA-BvalA, suggesting that their inactivity was not due to codon-usage bias during heterologous expression.

Returning to the native SagA precursor peptide, we noted a dramatic decrease in activity for $SagA^{1-35}$ (which could be considered a BvalA mimic; Figure B.10A), compared to $SagA^{1-36}$ (Figure B.3B). In case the hemolytic activity of SagA-BvalA was just below the limit of detection of this assay, we appended an alanine to the C-terminus of SagA-BvalA (SagA-BvalA+A is akin to $SagA^{1-36}$); however, hemolysis was again not observed (Figure B.10B). While the SagA core begins with a stretch of five contiguous cysteine residues, there is a serine residue at the fourth position of the comparable region of BvalA (Figure B.10A). We thus created SagA-BvalA-S27C, which partly rescued activity (Figure B.10B), causing hemolysis only directly beneath the colony compared to the zone of clearing observed in strains with high hemolytic activity. Synthetase reactions with reconstituted TOMM biosynthetic enzymes have established that cysteine is cyclized more rapidly than serine and threonine and several TOMM synthetases are known to be selective for cysteine.[40, 41] We speculate that replacing serine with cysteine partly ameliorates a bottleneck for SagBCD in processing an unnatural substrate.

The lack of activity from BvalA-derived substrates upon expression in *S. pyogenes ΔsagA* prompted us to assess cytolytic potential using purified SagBCD enzymes. Previously, SLS-like cytolysins with no detectable activity when expressed in *S. pyogenes ΔsagA* were converted to cytolysins by this approach.[8] Numerous factors contribute to this, but perhaps most importantly, high concentrations of reactants can be employed to increase the heterocyclization yield on non-optimal substrates.

Additionally, this cell-free system obviates the need for export/proteolysis machinery to generate lytic entities.[7, 8] Attempts to obtain sufficiently pure, full-length and active recombinant BvalBCD were unsuccessful; however, after treatment with the SagBCD homologs (Table B.1C), BvalA and SagA-BvalA were both converted into cytolysins with activity comparable to SagA and SagA$^{1\text{-}36}$ (Figure B.10C). Omission of either the peptide substrate or SagBCD resulted in no lysis, similar to the non-lytic C32A mutant of SagA.[8] These data indicated that the BvalA precursor peptide has the potential to be converted into a cytolysin under certain conditions.

The majority of non-cognate and mutant precursor peptides that become cytolytic upon treatment with purified SagBCD were also lytic when constitutively expressed in *S. pyogenes ΔsagA*.[8] However, there is precedent for the failure of *bona fide* SLS-like precursors to restore hemolysis in *S. pyogenes ΔsagA*, e.g. as was the case with listeriolysin S (Figure B.1B).[8, 19] The contrasting results for BvalA and SagA-BvalA (Figure B.10B & C) are reminiscent of the chimeric *L. monocytogenes* substrate, SagA-LlsA, which was inactive in *S. pyogenes ΔsagA* but was transformed into a cytolysin by purified SagBCD.[8] These results underscore the strict requirements for efficient enzymatic processing and toxin export in *S. pyogenes*. Nevertheless, the ability of purified SagBCD to convert BvalA into a cytolysin, coupled with the retention of activity by truncated SagA peptides upon functional expression, suggests that the short SagA homologs found in *Bbsl* may function as cytolysins in the context of their native producers. However, whether this activity is the primary function of the Bor TOMM (if produced) requires additional study.

Conclusions

The results of our SagA truncation series indicate that, of the 30 predicted to comprise the core peptide, the first 11 residues (SagA$^{24\text{-}34}$) are sufficient for detectable hemolytic activity of SLS when expressed in *S. pyogenes ΔsagA*. Thus, our data indicates that the C-terminus of SagA is largely dispensable and points to the NPH region as central to SLS functionality, consistent with the conservation of this motif in natural SLS variants as well as the greater TOMM cytolysin family. Similar regions consisting of contiguous heterocyclizable residues exist in the newly-described SLS-like precursors

encoded in multiple genera within *Firmicutes*, *Spirochaetes* and *Actinobacteria*. Although the ability of BvalA and SagA$^{1-35}$ to be matured into cytolysins by purified SagBCD suggests that the "minimal cytolytic unit" may be a general feature of TOMM cytolysins, additional study will be required to determine if this holds true for all family members, especially those with longer stretches of heterocyclizable residues (e.g. listeriolysin S). In order to comprehend our observation that much of the hemolytic activity of SLS can be attributed to a mere 11 residues, a complete chemical structure of SLS and greater molecular insight into the interaction of SLS with the cellular membrane will be required.

Our identification of novel SLS-like biosynthetic clusters greatly expands the TOMM cytolysin family. Of particular interest is our discovery that many pathogenic *Bbsl* harbor the genetic capacity to produce naturally truncated SLS-like toxins. Although confirmation is needed, the most probable function for the SLS-like biosynthetic clusters in *Bbsl* is to generate a cytolytic factor, akin to the virulence-enhancing SLS of *S. pyogenes*. Thus, the hypothetical "borreliolysin S" toxin we identify herein warrants further investigation as it would represent, to the best of our knowledge, the first posttranslationally modified natural product from any Spirochaete.

In summary, this work provides evidence that the core NPH motif represents the defining feature of the SLS-like TOMM family, as well as providing a foundation for studies of the potential utilization of cytolytic TOMMs by the diverse organisms described herein.

Experimental

Bacterial culture conditions

Strains and plasmids used in this study are listed in Table B.6, except *Bbsl* strains which are listed in Table B.3. *S. pyogenes* were grown at 37 ^{o}C in Todd–Hewitt broth or agar plates, or on plates of 5% sheep blood in tryptic soy agar (Hardy Diagnostics) when determining hemolytic activity by functional expression. For antibiotic selection, 5 μg ml^{-1} erythromycin was used. *E. coli* strains were grown at 37 ^{o}C with aeration in Luria–Bertani (LB) broth or on LB agar plates; antibiotic selection utilized 500 μg ml^{-1} erythromycin and 50 μg ml^{-1} kanamycin where relevant.

Strain/Plasmid	Relevant characteristics/Uses	References
Strains		
Escherichia coli DH5α	Plasmid construction/maintenance	Lab stock
Escherichia coli BL21(DE3)-RIPL	Protein overexpression for purification	Lab stock
Streptococcus pyogenes (GAS) M1	Group A Streptococcus	Lab stock
Streptococcus pyogenes (GAS) M1 Δ*sagA*	Group A Streptococcus; *sagA* allelic exchange mutant of *S. pyogenes* M1	[21]
Plasmids		
pDCerm-empty	Empty GAS expression vector; Erm^R	[21]
pDCerm-*sagA*	Constitutive expression of SagA wild-type precursor peptide; compliments GAS M1 *ΔsagA*	[21]
pDCerm-*sagA*$^{1-50}$	Constitutive expression of $SagA^{1-50}$ truncation	This study
pDCerm-*sagA*$^{1-44}$	Constitutive expression of $SagA^{1-44}$ truncation	This study
pDCerm-*sagA*$^{1-42}$	Constitutive expression of $SagA^{1-42}$ truncation	This study
pDCerm-*sagA*$^{1-40}$	Constitutive expression of $SagA^{1-40}$ truncation	This study
pDCerm-*sagA*$^{1-38}$	Constitutive expression of $SagA^{1-38}$ truncation	This study
pDCerm-*sagA*$^{1-37}$	Constitutive expression of $SagA^{1-37}$ truncation	This study
pDCerm-*sagA*$^{1-36}$	Constitutive expression of $SagA^{1-36}$ truncation	This study
pDCerm-*sagA*$^{1-35}$	Constitutive expression of $SagA^{1-35}$ truncation as BvalA mimic	This study
pDCerm-*sagA*$^{1-34}$	Constitutive expression of $SagA^{1-34}$ truncation	This study
pDCerm-*sagA*$^{1-33}$	Constitutive expression of $SagA^{1-33}$ truncation	This study
pDCerm-*bvalA*	Constitutive expression of BvalA wild-type precursor peptide	This study
pDCerm-*sagA*-*bvalA*	Constitutive expression of $SagA^{1-23}$ fused to $BvalA^{25-37}$ to yield chimeric precursor peptide consisting of SagA leader with BvalA core	This study
pDCerm-*sagA*-*bvalA*+A	Constitutive expression of SagA-BvalA chimera with alanine residue appended after terminal leucine residue	This study
pDCerm-*sagA*-*bvalA*-S27C	Constitutive expression of SagA-BvalA chimera with S27C substitution	This study
pET28b-MBP	Empty *E. coli* vector for high heterologous expression of proteins; Kan^R	[6]
pET28b-MBP-*sagA*	Overexpression of SagA plus N-terminal MBP tag	[6]
pET28b-MBP-*sagA*$^{1-35}$	Overexpression of $SagA^{1-35}$ truncation as BvalA mimic plus N-terminal MBP tag	This study
pET28b-MBP-*bvalA*	Overexpression of BvalA plus N-terminal MBP tag	This study
pET28b-MBP-*sagA*-*bvalA*	Overexpression of $SagA^{1-23}$ fused to $BvalA^{25-37}$ to yield chimeric precursor peptide consisting of SagA leader with BvalA core plus N-terminal MBP tag	This study
pET28b-MBP-*sagA*-C32A	Overexpression of SagA precursor peptide with C32A substitution plus N-terminal MBP tag	[7]
pET28b-MBP-*sagB*	Overexpression of SagB dehydrogenase plus N-terminal MBP tag	[6]
pET28b-MBP-*sagC*	Overexpression of SagC component of cyclodehydratase plus N-terminal MBP tag	[6]
pET28b-MBP-*sagD*	Overexpression of SagD component of cyclodehydratase plus N-terminal MBP tag	[6]

Table B.6. Strains and plasmids used in this study. In all cases, *bvalA*/BvalA refers to the TOMM gene/peptide from *B. valaisiana* VS116. The sequences of the relevant precursor peptides are shown in Figure B.3A & Figure B.10A.

Genome mining for novel SLS-like clusters

The majority of TOMM clusters encode a precursor peptide adjacent to SagB, -C and -D homologs,[9, 11] enabling their bioinformatic detection. Thus, novel SLS-like clusters were identified by BLAST-P of known SLS-like D proteins (SagD and homologs from *C. botulinum*, *S. aureus*, *L. monocytogenes*) as the query sequence. After identification of BvalD, BmurD, LactD, and EnduD, these sequences were used as a query sequence to find other homologs, with a BLAST e-value of 1e-30 as a cutoff. Subsequently, homologs of other *sag* genes were identified in the local genomic region of a single representative cluster from each genus. Manual searches were performed in the intergenic regions of the representative clusters for short, unannotated genes encoding recognizable TOMM precursors, which are usually overlooked by gene-identification algorithms. Because of inherent duplications in RefSeq, only a single representative sequence was used from organisms with no strain information. Genome clusters with low e-values were annotated manually for precursor peptides and genomic regions that resembled known cytolysins.

Maximum likelihood phylogenetic analysis

The D proteins of a diverse set of putative TOMM cytolysin producers were analyzed because it known that the B and C proteins can sometimes be omitted from TOMM biosynthetic gene clusters, whereas D proteins are always present.[11] Further, the B and C proteins are homologous to non-TOMM dehydrogenases and the E1-ubiquitin activating family of enzymes, and thus are prone to bioinformatic false-positivies. Thus, a D protein alignment was created using the standard parameters of ClustalW[42] and then a maximum likelihood tree was made using the standard parameters of Molecular Evolutionary Genetics Analysis (MEGA).[43] MLST phylogenetic analysis was generated in MEGA 5.1 using concatenated sequences of eight chromosomally-located housekeeping genes (*clpA*, *clpX*, *nifS*, *pepX*, *pyrG*, *recG*, *rplB*, and *uvrA*)[44] using the maximum likelihood method with 1,000 bootstrap repeats, the GTR model, uniform rates among sites and nearest-neighbor-interchange.

PCR screen for Bor TOMM biosynthetic genes

Genomic DNA of cultured *Bbsl* isolates was purified using either a Qiagen DNeasy blood and tissue kit or a High Pure PCR Template Preparation Kit (Roche Diagnostics).[45, 46] Total DNA from field-collected questing ticks was extracted using alkaline hydrolysis [47]. Genomic DNA from each *Bbsl* isolate was tested for the presence of *borB, -C,* and *-D* by *Taq* PCR with primers specific to VS116 (Table B.4). Amplification of a product of appropriate length as determined by agarose gel electrophoresis was considered a positive hit. Sanger sequencing confirmed the identity of PCR amplicons from each species. Primers that effectively amplified *Bbsl* 16S rRNA gene sequences were used to verify that sufficient DNA was present in all samples.

Preparation of plasmids for *S. pyogenes* transformation

The chemically synthesized genes (IDT) encoding BvalA and SagA-BvalA were subcloned into pDCerm as previously described.[8] Stop codons were introduced by site-directed mutagenesis. Mutant candidates were miniprepped and verified by DNA sequencing using pDCerm_SeqMCS_F Primers are listed in Table B.4. In order to obtain sufficient plasmid DNA for transformation of *S. pyogenes ΔsagA,* each pDCerm construct was maxiprepped using an E.Z.N.A. Plasmid Maxi Kit (Omega Bio-tek).

***S. pyogenes* transformation**

S. pyogenes ΔsagA[23] was rendered electrocompetent using a modified version of a previously published method.[48] Briefly, *S. pyogenes ΔsagA* was grown to an OD_{600} of 0.35 in 50 ml Todd–Hewitt broth with 0.6% (w/v) glycine. Cells were washed three times with cold 0.625 M sucrose (pH 4) and resuspended in 300 µl of the same buffer with 20% (v/v) glycerol. 50 µl cells were equilibrated with ~1 µg of plasmid DNA on ice for 30 min and electroporated at 1.65 kV with a BioRad Micropulser. Immediately, cells were added to 250 µl Todd–Hewitt broth with 0.25 M sucrose and recovered for 2 h at 37 °C, before plating. The *S. pyogenes*-nature of transformants[49] and presence of pDCerm were verified by PCR (Table B.4).

Reaction of SLS-like peptides with purified SagBCD

All proteins were expressed with maltose-binding protein (MBP) fused to the N-terminus. SagA, SagA-C32A,[8] SagB, SagC and SagD were prepared as previously described.[7] The genes encoding BvalA

and SagA-BvalA were subcloned into pET28b-MBP as previously described.[7] Site-directed mutagenesis was performed to yield pET28b-MBP-SagA$^{1-35}$ (Table B.4). Precursor peptides were expressed from *E. coli* BL21(DE3)-RIPL (Strategene) for 1 h at 30 °C after induction with 1 mM IPTG at OD_{600} of 0.6. Following standard lysis procedures, each protein was purified by amylose affinity chromatography.[7] MBP-SagB, -C and -D (10 μM each) were combined with the relevant MBP-tagged precursor peptide (40 μM) and allowed to react in synthetase buffer as previously described.[7] The products of the synthetase reactions were applied to washed erythrocytes and lysis assessed as previously described.[7] Hemoglobin absorbance was measured at 410 nm on a Tecan Infinite® 200 PRO plate reader. Values were normalized to a Triton X-100 positive control. Phosphate buffered saline (PBS) served as a negative control.

References

1. Molloy, E. M., Casjens, S. R., Cox, C. L., Maxson, T., Ethridge, N. A., Margos, G., Fingerle, V., and Mitchell, D. A. (2015) Identification of the minimal cytolytic unit for streptolysin S and an expansion of the toxin family, *BMC Microbiol. 15*, 141.

2. Cunningham, M. W. (2000) Pathogenesis of group A streptococcal infections, *Clin. Microbiol. Rev. 13*, 470-511.

3. Marmorek, A. (1895) Le streptocoque et le sérum antistreptococcique, *Ann. Inst. Pasteur*, 593–620.

4. Todd, E. W. (1938) The differentiation of two distinct serologic varieties of streptolysin, streptolysin O and streptolysin S., *J. Pathol. Bacteriol.*, 423–445.

5. Molloy, E. M., Cotter, P. D., Hill, C., Mitchell, D. A., and Ross, R. P. (2011) Streptolysin S-like virulence factors: the continuing sagA, *Nat. Rev. Microbiol. 9*, 670-681.

6. Nizet, V., Beall, B., Bast, D. J., Datta, V., Kilburn, L., Low, D. E., and De Azavedo, J. C. (2000) Genetic locus for streptolysin S production by group A streptococcus, *Infect. Immun. 68*, 4245-4254.

7. Lee, S. W., Mitchell, D. A., Markley, A. L., Hensler, M. E., Gonzalez, D., Wohlrab, A., Dorrestein, P. C., Nizet, V., and Dixon, J. E. (2008) Discovery of a widely distributed toxin biosynthetic gene cluster, *Proc. Natl. Acad. Sci. U. S. A. 105*, 5879-5884.

8. Mitchell, D. A., Lee, S. W., Pence, M. A., Markley, A. L., Limm, J. D., Nizet, V., and Dixon, J. E. (2009) Structural and functional dissection of the heterocyclic peptide cytotoxin streptolysin S, *J. Biol. Chem. 284*, 13004-13012.

9. Melby, J. O., Nard, N. J., and Mitchell, D. A. (2011) Thiazole/oxazole-modified microcins: complex natural products from ribosomal templates, *Curr. Opin. Chem. Biol. 15*, 369-378.

10. Oman, T. J., and van der Donk, W. A. (2010) Follow the leader: the use of leader peptides to guide natural product biosynthesis, *Nat. Chem. Biol. 6*, 9-18.

11. Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P. C., Entian, K. D., Fischbach, M. A., Garavelli, J. S., Goransson, U., Gruber, C. W., Haft, D. H., Hemscheidt, T. K., Hertweck, C., Hill, C., Horswill, A. R., Jaspars, M., Kelly, W. L., Klinman, J. P., Kuipers, O. P., Link, A. J., Liu, W., Marahiel, M. A., Mitchell, D. A., Moll, G. N., Moore, B. S., Muller, R., Nair, S. K., Nes, I. F., Norris, G. E., Olivera, B. M., Onaka, H., Patchett, M. L., Piel, J., Reaney, M. J., Rebuffat, S., Ross, R. P., Sahl, H. G., Schmidt, E. W., Selsted, M. E., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., Sussmuth, R. D., Tagg, J. R., Tang, G. L., Truman, A. W., Vederas, J. C., Walsh, C. T., Walton, J. D., Wenzel, S. C., Willey, J. M., and van der Donk, W. A. (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature, *Nat. Prod. Rep. 30*, 108-160.

12. Pei, J., Mitchell, D. A., Dixon, J. E., and Grishin, N. V. (2011) Expansion of type II CAAX proteases reveals evolutionary origin of gamma-secretase subunit APH-1, *J. Mol. Biol. 410*, 18-26.

13. Maxson, T., Deane, C. D., Molloy, E. M., Cox, C. L., Markley, A. L., Lee, S. W., and Mitchell, D. A. (2015) HIV Protease Inhibitors Block Streptolysin S Production, *ACS Chem. Biol.*

14. Humar, D., Datta, V., Bast, D. J., Beall, B., De Azavedo, J. C., and Nizet, V. (2002) Streptolysin S and necrotising infections produced by group G streptococcus, *Lancet 359*, 124-129.

15. Fuller, J. D., Camus, A. C., Duncan, C. L., Nizet, V., Bast, D. J., Thune, R. L., Low, D. E., and De Azavedo, J. C. (2002) Identification of a streptolysin S-associated gene cluster and its role in the pathogenesis of Streptococcus iniae disease, *Infect. Immun. 70*, 5730-5739.

16. Flanagan, J., Collin, N., Timoney, J., Mitchell, T., Mumford, J. A., and Chanter, N. (1998) Characterization of the haemolytic activity of Streptococcus equi, *Microb. Pathog. 24*, 211-221.

17. Tabata, A., Nakano, K., Ohkura, K., Tomoyasu, T., Kikuchi, K., Whiley, R. A., and Nagamune, H. (2013) Novel twin streptolysin S-like peptides encoded in the sag operon homologue of beta-hemolytic Streptococcus anginosus, *J. Bacteriol. 195*, 1090-1099.

18. Gonzalez, D. J., Lee, S. W., Hensler, M. E., Markley, A. L., Dahesh, S., Mitchell, D. A., Bandeira, N., Nizet, V., Dixon, J. E., and Dorrestein, P. C. (2010) Clostridiolysin S, a post-translationally modified biotoxin from Clostridium botulinum, *J. Biol. Chem. 285*, 28220-28228.

19. Cotter, P. D., Draper, L. A., Lawton, E. M., Daly, K. M., Groeger, D. S., Casey, P. G., Ross, R. P., and Hill, C. (2008) Listeriolysin S, a novel peptide haemolysin associated with a subset of lineage I Listeria monocytogenes, *PLoS Pathog. 4*, e1000144.

20. Clayton, E. M., Daly, K. M., Guinane, C. M., Hill, C., Cotter, P. D., and Ross, P. R. (2014) Atypical Listeria innocua strains possess an intact LIPI-3, *BMC Microbiol. 14*, 58.

21. Clayton, E. M., Hill, C., Cotter, P. D., and Ross, R. P. (2011) Real-time PCR assay to differentiate Listeriolysin S-positive and -negative strains of Listeria monocytogenes, *Appl. Environ. Microbiol. 77*, 163-171.

22. Carr, A., Sledjeski, D. D., Podbielski, A., Boyle, M. D., and Kreikemeyer, B. (2001) Similarities between complement-mediated and streptolysin S-mediated hemolysis, *J. Biol. Chem. 276*, 41790-41796.

23. Datta, V., Myskowski, S. M., Kwinn, L. A., Chiem, D. N., Varki, N., Kansal, R. G., Kotb, M., and Nizet, V. (2005) Mutational analysis of the group A streptococcal operon encoding streptolysin S and its virulence role in invasive infection, *Mol. Microbiol. 56*, 681-695.

24. Radolf, J. D., Caimano, M. J., Stevenson, B., and Hu, L. T. (2012) Of ticks, mice and men: understanding the dual-host lifestyle of Lyme disease spirochaetes, *Nat. Rev. Microbiol. 10*, 87-99.

25. Stanek, G., Wormser, G. P., Gray, J., and Strle, F. (2012) Lyme borreliosis, *Lancet 379*, 461-473.

26. Dunbar, K. L., Chekan, J. R., Cox, C. L., Burkhart, B. J., Nair, S. K., and Mitchell, D. A. (2014) Discovery of a new ATP-binding motif involved in peptidic azoline biosynthesis, *Nat. Chem. Biol. 10*, 823-829.

27. Letzel, A. C., Pidot, S. J., and Hertweck, C. (2014) Genome mining for ribosomally synthesized and post-translationally modified peptides (RiPPs) in anaerobic bacteria, *BMC Genomics 15*, 983.

28. Casjens, S. R., Mongodin, E. F., Qiu, W. G., Dunn, J. J., Luft, B. J., Fraser-Liggett, C. M., and Schutzer, S. E. (2011) Whole-genome sequences of two Borrelia afzelii and two Borrelia garinii Lyme disease agent isolates, *J. Bacteriol. 193*, 6995-6996.

29. Schutzer, S. E., Fraser-Liggett, C. M., Qiu, W. G., Kraiczy, P., Mongodin, E. F., Dunn, J. J., Luft, B. J., and Casjens, S. R. (2012) Whole-genome sequences of Borrelia bissettii, Borrelia valaisiana, and Borrelia spielmanii, *J. Bacteriol. 194*, 545-546.

30. Mongodin, E. F., Casjens, S. R., Bruno, J. F., Xu, Y., Drabek, E. F., Riley, D. R., Cantarel, B. L., Pagan, P. E., Hernandez, Y. A., Vargas, L. C., Dunn, J. J., Schutzer, S. E., Fraser, C. M., Qiu, W. G., and Luft, B. J. (2013) Inter- and intra-specific pan-genomes of Borrelia burgdorferi sensu lato: genome stability and adaptive radiation, *BMC Genomics 14*, 693.

31. Biskup, U. G., Strle, F., and Ruzic-Sabljic, E. (2011) Loss of plasmids of Borrelia burgdorferi sensu lato during prolonged in vitro cultivation, *Plasmid 66*, 1-6.

32. Stewart, P. E., Byram, R., Grimm, D., Tilly, K., and Rosa, P. A. (2005) The plasmids of Borrelia burgdorferi: essential genetic elements of a pathogen, *Plasmid 53*, 1-13.

33. Casjens, S. (2001) Borrelia genomes, In *The spirochetes: molecular and cellular biology* (Saier, M., and Garcia-Lara, G., Ed.), pp 75-85, Horizon Scientific Press, Norfolk UK.

34. Schutzer, S. E., Fraser-Liggett, C. M., Casjens, S. R., Qiu, W. G., Dunn, J. J., Mongodin, E. F., and Luft, B. J. (2011) Whole-genome sequences of thirteen isolates of Borrelia burgdorferi, *J. Bacteriol. 193*, 1018-1020.

35. Williams, L. R., and Austin, F. E. (1992) Hemolytic activity of Borrelia burgdorferi, *Infect. Immun. 60*, 3224-3230.

36. Shaw, D. K., Hyde, J. A., and Skare, J. T. (2012) The BB0646 protein demonstrates lipase and haemolytic activity associated with Borrelia burgdorferi, the aetiological agent of Lyme disease, *Mol. Microbiol. 83*, 319-334.

37. Hyde, J. A., Weening, E. H., and Skare, J. T. (2011) Genetic transformation of Borrelia burgdorferi, *Current Protocols Microbiology Chapter 12*, Unit 12C 14.

38. Chaconas, G., and Norris, S. J. (2013) Peaceful coexistence amongst Borrelia plasmids: getting by with a little help from their friends?, *Plasmid 70*, 161-167.

39. Lee, S. F., Li, Y. J., and Halperin, S. A. (2009) Overcoming codon-usage bias in heterologous protein expression in Streptococcus gordonii, *Microbiology 155*, 3581-3588.

40. Melby, J. O., Dunbar, K. L., Trinh, N. Q., and Mitchell, D. A. (2012) Selectivity, directionality, and promiscuity in peptide processing from a Bacillus sp. Al Hakam cyclodehydratase, *J. Am. Chem. Soc. 134*, 5309-5316.

41. Belshaw, P. J., Roy, R. S., Kelleher, N. L., and Walsh, C. T. (1998) Kinetics and regioselectivity of peptide-to-heterocycle conversions by microcin B17 synthetase, *Chem. Biol. 5*, 373-384.

42. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0, *Bioinformatics 23*, 2947-2948.

43. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, *Mol. Biol. Evol. 28*, 2731-2739.

44. Margos, G., Vollmer, S. A., Cornet, M., Garnier, M., Fingerle, V., Wilske, B., Bormane, A., Vitorino, L., Collares-Pereira, M., Drancourt, M., and Kurtenbach, K. (2009) A new Borrelia species defined by multilocus sequence analysis of housekeeping genes, *Appl. Environ. Microbiol. 75*, 5410-5416.

45. Margos, G., Gatewood, A. G., Aanensen, D. M., Hanincova, K., Terekhova, D., Vollmer, S. A., Cornet, M., Piesman, J., Donaghy, M., Bormane, A., Hurn, M. A., Feil, E. J., Fish, D., Casjens, S., Wormser, G. P., Schwartz, I., and Kurtenbach, K. (2008) MLST of housekeeping genes captures geographic population structure and suggests a European origin of Borrelia burgdorferi, *Proc. Natl. Acad. Sci. U. S. A. 105*, 8730-8735.

46. Schulte-Spechtel, U., Fingerle, V., Goettner, G., Rogge, S., and Wilske, B. (2006) Molecular analysis of decorin-binding protein A (DbpA) reveals five major groups among European Borrelia burgdorferi sensu lato strains with impact for the development of serological assays and indicates lateral gene transfer of the dbpA gene, *Int. J. Med. Microbiol. 296 Suppl 40*, 250-266.

47. Vollmer, S. A., Bormane, A., Dinnis, R. E., Seelig, F., Dobson, A. D., Aanensen, D. M., James, M. C., Donaghy, M., Randolph, S. E., Feil, E. J., Kurtenbach, K., and Margos, G. (2011) Host migration impacts on the phylogeography of Lyme Borreliosis spirochaete species in Europe, *Environ. Microbiol. 13*, 184-192.

48. Framson, P. E., Nittayajarn, A., Merry, J., Youngman, P., and Rubens, C. E. (1997) New genetic techniques for group B streptococci: high-efficiency transformation, maintenance of temperature-

sensitive pWV01 plasmids, and mutagenesis with Tn917, *Appl. Environ. Microbiol. 63*, 3539-3547.

49. Liu, D., Hollingshead, S., Swiatlo, E., Lawrence, M. L., and Austin, F. W. (2005) Rapid identification of Streptococcus pyogenes with PCR primers from a putative transcriptional regulator gene, *Res. Microbiol. 156*, 564-567.

50. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Mol. Syst. Biol. 7*, 539.

51. Casjens, S. R., Mongodin, E. F., Qiu, W. G., Luft, B. J., Schutzer, S. E., Gilcrease, E. B., Huang, W. M., Vujadinovic, M., Aron, J. K., Vargas, L. C., Freeman, S., Radune, D., Weidman, J. F., Dimitrov, G. I., Khouri, H. M., Sosa, J. E., Halpin, R. A., Dunn, J. J., and Fraser, C. M. (2012) Genome stability of Lyme disease spirochetes: comparative genomics of Borrelia burgdorferi plasmids, *PLoS One 7*, e33280.

52. Casjens, S., Eggers, C., and Schwartz, I. (2010) Borrelia genomics: chromosome, plasmids, baceriophages and genetic variation., In *Borrelia: Molecular biology, host interaction and pathogenesis* (Samuels, S., and Radolf, J., Ed.), pp 27-52, Horizon Scientific Press, Norwich.

53. Casjens, S. R., Fraser-Liggett, C. M., Mongodin, E. F., Qiu, W. G., Dunn, J. J., Luft, B. J., and Schutzer, S. E. (2011) Whole genome sequence of an unusual Borrelia burgdorferi sensu lato isolate, *J. Bacteriol. 193*, 1489-1490.

54. Margos, G., Vollmer, S. A., Ogden, N. H., and Fish, D. (2011) Population genetics, taxonomy, phylogeny and evolution of Borrelia burgdorferi sensu lato, *Infect. Genet. Evol. 11*, 1545-1563.

55. Ivanova, L. B., Tomova, A., Gonzalez-Acuna, D., Murua, R., Moreno, C. X., Hernandez, C., Cabello, J., Cabello, C., Daniels, T. J., Godfrey, H. P., and Cabello, F. C. (2014) Borrelia chilensis, a new member of the Borrelia burgdorferi sensu lato complex that extends the range of this genospecies in the Southern Hemisphere, *Environ. Microbiol. 16*, 1069-1080.

56. Wang, D., Botkin, D. J., and Norris, S. J. (2003) Characterization of the vls antigenic variation loci of the Lyme disease spirochaetes Borrelia garinii Ip90 and Borrelia afzelii ACAI, *Mol. Microbiol. 47*, 1407-1417.

B.2 Plantazolicin is an ultra-narrow spectrum antibiotic that targets the *Bacillus anthracis* membrane

Katie J. Molohon, Patricia M. Blair, Seongjin Park, James R. Doroghazi, Tucker Maxson, Jeremy R. Hershfield, Kristen M. Flatt, Nathan E. Schroeder, Taekjip Ha, and Douglas A. Mitchell

This chapter was reproduced from Molohon et al., currently in review at ACS Infectious Diseases.

I generated genetic deletions in *Bacillus anthracis*.

Abstract

Plantazolicin (PZN) is a ribosomally synthesized and post-translationally modified natural product from *Bacillus amyloliquefaciens* FZB42 and *Bacillus pumilus*. Extensive tailoring to twelve of the fourteen amino acid residues in the mature natural product endows PZN with not only a rigid, polyheterocyclic structure, but also antibacterial activity. Here we report on the remarkably discriminatory activity of PZN toward *Bacillus anthracis*, which rivals a previously-described gamma (γ) phage lysis assay in distinguishing *B. anthracis* from other members of the *Bacillus cereus* group. We evaluate the underlying cause of this selective activity by measuring the RNA expression profile of PZN-treated *B. anthracis*, which revealed significant upregulation of genes within the cell envelope stress response. PZN depolarizes the *B. anthracis* membrane like other cell envelope-acting compounds but uniquely localizes to distinct foci within the envelope. Selection and whole-genome sequencing of PZN-resistant mutants of *B. anthracis* implicate a relationship between the action of PZN and cardiolipin (CL) within the membrane. Exogenous CL increases the potency of PZN in wild type *B. anthracis* and promotes the incorporation of fluorescently tagged PZN in the cell envelope. We propose that PZN localizes to and exacerbates structurally compromised regions of the bacterial membrane, which ultimately results in cell lysis.

Introduction

The current practice of employing broad-spectrum antibiotics to treat bacterial infections contributes to the rise of antibiotic resistance.[1] As a countermeasure, species-selective and narrow-spectrum antibacterial compounds are garnering increased attention in the medical community for their

potential as therapeutics and diagnostics.[2, 3] Plantazolicin (PZN) is a polyheterocyclic, linear compound of the ribosomally synthesized and post-translationally modified peptide (RiPP) natural product family with narrow-spectrum antibiotic activity (Figure S1).[4] More specifically, PZN is a member of the thiazole/oxazole-modified microcins (TOMMs), a recently grouped and rapidly expanding RiPP class with ~1,500 identified gene clusters.[5, 6] Previously, PZN was described as an antibiotic compound that inhibits Gram-positive organisms closely related to its producing organism, *Bacillus amyloliquefaciens* FZB42.[4] In 2011, by screening a small panel of microorganisms, we described PZN as having potent activity towards *Bacillus anthracis*, but not other Gram-positive pathogens.[7] Several additional PZN-like gene clusters have been identified in six distinct bacterial genera (from three phyla) through genome mining, but experimental data on antibiotic specificity has so far been limited to PZN.[4, 7] Although PZN has been the subject of total synthesis,[8, 9] heterologous expression,[10] and enzymological studies,[11-13] insight into the mode of action (MOA) of PZN has not been reported in the seven years since the discovery of its biosynthetic gene cluster.[14]

B. anthracis, the causative agent of anthrax and a category A priority pathogen, is a Gram-positive bacterium and is a member of the *B. cereus sensu lato* group, which includes *B. cereus*, *B. anthracis*, *B. thuringiensis*, and *B. mycoides*.[15] Microbiologists have debated whether these organisms should be considered as one species, given that some strains share greater than 99% DNA sequence identity. Despite being grouped with other *Bacillus* species, *B. anthracis* harbors a number of features that set it apart from other members of the *B. cereus* group. Fully virulent *B. anthracis* contains two conserved plasmids, pXO1 and pXO2, which harbor the genes responsible for producing the anthrax toxin and poly-D-glutamic acid capsule, respectively. However, homologous plasmids are also found in certain *B. cereus* strains.[16] Beyond characteristic plasmid content, *B. anthracis*, unlike other members of the *B. cereus* group, harbors a nonsense mutation in *plcR* (phospholipase C regulator), rendering *B. anthracis* non-motile and non-hemolytic.[16] Nonetheless, differentiation of *B. anthracis* from other nonpathogenic *B. cereus* species is necessary due to concerns in the agricultural and public health communities regarding bioterrorism.

Other defining features of *B. anthracis* that may facilitate species selectivity are exterior to the cell wall. *B. anthracis* displays a two-dimensional protein lattice called the surface layer (S-layer). Decorated with surface-associated proteins in a *csaB* (cell surface attachment)-dependent manner,[17] the S-layer is non-covalently attached to the secondary cell wall polysaccharide (SCWP),[18] which is covalently tethered to the peptidoglycan. The *B. anthracis* SCWP is structurally unique[19] and serves as the binding site for gamma (γ) phage[20, 21] and previously described *B. anthracis* typing antibodies.[22] γ phage produce a peptidoglycan hydrolase, PlyG, which specifically recognizes the terminal galactoses of the *B. anthracis* SCWP and subsequently hydrolyzes the cell wall.[20] Similarly, typing methods using monoclonal antibodies to the SCWP also exploit differences in the terminal sugar unit. However, there exist atypical *B. anthracis* strains that would constitute false-negatives in any diagnostic assay based on these methods.[22, 23] Wip1, another *B. anthracis*-specific phage, is even more selective than γ phage, yet certain *B. cereus* strains remain sensitive.[24] Thus, the species specificity of PZN is intriguing not just from a MOA standpoint, but also as a means to distinguish *B. anthracis* from other *B. cereus sensu lato* members.

Here we describe PZN as a remarkably selective small molecule antibiotic towards *B. anthracis*. The intriguing specificity was first examined by gene expression profiling, which yielded an expression signature distinct from broader spectrum antibiotics. We have identified and characterized a set of resistant mutants and evaluated their role in PZN resistance, which led to us to further investigate the bacterial membrane as the probable target of PZN. Using fluorescence-based approaches, we confirmed that PZN localizes to the cell envelope and rapidly causes membrane depolarization. Taken together with the observation that PZN interacts synergistically with the negatively charged phospholipid, cardiolipin (CL), we propose that PZN localizes to and aggravates transient weaknesses present in the *B. anthracis* cell membrane.

Results and Discussion

B. anthracis, the causative agent of anthrax, can often be mistaken for other members of the *B. cereus* group. As a bacterium associated with bioterrorism, there is an increasing need for accurate and

rapid identification of this dangerously pathogenic strain. We therefore set out to characterize and understand the selectivity of PZN.

Defining the species selectivity of PZN

PZN was originally described as a Gram-positive antibiotic, inhibiting the growth of *B. subtilis*, *B. cereus*, and *B. megaterium*.[4] It is important to note, however, that the spot on lawn assay employed to reach this conclusion used an exorbitant amount of purified PZN (1 mg per spot). Using a microbroth dilution assay, the activity of PZN was later revealed to be considerably more selective, in that antibacterial activity was only detected toward *B. anthracis* upon screening a small panel of human pathogens.[7] We continued to establish this unusually narrow-spectrum of activity by screening a larger panel of strains with varying degrees of genetic similarity (Table S1). PZN was found to be selective for vegetative *B. anthracis*, including fully virulent biosafety level 3 strains, with minimum inhibitory concentrations (MICs) between 1-16 μg/mL (0.75-12 μM). Endospores, the dormant phase of the *B. anthracis* life cycle, were resistant to PZN until germination was initiated (Table S2). By microbroth dilution, *B. subtilis* and *B. cereus* were not susceptible to PZN at concentrations up to 64 μg/mL, which is in contrast to the previously reported spot on lawn assay.[4]

To further investigate the selectivity of PZN towards *B. anthracis*, we conducted a head-to-head comparison using the γ phage assay. Prior to modern genomic methods, γ phage sensitivity and other phenotype testing were popular methods for identifying *B. anthracis*.[23] Notwithstanding the reported 96% positive accuracy, non-*B. anthracis* strains that are sensitive to γ phage and true *B. anthracis* strains that are insensitive have been reported.[23, 25, 26] We obtained a panel of atypical *B. cereus* strains that are sensitive to γ phage and tested them for PZN susceptibility (Table B.7). *B. cereus* strains that generated a false positive in the γ phage assay were not susceptible to PZN.[24, 26, 27]

Table B.7. γ phage sensitivity and PZN susceptibility of *Bacillus* strains

Strain	γ Phage Sensitivity[a]	PZN MIC[b]	Source[c]
B. anthracis Sterne 7702	+++	1	USDA
B. anthracis Sterne 34F2 A0517-1[d]	+++	2	BEI
B. cereus 2002013145	+++	>64	CDC[e]
B. cereus 2002013146	+++	>64	CDC
B. cereus 2002013100	++	>64	CDC
B. cereus 2000031002	+++	>64	CDC
B. cereus ATCC 4342	+	>64	ATCC
B. cereus ATCC 7064	+	>64	ATCC
B. cereus CDC 32805	+	>64	Ref 25
B. cereus G9241	-	8	BEI
B. megaterium 899	-	32	BGSC
B. mycoides 96/3308	-	>64	BGSC

[a] Plus signs indicate the level of phage sensitivity, with +++ representing the most sensitive
[b] MIC (minimum inhibitory concentration) as determined by microbroth dilution, units in μg/mL
[c] Abbreviations: USDA, United States Department of Agriculture; BEI, Biodefense and Emerging Infections Research Resources Repository; CDC, United States Centers for Disease Control and Prevention; ATCC, American Type Culture Collection; BGSC, Bacillus Genetic Stock Center
[d] LLNL A0517 was obtained from BEI as a mixture of two colony types. A0517_1 was confirmed to be devoid of pXO1 by PCR (Figure S2)
[e] Strains identified by multilocus sequence typing analysis[25]

In an attempt to further define the attributes giving rise to the species selectivity of PZN, we procured various bacterial strains that address key differences between *B. anthracis* and *B. cereus*. *plcR*, encoding the phospholipase C regulator, is nonfunctional in *B. anthracis* but is intact in *B. cereus.*[16] Deletion of *plcR* in *B. cereus* did not increase its susceptibility to PZN (Table S1). Additionally, sortase-deficient strains of *B. anthracis*, which lack the ability to anchor various proteins to the cell wall, remain susceptible to PZN.[28] Similarly, PZN activity was not dependent on the presence or composition of the *B. anthracis* S-layer, as strains deficient in S-layer assembly or decoration, namely those harboring mutations in *csaB*, *sap*, and *eag*, are equally susceptible to PZN.[17] We further confirmed that susceptibility to PZN is plasmid-independent given that the susceptibility of a plasmid-deficient strain, *B. anthracis* LLNL A0517-1. (Table S1; Figure S2). Wip1 phage and antibody typing have also been used to discriminate *B. cereus sensu lato* strains, but also have known exceptions to their specificity for *B.*

anthracis.[22, 24] We obtained a "false-positive for *B. anthracis*" strain for each marker: *B. cereus* CDC32805 for Wip1 and *B. cereus* ATCC 7064 for antibody typing. We again observed no measurable PZN susceptibility for either strain (Table S1).

After extensive susceptibility testing, the only notable exception to the *B. anthracis* selectivity of PZN was *B. cereus* G9241 (MIC of 8 μg/mL). *B. cereus* G9241 encodes the genes for an anthrax-like toxin on its pBCXO1 plasmid, which is named for its homology to the *B. anthracis* pXO1 plasmid.[29] As *B. cereus* G9241 is encapsulated and toxigenic, it causes an anthrax-like disease but is undetectable in the γ phage assay.[25] From a pathogen detection perspective, the action of PZN towards G9241 could be considered fortuitous if it were to be further developed as a diagnostic test. Together, these data not only highlight the species discrimination of PZN but also rule out *plcR*-related effects, sortase-mediated proteins, the SCWP, the S-layer, and plasmid-borne entities as targets of PZN.

The spectrum of PZN activity calls into question whether bacteria are the naturally intended target. The canonical PZN-producing strain, *B. amyloliquefaciens* FZB42, has been described to have antifungal and nematicidal activity and is a prolific producer of other natural products.[30] Liu *et al.* assigned a nematicidal activity to PZN, derived from experiments showing that PZN-deficient FZB42 strains exhibit reduced nematicidal activity against *Caenorhabditis elegans*.[31] Since these experiments employed crude cellular extracts, we evaluated purified PZN in a similar manner, embedding the compound in agar (slow killing assay) or providing PZN in a liquid suspension (liquid fast killing assay). PZN was found to be no more toxic to *C. elegans* than a vehicle control and is clearly not nematicidal in its own right (Figure S3). Purified PZN was also not responsible for the antifungal activity of the native producer, leaving the ecological function of PZN unknown (Table S1).

After observing the specificity of PZN under one growth medium condition (Luria-Bertani broth, LB), we re-assessed specificity against a smaller but representative panel of strains in two additional growth media (Mueller-Hinton and brain-heart infusion broths, Table S3). All tested strains of *B. anthracis* remained equally susceptible, but unexpectedly, some *Staphylococcus aureus* strains were susceptible to PZN under alternative growth media (MICs from 8 – 32 μg/mL). Only *S. aureus* showed

media-dependent susceptibility to PZN; all other tested strains remained non-susceptible to PZN. The broadened spectrum of PZN activity in rich media convolutes the search for PZN's precise MOA.

Assessing potential macromolecules as the target of PZN

In an attempt to identify the molecular target(s) of PZN, two biotinylated PZN probes were prepared for affinity purification, a standard technique for identifying targets of bioactive small molecules.[32] First, an *N*-terminally biotinylated PZN was prepared from desmethylPZN, which is a biosynthetic intermediate obtainable by fermentation of an expression strain missing the PZN *N*-methyltransferase (Figure S1; Figure S4).[7] Unfortunately, *N*-biotinylated PZN was devoid of antibiotic activity at the concentrations tested; however, *C*-terminally biotinylated PZN retained bioactivity, albeit with a 16-fold reduction in potency (Figure S1; Figure S5). For this reason, *C*-terminally labeled PZN-biotin was chosen for affinity-based target identification. Despite numerous attempts, we were unable to identify interactions unique to PZN-biotin compared to the control (data not shown). A photoaffinity tagged derivative of PZN was then prepared on the notion that PZN-biotin may weakly interact with its target, (Figure S1, Figure S6). To minimally perturb potential interactions, a minimalist, bifunctional analog containing a diazirine and alkyne was employed to enable crosslinking prior to biotinylation and enrichment.[33] Again, no binding partners were identified by either mass spectrometry or western blot.

Because affinity purification-based strategies to identify small molecule targets is most successful when the interaction is proteinaceous and of high-affinity,[34] we next considered the possibility that PZN interacted with a non-protein macromolecule. We used radiolabeled precursors of the cell wall, fatty acids, and RNA (as well as protein), to assay for their biosynthesis in the presence of PZN. Similar to the nisin-like lanthipeptide Pep5, PZN extensively disrupted macromolecular biosynthesis (Figure S7).[35] Interestingly, PZN did not affect cell wall biosynthesis compared to the vancomycin control.

The gene expression signature of PZN

Sub-lethal antibiotic treatment stimulates rapid transcriptional responses in bacteria and the induced/repressed genes may be indicative of MOA.[36] We thus performed RNA-Seq to evaluate the

transcriptional response of *B. anthracis* following exposure to 0.25 µg/mL (0.25 × MIC) PZN for 10 min.[37] A total of 74 genes were differentially regulated, including 63 upregulated and 11 downregulated genes, with an adjusted false discovery rate (q-value) of 0.01 (Figure B.11; Table S4). The expression of a subset of these genes was validated by qRT-PCR (Tables S5-S6). Fourteen of the upregulated genes were transporter subunits, a common stress response upon antibiotic treatment.[38] Conversely, PZN treatment led to the downregulation of genes associated with L-lactate metabolism, for which the implications remain unclear.

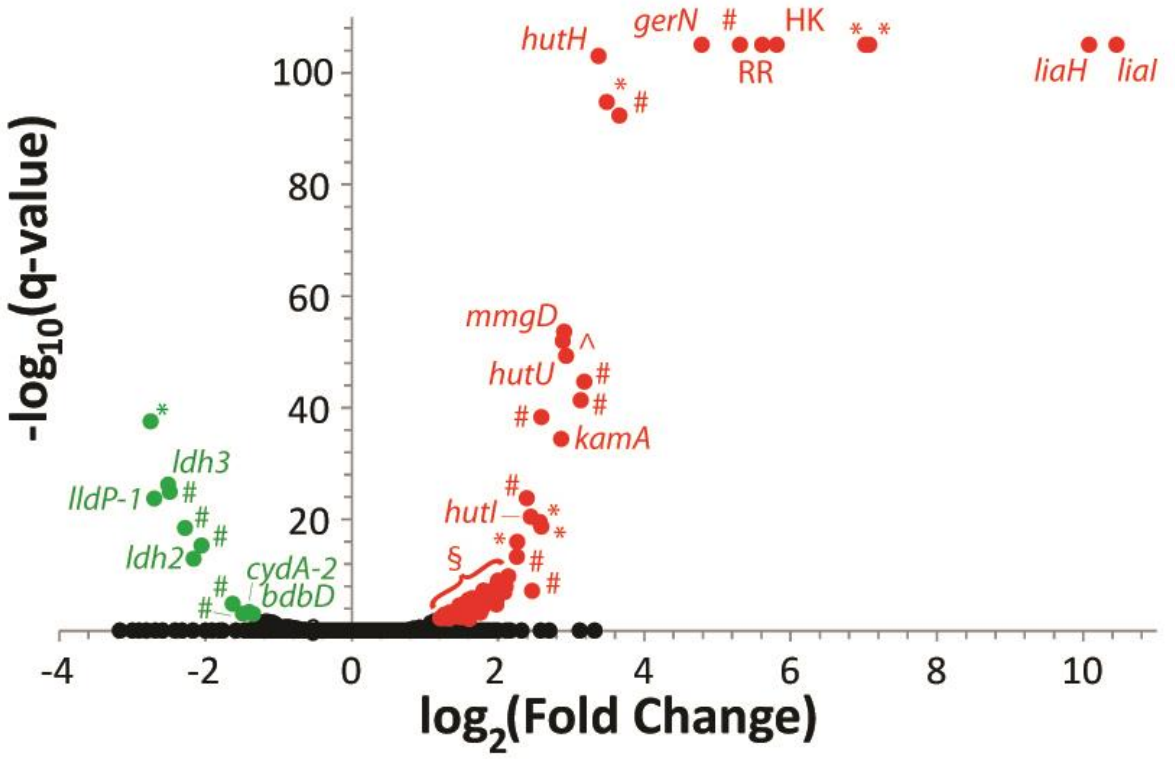


Figure B.11. *B. anthracis* gene expression profile when treated with PZN. A volcano plot represents the set of differentially regulated genes in response to treatment with 0.25 × MIC of PZN. Red (upregulated) and green (downregulated) points are genes with significantly altered expression in response to PZN treatment, while black genes did not meet the q--value (< 0.01) threshold. Genes where the q-value = 0 were given a value of 1 × 10-105. Abbreviations: # – hypothetical, * – transporter, ^ – transcriptional regulator, *bdbD* – *B. subtilis* hypothetical homolog, *lldP-1* – L-lactate permease, *ldh2/3* – L-lactate dehydrogenase, *cydA-2* – cytochrome d ubiquinol oxidase, *mmgD* – citrate synthase 3, *hutHIU* – histidine utilization genes, *gerN* – germination protein, *kamA* – L-lysine 2,3-aminomutase, § – remaining upregulated genes. All differentially expressed genes are summarized in Table S4.

The most highly upregulated *B. anthracis* genes upon PZN treatment were *bas1344* and *bas1345*, which encode a hypothetical protein and a predicted member of the PspA/IM30 family, respectively (Table S5). These genes are homologous to the *B. subtilis* genes *liaI* and *liaH* (lipid II cycle interfering antibiotics), which are involved in the cell envelope stress response. Induction of these genes upon antibiotic treatment is well documented in *B. subtilis*, specifically to antibiotics interacting with lipid II in some capacity (*e.g.*, nisin, vancomycin, and bacitracin).[39] Induction of *lia* is also seen upon treatment of *B. subtilis* with daptomycin, despite the lack of direct interaction between daptomycin and lipid II.[40] PZN treatment also results in massive upregulation of *bas5200* and *bas5201*, which are homologous to a *B. subtilis* thermosensor two-component system that regulates the lipid desaturase, *desRK*.[41] To our knowledge, PZN is the first compound known to alter the expression of *desRK*, further suggestive of a unique MOA.

Recently, we reported on the synthesis of a PZN derivative, Me_2-Arg-Az_5 (Figure S1).[13] Chemically, Me_2-Arg-Az_5 represents the *N*-terminal half of PZN, but the activity spectrum of Me_2-Arg-Az_5 is broader and includes other *Bacillus* species as well as *S. aureus* (Table S7). Additionally, mutations conferring PZN resistance did not confer resistance to other antibiotics or to Me_2-Arg-Az_5. To investigate their differing spectra of activity, we recorded the gene expression profile of *B. anthracis* treated with Me_2-Arg-Az_5 under otherwise identical conditions (0.25 × MIC, 10 min) by RNA-Seq. The two compounds shared a minor portion of their expression profiles, but each profile was largely independent (Figure S8; Tables S8-S9). For example, sub-lethal Me_2-Arg-Az_5 treatment also induced the *desRK* two-component system, but expression of *liaIH* remained unchanged. Additionally, Me_2-Arg-Az_5 failed to induce *B. anthracis* lysis, in contrast to PZN (Figure S9). A possible explanation for the observed differences between PZN and Me_2-Arg-Az_5 is that the *C*-terminal portion of the molecule is responsible for the species selectivity of the mature molecule and the *N*-terminal portion harbors the antibiotic activity, although this remains to be more extensively investigated. The expression profiles of PZN and Me_2-Arg-Az_5, together with strain susceptibility, suggest that PZN and Me_2-Arg-Az_5 pursue

independent, but possibly related, targets. Thus, while Me_2-Arg-Az_5 is not useful as a mimic for the full-length natural product, it may represent a way to broaden the spectrum of PZN activity.

PZN depolarizes the B. anthracis membrane

We examined the activity of PZN towards *B. anthracis* in the presence of the membrane disrupting agents nisin and daptomycin. Based on the resulting isobolograms, both compounds elicited strong synergistic activity with PZN (Figure B.12).[42] Both nisin and daptomycin disrupt membrane potential in Gram-positive organisms,[43, 44] and their synergism with PZN suggests that PZN might also depolarize the *B. anthracis* membrane. Thus, $DiOC_2(3)$ (3,3'-diethyloxacarbocyanine iodide) was used to measure the membrane potential of PZN-treated *B. anthracis* cells. As expected, PZN treatment, at both inhibitory and sub-inhibitory concentrations, correlated with a decrease in membrane potential, suggesting that PZN exerts its action by disrupting the integrity of the cell membrane (Figure B.12).

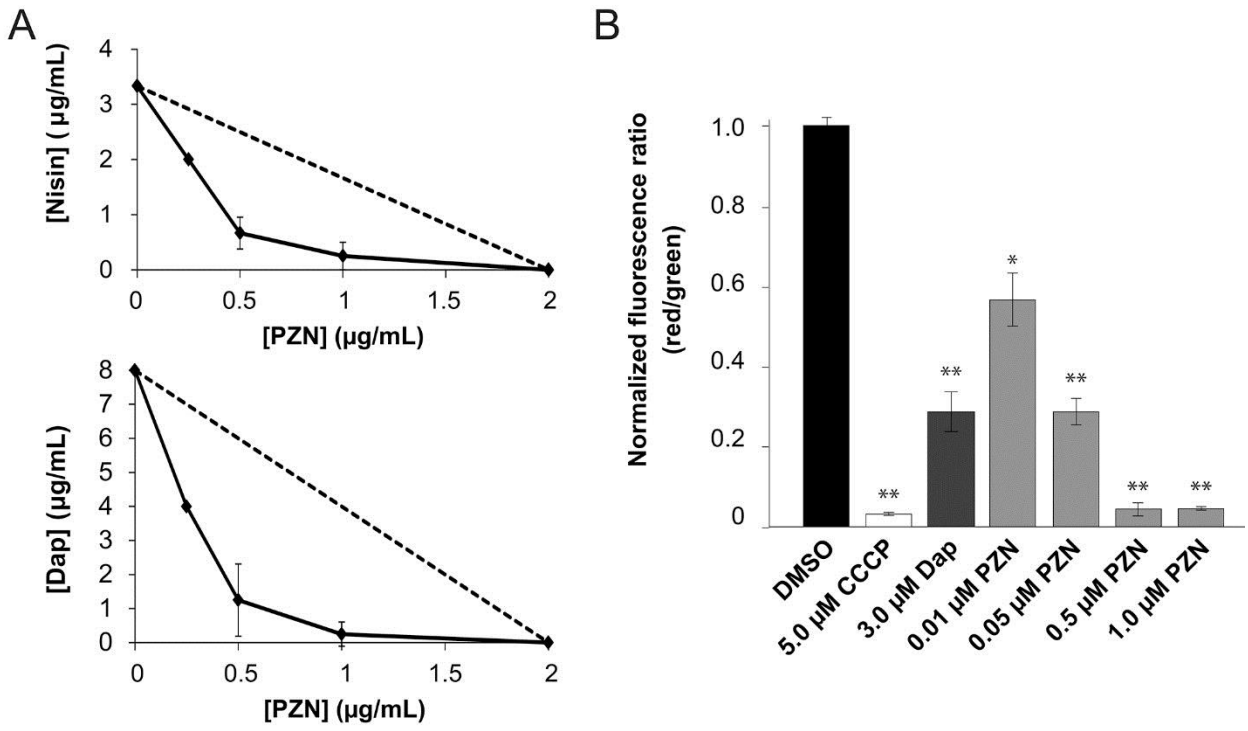


Figure B.12. PZN synergizes with cell envelope-acting antibiotics and depolarizes the *B. anthracis* membrane. (A) Isobolograms of the minimum inhibitory concentrations (µg/mL) of PZN with nisin (top) and daptomycin (bottom). Interactions taking place below the dotted line represent synergistic behavior. (B) Detection of membrane potential in *B. anthracis*. Red/green ratios were calculated using mean fluorescence intensities of cells treated for 30 min at RT with 0.1 µM DiOC2(3) and the vehicle of DMSO (negative control), 0.1 µM DiOC2(3) and 5.0 µM CCCP (positive control), 0.1 µM DiOC2(3) and 0.5 µM

PZN, and 0.1 μM DiOC2(3) and 1.0 μM PZN. Data were normalized to the positive control sample of DiOC2(3) and vehicle (DMSO). Error is given as standard deviation with n = 3. P-values are given relative to the DMSO control with * indicating < 0.0005 and ** indicating < 0.0001. Abbreviations: Dap, daptomycin; DiOC2(3), 3,3'- diethyloxacarbocyanine iodide; CCCP, carbonyl cyanide *m*-chlorophenyl hydrazone.

Subcellular Localization of PZN

With mounting evidence that PZN targets the cell membrane, we aimed to determine the subcellular localization of PZN by confocal microscopy. Antibiotics derivatized with fluorescent probes have previously been used to shed light on their MOAs.[45, 46] Localization of PZN was established by employing a Cy5-labeled PZN derivative (PZN-Cy5) (Figs. S1, S10) that retained much of its anti-*B. anthracis* activity (MIC of 4 μg/mL, 2 μM). PZN-Cy5 localized to distinct ~200 nm wide foci in *B. anthracis* Sterne (Figure B.13a). To confirm that PZN-Cy5 behaved in a matter identical to unlabeled PZN, we carried out a competition assay where and excess of PZN was applied to *B. anthracis* Sterne followed by addition of PZN-Cy5. Due to the extensive cell lysis elicited by PZN, we employed a later-described spontaneously PZN-resistant Sterne mutant (PR06) for this competition assay. Just as in *B. anthracis* Sterne, PZN-Cy5 localized to distinct foci in *B. anthracis* PR06 (Figure B.13b). PZN-Cy5 failed to label strain PR06 when an excess of unlabeled PZN was administered first, demonstrating that PZN and the PZN-Cy5 probe identically interact with *B. anthracis* (Figure B.13c). Due to the photoswitching properties of Cy5, we were able to further investigate PZN-Cy5 localization using stochastic optical reconstruction microscopy (STORM, Figure B.14).[47] Using this super-resolution imaging technique, *B. anthracis* Sterne cells were again confirmed to accumulate PZN-Cy5 at the foci described above. These foci were clearly localized to the surface of the cells, further implicating a component of the cell envelope as the target of PZN (Figure S11, Videos S1-S2). *B. anthracis* cells contain 16 ± 2 foci per cell, each with a diameter of 181 ± 7 nm, as determined by analysis of 14 cells treated with PZN-Cy5 (Figure S11). Notably, the labeling pattern of PZN-Cy5 is different from BODIPY-vancomycin, which has been shown to localize to bacterial septa at the site of peptidoglycan synthesis.[48] If PZN were acting on the cell wall, sites of active PG synthesis or the entire cell wall would be labeled with PZN-Cy5. The non-septal, punctate labeling of PZN-Cy5 suggests that the target of PZN is neither nascent nor existing

peptidoglycan, which is congruent with the observation that PZN did not block cell wall biosynthesis (Figure S7).

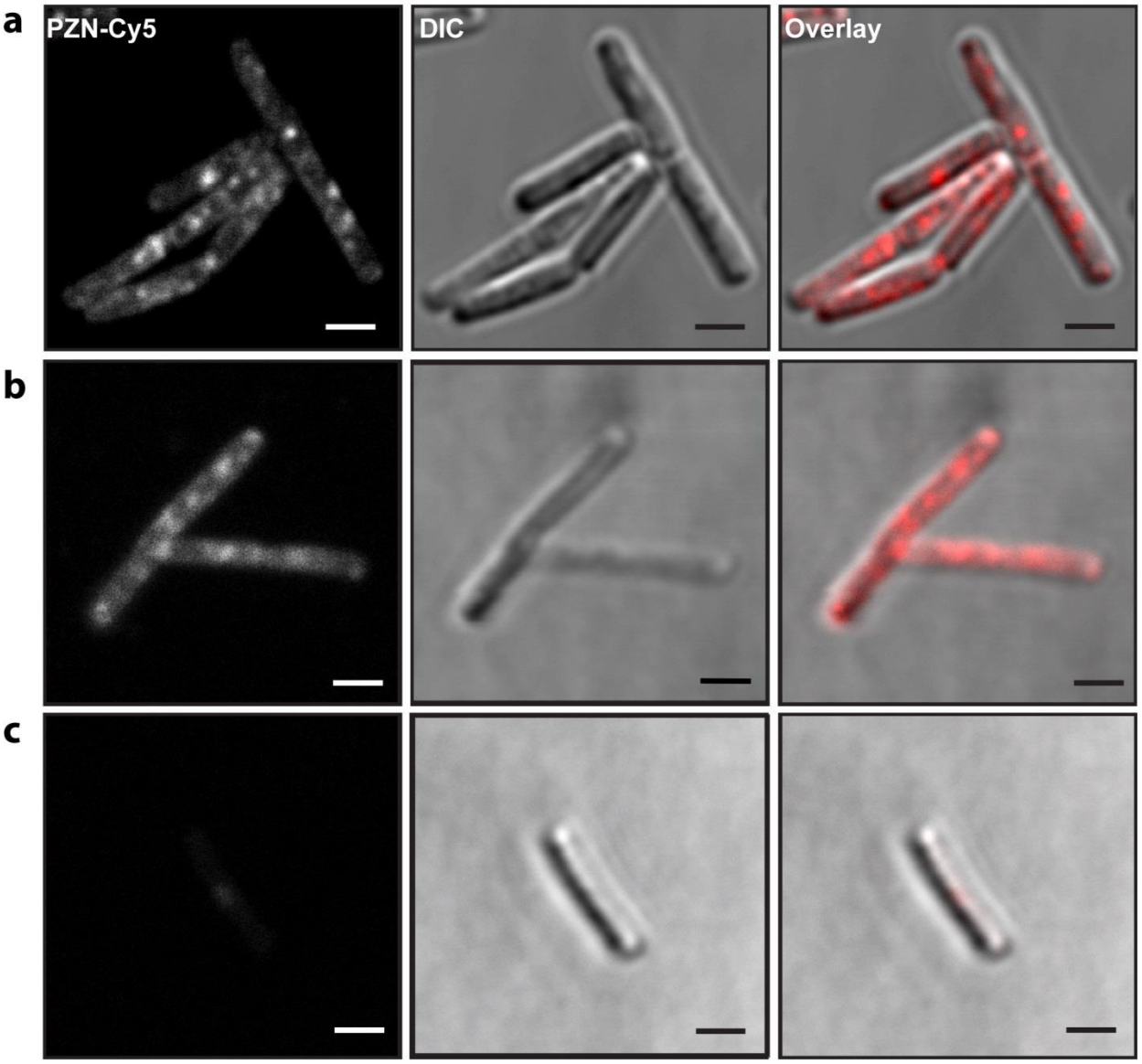


Figure B.13. PZN-Cy5 localizes to distinct foci on *B. anthracis*. (A) Representative fluorescence microscopy images are shown for the Cy5, DIC, and merged channels of *B. anthracis* Sterne treated with 0.1 μM PZN-Cy5 (0.05 × MIC) for 30 min. Competition experiments in a PZN-resistant *B. anthracis* strain (PR06, *vide infra*) show (B) robust labeling with 0.05 μM PZN-Cy5 in the absence of unlabeled PZN and (C) significantly decreased labeling when cells are pretreated with 1 μM PZN (0.016 × MIC, resistant strain PR06) for 20 min. Scale bars, 2 μm.

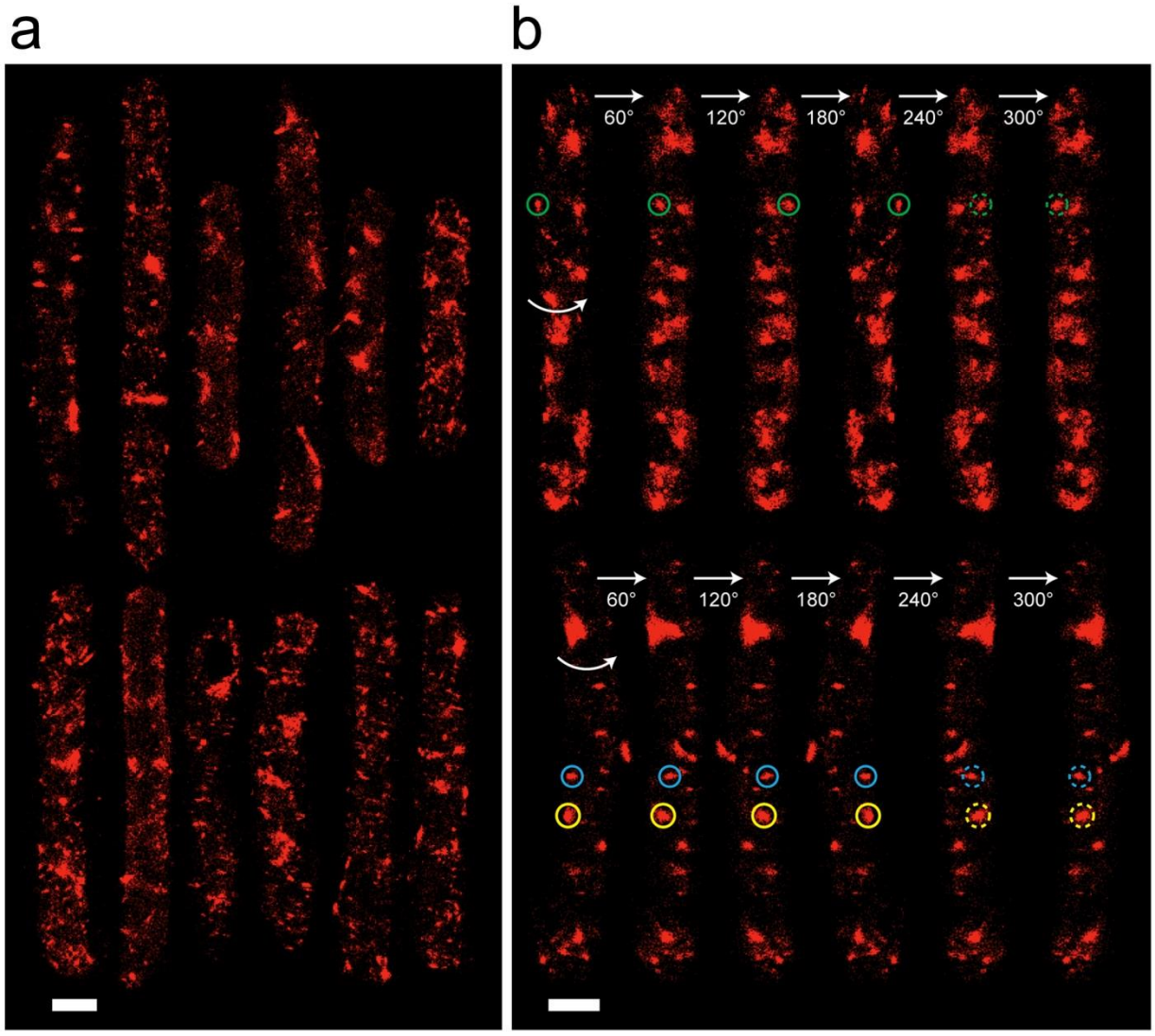


Figure B.14. STORM images of PZN-Cy5 labeled *B. anthracis*. (A) 3-D super-resolution images of 12 representative *B. anthracis* treated with PZN-Cy5. (B) Two representative cells rotated about the z-axis show distinct, non--septal localization of PZN-Cy5. Green, blue, and yellow circles mark three individual foci in each rotated view. Scale bars, 1 μm.

Isolation and characterization of PZN-resistant mutants

An orthogonal strategy for obtaining antibiotic MOA information involves the selection and mapping of resistance-conferring polymorphisms.[49] The mutated gene(s) can either be involved directly in the MOA of the antibiotic or in a target-unrelated mechanism of immunity. We isolated PZN-resistant *B. anthracis* by growing the Sterne strain on agar plates containing PZN at 4 × MIC. The resistance frequency was determined to be 2.3×10^{-7}, and the resulting mutants exhibited MICs that were ≥32

μg/mL. Genomic DNA was isolated and sequenced for six independently-selected PZN-resistant strains (PR01 through PR06) and the parent Sterne strain. Comparison of PR01 through PR06 to the parent revealed that all six polymorphisms were confined to a 50-nucleotide section of a single gene, *bas4114,* which is annotated as an AcrR transcriptional repressor (Table S10).[50] This particular AcrR protein is predicted to contain a single transmembrane domain near the *C*-terminus (Figure S12), which is precisely where the PZN-resistance conferring mutations were found, all resulting in premature stop codons. Directly downstream of *bas4114* are two EmrE-type efflux pumps, encoded by *bas4115-4116.* We hypothesized that as an AcrR transcriptional repressor, BAS4114 would negatively regulate *bas4115-4116*, and mutations near the C-terminus of BAS4114 would result in regulator dysfunction and derepression of the efflux pumps. This in turn would increase resistance to PZN. Therefore, RNA-Seq was employed to compare the mRNA expression profiles of PR06 to the parent Sterne strain. This analysis revealed a significant upregulation of *bas4114-4117* (Table S11). PR06 and the Sterne parent were equally susceptible to Me_2-Arg-Az_5 (Table S7), again underscoring differences between PZN and Me_2-Arg-Az_5. The susceptibility of PR06 to a panel of mechanistically-diverse antibiotics, including daptomycin, was also assessed (Table S12). The PZN resistance-conferring mutation did not alter the susceptibility towards any tested antibiotic.

Mutation of *bas4114* is clearly the favored route for generating PZN-resistance in *B. anthracis*, as shown by the occurrence of multiple independent mutations within the same gene. In order to subvert this resistance mechanism, and to obtain more insightful information about PZN MOA, we deleted *bas4114-4117* from the parental strain by homologous recombination (Figure S13). *B. anthracis* Sterne *Δbas4114-4117* thus became the new parental strain for isolating second-generation PZN-resistant mutants, as the removal of *bas4114-4117* rendered the strain resensitized to PZN at wild type levels (1 μg/mL). This time, two routes were pursued for obtaining PZN-resistant strains. First, we selected spontaneous PZN-resistant mutants by challenging with 4 × MIC. Isolation of the spontaneous mutants resulted in a mutation frequency an order of magnitude lower than before (1.3×10^{-8}). Two independently-selected resistant strains (PR07 and PR08) were subjected to whole-genome sequencing, revealing single point

mutations within *ftsE* (Table B.8). FtsE is an ATP-binding protein that associates with its cognate permease FtsX, which together comprise an ABC transporter that functions during cell wall elongation and septum formation.[51]

Table B.8. PZN-resistant mutants of *B. anthracis* Sterne Δ*bas4114-4117*

Strain	Mutation	Gene	Amino Acid Consequence	MIC[a]
PR07	*bas5034* A425G	cell division ABC transporter, FtsE	G142E	8
PR08	*bas5034* G270T	cell division ABC transporter, FtsE	F90L	32
PR09-1	*bas1659* G190C	CitB RR[b]/luxR family transcriptional regulator	V64L	16
PR09-4	*bas1659* G190C	CitB RR[b]/luxR family transcriptional regulator	V64L	>64
	bas1662 A638G	ABC transporter permease	H213R	
PR10-4	*bas1659* C248T	CitB RR[b]/luxR family transcriptional regulator	T83M	64
	bas1663 C1127T	ABC transporter permease	A376V	
	bas1842 A43G	petrobactin biosynthesis *asbE*	S15G	

[a] MIC as determined by microbroth dilution, measured in μg/mL
[b] Response regulator

In a second strategy to obtain PZN-resistant mutants, we cultured *B. anthracis* Sterne Δ*bas4114-4117* in the presence of a sub-lethal concentration of PZN. The concentration of PZN was gradually increased with the number of passages.[52] We isolated genomic DNA from a 1st passage strain (PR09-1, MIC 16 μg/mL) in addition to two independent 4th passage strains (PR09-4, PR10-4, MICs ≥64 μg/mL) for whole genome sequencing. PR09-1 contained a point mutation in *bas1659*, which is annotated as a CitB-like response regulator (Table B.8). Downstream of *bas1659* are genes encoding a predicted histidine kinase (*bas1660*), ABC transporter subunits (*bas1661-1663*), and a cardiolipin (CL) synthase gene (*bas1664*). PR09-4 is a descendent of PR09-1, and as such, PR09-4 contained the same *bas1659* mutation as PR09-1 in addition to another point mutation in *bas1662* (the permease domain of a locally-encoded ABC transporter). PR10-4 contained a similar mutation series (*bas1663*, a second permease gene for what is presumably a trimeric ABC transporter) but had an additional mutation in *bas1842*, which is implicated in

petrobactin biosynthesis.[53] However, deletion of the petrobactin biosynthetic gene cluster did not decrease susceptibility to PZN (Table S1).

Cardiolipin increases sensitivity to PZN

We hypothesized that the regulatory- and transport-related mutations upstream of the gene encoding CL synthase could alter CL concentrations and thus, CL may be playing a role in PZN's activity. We first examined the effect of exogenous CL on the interaction of PZN with the *B. anthracis* cell membrane. *B. anthracis* cells were treated with PZN-Cy5 (1 nM) in the presence and absence of exogenous CL (up to 100 µg/mL). Cell-associated PZN was then quantified by flow cytometry. The extent of PZN-Cy5 binding to *B. anthracis* was significantly increased when cells were co-treated with CL (Figure B.15). This result is in contrast to that with daptomycin, which acts on the bacterial membrane but exhibits an antagonistic relationship with CL in Enterococci.[54, 55] As predicted, co-administration of CL did not increase the labeling efficiency of daptomycin-Cy5 on *B. anthracis* cells (Figs. S1, 5, S14).[54] Congruent with these data was the observation that CL potentiated the killing activity of PZN towards *B. anthracis* without being cytotoxic at the concentrations investigated here. Indeed, the strongly synergistic behavior with CL enhanced the potency of PZN upwards of 16-fold while CL alone had no antibiotic activity at the concentrations tested (Figure B.15).

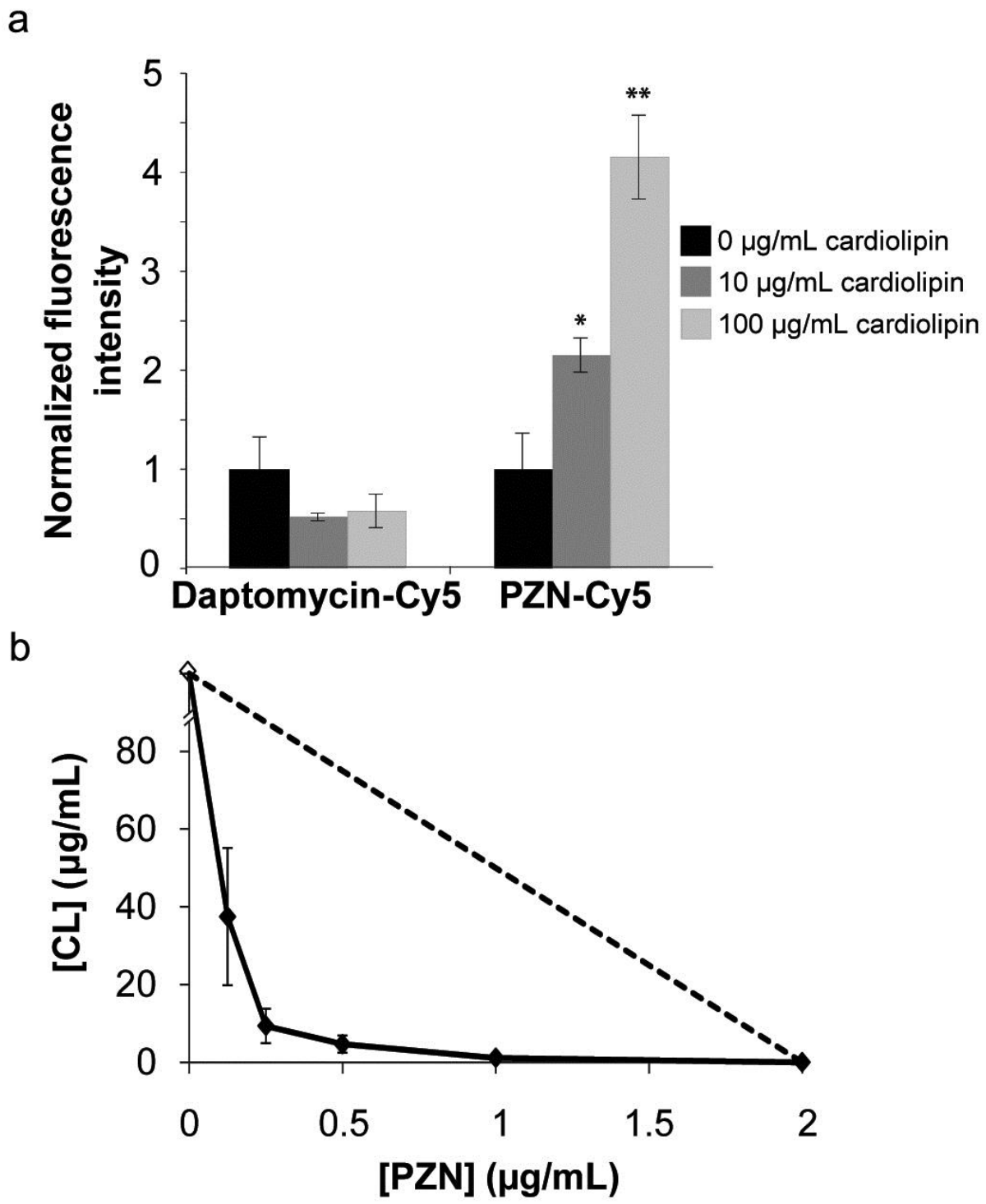


Figure B.15. Cardiolipin increases PZN-Cy5 interaction with bacterial cells. (A) Mean fluorescence intensities of *B. anthracis* cell populations treated with daptomycin-Cy5 or PZN-Cy5 in the presence or absence of exogenous CL were determined by flow cytometry and normalized to cells treated only with the Cy5-labeled compound and vehicle. Error is reported as standard deviation with n = 3. The PZN-Cy5 p-values are given relative to the 0 µg/mL control with * indicating <0.01 and ** indicating <0.001. The

p-values for daptomycin-Cy5 were both >0.01. (B) Isobologram of the minimum inhibitory concentrations (μg/mL) of PZN and CL. Interactions taking place below the dotted line represent synergistic behavior.

The proportion of CL in cell membranes has been reported to increase during growth in high osmolarity medium, especially for *B. subtilis*.[56] We thus tested whether increasing the osmolarity of the *B. subtilis* medium (and thus the CL content) would induce susceptibility to PZN. When grown in standard LB supplemented with an additional 1.5 M NaCl (final 1.67 M NaCl), PZN was weakly growth-suppressive towards *B. subtilis* (Table B.9). By measuring CL from total lipid extractions, CL levels did increase compared to standard growth in LB (Table B.9). However, exogenous cardiolipin alone does not induce PZN susceptibility in *B. subtilis* (data not shown). Unfortunately, members of *B. cereus sensu lato*, including *B. anthracis*, are not tolerant of high osmolarity media.[57] Therefore, induced susceptibility to PZN under high osmolarity conditions could not be evaluated for these strains.

Table B.9. Effect of *B. anthracis* Sterne mutations and *B. subtilis* growth conditions on PZN susceptibility and cardiolipin content of bacterial membranes

Strain	**MIC (μg/mL)**	**% CL[a]**
B. anthracis Sterne 7702	1	4.6 ± 0.1
B. anthracis Sterne 7702 *Δbas4114-4117*	1	3.6 ± 0.5
B. anthracis PR09-4	>64	6.1 ± 2.5
B. anthracis PR10-4	64	24.2 ± 1.6*
B. subtilis 168	>128	27.5 ± 2.7
B. subtilis 168[b]	32	37.0 ± 6.1

[a] Error is given as standard deviation with n = 3. P-values were compared to either a parental strain (Δ*bas4114-4117* for PR10-4) or the same strain grown in a normal osmolarity medium with * indicating <0.0001

[b] Luria-Bertani broth supplemented with 1.5 M NaCl

Cell survival in conditions of increased osmolarity is dependent on membrane fluctuations with an increase in unsaturated fatty acid composition.[56] Increased CL levels are characteristic of osmotic stress in *B. subtilis*, *Escherichia coli*, and *Lactococcus lactis*, among others.[56] Furthermore, excess CL within the membrane results in increased fluidity and lipid bilayer deformation.[58] Regulation of the lipid desaturase, Des, influences the composition of phospholipids within the membrane.[59] We predict that

PZN induces stress in the cell membrane and induces the *des* two-component system, potentially affecting membrane fluidity.

As stated previously, the stepwise-selected PZN-resistant mutants accumulated mutations in genes upstream of the CL synthase gene. We next analyzed the transcriptional response within the *cls* locus, including the upstream regulators/transporters and observed a dramatic increase in the expression of several *cls* genes in PR09-1 and PR09-4, including CL synthase itself and the transporters, but not the predicted response regulator and histidine kinase genes (Table S13). In contrast, there was no differential expression of any tested gene in the PR10-4 strain. Upon CL quantification, we did not detect differences in CL content between *B. anthracis* and the Δ*bas4114-4117* strain; however, only mutant PR10-4 contained a proportionally increased level of CL (Table B.9). This suggests that CL does play a role in the activity of PZN, although there is not a straightforward relationship between CL content composition and PZN susceptibility. Based on our findings, we expect that in addition to CL, other membrane-associated biomolecules contribute to the ability of PZN to destabilize *B. anthracis* cell membranes.

PZN colocalizes with cardiolipin and regions of increased fluidity

The genetic and functional association with CL implicates the membrane as the most probable target for PZN. The lipid dye 10-N-nonyl acridine orange (NAO) approximates regions of the cell membrane enriched in CL.[60] In *B. subtilis*, NAO organizes into distinct foci at the septa and the poles,[61] but it appears that in *B. anthracis* Sterne, NAO labels distinct foci throughout the entirety of the cell membrane (Figure B.16). CL has the potential to dramatically alter membrane architecture and may contribute to the susceptibility of *B. anthracis* through punctate localization throughout the cell. We therefore equilibrated *B. anthracis* cells with PZN-Cy5 and NAO to investigate if PZN localized to CL-rich regions in the cell membrane. There existed a clear but imperfect colocalization of the two dyes, suggesting a possible interaction with CL in the bacterial membrane (Figure B.16). Thus, *B. anthracis* appears to have a unique distribution of CL that facilitates an interaction with PZN and leads to cell death, whereas CL localization within other species may not facilitate the lytic activity of PZN. Additionally, 1,1'-Didodecyl-3,3,3',3'-tetramethylindocarbocyanine perchlorate, (DilC12(3)), is a dye reported to

associate with regions of increased fluidity (RIF) within cell membranes of *B. subtilis*, and may also be indicative of CL localization.[62] RIFs are transiently weakened regions within the bacterial membrane that affect lipid homeostasis and membrane fluidity. We observed colocalization of DilC12(3) and PZN-Cy5, consistent with PZN and CL co-associating with *B. anthracis* RIFs (Figure B.16).

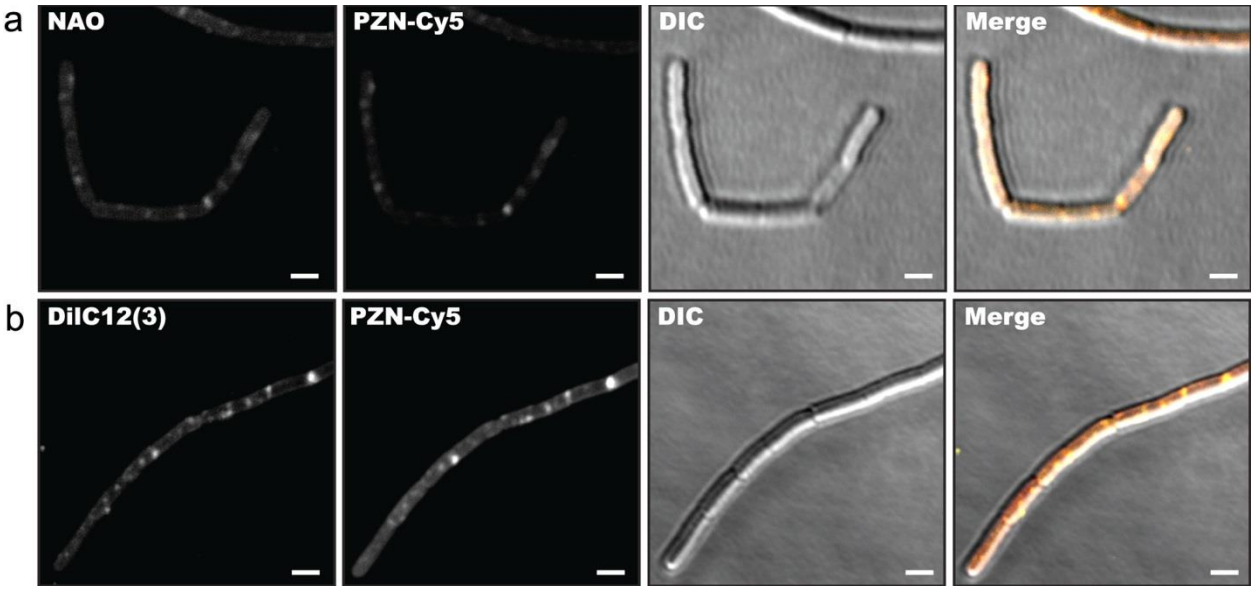


Figure B.16. PZN colocalizes with CL- and RIF-specific dyes. (A) From left to right, the channels for NAO, Cy5, DIC, and a merged image are shown to illustrate the co-localization of NAO (yellow) and PZN (red). (B) Same as panel A but with DilC12(3) replacing NAO. Scale bars, 2 μm.

In this paper, we report a MOA for PZN bioactivity involving cell membrane depolarization in a CL-enhanced manner. Due to the advance in genomics, *B. anthracis*, the causative agent of anthrax, can be distinguished from the other members of the *B. cereus sensu lato* group by whole genome sequencing, multilocus sequence typing, the presence of chromosomal lambdoid prophages, and the presence of a characteristic nonsense mutation in *plcR*.[16, 27] An alternative approach to *B. anthracis* identification now includes susceptibility to PZN, which is a natural product exhibiting potent and specific activity for *B. anthracis* under standard laboratory conditions. Our data demonstrate that the species selectivity of PZN is even more discriminating than that of the reputedly selective γ phage.[23, 24, 26, 27] Additionally, the compound on its own does not contribute to *B. amyloliquefaciens* FZB42 antifungal or nematicidal activity. *B. anthracis* is very similar to other members of the *B. cereus sensu lato* family as described, but strains of *B. anthracis* designed to analyze key genetic differences retain their respective susceptibility to

PZN. Gene expression analysis, together with confocal and STORM microscopy, reveals that PZN operates by a different MOA than previously described cell envelope-targeting antibiotics. Thus, we present a model for PZN activity wherein PZN takes advantage of a locally weakened cell membrane, whether due to RIFs, CL-dependent membrane deformation, or some combination thereof. PZN accumulates to such membrane defects, which results in membrane depolarization and eventual lysis of *B. anthracis* in a species-specific manner. The activity of PZN suggests homeland security applications, where it could be further developed as a rapid *B. anthracis* detection test.

Methods

Strain and growth conditions

All strain references are displayed in Table S1. All strains were grown in Luria-Bertani (LB) broth unless otherwise described (10 g/L tryptone, 5 g/L yeast extract, 10 g/L (0.34 M) NaCl). Biosafety level 3 strains of *B. anthracis* were grown on Mueller-Hinton agar. *Neisseria* strains were grown in gonococcal medium base supplemented with Kellogg's I and II.[63] *Streptomyces* endospores were isolated on mannitol soybean flour agar (20 g/L mannitol, 20 g/L soybean flour, 1.5% agar) and used to determine PZN susceptibility in ISP2 (4 g/L yeast extract, 10 g/L malt extract, 4 g/L dextrose). Yeast strains were grown in YPD medium (10 g/L yeast extract, 20 g/L peptone, 20 g/L dextrose). *C. elegans* was cultured on nematode growth medium with *E. coli* OP50. Cultures were supplemented with 1.25 mM $CaCl_2$ when assaying daptomycin susceptibility. In cases where increased osmolarity was desired, the LB was supplemented with additional NaCl (final concentration of 1.84 M).

PZN Bioactivity

PZN and Me_2-Arg-Az_5 bioactivity was determined via microbroth dilution assay as described in the Clinical and Laboratory Standards Institute manual.[64] The optical density (OD_{600}) of a stationary phase culture was adjusted to 0.01 and added to a microtiter plate containing serially diluted PZN. Wells were visually inspected for turbidity, and the minimum inhibitory concentration (MIC) was determined as the lowest compound concentration that incurred no growth after 16 h. MICs were determined in LB unless growth conditions required an alternative medium (see above). The *S. aureus* media-dependent

PZN susceptibility was analyzed using LB, Brain-Heart Infusion (BHI, Bacto), and Mueller-Hinton (BBL) broths. When indicated, cardiolipin (CL) was added to the medium at 100 μg/mL.

A *B. anthracis* growth curve was generated using Tecan Infinite M200 Pro. *B. anthracis* Sterne 7702 cultures were grown in LB to stationary phase at 37 ˚C. Cultures were diluted to OD_{600} of 0.05 with fresh LB and allowed to recover to an OD_{600} of 0.35. Cultures were aliquoted into 96-well plates containing PZN and incubated at 37 ˚C with orbital shaking. OD_{600} was measured every 2 min. Values were normalized to an initial OD_{600} of 0.35 and adjusted to a 1 cm path length. Error bars represent standard deviation of two independent experiments.

A growth curve in the presence of Me_2-Arg-Az_5 was generated as described above with the following differences: *B. anthracis* Sterne 7702 and *S. aureus* USA300 cultures were grown in duplicate to OD_{600} 1.0 and aliquoted into 96-well plates. Wells were treated with 1:1 dilutions of Me_2-Arg-Az_5 at a maximum concentration of 12 μM. The plate was incubated at 37 °C with orbital shaking and OD_{600} was measured every 2 min. Values were normalized to an initial OD_{600} of 1.0, adjusted to a 1 cm path length, and averaged at each time point.

Gamma (γ) phage sensitivity

γ phage were propagated as described previously[23] using *B. anthracis* Sterne 7702 cells on brain heart infusion (BHI) agar plates, with no visible loss in infectivity. Phage infectivity was tested against a panel of *B. cereus* and *B. anthracis* strains using a serial dilution assay. Stationary phase cultures were adjusted to an OD_{600} of 0.1 and 100 μL was plated on BHI plates. 5 μL of phage stock (2.6×10^8 plaque forming units/mL) was serially diluted (2-fold) and spotted onto the plates and allowed to dry. After incubation at 37 °C for 16 h, plates were removed and visually inspected for plaques.

RNA isolation and transcriptional profiling of PZN-treated Sterne cells

For the compound-treated samples, independent 3 mL cultures of *B. anthracis* Sterne 7702 cells were grown to an OD_{600} of 0.4, and 0.25 × MIC of PZN, 0.25 × MIC Me_2-Arg-Az_5,[13] or an equivalent volume of DMSO was added and allowed to incubate for 10 min at 37 °C. Together with resistant mutant PR06, RNA was isolated and prepared as described previously [4]. RNA-Seq libraries were created using

the TruSeq Stranded RNA Sample Prep kit (Illumina, San Diego, CA) after rRNA depletion using the RiboZero Bacteria kit (Epicentre, Madison, WI). Sequencing was performed for 1 × 100 cycles on a HiSeq 2000 with Version 3 Chemistry. Transcriptomic data was processed with the Rockhopper version 1.30 pipeline[65] using *B. anthracis S*terne and *B. anthracis* Ames Ancestor plasmid pXO1 (NC_007322.2) as references. Default values (allowed mismatches 0.15, minimum seed length 0.33, minimum expression of UTRs and ncRNAs 0.5) were used, with the exception that reverse complement reads were used for mapping. The RNA-Seq data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and is accessible through GEO Series accession number GSE73343 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE73343).

Membrane depolarization

Three independent stationary phase cultures of *B. anthracis* Sterne 7702 were used to inoculate fresh LB and grown to OD_{600} 0.5 at 37 ˚C with shaking. Aliquots (10 µL) were diluted to 1 mL in PBS containing 3 mM $DiOC_2(3)$ (3,3'-diethyloxacarbocyanine iodide) and compounds (DMSO - vehicle, 5 µM carbonyl cyanide *m*-chlorophenyl hydrazone (CCCP), 3.0 µM daptomycin, 0.5 µM PZN, 1.0 µM PZN). Cells were mixed at 21 °C for 30 min prior to analysis by flow cytometry (BD LSR II Flow Cytometry Analyzer). Voltages for fluorescein isothiocyanate (FITC) and propidium iodide (PI) fluorescence were set so that average counts per cell were between 10^3 and 10^4. Geometric means for fluorescence ratios were normalized to the control $DiOC_2(3)$ samples.

Confocal microscopy

In general, cells were prepared by inoculating 5 mL LB with 200 µL of a stationary phase culture. After growing to an OD_{600} of 0.5 at 37 °C with shaking, 1 mL aliquots were centrifuged (3 min, 8000 × *g*), decanted, and resuspended in sterile PBS. Slides were prepared by mixing 1:1 (v/v) cell suspensions in PBS and liquefied low gelling temperature agarose (Sigma-Aldrich, 2% w/v in water). All microscopy images were obtained using a Zeiss LSM 700 Confocal microscope with a 63×/1.4 Oil DIC objective and processed using Zen 2012 software. Laser intensity and gain were kept at a minimum and held constant for all experiments. Linear contrast was equally applied during image processing. To localize PZN, *B.*

anthracis Sterne 7702 was treated in PBS with 0.2 μM PZN-Cy5 for 30 min at 22 °C. After washing in PBS (3 × 500 μL), cells were resuspended in a final volume of 250 μL PBS. Competition experiments were performed using PR06 (PZN-resistant) in PBS treated with DMSO (vehicle) or 1 μM PZN for 20 min at 22 °C before the addition of 0.05 μM PZN-Cy5. After 20 min at 21 °C, the cells were washed in PBS (3 × 500 μL) and resuspended in a final volume of 250 μL PBS. Sterne underwent co-treatment in PBS with 0.2 μM PZN-Cy5 for 25 min before the addition of other fluorescent compounds. After 5 min additional treatment, cells were washed in PBS (5 × 500 μL) and resuspended in a final volume of 250 μL PBS. Concentrations used: NAO (Sigma-Aldrich), 1 μM; Dil-C12, 1 μM; BODIPY-vancomycin, 1 μM; Bocillin-FL, 1 μM. For CL experiments, cells were treated with EtOH (vehicle), 10 μg/mL CL, or 100 μg/mL CL in addition to 0.2 μM PZN-Cy5 for 30 min.

Super-resolution microscopy (STORM)

Cells for 3D super-resolution microscopy were grown and treated with PZN-Cy5 as described for confocal microscopy. The cells were immobilized on Nunc® Lab-Tek® 8-well chambered coverglass (Sigma-Aldrich) coated with 0.1% (w/v) poly-L-lysine (Sigma-Aldrich). After 10 min incubation, unattached cells were removed by washing chambers with sterile PBS. Chambers were filled with 500 μL imaging buffer (10 mM NaCl, 50 mM Tris-HCl (pH 8.5), 10% w/v glucose). Immediately prior to imaging, cysteamine (Sigma-Aldrich, 10 mM final concentration), catalase (EMD Millipore, 909 U/mL), and pyranose oxidase (Sigma-Aldrich, 4.44 U/mL) were added to the imaging buffer. 3D super-resolution microscopy was performed as described previously.[66, 67] Briefly, samples were imaged using an Olympus IX-71 inverted microscope outfitted with a 100× NA 1.4 SaPo oil objective. Mechanical shutters (LS6T2, Unibliz) were used to alternatively excite the sample with a red laser (DL640-100-AL-O, Crystalaser) and reactivate Cy5 with a violet laser (405 nm, 20 mW, Spectra Physics Excelsor). The lasers were expanded by 7.5×, reflected by a dichoroic mirror (Semrock FF408/504/581/667/762-Di01-25X36), and sent to the sample chamber with a focusing lens that also creates an incidental angle slightly smaller than the total internal reflection angle, reducing the background signal while allowing illumination of several hundred nm along the z-axis. The emission signal from the sample was passed through an emission filter (Semrock

FF01-594/730-25) and two additional notch filters (Semrock NF01-568/647-25X5.0 and NF01-568U-25), and was imaged on an EMCCD camera (DV887ECS-BV, Andor Tech). A cylindrical lens (SCX-50.8-1000.0-UV-SLMF-520-820, CVI Melles Griot, 2 m focal length) in the emission beam path induced astigmatism for 3D detection.[47] ASI CRISP (Applied Scientific Instrumentation) and a piezo-objective (PI P-721.10) were used to compensate for vertical drift during data collection. The horizontal drift was corrected in the post data acquisition step by the analysis software utilizing the correlation function.[68] The data analysis software was provided by Xiaowei Zhuang[66] and modified for 3D imaging.

Selection of spontaneous PZN-resistant mutants

Spontaneous PZN-resistant mutants were generated by plating 2×10^8 *B. anthracis* Sterne 7702 cells grown to stationary phase onto a PZN plate containing 4 × PZN MIC. Surviving colonies were tested for sustained PZN resistance via microbroth dilution as described above. Resistant mutants PR01, PR02, PR05, and PR06 were subjected to genomic DNA isolation as follows: 3 × 10 mL cultures of each strain were grown to stationary phase, harvested, and resuspended in 400 µL of water. Cells were lysed with 50 µL of 10% SDS and 5 µL of 20 mg/mL RNase solution at 22 °C for 5 min. DNA was isolated via 25:24:1 phenol/chloroform/isoamyl alcohol extraction, followed by addition of 24:1 chloroform/isoamyl alcohol. DNA precipitation via cold isopropyl alcohol and a subsequent 70% ethyl alcohol wash resulted in purified gDNA.

After genetic deletion of *bas4114-4117* (described in the supporting methods), a second round of spontaneously resistant mutants to PZN were selected and isolated as above. Serial-passage mutants were isolated as previously described,[52] starting with three independent cultures of an OD_{600} of 0.1 *B. anthracis* Sterne 7702 Δ*bas4114-4117* in 0.25 µg/mL (0.25 × MIC) PZN LB. Cultures that grew were diluted to an OD_{600} of 0.1 and subjected to increased concentrations of PZN, until cultures were resistant to 64 µg/mL. Cultures were serially passaged onto PZN-free medium to confirm mutant stability. Genomic DNA was isolated as described above. All mutants derived from the *Δbas4114-4117* deletion strain were sequenced as described and assembled via CLC Genomics Workbench and SNP analysis was performed with Mauve version 2.3.1.

Whole genome sequencing and assembly

Genomic libraries for resequencing were prepared using the TruSeq DNAseq Sample prep kit (Illumina, San Diego, CA). Sequencing was performed on a HiSeq 2000 with Version 3 Chemistry for 1×100 cycles. SNP and DIP discovery was performed with two different methods. Regarding PR02, PR05, and PR06, CLC Genomics Workbench SNP and DIP discovery pipelines were employed using with the publicly available *B. anthracis* str. Sterne genome NC_005945.1 as a reference. PR01 required de novo assembly with IDBA UD version 1.0.9, followed by whole genome alignment and SNP discovery using Mauve version 2.3.1. Resistant mutants PR03 and PR04 were selected separately and underwent Sanger sequencing after PCR amplification of *bas4114* and sequencing with the BamHI-BAS4114-f primer (Table S2). The WGS data discussed in this publication have been deposited in NCBI's GenBank and are accessible via BioProject accession number PRJNA295544. Within this BioProject are individual accession numbers for each *B. anthracis* strain (taxId:1392) for which whole genome sequencing was performed: CP012720, PR01; CP012721, PR02; CP012722, PR05; CP012723, PR06; CP012724, PR07; CP012725, PR08; CP012726, PR09-1; CP012727, PR09-4; CP012728, PR10-4; CP012730, Parent1 (for PR01 through PR06); CP012729, Parent2 (for PR07 through PR10-4.

Effect of cardiolipin on fluorescence intensity

Three independent stationary phase cultures of *B. anthracis* Sterne 7702 were used to inoculate fresh LB (200 µL into 5 mL LB) and the new cultures were grown to OD_{600} 0.5 at 37 ˚C with shaking. Samples were prepared by diluting 10 µL aliquots of culture to 1 mL in PBS containing 1 nM PZN-Cy5 and vehicle (EtOH), 10 µg/mL CL (Sigma Aldrich), or 100 µg/mL CL. After mixing at 22 °C for 30 min, cells were analyzed by flow cytometry as described above for differences in PZN-Cy5 fluorescence intensity. Geometric means were normalized to the control samples.

Cardiolipin quantification from total lipid extracts

Cultures of *B. anthracis* Sterne 7702, *B. anthracis* Δ*bas4114-4117*, *B. anthracis* PR09-4, *B. anthracis* PR10-4, *B. subtilis* 168, *E. faecium* U503, and *S. aureus* USA300 (three independent 10 mL cultures for each strain) were grown for 20 h at 37 ˚C. LB containing an additional 1.5 M NaCl was

inoculated with 200 µL aliquots of stationary phase cultures of *B. subtilis* 168, *E. faecium* U503, and *S. aureus* USA300 (three independent 10 mL cultures for each strain) and grown for 40 h at 37 °C. The cells were harvested by centrifugation (4000 × *g*, 10 min, 4 °C) and resuspended in 5 mL 2:1 $CHCl_3$:MeOH and 1.25 mL PBS, and then extracted for 1 h at 22 °C. The supernatant was removed after centrifugation (4000 × *g*, 10 min, 4 °C) and layers were washed with 1 mL $CHCl_3$ and 1 mL PBS. The organic layer was removed and dried by speed vacuum. The crude lipids were redissolved in 200 µL $CHCl_3$ and transferred to microfuge tubes, then dried again. The lipids were then dissolved in 20 µL $CHCl_3$ and separated by preparative TLC (Analtech Silica Gel G) using 2:1 $CHCl_3$:MeOH. The CL-containing spots were removed and extracted with 200 µL 2:1 $CHCl_3$:MeOH for 15 min with shaking. The extracts were dried by speed vacuum and redissolved in 20 µL $CHCl_3$. Each sample was spotted (0.5 µL) onto Merck Silica Gel 60 F_{254} analytical TLC plates and separated using 80:20:5 $CHCl_3$:MeOH:AcOH. The plates were developed in iodine and imaged using a Bio-Rad ChemiDoc XRS+. ImageJ was used to subtract background and measure spot density to determine percent CL out of total lipid content.

Additional Methods

PZN production

Stationary phase cultures of RSpMarA2 (Δ*sfp*, *yczE*, *degU*)[4] were grown in LB with 7 µg/mL kanamycin and 7 µg/mL chloramphenicol. Sterilized aluminum sheet cake pans were prepared with M9 agar medium supplemented with BME vitamin mix and ATCC trace mineral solution and the aforementioned antibiotics. Cake pans were inoculated with 1.5 mL stationary phase RSpMarA2 and incubated 48 h at 37 °C. Bacterial lawns were loosened from the agar and resuspended in Tris-buffered saline (TBS) (150 mL/pan). The bacteria were then harvested via centrifugation (11,000 x *g*, 20 min). The supernatant was decanted and the cell pellets were stored at -20 °C until extraction.

PZN purification

PZN was extracted from bacterial pellets by resuspending in 150 mL MeOH/tray of cells with intermittent vortexing for 15 min at 22 °C. The cells were harvested by centrifugation as above. The supernatant was vacuum filtered using Whatman filter paper, rotary evaporated, and subsequently

lyophilized to dryness. The dried, crude PZN was resuspended in MeOH (1 mL/tray of extract) and centrifuged to remove insoluble debris (4,000 × *g*, 10 min). The supernatant was then injected onto a RediSep R_f High Performance 15.5 g HP C_{18} cartridge (Teledyne Isco) and purified by MPLC using a Combiflash Rf 200 system (25-100% MeOH/10 mM aqueous NH_4HCO_3 over 120 column volumes). The fractions containing PZN were pooled, rotary evaporated, and lyophilized to dryness, yielding roughly 5 mg/tray. PZN was dissolved in DMSO at concentrations of 30-50 mg/mL and stored at -80 ˚C for later use.

Bacterial endospore preparation and susceptibility screening

B. anthracis Sterne 7702[2] endospores were prepared as described previously.[3] Briefly, *B. anthracis* Sterne 7702 cells were incubated overnight in LB at 37 °C. Difco Sporulation Media agar plates (8 g Nutrient Broth, 1 g KCl, 0.25 g $MgSO_4 + 7H_2O$, and 17 g agar, per liter) was inoculated with 150 µL of the stationary phase culture and incubated at 30 °C for 5 days. Lawns were recovered by resuspension in sterile H_2O and filtered using 3.1 and 1.2 µm filters to remove vegetative cells and aggregated endospores. Endospore filtrate was incubated at 65 °C for 1 h to remove any remaining vegetative cells. Endospores were harvested by centrifugation at 8000 × *g* for 25 min at 4 °C. After decanting the supernatant, the endospores were washed 3 times with 40 mL of H_2O, harvesting by centrifugation. Endospores were stored at 4 °C in H_2O and quantified by serial dilution. Endospores were screened for PZN activity by incubating 1×10^9 endospores with PZN (0-16 µg/mL) in H_2O for 20 h at 37 °C. Endospores were subsequently incubated at 65 °C for 1 h to destroy any contaminating vegetative cells, and then 4 °C for 4 h to recover, and followed by serial dilution onto LB agar plates to assess germination. Plate counts represent the average of two trials.

***C. elegans* nematicidal assays**

Wild-type N2 *C. elegans* were cultured on *E. coli* OP50 using standard techniques.[4] Eggs were isolated from gravid hermaphrodites using standard bleaching protocols[5] and incubated for 18 h at 20 °C to obtain synchronized L1 larvae. Cultures were then transferred to new NGM plates seeded with OP50 *E. coli* and incubated for 48 h at 25 °C. Resulting L4 larvae were used for assays adapted from Liu *et al*.[4]

Slow killing assays were performed using 12-well cell culture dishes with wells containing 1 mL NGM agar amended with 64 μg/mL PZN dissolved in DMSO or an equivalent volume (2 μL) of DMSO alone. One half of the slow killing assay wells were also seeded with 10 μL of OP50 *E. coli*. The effect of DMSO alone on mortality was also tested and found to be negligible. Subsequently, 40-60 L4 hermaphrodites were transferred to each well, incubated at 25 °C and checked every 24 h for 3 days. Each treatment was performed with two biological replicates, each having three technical replicates. Liquid fast killing assays were also performed using 12-well cell culture dishes with wells containing 1 mL M9 buffer combined with 64 μg/mL PZN dissolved in DMSO or DMSO alone. 40-60 L4 larvae were transferred into each well and incubated at 25 °C for 24 h. Fast killing assays were replicated similarly to the slow killing assay (two biological, three technical replicates). For the duplicate assay, an additional control of M9 buffer was included. In both fast and slow killing assays, animals were scored according to criteria shown in Liu *et al*. Mortality rates were also calculated using identical ratios.

cDNA construction and qRT-PCR analysis

cDNA was generated using the published protocol from the Promega Improm-II system, replacing the reverse transcriptase with Invitrogen m-MLV reverse transcriptase. qRT-PCR was performed with iTaq universal SYBR Green Supermix and following manufacturer's instructions using a Roche Lightcycler 480 System. Fold changes upon PZN treatment were calculated as described previously.[7]

Compound preparation

Me_2-Arg-Az_5. Me_2-Arg-Az_5 was synthesized as previously described.[7]

PZN-Cy5. PZN (500 μg, 3.74 x 10^{-7} mol) in 20 μL dimethylformamide (DMF) was mixed with 5 μL *N,N'*-diisopropylethylamine (DIPEA), 10 μL of a 190 mM solution of 1-hydroxybenzotriazole (HOBt) in DMF, and 10 μL of a 190 mM solution of 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) in DMF for 15 min at 21 °C. Then 29.3 μL of a 10 μg/μL solution of Cy5 amine (Lumiprobe) in DMF was added. The reaction was protected from light and stirred at 22 °C for 24 h. The sample was dried by speed vacuum, redissolved in 5% MeOH, and purified on a Sep-Pak C_{18} cartridge (Waters) on a

gradient from 0 – 100% MeOH. Fractions containing PZN-Cy5, as determined by MALDI-TOF mass spectrometry, were combined and their purity assessed by analytical HPLC (Thermo BetaSil C_{18} column [250 mm × 4.6 mm], 40 – 95% MeOH). Concentration was determined using the extinction coefficient of Cy5 in DMSO (ε_{646} = 250,000 L cm^{-1} mol^{-1}).

Daptomycin-Cy5. A 1 mg/mL solution of daptomycin in DMSO was prepared and 33 µL was mixed with 2 µL of 10 mM Cy5 NHS ester (Lumiprobe) and stirred at 22 °C, protected from light, for 4 h. Labeled compound was purified by reverse-phase chromatography on a Perkin Elmer Flexar HPLC outfitted with a Thermo Scientific BetaSil C_{18} column (250 mm × 4.6 mm I.D., 5 µm particle size) using 10 mM NH_4HCO_3 as the aqueous phase on a 45 min gradient from 25 – 95% MeCN at a flow rate of 1 mL/min. Absorbance monitoring at 647 nm was used to identify the fraction containing daptomycin-Cy5, which was confirmed by MALDI mass spectrometry. Purity was determined by a second analytical HPLC run using the same solvents on a 30 min gradient from 40 – 95% MeOH. Concentration was determined using the extinction coefficient of Cy5 in DMSO.

Biotin-PZN (N-terminal label). A 100 mM solution of EZ-link *N*-hydroxysuccinimide (NHS) Biotin (Thermo Scientific) in 80% MeCN, 10 mM MOPS (pH 8.0) was prepared. DesmethylPZN (200 µg, 1.53 × 10^{-7} mol) was dissolved in 24 µL of the same buffer and treated with 6 µL EZ-link NHS biotin solution for 4 h at 22 °C. The solvent was removed by speed vacuum and the sample was dissolved in MeOH for purification by HPLC (BetaSil C_{18} column, 250 mm × 4.6 mm, 1 mL/min, 50 – 95% MeOH over 60 min, 10 mM NH_4HCO_3 aqueous phase, monitored at 254 nm). Fractions containing purified biotin-PZN were combined, dried, and redissolved in DMSO. *B. anthracis* Sterne 7702 susceptibility was determined to be >32 µg/mL using a standard microbroth dilution assay.

PZN-Biotin (C-terminal label). PZN (500 µg, 3.74 × 10^{-7} mol) was dissolved in 10 µL DMF and 1 µL DIPEA was added to the sample. A solution of 2.0 mM EDC and 2.0 mM HOBt in DMF was prepared and 2 µL was added to the PZN. After 5 min stirring, 1.0 mg biotin cadaverine was added to the reaction and allowed to proceed at 22 °C for 30 h. The reaction was dried by speed vacuum, redissolved in MeOH, and purified by HPLC (BetaSil C_{18} column, 250 mm × 4.6 mm, 1 mL/min, 50 – 95% MeOH

over 60 min, 10 mM NH_4HCO_3 aqueous phase, monitored at 254 nm). Fractions containing purified PZN-biotin were combined, dried, and redissolved in DMSO. *B. anthracis* Sterne 7702 susceptibility was determined to be 32 µg/mL by a standard microbroth dilution assay.

PZN-diazirine-alkyne. PZN (6.68 mg, 5.0×10^{-6} mol) was dissolved in 75 µL DMF and 1.4 µL Et_3N was added with stirring. The sample was cooled in an ice bath to 0 ˚C. HOBt (0.95 mg, 0.007 mmol) and EDC (1.1 mg, 0.007 mmol) were dissolved in 100 µL DMF and added to the PZN. 2-(2-azidoethyl)-2-(but-3-ynyl)-1,3-dioxolane (0.96 mg, 0.007 mmol)[50] was dissolved in 200 µL DMF and 24 µL was added after 10 min at 0 ˚C with stirring, protected from light. The reaction was allowed to warm to 22 ˚C and allowed to proceed for 24 h. The solvent was removed by speed vacuum and the sample was redissolved in MeOH. The sample was purified by HPLC (BetaSil C_{18} column, 250 mm × 4.6 mm, 1 mL/min, 50 – 95% MeOH over 60 min, 10 mM NH_4HCO_3 aqueous phase, monitored at 254 nm). Fractions containing purified PZN-diazirine-alkyne were combined, dried, and redissolved in DMSO. *B. anthracis* Sterne 7702 susceptibility was determined to be 32 µg/mL using a standard microbroth dilution assay.

Structural characterization of PZN derivatives

Purified samples were dried by speed vacuum and dissolved in 50% MeCN supplemented with 1% (v/v) acetic acid. The diluted samples were directly infused using an Advion Nanomate 100 to an LTQ-FT hybrid linear ion trap-FTMS system (ThermoFisher) operating at 11 T. The MS was calibrated weekly using calibration mixture following the manufacturer's instructions, and tuned daily with Pierce LTQ Velos ESI Positive Ion Calibration Solution (ThermoFisher). Spectra were collected in profile mode with a resolution of 100,000. The singly charged ions were targeted for CID using an isolation width of 5 *m/z*, a normalized collision energy of 35, an activation q value of 0.4, and an activation time of 30 ms. Data analysis was performed using Thermo Xcalibur software.

Affinity purification using PZN-Biotin

LB (200 mL) was inoculated with 10 mL stationary phase *B. anthracis* Sterne 7702. The culture was grown with shaking at 37 °C to OD_{600} 0.5. The sample was divided into 100 mL aliquots,

centrifuged, and individually resuspended in 1.5 mL lysis buffer (50 mM Tris, 500 mM NaCl, 2.5% v/v glycerol, 0.1% v/v Triton X-100, pH 7.5) with 500 mU mutanolysin. After equilibrating for 30 min at 22 °C, the samples were sonicated (4 × 30 s, 50% amplitude) and harvested by centrifugation (17,000 × *g*, 10 min). The insoluble fractions were resuspended in 1.5 mL 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) buffer (50 mM Tris-HCl, 150 mM NaCl, 1% w/v CHAPS, pH 7.5) and sonicated (10 s, 50% amplitude). One sample was treated with 1.0×10^{-7} mol PZN-biotin and the other was treated with vehicle (DMSO). The samples were treated for 30 min at 22 °C before the addition of 100 µL streptavidin-agarose resin suspension (pre-equilibrated with Tris buffer, 50 mM Tris-HCl, 150 mM NaCl, pH 7.5). After mixing for 3 h at 22 °C, samples were applied to Bio-Rad spin columns and the flow through was collected. The resin was washed (5 × 4 mL Tris buffer). The resin was then resuspended in 200 µL 1% SDS in PBS and boiled for 10 min. After cooling, the eluent was collected and analyzed by SDS-PAGE for the presence of unique bands in the sample containing PZN-biotin. Unique bands, as visualized by Coomassie and silver stain, were not found compared to the vehicle-treated sample (data not shown). Samples were also analyzed by Western blot using the following procedure: proteins were transferred by electroblot to a polyvinylidene fluoride (PVDF) transfer membrane (EMD Millipore). The membrane was blocked overnight at 4 °C in TBS containing 0.1% (v/v) Tween-20 (TBST buffer), treated for 1 h at 4 °C with 1/3,000 dilution of streptavidin-HRP in TBST, and then washed (2 × 30 s then 5 × 5 min) with TBST. After washing once in PBS (1 × 30 s), the membrane was treated with 1:1 hydrogen peroxide solution:luminol solution (Bio-Rad) for imaging.

Photoaffinity purification using PZN-diazirine-alkyne

The procedure used for photoaffinity purification studies was similar to that of affinity purification experiments, except that the samples were incubated with PZN-diazirine-alkyne or vehicle (DMSO) for 4 h at 4 °C, protected from light. The cells were then exposed to UV (4 W, 365 nm) for 15 min at 22 °C to induce crosslinking. The samples were pelleted and resuspended in PBS containing 1 mM $CuSO_4$, 128 µM tris(3-hydroxypropyltriazolylmethyl)amine, 1.2 mM sodium ascorbate, and 50 µM biotin-azide for 1 h at 22 °C. The cells were then centrifuged (4,000 × *g*, 2 min), the supernatant was

decanted, and the cells were resuspended in PBS treated with 200 U/mL mutanolysin. After incubation for 1 h at 37 °C, cells were sonicated as previously described. The insoluble fraction was separated by centrifugation and resuspended as previously described. Clarified lysates and insoluble fractions were all enriched with streptavidin-agarose resin as previously described, and eluents were compared by SDS-PAGE followed by Western blot and mass spectrometric analysis.

Macromolecular synthesis assay

Radiolabel incorporation into cellular macromolecules was carried out as previously described.[8] Briefly, identical *B. anthracis* Sterne 7702 cultures were grown to an OD_{600} of 0.6 in Luria Broth and diluted to a final OD_{600} of 0.3. Radiolabelled precursor was added to a final concentration of 0.1 μCi, and compound (either PZN or control) was added to 1×, or 2× the MIC. Samples were taken at 60 min post-compound addition. OD_{600} at each time point was determined using identically treated cultures lacking radioactive compound. The macromolecules of the cultures treated with radiation were precipitated using trichloroacetic acid on glass filter disks and washed with dilute acid and ethanol successively. The filters were placed in scintillation vials with Ultima Gold scintillation cocktail. Radioactive counts were determined using a liquid scintillation analyzer (PerkinElmer Tri-Carb 2910 TR). The macromolecules, radiolabelled precursors, and their corresponding control antibiotics were as follows:

cell wall: [1,6-^{3}H (N)] N-acetyl-D-glucosamine (ARC, ART 0142), vancomycin control protein: [U-^{14}C] L-amino acid mixture (MP, 1014750), chloramphenicol control

RNA: [^{14}C(U)] uridine (ARC, ARC 0154), rifampicin control

fatty acid: acetic acid [1-^{14}C] sodium salt (ARC, ARC 0101A), triclosan control

2D Projection analysis

2D projection analysis was used to determine spatial localization of PZN on *B. anthracis* cells. A total of 11 cells imaged by super-resolution microscopy were aligned lengthwise along the y-axis, and then sectioned to remove poles and septa. Cells were projected onto the XZ plane and divided into units of 30 × 30 nm^2. Within each unit, spot density was determined and color-mapped. The probability of

finding a dye molecule a certain distance R from the y-axis was calculated for radial windows with a 20 nm bin size and normalized to the area of the radial window.

PZN-Cy5 cluster analysis

A density-based clustering analysis algorithm, DBSCAN,[9] was used as previously reported[10, 11] to analyze super-resolution images of PZN-Cy5-treated *B. anthracis*. Briefly, spots in super-resolution images of 14 cells were grouped into clusters based on spatial density. The required *Npts* and *Eps* parameters were set to 19 and 40 nm, respectively, and used to identify core points in high density spots. Parameter values empirically set such that core points within a cluster were within *Eps* distance of each other and surrounded by at least *Npts* points. Cluster borders were defined by points located with *Eps* distance to any core point. Cluster size was calculated as twice the average distance between the cluster center and every point in the cluster.

Genetic deletion of *bas4114-bas4117*

Markerless genetic deletions were created in *B. anthracis* Sterne 7702 as described previously.[12] The 500 base pairs upstream and downstream the bas4114-4117 gene cluster were cloned into the homologous recombination vector pBKJ236 using Gibson cloning (New England Biolabs) with the BamHI and NotI restriction sites. The pBKJ236 constructs were transformed into the *E. coli dam dcm* strain SCS110. Overnight cultures of the vector-containing SCS110 strain grown in LB containing 500 µg/mL erythromycin were used for conjugation along with overnight cultures of the conjugation helper strain *E. coli* SS1827 grown in LB with 200 µg/mL ampicillin and *B. anthracis* Sterne 7702 grown in BHI. From the stationary phase cultures, 400 µL were removed and washed twice by pelleting and suspending in 500 µL of LB to remove residual antibiotic. The pellets were resuspended in 200 µL of LB, then mixed thoroughly in equal volumes, and 150 µL of the mixture was inoculated onto a BHI agar plate without spreading. The inoculum was allowed to dry and the plate was incubated at 22 °C for 24 h. The entire growth was then carefully removed from the plate, resuspended in 200 µL of LB, and spotted onto a BHI agar plate containing 5 µg/mL erythromycin and 60 U/mL polymyxin B (BHIep). The culture was allowed to dry and then subsequently plated to achieve single colonies. The plate was incubated at 22 °C

until single colonies of *B. anthracis* Sterne 7702 were visible (~48 h). A single colony was used to inoculate a 2 mL culture of BHI containing 5 µg/mL erythromycin, and the culture was incubated with shaking at 22 °C overnight. The saturated culture was used to inoculate a fresh culture of BHI containing 5 µg/mL at a 1:1000 dilution. This culture was incubated at 37 °C with shaking until saturation (~8 h), then 150 µL of the culture was spotted onto a BHIep plate and allowed to dry. The spot was streaked for single colonies and the plate was incubated at 37 °C overnight. A single colony was picked and used to make competent cells as previously described.[13] Briefly, a colony was used to inoculate a 1 mL culture of LB containing 0.1% glucose (LBG). The culture was incubated at 37 °C without shaking for 10 min, and then used to inoculate 25 mL of LBG in a 250 mL sealable Erlenmeyer flask. The culture was incubated with shaking at 100 rpm at 37 °C until it reached an OD_{600} of 0.20, at which point it was transferred to a 50 mL conical tube and pelleted at 4000 × *g* at 4 °C for 10 min. The spent media was discarded and the pellet was washed with ice cold electroporation buffer (EB; 10% w/v sucrose, 15% v/v glycerol, 2 mM potassium phosphate, pH 7.8) twice. The cells were resuspended in 400 µL of cold EB, transferred to a 0.4 cm gap electroporation cuvette (USA Scientific), and incubated on ice for 10 min. A plasmid encoding the I-SceI restriction enzyme, pSS4332,[14] was added as 10 µL of a 500 ng/µL stock and the cells were pulsed one time at 2.5 kV, 25 µF, 200 Ω (mean time constant of 4.4 ms). The cells were then placed on ice for 10 min before being diluted with 600 µL LBG. The cells were recovered at 37 °C for 2 h and then plated on LBG agar plates containing 25 µg/mL kanamycin. Colonies were pooled and passaged repeatedly on BHI agar plates containing 25 µg/mL kanamycin until erythromycin sensitive colonies could be isolated. Sensitive colonies with the desired genes deleted were determined by PCR amplification with the cloning primers and were then repeatedly streaked on antibiotic free BHI agar plates until kanamycin sensitive colonies were isolated. Gene deletion was confirmed by PCR amplification with gene specific primers (Fig. S13) and the strains were confirmed to be sensitive to both kanamycin and erythromycin by plating on the appropriate antibiotic containing BHI agar plates.

Supporting Tables

Table S1. Minimum inhibitory concentrations reveal the species specificity of plantazolicin (PZN)

Strain	MIC (μg/mL)	MIC (μM)	Source
***B. cereus* group**			
B. anthracis Sterne 34F2	1	0.75	S. Blanke (UIUC)
B. anthracis Sterne 7702	1	0.75	USDA
B. anthracis Sterne 7702 *Δbas4114-4117*	1	0.75	This study
B. anthracis Sterne 7SBON30	1	0.75	[15]
B. anthracis Sterne 7SBON40	1	0.75	[16]
B. anthracis Sterne 7SDG30	1	0.75	[16]
B. anthracis Sterne 7SBTR30	1	0.75	[17]
B. anthracis Sterne 7SBONTO	1	0.75	[18]
B. anthracis Sterne 7SBTRTO	1	0.75	[19]
B. anthracis Sterne 7SBTO30	1	0.75	[20]
B. anthracis BSL3 strains	2 – 16	1.5 – 12	USAMRIID
B. anthracis CDC 684	2	1.5	USAMRIID
B. anthracis Sterne 34F2 A0517-1	1	0.75	BEI
B. anthracis Sterne 34F2 A0517-2	2	1.5	BEI
B. anthracis Sterne 34F2 Δ*blsO*	1	0.75	[21]
B. anthracis Sterne 34F2 Δ*csaB*	1	0.75	[22]
B. anthracis Sterne 34F2 Δ*sap*	1	0.75	[20]
B. anthracis Sterne 34F2 Δ*eag*	1	0.75	[22]
B. anthracis Sterne 34F2 Δanthrose	1	0.75	C. Turnbough (U. Alabama Birmingham)
B. anthracis Sterne 34F2 Δ*bclA*	1	0.75	[23]
B. anthracis Sterne BA850 (Δpetrobactin)	1	0.75	[24]
B. anthracis Sterne BA781	1	0.75	BEI
B. anthracis Weybridge	1	0.75	BEI
B. cereus 14579	>64	>48	USDA
B. cereus 541 (Δ*plcR*)	>64	>48	[4]
B. cereus ATCC 7064	>64	>48	ATCC
B. cereus ATCC 13472	>64	>48	BGSC
B. *cereus* BAG4x2-1	>64	>48	BEI
B. *cereus* E33L	>64	>48	BEI
B. cereus G9241	8	6	BEI
B. cereus GP7	>64	>48	BGSC
B. cereus Rock3-44	64	48	NMRS

B. *cereus* VD014	>64	>48	BEI
B. *cereus* VD115	>64	>48	BEI
B. mycoides 96/3308	>64	>48	BGSC
B. samanii C1	>64	>48	BEI
B. thuringiensis Konkukian 97-27	>64	>48	BEI
B. thuringiensis israelensis ATCC 35646	>64	>48	BEI
B. thuringiensis subsp. *thuringiensis*	>64	>48	BGSC
Atypical gamma phage sensitivity			
B. cereus ATCC 4342	>64	>48	BEI
B. cereus CDC 32805	>64	>48	[7]
B. cereus 2002013145	>64	>48	CDC
B. cereus 2002013146	>64	>48	CDC
B. cereus 2002013100	>64	>48	CDC
B. cereus 2000031002	>64	>48	CDC
Non *B. cereus* group			
B. amyloliquefaciens FZB42	>64	>48	BGSC
B. amyloliquefaciens B-14393T	>64	>48	BGSC
B. megaterium 899	32	24	BGSC
B. pumilus SAFR-032	>64	>48	BGSC
B. subtilis			
AKP4 (Δ*des*)	>64	>48	D. de Mendoza (U. Nacional de Rosario)
AKP21 (Δ*desRK*)	>64	>48	D. de Mendoza (U. Nacional de Rosario)
BSU-LIKE1 (Δ*liaIH*)	>64	>48	BGSC
HB0042	>64	>48	Personal Collection
HB0934 (Δ*liaGFSR*)	>64	>48	J. Helmann (Cornell)
HB5126 (Δ*liaIH*)	>64	>48	J. Helmann (Cornell)
str. 168	>64	>48	J. Wells (USCF)
Brevibacillus formosus SS 86-3	>64	>48	BGSC
Brevibacillus laterosporus ATCC 9141	>64	>48	BGSC
Neisseria meningititis serotype C	>64	>48	[4]
Neisseria sicca	>64	>48	ATCC
Staphylococcus aureus			
12608	>64	>48	P. Hergenrother (UIUC)
29213	64	48	P. Hergenrother (UIUC)
33591	64	48	P. Hergenrother (UIUC)

USA300	>64	>48	P. Hergenrother (UIUC)
Streptomyces coelicolor	>64	>48	USDA
Streptomyces lividans	>64	>48	USDA
Eukaryotic organisms			
Caenorhabditis elegans N2	>64	>48	CGC
Saccharomyces cerevisiae	>64	>48	H. Zhao (UIUC)
Talaromyces stipitatus	>64	>48	H. Zhao (UIUC)

[a]n ≥ 3 replicates
[b]Highlighted strains were tested for gamma phage susceptibility and are displayed in the main text

Table S2. PZN is not antibacterial against *B. anthracis* endospores

[PZN] (µg/mL)	CFU/mL[a]
Vehicle	9.8E+08
0.5	4.3E+08
1	8.3E+08
2	6.8E+08
4	3.4E+08
8	6.5E+08
16	3.5E+08

[a]Colony forming units per mL of endospore suspension

Table S3. MICs of representative strains demonstrating PZN activity in various rich media

	MIC (μg/mL)		
Strain	**LB[a]**	**MH**	**BHI**
Bacillus anthracis Sterne	1	1	1
Bacillus cereus CDC32805	>64	>64	>64
Bacillus cereus ATCC 4242	>64	>64	>64
Bacillus sp. Al Hakam	>64	>64	>64
Bacillus amyloliquefaciens FZB42	>64	>64	>64
Bacillus subtilis strain 168	>64	>64	>64
Enterococcus faecium U503	>64	>64	>64
Listeria monocytogenes 4b	>64	>64	>64
Staphylococcus aureus ATCC 12608	>64	16	32
Staphylococcus aureus ATCC 29213	64	8	16
Staphylococcus aureus ATCC 33591	64	8	16
Staphylococcus aureus USA300	>64	>64	>64

[a]Luria-Bertani (LB), Mueller-Hinton (MH), Brain Heart Infusion (BHI)

Table S4. Complete RNA-Seq analysis of differentially expressed *B. anthracis* genes after PZN treatment

Locus Tag	Gene	Description	Fold Change	q-value
BAS1222	*ywcJ*[a]	formate/nitrite transporter family protein	-7	3.19E-38
BAS0577	*lldp-1*	L-lactate permease	-7	2.14E-24
BAS4869	*ldh3*	L-lactate dehydrogenase	-6	6.40E-27
BAS0169		hypothetical protein	-6	1.30E-25
BAS4146		hypothetical protein	-5	3.66E-19
BAS4762	*ldh2*	L-lactate dehydrogenase	-4	1.19E-13
BAS1089	*yjzC*[a]	hypothetical protein	-4	5.38E-16
BAS1615	*yfmQ*[a]	hypothetical protein	-3	1.49E-05
BAS3917		hypothetical protein	-3	9.12E-04
BAS4690	*cydA-2*	cytochrome d ubiquinol oxidase subunit I	-3	4.40E-04
BAS0513	*bdbD*	hypothetical protein	-3	1.00E-03
BAS1942	*sdpI*[a]	hypothetical protein	2	5.98E-03
BAS3405		ahpC/TSA family protein	2	6.00E-03
BAS2363		hypothetical protein	2	1.27E-03
BAS2776	*yetG*[a]	hypothetical protein	2	9.17E-04
BAS1163		hypothetical protein	3	2.06E-03
BAS0730	*yfhC*[a]	nitroreductase family protein	3	6.85E-03
BAS4453		hypothetical protein	3	4.90E-04
BAS2568		TetR family transcriptional regulator	3	1.43E-03
BAS2566	*kynB*	hypothetical protein	3	2.83E-03
BAS0683	*yqgI*[a]	phosphate ABC transporter permease	3	7.44E-04
BAS4452		hypothetical protein	3	2.81E-05
BAS2565	*kynA*	tryptophan 2,3-dioxygenase family protein	3	3.85E-03
BAS3439	*hutG*[a]	formimidoylglutamase	3	1.17E-03
BAS3871		hypothetical protein	3	1.35E-05
BAS0214	*lagB*	invasion protein LagB	3	2.21E-04
BAS1691	*fabH2*	3-oxoacyl-ACP synthase	3	6.35E-04
BAS4502	*ytpI*[a]	hypothetical protein	3	2.71E-06
BAS1573		hypothetical protein	3	8.18E-03
BAS0849		hypothetical protein	3	1.49E-06
BAS1346	*liaF*[a]	hypothetical protein	3	1.77E-04
BAS0525	*yuaG*[a]	hypothetical protein	3	6.06E-04
BAS0610	*rocE*[a]	amino acid ABC transporter permease	3	1.77E-04

BAS4777	*sodC*	superoxide dismutase, Cu-Zn	3	6.63E-08
BAS0524	*yuaF*[a]	hypothetical protein	4	2.77E-07
BAS0848	*yhaR*[a]	enoyl-CoA hydratase	4	1.25E-07
BAS4173	*pstC*	phosphate ABC transporter permease	4	9.53E-06
BAS4172	*pstA*	phosphate ABC transporter permease	4	2.15E-07
BAS4171	*pstB*	phosphate transporter ATP-binding protein	4	7.63E-07
BAS1307	*ilvE1*	branched-chain amino acid aminotransferase	4	1.40E-08
BAS4560	*acsA*	acetyl-CoA synthetase	4	5.09E-06
BAS3772	*ylbP*[a]	hypothetical protein	4	2.05E-05
BAS0464	*rocR-1*	arginine utilization regulatory protein RocR	4	1.01E-09
BAS4877	*fadN*[a]	3-hydroxyacyl-CoA dehydrogenase	4	1.18E-09
BAS0624	*dppC*[a]	oligopeptide ABC transporter permease	4	7.90E-08
BAS4174	*phoX*	phosphate ABC transporter substrate-binding protein	4	1.87E-08
BAS5194	*fadF*[a]	ferredoxin, 4Fe-4S	4	1.32E-07
BAS1894	*dppE*[a]	oligopeptide ABC transporter substrate-binding protein	4	1.46E-08
BAS1070		hypothetical protein	4	1.73E-10
BAS4778		hypothetical protein	5	5.76E-14
BAS0681	*pstS*[a]	phosphate ABC transporter substrate-binding protein	5	1.12E-16
BAS4512		hypothetical protein	5	1.76E-24
BAS3440	*hutI*	imidazolonepropionase	5	3.49E-21
BAS2288		hypothetical protein	6	7.47E-08
BAS0625	*appB, oppB*[a]	oligopeptide ABC transporter permease	6	3.26E-20
BAS2146	*yokU*[a]	hypothetical protein	6	5.30E-39
BAS0626	*appF, oppF*[a]	oligopeptide ABC transporter ATP-binding protein	6	2.29E-19
BAS2145	*kamA*	L-lysine 2,3-aminomutase	7	4.12E-35
BAS3024		arsR family transcriptional regulator	7	1.17E-52
BAS2188	*mmgD*	citrate synthase 3	8	2.67E-54
BAS3441	*hutU*	urocanate hydratase	8	5.49E-50
BAS2649		hypothetical protein	9	5.16E-42
BAS2650		hypothetical protein	9	2.30E-45
BAS3442	*hutH*	histidine ammonia-lyase	10	9.17E-104
BAS0627	*appD, oppD*[a]	ABC transporter nucleotide-binding protein	11	1.62E-95

BAS2287	*yaaN*[a]	hypothetical protein	13	4.38E-93
BAS1521	*gerN*	germination protein gerN	28	0
BAS3456		hypothetical protein	40	0
BAS5200	*yvfU*, *desR*[a]	DNA-binding response regulator	49	0
BAS5201	*yvfT*, *desK*[a]	sensor histidine kinase	56	0
BAS5202	*yvfS*[a]	ABC transporter permease	130	0
BAS5203	*yvfR*[a]	ABC transporter ATP-binding protein	135	0
BAS1345	*liaH*[a]	PspA/IM30 family protein	1084	0
BAS1344	*liaI*[a]	membrane protein	1407	0

[a]Genes are annotated by sequence homology to the corresponding *Bacillus subtilis* gene

The q-value is an adjusted p-value, taking in to account the false discovery rate

Table S5. RNA-Seq and qRT-PCR analysis of PZN-treated *Bacillus anthracis*

			Fold Change	
Locus Tag	**Gene**	**Description**	**RNA-Seq**	**qRT-PCR**[a]
BAS0577	*lldp-1*	L-lactate permease	-7	-7 ± 2
BAS4869	*ldh3*	L-lactate dehydrogenase	-6	-24 ± 14
BAS4762	*ldh2*	L-lactate dehydrogenase	-4	-3 ± 1
BAS3439	*hutG*	formimidoylglutamase	3	3 ± 1
BAS3440	*hutI*	imidazolonepropionase	5	4 ± 0
BAS3441	*hutU*	urocanate hydratase	8	3 ± 0
BAS3442	*hutH*	histidine ammonia-lyase	10	5 ± 2
BAS0627	*appD, oppD*[b]	ABC transporter nucleotide-binding protein	11	6 ± 3
BAS1521	*gerN*	germination protein	28	14 ± 9
BAS5200	*yvfU, desR*[b]	DNA-binding response regulator	49	37 ± 30
BAS5201	*yvfT, desK*[b]	sensor histidine kinase	56	94 ± 36
BAS1345	*liaH*[b]	PspA/IM30 family protein	1084	138 ± 10
BAS1344	*liaI*[b]	membrane protein	1407	352 ± 191

[a] Error is given as standard deviation with $n \geq 3$ replicates
[b] Gene annotations are derived from sequence homology to the given *B. subtilis* gene

Table S6. Primers used in this study

Confirmation of LLNL A0517_1 pXO1 plasmid loss	
pXO1_0172 lef RT-r	GATGCGAAAGTAGTGCCAAAGA
pXO1_0172 lef RT-r	CCACAGCATGTCCAAATTCG
Confirmation of spontaneous mutants	
BamHI-BAS4114-f	AAAAGGATCCATGACAGCAAACCGCATTAAAG
NotI-BAS4114-r	AAAAGCGGCCGCTCAGTTTGAAAGGCCTCGC
Generation and confirmation of the bas4114-4117 deletion strain	
G236Tet-up-vec-f	TGGAGCTCCACCGCGGTGGCAGGCGTTTGCTGATACAC
G236Tet-up-down-r	TAAAAAAGGACAGTTTCATCCCCTACCTACC
G236Tet-down-up-f	GGGGATGAAACTGTCCTTTTTTATATTCATTCAGAC
G236Tet-down-vec-r	TCGACCTGCAGCCAAGGACGCATCTGTCTTTGTTTCAGTG
TetR-f	GCAAACCGCATTAAAGCTGTAGC
TetR-r	CAGTTTGAAAGGCCTCGCC
SugE4115-f	GGCATGGATTTATGTAATCTTAGCTGG
SugE4115-r	GCCTCCTTCGCTTCTTTTTCTTC
SugE4116-f	GGCTTGGGTATTTTTAATTCTAGCTGG
SugE4116-r	CTTAATAGTTTTAAGCCAACAGCGCC
BAS4117-f	GTGGAAAGAAAAAGGGAAGCAACTG
BAS4117-r	GTTCAGAAGAACTGTCCTTTTTAAATAACTTATTCC
RT-qPCR	
Banth 16S qRT-f	CGGAATTATTGGGCGTAAAG
Banth 16S qRT-r	TCTCCCAGTTTCCAATGACC
BAS0577 lldp-1 qRT-f	TGGTTCACTATTCGCACCAC
BAS0577 lldp-1 qRT-r	TTTGCTATTGTGCCACCAAC
BAS0627 appD qRT-f	ACGAATTATCTGGCGGAATG
BAS0627 appD qRT-r	AGCCGTTGTTGGCTCATC
BAS1344 liaI qRT-f	GGAGCAGGAGTTGTGTACTGG
BAS1344 liaI qRT-r	GATGGACAAGCCGATTAAACC
BAS1345 liaH qRT-f	ATCAAAGCAAGCGCTTATCG
BAS1345 liaH qRT-r	TTCTAATCGAGTTACTTGCCCTTC
BAS1521 gerN qRT-f	ACGAATGACTGGATTTGATGC
BAS1521 gerN qRT-r	GAAAGTCCTGTTCCTGCAATG
BAS1659 CitB RR qRT-f	AACGACGTTCGATGATGATG
BAS1659 CitB RR qRT-r	TTATCGCATCACGAATACGC
BAS1660 Sensor HK qRT-f	TGAAAGCATTCGGATTACATTG
BAS1660 Sensor HK qRT-r	TCATGCTTACTGGCAAATTCC
BAS1661 ABC trans qRT-f	TGCGAAGATGAATATTGGTGTC

BAS1661 ABC trans qRT-r	ACCCGTACAGTCCAGCAAAG
BAS1662 ABC trans qRT-f	CGGTTGTAATGACGACGATG
BAS1662 ABC trans qRT-r	TAGGCGCAGCAATAAGGTG
BAS1663 ABC trans qRT-f	TATGGGAATGTAGGGCAAGG
BAS1663 ABC trans qRT-r	AGAAATTTCACCGTATCATTTGC
BAS1664 minor cls ClsB qRT-f	TGGCACAACAAACTTTACTTCG
BAS1664 minor cls ClsB qRT-r	AACGCCTTATTAACTTGCCCTAC
BAS1842 asbE qRT-f	GGGTATTTGCTTTCTGGTCTTG
BAS1842 asbE qRT-r	TTTCACGGAAGTATGCAAAGG
BAS1843 asbF qRT-f	CAGATCCAGTTGACAGCTTCC
BAS1843 asbF qRT-r	CGGTTCGAACACATGTAAATAATC
BAS3439 hutG qRT-f	TAACGGGCTTTGCAAACAG
BAS3439 hutG qRT-r	CATTTGACGGACCACCATC
BAS3440 hutI qRT-f	TGACCCGCATACTCATCTTG
BAS3440 hutI qRT-r	AAGAATACCTCCGCCTTGTTC
BAS3441 hutU qRT-f	ATTTGTTGGCTTGGTTACGG
BAS3441 hutU qRT-r	CACGACCGATAACGATTGG
BAS3442 hutH qRT-f	TGCGATGGTTGCTCTTACAG
BAS3442 hutH qRT-r	AAAGAACTCCTTGCGCTGTC
BAS4762 ldh2 qRT-f	AATCGTTCGCGGTATTATGG
BAS4762 ldh2 qRT-r	TGGTAGACCAGATTCTTTCCAAG
BAS4869 ldh3 qRT-f	AATCATGAACGTGCAGTTGG
BAS4869 ldh3 qRT-r	TTGCAGTCTTCATAGCTTCCTG
BAS5033 ftsX qRT-f	CGGAAAGACGTTTGAGTTATTTG
BAS5033 ftsX qRT-r	TCGCAATTGTTGCTGTATCTG
BAS5034 ABC ftsE qRT-f	TCTTGAAGATCGTGCAGACG
BAS5034 ABC ftsE qRT-r	ATCGGCAATTACGACTTTCG
BAS5200 desR qRT-f	TTGAAGTAATTGGGCAAGCTG
BAS5200 desR qRT-r	TCTAACCCGCTTTGAATTGG
BAS5201 desK qRT-f	GCAGTGACGAATGTTGTAAAGC
BAS5201 desK qRT-r	CCAATTCCGTTATCTTCTACCG
BAS5288 TetR reg qRT-f	AAGATGAGGAATTGCTTGTTACG
BAS5288 TetR reg qRT-r	CTTGTATAACCGGAATCCTTGG
BAS5289 transporter qRT-f	CGGAGCTCTTGTTGCCTTAC
BAS5289 transporter qRT-r	AAGCACGATTGCGTTTGTTAC

Table S7. Minimum inhibitory concentrations (μM) of Me_2-Arg-Az_5 reveal a broader spectrum of activity than PZN

Strain	**MIC (μM)[a]**
Bacillus anthracis Sterne 7702	3
Bacillus anthracis PR06	3
Bacillus anthracis A0517-1	3
Bacillus cereus ATCC 4342	6
Bacillus megaterium 899	3
Bacillus subtilis 168	6
Escherichia coli DH5a	>48
Enterococcus faecium U503	48
Listeria monocytogenes 4b	>48
Neisseria sicca ATCC 29256	>48
Staphylococcus aureus USA300	3
Streptococcus pyogenes M1	48
Pseudomonas putida	>48

[a]Minimum inhibitory concentration, determined my microbroth dilution assay, measured in μM

Table S8. Common differentially expressed genes after PZN and Me_2-Arg-Az_5 treatment

			Fold Change	
Locus Tag	**Gene**	**Description**	**PZN**	**Me_2-Arg-5Az**
BAS0169		hypothetical protein	-6	-4
BAS0513	*bdbD*	hypothetical protein	-3	-2
BAS0577	*lldp-1*	L-lactate permease	-7	-7
BAS1089		hypothetical protein	-4	-4
BAS1222	*ywcJ*	formate/nitrite transporter family protein	-7	-7
BAS1615		hypothetical protein	-3	-3
BAS3917		hypothetical protein	-3	-2
BAS4146		hypothetical protein	-5	-5
BAS4690	*cydA-2*	cytochrome d ubiquinol oxidase subunit I	-3	-2
BAS4762	*ldh2*	L-lactate dehydrogenase	-4	-4
BAS4869	*ldh3*	L-lactate dehydrogenase	-6	-4
BAS5200	*ycfU,desR*	DNA-binding response regulator	49	6
BAS5201	*yvfT, desK*	sensor histidine kinase	56	5
BAS5202	*yvfS*	ABC transporter permease	130	13
BAS5203	*yvfR*	ABC transporter ATP-binding protein	135	12

Table S9 Complete RNA-Seq analysis of differentially expressed *B. anthracis* genes after Me_2-Arg-Az_5 treatment

Locus Tag	Gene	Description	Fold Change	qValue
BAS0577	*lldp-1*	L-lactate permease	-7	5.95E-90
BAS1222	*ywcJ*[a]	formate/nitrite transporter family protein	-7	1.30E-59
BAS4134		hypothetical protein	-7	1.61E-116
BAS4146		hypothetical protein	-5	6.09E-31
BAS4711		hypothetical protein	-5	3.53E-31
BAS0169		hypothetical protein	-4	9.57E-21
BAS1089	*yjzC*[a]	hypothetical protein	-4	2.76E-23
BAS1357	*hmp*	nitric oxide dioxygenase	-4	6.57E-12
BAS4762	*ldh2*	L-lactate dehydrogenase	-4	1.24E-16
BAS4869	*ldh3*	L-lactate dehydrogenase	-4	3.46E-17
BAS1023		hypothetical protein	-3	3.80E-06
BAS1349		hypothetical protein	-3	1.12E-08
BAS1615	*yfmQ*[a]	hypothetical protein	-3	1.28E-06
BAS4990	*cggR*	gapA transcriptional regulator CggR	-3	0.004179
BAS0070	*pabC*	4-amino-4-deoxychorismate lyase	-2	0.003344
BAS0513	*bdbD*	hypothetical protein	-2	8.87E-04
BAS0547	*maeN*[a]	citrate cation symporter family protein	-2	0.002003
BAS0566	*rapD, rapI*[a]	transcriptional regulator	-2	0.004179
BAS0631	*rbsR*	ribose operon repressor	-2	6.17E-05
BAS1024		hypothetical protein	-2	8.87E-04
BAS1185		hypothetical protein	-2	0.007387
BAS1273		ABC transporter permease	-2	8.71E-05
BAS1666	*gltT*[a]	proton/sodium-glutamate symport protein	-2	5.37E-04
BAS2706		hypothetical protein	-2	0.003549
BAS2707		ABC transporter ATP-binding protein	-2	8.05E-04
BAS3173		transcriptional regulator	-2	8.18E-07
BAS3917		hypothetical protein	-2	8.14E-04
BAS4414		succinate dehydrogenase, cytochrome b558 subunit	-2	0.003007
BAS4690	*cydA-2*	cytochrome d ubiquinol oxidase subunit I	-2	0.004179
BAS4922	*nupC*[a]	NupC family nucleoside transporter	-2	0.004085
BAS5002	*whiA*	hypothetical protein	-2	0.007026
BAS5316	*yycI*[a]	YycI protin	-2	7.74E-04
BAS0982	*fadR, yvdT*[a]	TetR family transcriptional regulator	3	8.50E-04
BAS4115	*sugE-1*	sugE protein	4	4.42E-10
BAS5201	*yvfT, desK*[a]	sensor histidine kinase	5	2.03E-21
BAS5200	*yvfU, desR*[a]	DNA-binding response regulator	6	4.21E-26
BAS0375	*yuxN, yfiR*[a]	TetR family transcriptional	8	2.92E-37

		regulator		
BAS0902	*atpI*[a]	ATP synthase protein I	8	2.99E-46
BAS5203	*yvfR*[a]	ABC transporter ATP-binding protein	12	1.66E-119
BAS5202	*yvfS*[a]	ABC transporter permease	13	2.82E-151
BAS0376	*ykuC, yfiS*[a]	major facilitator family transporter protein	181	0

[a]Genes are annotated by sequence homology to the corresponding *Bacillus subtilis* gene
The q-value is an adjusted p-value, taking in to account the false discovery rate

Table S10. *B. anthracis* Sterne mutations conferring PZN-resistance accumulate in *bas4114*, an AcrR family transcriptional regulator

Strain	**Mutation**	**Consequence**
Sterne	Wild type	D**ALEAFLCLLDGLMVELLFAGL**NRFETRLNASWQVFWRGLSN[a]
PR01[c]	457G→T	DAL*[b]
PR02	492^T	DALEAFLCLLDGLMV*
PR03	495^AG	DALEAFLCLLDGLMVESYYSQV*
PR04	504^C	DALEAFLCLLDGLMVELLFRRFKSL*
PR05	506^GC	DALEAFLCLLDGLMVELLFAQV*
PR06	507^ATTCGCA	DALEAFLCLLDGLMVELLFAIRRFKSL*

[a] Amino acids 150-191 of BAS4114. Bold residues represent the predicted transmembrane region, as defined by Spoctopus (Figure S12)[23]
[b] Asterisk (*) represents the termination of the protein sequence due to the nonsense mutation.
[c] MICs for the PZN-resistant (PR) mutants are ≥ 32 μg/mL PZN, compared to 1 μg/mL for the wild type.

Table S11. Upregulation of *bas4114-4117* in PZN-resistant mutant, *B. anthracis* PR06

Gene	Annotation	Fold change
BAS4114	*acrR* family transcriptional regulator	34
BAS4115	*emrE* drug efflux pump	152
BAS4116	*emrE* drug efflux pump	47
BAS4117	hypothetical protein	62

Table S12. MICs (μg/mL) of selected antibiotics against PZN-resistant mutant, *B. anthracis* PR06

Compound	***B. anthracis* Sterne**	***B. anthracis* PR06**
Plantazolicin	1	64
Tetracycline	0.125	0.125
Vancomycin	1	1
Triclosan	1	1
Nisin	2	2
Kanamycin	4	4
Cerulenin	4	8
Spectinomycin	8	4
Chloramphenicol	8	4
Daptomycin	8	8
Ethidium Bromide	16	16

Table S13. qRT-PCR of *cls* locus in step-wise evolved PZN-resistant mutants

	Fold change		
Gene	**PR09-1**	**PR09-4**	**PR10-4**
BAS1659	NS	NS	NS[a]
BAS1660	NS	NS	NS
BAS1661	21 ± 7	48 ± 26	NS
BAS1662	23 ± 13	NS	NS
BAS1663	24 ± 9	48 ± 26	NS
BAS1664 (*cls*)	39 ± 7	74 ± 26	NS
BAS1842 (*asbE*)	-	-	NS
BAS1843 (*asbF*)	-	-	NS

[a]NS not significant

Supporting Figures

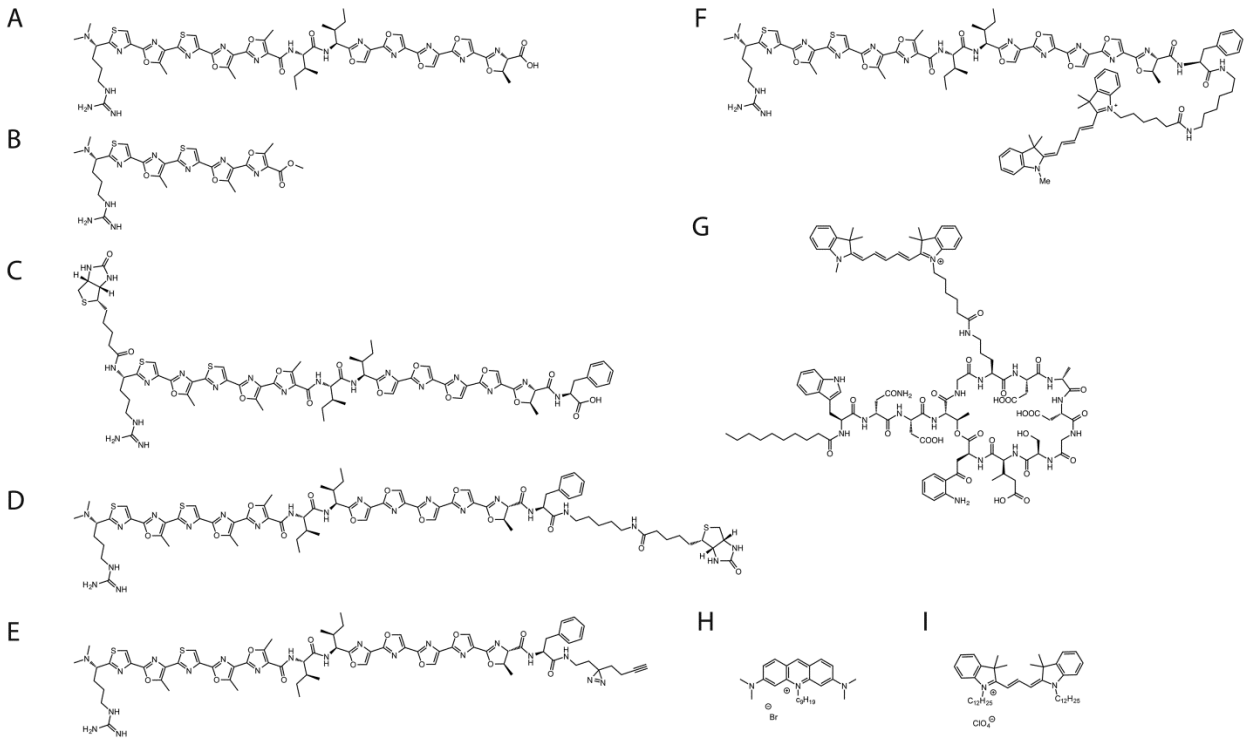


Figure S1. Chemical structures of plantazolicin (PZN) (A), Me_2-Arg-Az_5 (B), Biotin-PZN (C), PZN-Biotin (D), PZN-photoaffinity probe (E), PZN-Cy5 (F), daptomycin (Dap)-Cy5 (G) NAO, nonyl acridine orange (H), and DilC12(3) perchlorate (I).

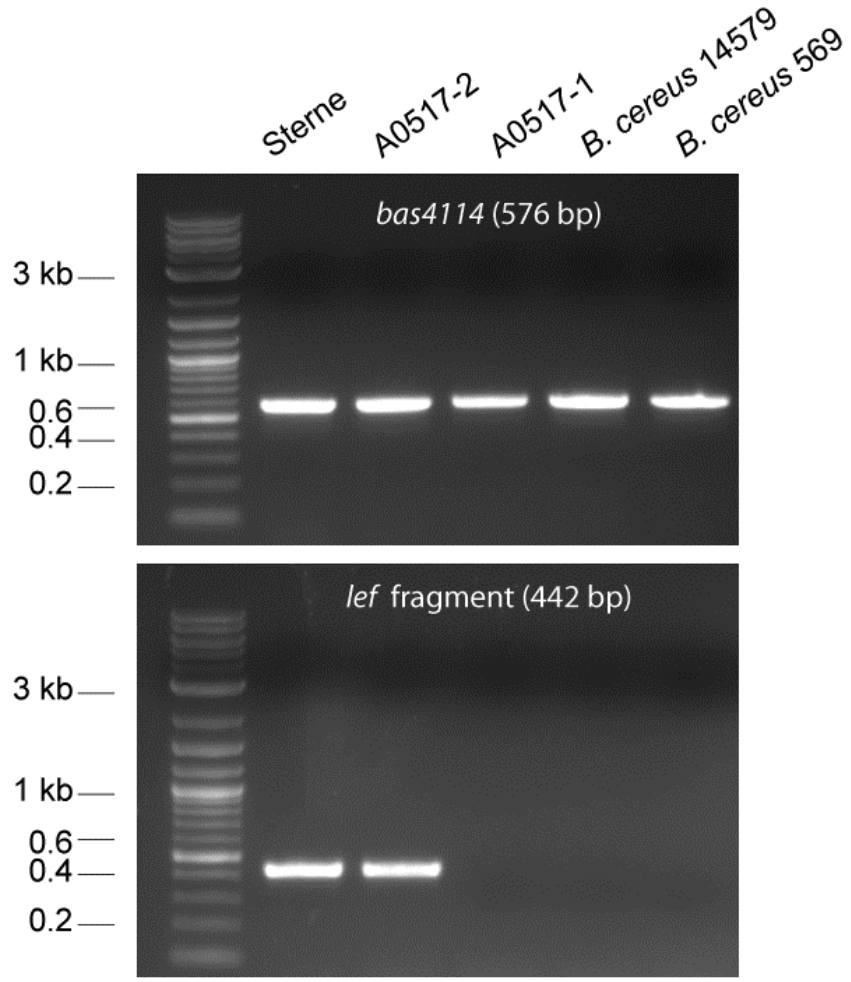


Figure S2. Confirmation of plasmid loss in *B. anthracis* 34F2 LLNL A0517-1. (Top) PCR of a conserved chromosomal gene, *bas4114*. (Bottom) pXO1 encoded gene *lef* is not present in A0517-1 or the select *B. cereus* strains.

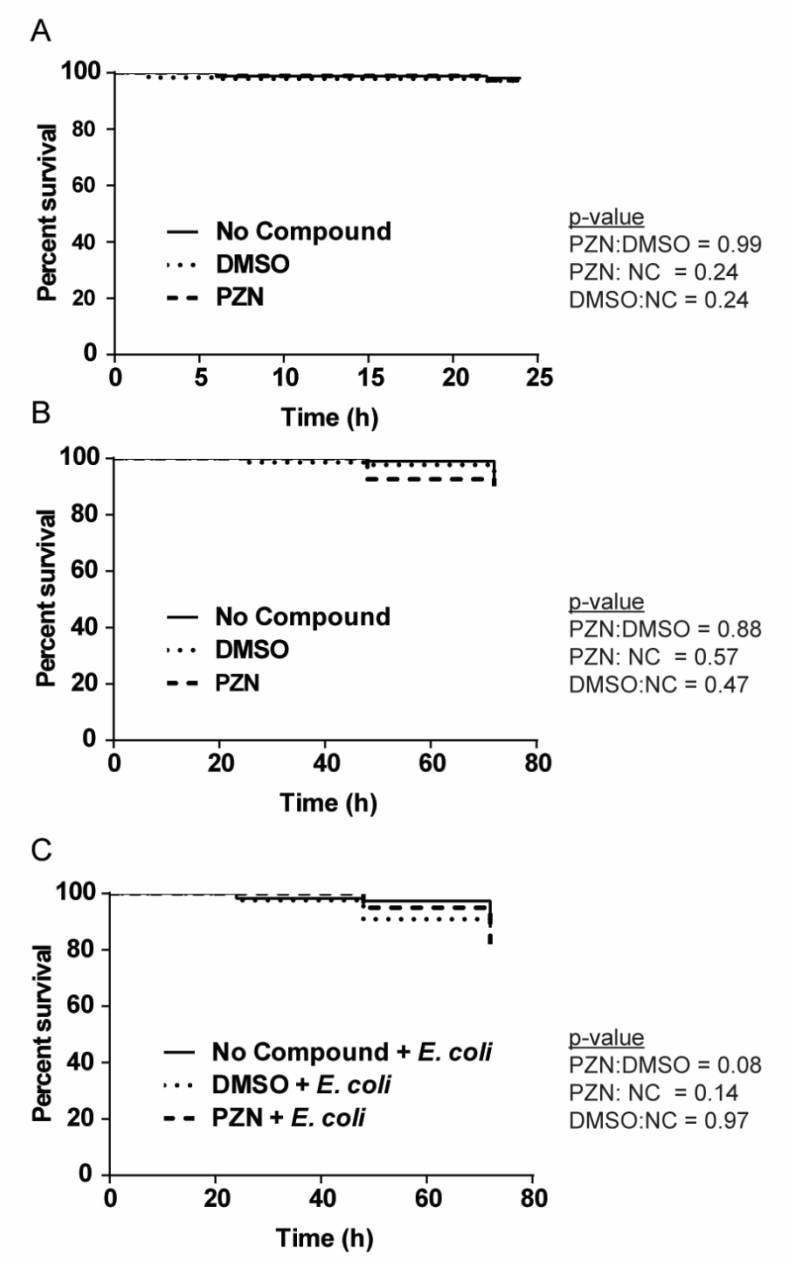


Figure S3. PZN is not toxic to *Caenorhabditis elegans* using two different experiments. PZN was subjected to purified PZN during both liquid fast killing (A) and slow killing assays (B-C). Animals were treated with 64 μg/mL PZN and analyzed by live/dead screening over the course of the experiments. P-values were generated using the Log-rank (Mantel-Cox) test.

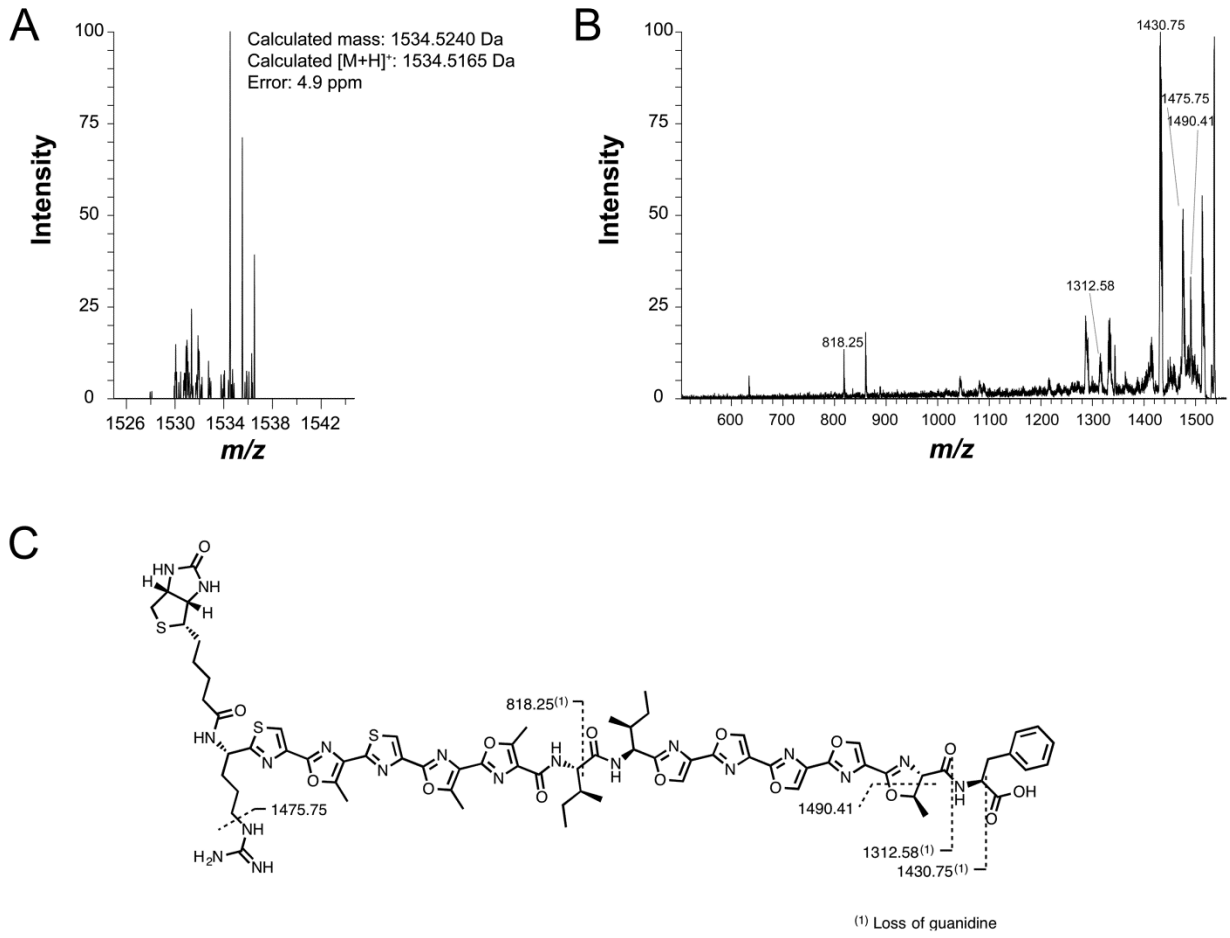


Figure S4. Mass spectrometry structural characterization of biotin-PZN. (A) Mass spectrum of the $[M+H]^{2+}$ species by LTQ-FT-MS, which was used to calculate exact mass. (B) CID spectrum of the singly charged ion acquired by LTQ-FT-MS. Labeled peaks correspond to identified fragments of biotin-PZN, shown on the structure in panel (C).

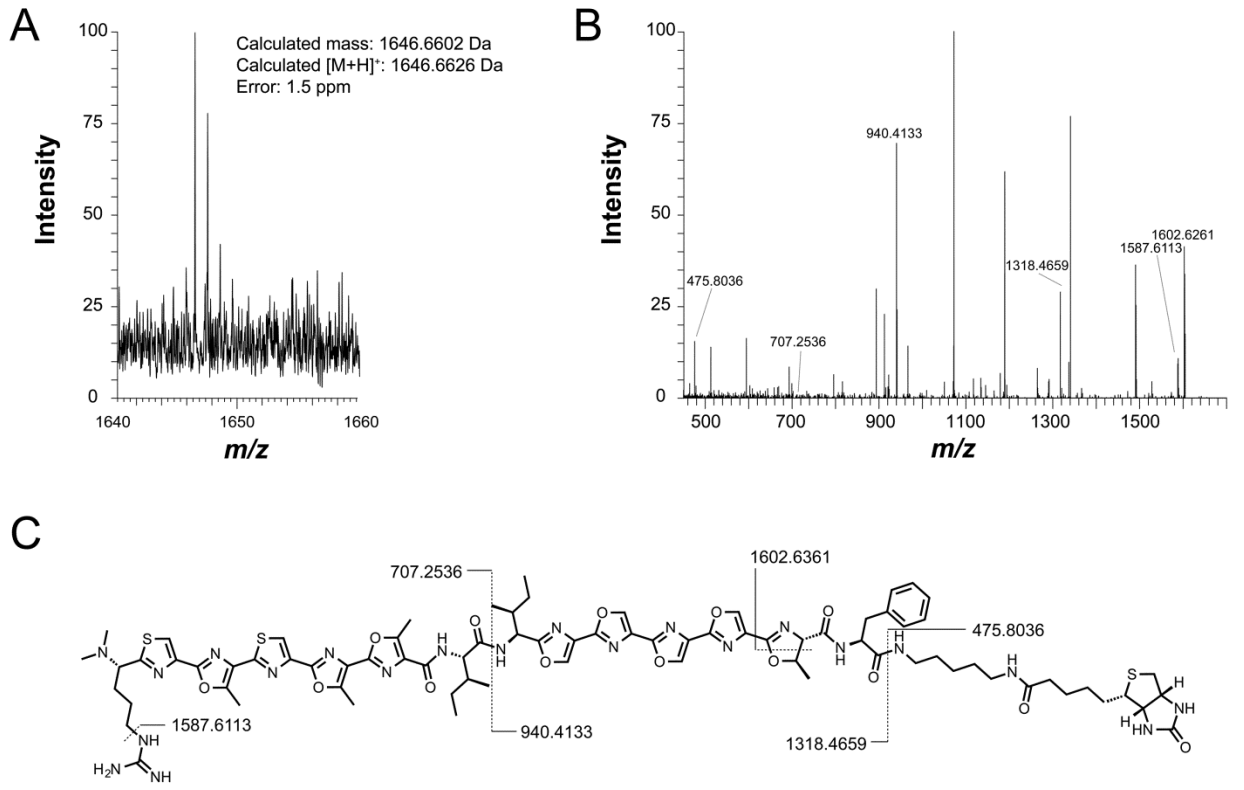


Figure S5. | Mass spectrometry structural characterization of PZN-biotin. (A) Mass spectrum of the $[M+H]^{2+}$ species by LTQ-FT-MS, which was used to calculate exact mass. (B) CID spectrum of the singly charged ion acquired by LTQ-FT-MS. Labeled peaks correspond to identified fragments of PZN-biotin, shown on the structure in panel (C).

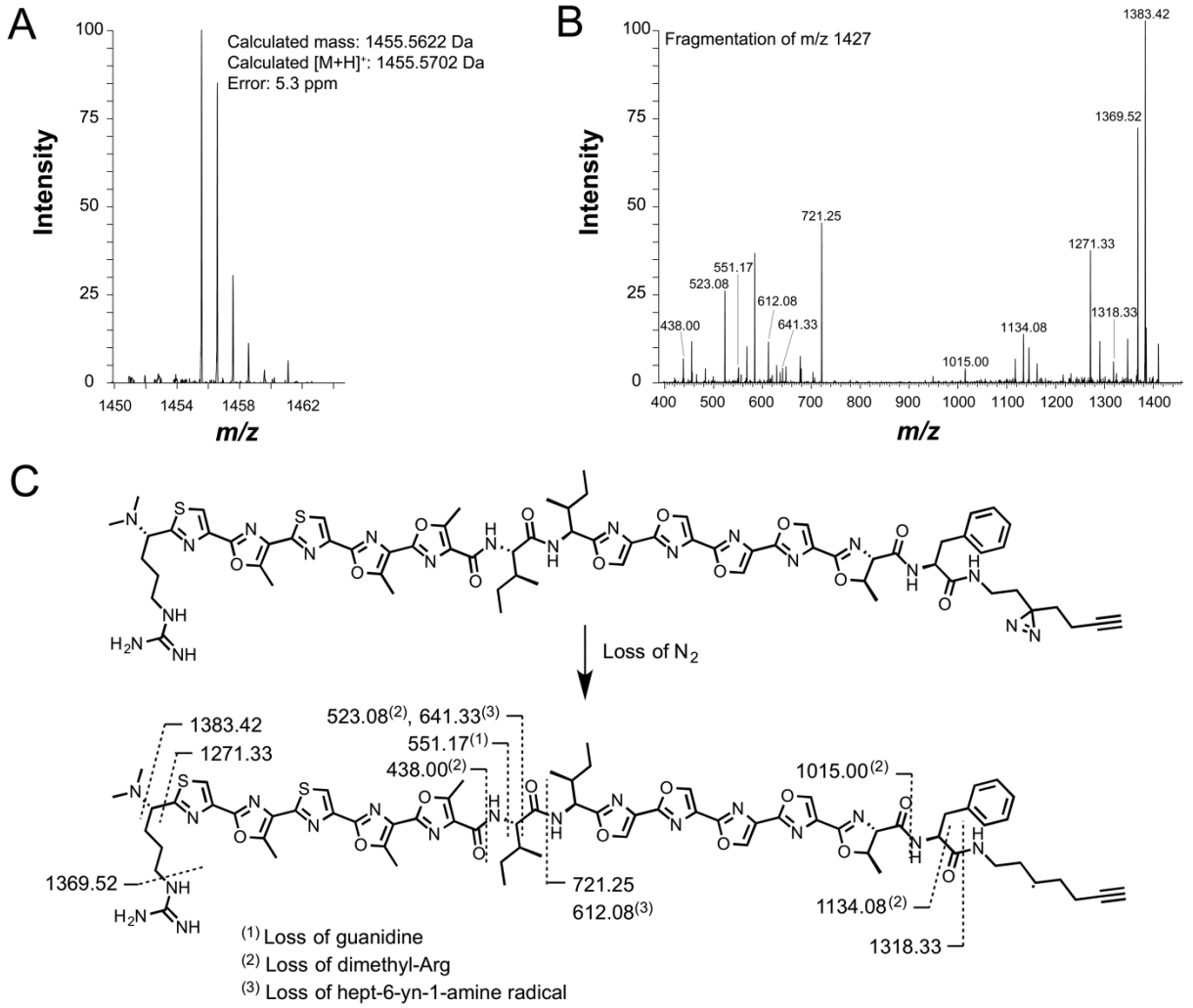


Figure S6. Mass spectrometry structural characterization of PZN-photoaffinity probe. (A) Mass spectrum of the $[M+H]^{2+}$ species by LTQ-FT-MS, which was used to calculate exact mass. (B) CID spectrum of the *m/z* 1427 species, corresponding to the loss of N_2 from the original molecule. Labeled peaks correspond to identified fragments, shown on the structure in panel (C).

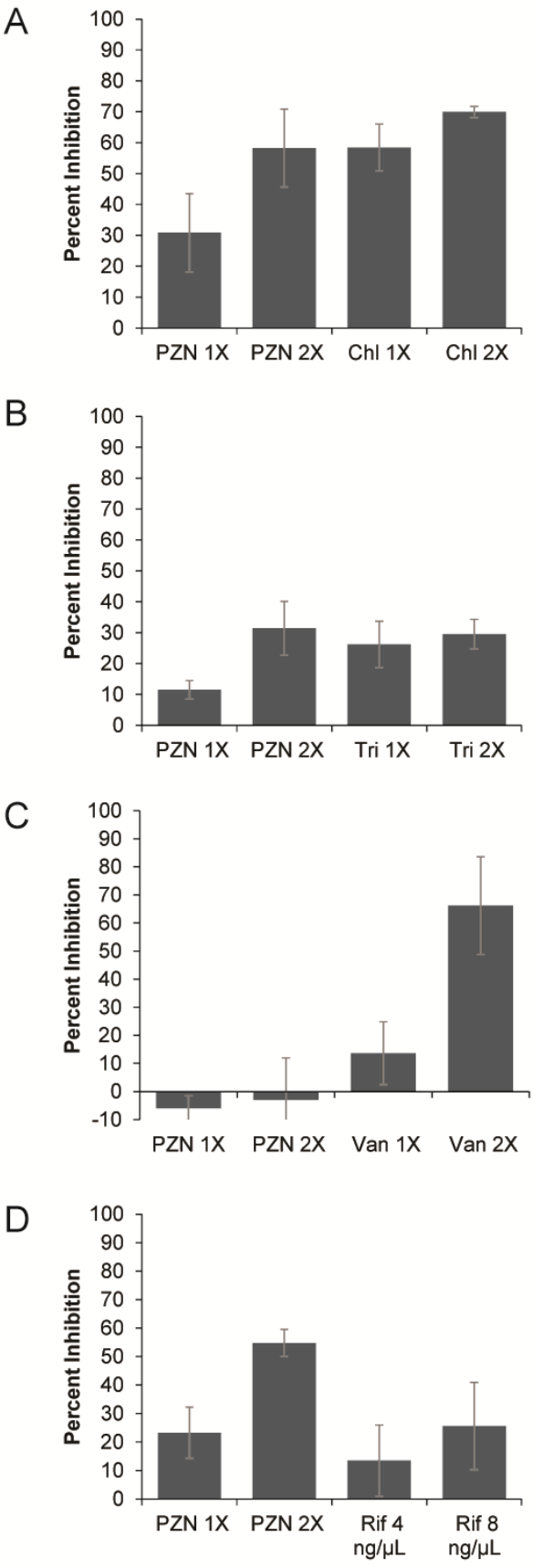


Figure S7. PZN treatment results in the inhibition of three out of four analyzed macromolecular pathways. *B. anthracis* cells were treated with 1 × or 2 × MIC PZN and the appropriate radiolabelled precursor to the macromolecule of interest. After incubation for 1 h, macromolecules were precipitated and radiolabelled precursor incorporation into protein (A), fatty acid (B), peptidoglycan (C), or RNA (D) was measured. Each experiment included samples treated with a control compound that is known to affect a particular pathway (protein: chloramphenicol; fatty acid: triclosan; peptidoglycan: vancomycin; RNA: rifampicin. All compounds were added at 1 × or 2 × MIC, except rifampicin, which required significantly lower concentrations to achieve medial inhibition. Percent inhibition was determined by comparing radioactive incorporation between cells containing compound vs. vehicle. Error is reported as standard deviation with n = 3.

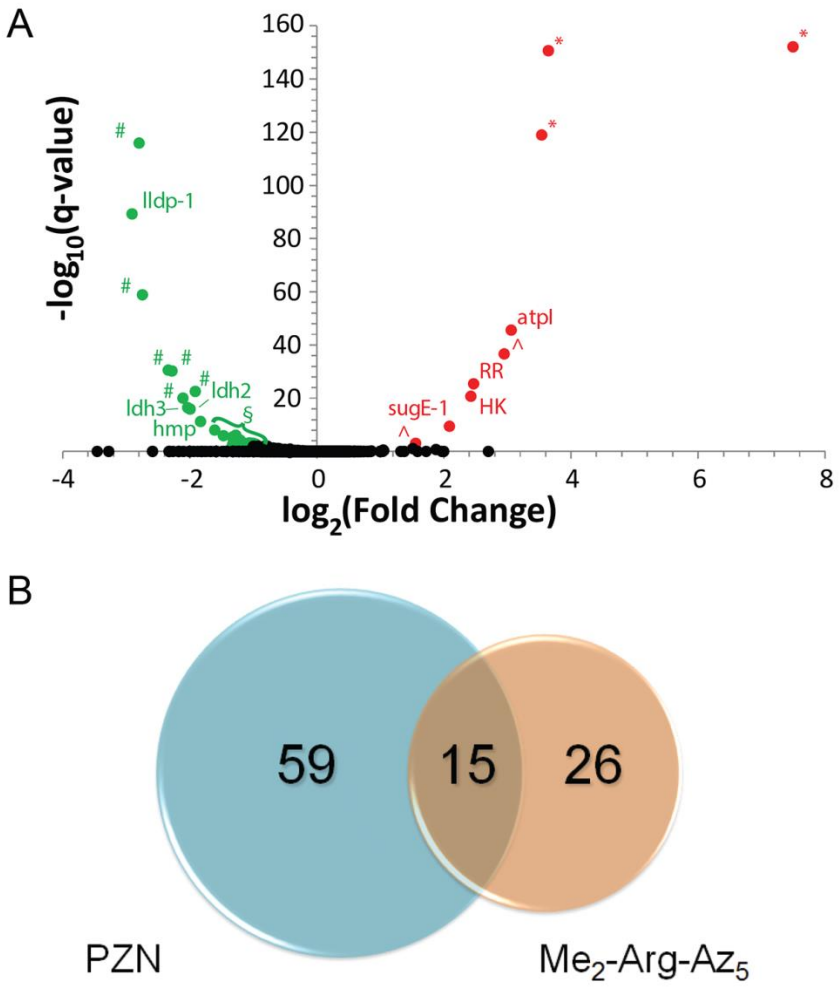


Figure S8. The *B. anthracis* gene expression profile with Me_2-Arg-Az_5 differs from PZN treatment. (A) A volcano plot represents the set of differentially regulated genes in response to treatment with 0.25 × MIC of Me_2-Arg-Az_5. Red (upregulated) and green (downregulated) points are significantly expressed genes in response to Me2-Arg-Az5, while black genes did not meet the q-value (< 0.01) threshold. Genes where the q-value = 0 were given a value of E-155. Abbreviations: # – hypothetical, * – transporter, ^ – transcriptional regulator, atpI – ATP synthase protein I, hmp – nitric oxide dioxygenase, lldP-1 – L-lactate permease, ldh2/3 – L-lactate dehydrogenase, sugE-1 – EmrE protein (cationic drug membrane transporter), RR – response regulator, HK: histidine kinase , § – remaining upregulated. All genes are summarized in Table S9. (B) Venn diagram demonstrating the commonality of the PZN and Me_2-Arg-Az_5 expression profiles. The compilation of common genes is located in Table S8.

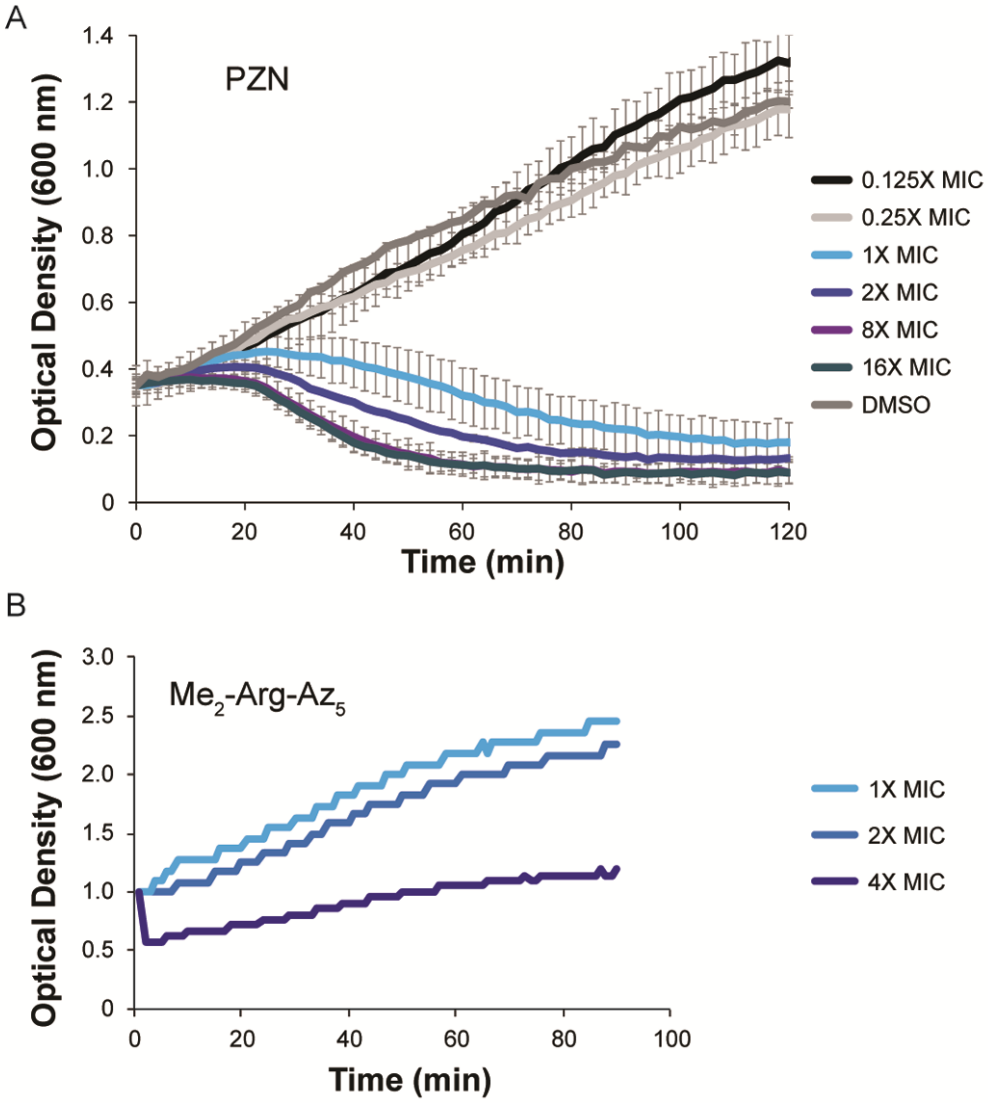


Figure S9. Growth curves of PZN and Me_2-Arg-Az_5 reveal different phenotypes. (A) *B. anthracis* cells treated with PZN undergo rapid decrease in optical density with concentrations above the minimum inhibitory concentration (MIC, determined from microbroth dilution assays). Time points were collected every two minutes and adjusted to a 1-cm path length. Error is reported as standard deviation with n = 3. (B) *B. anthracis* treated with Me_2-Arg-Az_5. Reduction in optical density is observed briefly at 4 × MIC. Both growth curves were measured in duplicate and normalized to OD_{600} 0.35 (A) and 1.0 (B).

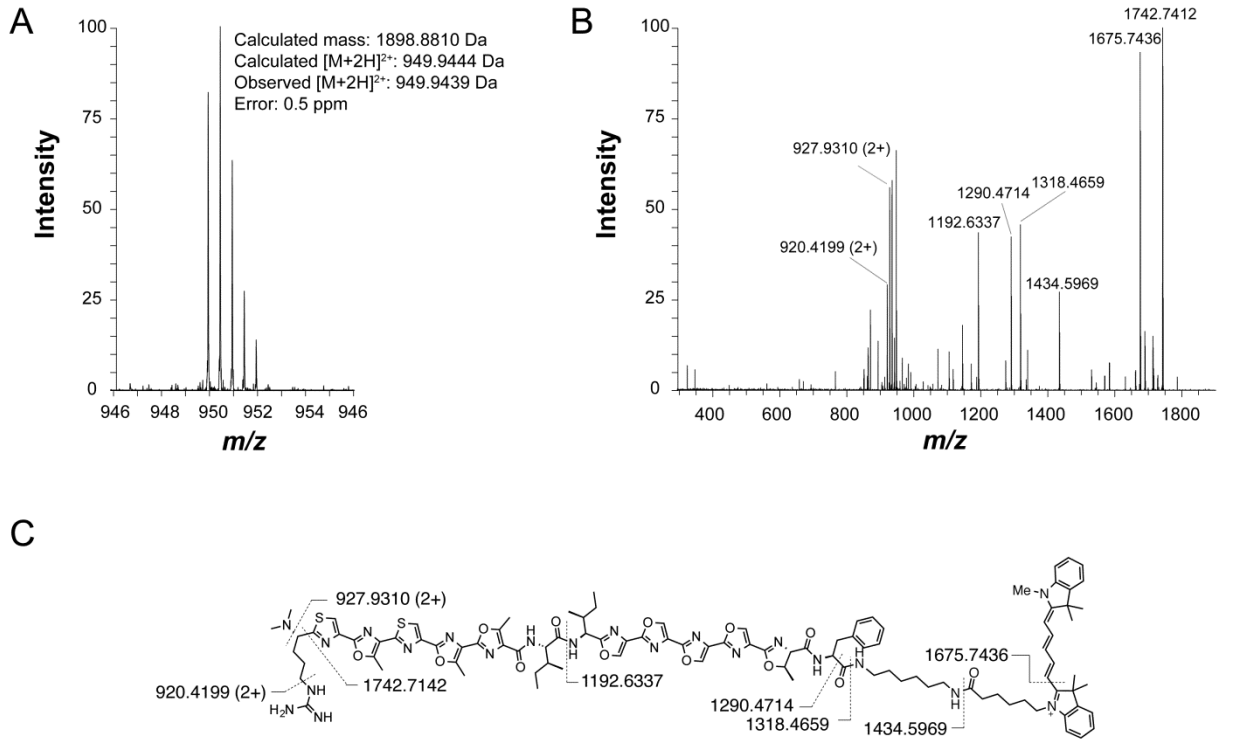


Figure S10. Mass spectrometry characterization of PZN-Cy5. (A) Mass spectrum of the $[M+H]^{2+}$ species by LTQ-FT-MS, which was used to calculate exact mass. (B) CID spectrum of the singly charged ion acquired by LTQ-FT-MS. Labeled peaks correspond to identified fragments of PZN-Cy5, shown on the structure in panel (C).

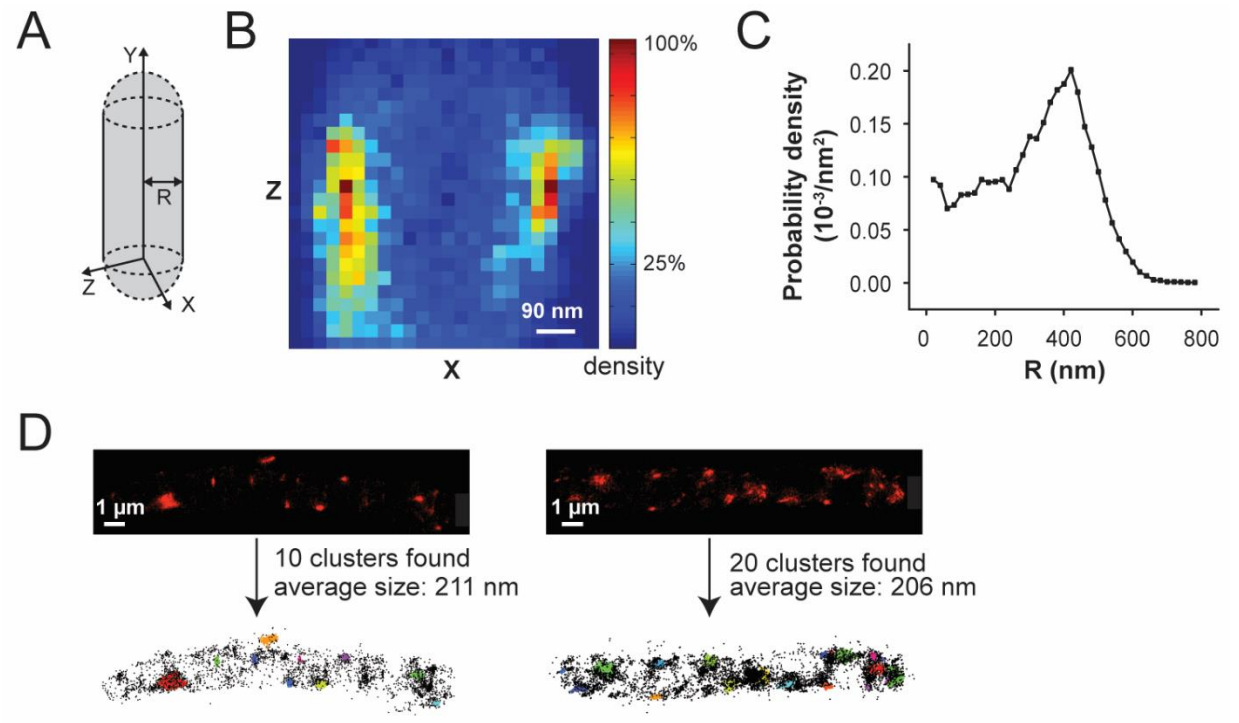


Figure S11. Cluster analysis of *B. anthracis* treated with PZN-Cy5 and imaged by super-resolution microscopy. (A) Representation of cell alignment used in subsequent analyses. (B) PZN-Cy5 color density map of a representative cell projected onto the XZ plane. (C) Probability density of PZN-Cy5 increases as distance from the center of the cell (R) increases, reaching a maximum at the cell membrane. (D) Cluster analysis of two representative *B. anthracis* cells (top) shows identified clusters (bottom) used in subsequent size calculations.

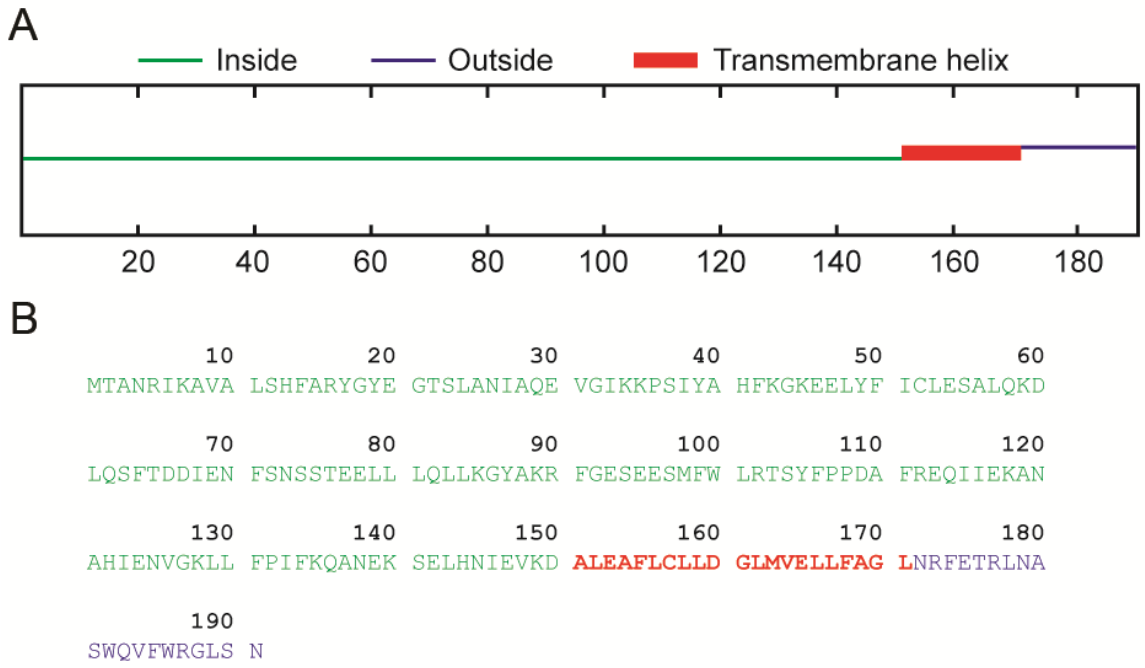


Figure S12. Hydropathy plot of BAS4114 reveals a single transmembrane helix. (A) The amino acid sequence of BAS4114 was analyzed using SPOCTOPUS, which predicted a transmembrane domain from residues 151-171. (B) Corresponding amino acids for each predicted membrane association. Green, inside the membrane; Red, transmembrane; Blue, outside the membrane.

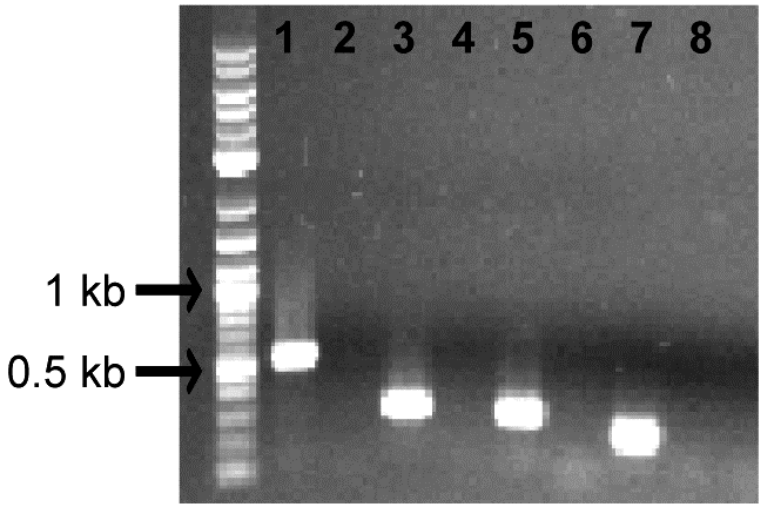


Figure S13. Confirmation of *bas4114-4117* deletion in *B. anthracis* Sterne. Specific primers for each deleted gene were used to confirm their absence by PCR amplification. 1, *bas4114* in wild type Sterne (WT); 2, *bas4114* in deletion strain; 3, *bas4115* in WT; 4, *bas4115* in deletion strain; 5, *bas4116* in WT; 6, *bas4116* in deletion strain; 7, *bas4117* in WT; 8, *bas4117* in deletion strain.

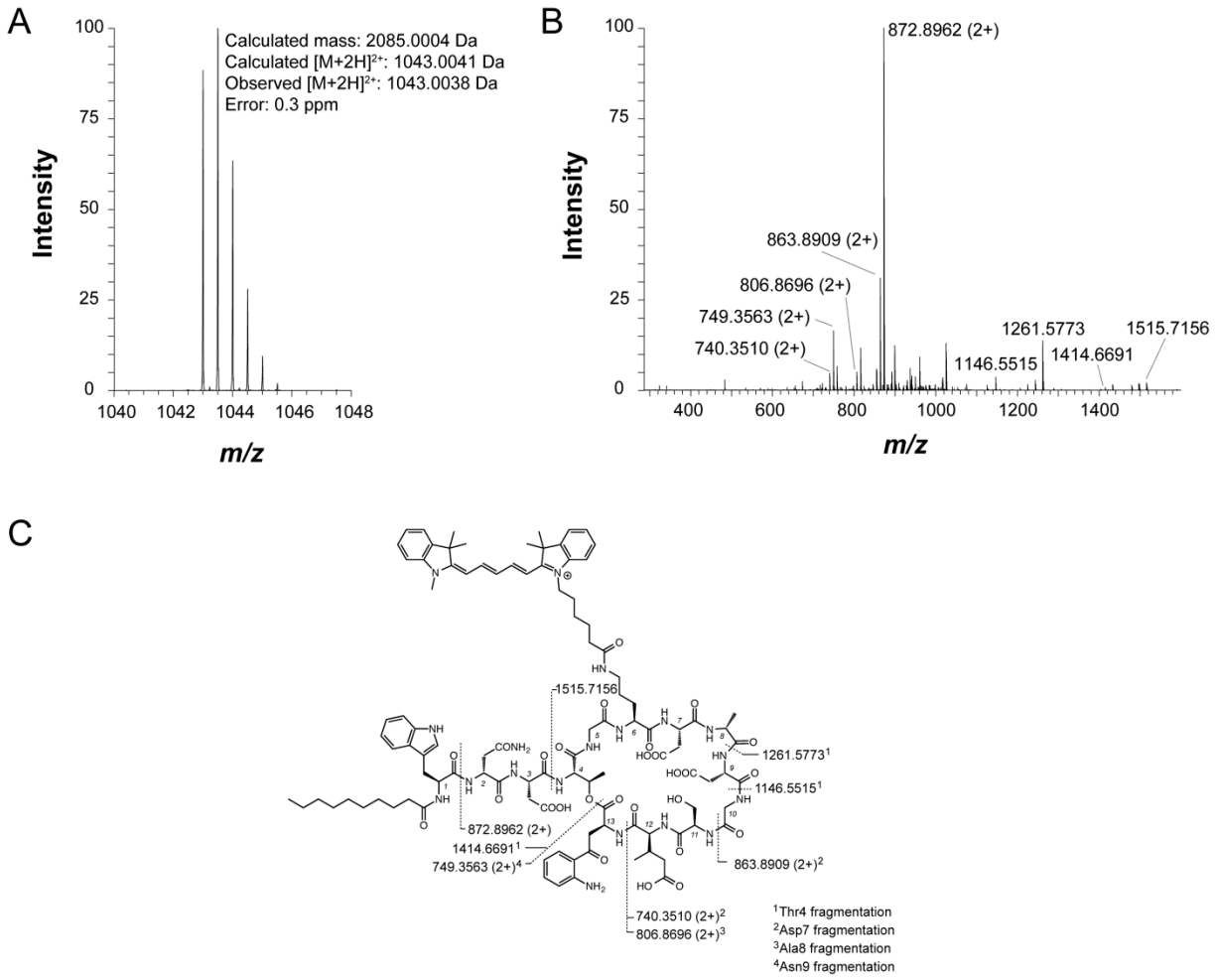


Figure S14. Mass spectrometry characterization of daptomycin-Cy5. (A) Mass spectrum of the $[M+2H]^{2+}$ species by LTQ-FT-MS, which was used to calculate exact mass. (B) CID spectrum of the doubly charged ion acquired by LTQ-FT-MS. Labeled peaks correspond to identified fragments of daptomycin-Cy5, shown on the structure in panel (C).

References

1. de Man, P., Verhoeven, B. A., Verbrugh, H. A., Vos, M. C., and van den Anker, J. N. (2000) An antibiotic policy to prevent emergence of resistant bacilli, *Lancet 355*, 973-978.

2. Wilson, D. N., Harms, J. M., Nierhaus, K. H., Schlunzen, F., and Fucini, P. (2005) Species-specific antibiotic-ribosome interactions: implications for drug development, *Biol. Chem. 386*, 1239-1252.

3. Payne, D. J. (2008) Desperately seeking new antibiotics, *Science 321*, 1644-1645.

4. Scholz, R., Molohon, K. J., Nachtigall, J., Vater, J., Markley, A. L., Sussmuth, R. D., Mitchell, D. A., and Borriss, R. (2011) Plantazolicin, a novel microcin B17/streptolysin S-like natural product from *Bacillus amyloliquefaciens* FZB42, *J. Bacteriol. 193*, 215-224.

5. Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J., Challis, G. L., Clardy, J., Cotter, P. D., Craik, D. J., Dawson, M., Dittmann, E., Donadio, S., Dorrestein, P. C., Entian, K. D., Fischbach, M. A., Garavelli, J. S., Goransson, U., Gruber, C. W., Haft, D. H., Hemscheidt, T. K., Hertweck, C., Hill, C., Horswill, A. R., Jaspars, M., Kelly, W. L., Klinman, J. P., Kuipers, O. P., Link, A. J., Liu, W., Marahiel, M. A., Mitchell, D. A., Moll, G. N., Moore, B. S., Muller, R., Nair, S. K., Nes, I. F., Norris, G. E., Olivera, B. M., Onaka, H., Patchett, M. L., Piel, J., Reaney, M. J., Rebuffat, S., Ross, R. P., Sahl, H. G., Schmidt, E. W., Selsted, M. E., Severinov, K., Shen, B., Sivonen, K., Smith, L., Stein, T., Sussmuth, R. D., Tagg, J. R., Tang, G. L., Truman, A. W., Vederas, J. C., Walsh, C. T., Walton, J. D., Wenzel, S. C., Willey, J. M., and van der Donk, W. A. (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature, *Nat. Prod. Rep. 30*, 108-160.

6. Cox, C. L., Doroghazi, J. R., and Mitchell, D. A. (2015) The genomic landscape of ribosomal peptides containing thiazole and oxazole heterocycles, *BMC Genomics 16*, 778.

7. Molohon, K. J., Melby, J. O., Lee, J., Evans, B. S., Dunbar, K. L., Bumpus, S. B., Kelleher, N. L., and Mitchell, D. A. (2011) Structure determination and interception of biosynthetic intermediates for the plantazolicin class of highly discriminating antibiotics, *ACS Chem. Biol. 6*, 1307-1313.

8. Banala, S., Ensle, P., and Sussmuth, R. D. (2013) Total synthesis of the ribosomally synthesized linear azole-containing peptide plantazolicin A from *Bacillus amyloliquefaciens*, *Angew. Chem. Int. Ed. 52*, 9518-9523.

9. Wilson, Z. E., Fenner, S., and Ley, S. V. (2015) Total syntheses of linear polythiazole/oxazole plantazolicin A and its biosynthetic precursor plantazolicin B, *Angew. Chem. Int. Ed. 54*, 1284-1288.

10. Deane, C. D., Melby, J. O., Molohon, K. J., Susarrey, A. R., and Mitchell, D. A. (2013) Engineering unnatural variants of plantazolicin through codon reprogramming, *ACS Chem. Biol. 8*, 1998-2008.

11. Lee, J., Hao, Y., Blair, P. M., Melby, J. O., Agarwal, V., Burkhart, B. J., Nair, S. K., and Mitchell, D. A. (2013) Structural and functional insight into an unexpectedly selective N-

methyltransferase involved in plantazolicin biosynthesis, *Proc. Natl. Acad. Sci. U. S. A. 110*, 12954-12959.

12. Sharma, A., Blair, P. M., and Mitchell, D. A. (2013) Synthesis of plantazolicin analogues enables dissection of ligand binding interactions of a highly selective methyltransferase, *Org. Lett. 15*, 5076-5079.

13. Hao, Y., Blair, P. M., Sharma, A., Mitchell, D. A., and Nair, S. K. (2015) Insights into methyltransferase specificity and bioactivity of derivatives of the antibiotic plantazolicin, *ACS Chem. Biol. 10*, 1209-1216.

14. Lee, S. W., Mitchell, D. A., Markley, A. L., Hensler, M. E., Gonzalez, D., Wohlrab, A., Dorrestein, P. C., Nizet, V., and Dixon, J. E. (2008) Discovery of a widely distributed toxin biosynthetic gene cluster, *Proc. Natl. Acad. Sci. U. S. A. 105*, 5879-5884.

15. Rasko, D. A., Altherr, M. R., Han, C. S., and Ravel, J. (2005) Genomics of the *Bacillus cereus* group of organisms, *FEMS Microbiol. Rev. 29*, 303-329.

16. Kolsto, A. B., Tourasse, N. J., and Okstad, O. A. (2009) What sets *Bacillus anthracis* apart from other Bacillus species?, *Annu. Rev. Microbiol. 63*, 451-476.

17. Mesnage, S., Fontaine, T., Mignot, T., Delepierre, M., Mock, M., and Fouet, A. (2000) Bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation, *Embo J. 19*, 4473-4484.

18. Kern, J., Ryan, C., Faull, K., and Schneewind, O. (2010) *Bacillus anthracis* surface-layer proteins assemble by binding to the secondary cell wall polysaccharide in a manner that requires *csaB* and *tagO*, *J. Mol. Biol. 401*, 757-775.

19. Weidenmaier, C., and Peschel, A. (2008) Teichoic acids and related cell-wall glycopolymers in Gram-positive physiology and host interactions, *Nat. Rev. Microbiol. 6*, 276-287.

20. Ganguly, J., Low, L. Y., Kamal, N., Saile, E., Forsberg, L. S., Gutierrez-Sanchez, G., Hoffmaster, A. R., Liddington, R., Quinn, C. P., Carlson, R. W., and Kannenberg, E. L. (2013) The secondary cell wall polysaccharide of *Bacillus anthracis* provides the specific binding ligand for the C-terminal cell wall-binding domain of two phage endolysins, PlyL and PlyG, *Glycobiology 23*, 820-832.

21. Schuch, R., Pelzek, A. J., Raz, A., Euler, C. W., Ryan, P. A., Winer, B. Y., Farnsworth, A., Bhaskaran, S. S., Stebbins, C. E., Xu, Y., Clifford, A., Bearss, D. J., Vankayalapati, H., Goldberg, A. R., and Fischetti, V. A. (2013) Use of a bacteriophage lysin to identify a novel target for antimicrobial development, *PLoS One 8*, e60754.

22. Ezzell, J. W., Abshire, T. G., Little, S. F., Lidgerding, B. C., and Brown, C. (1990) Identification of *Bacillus anthracis* by using monoclonal antibody to cell wall galactose-*N*-acetylglucosamine polysaccharide, *J. Clin. Microbiol. 28*, 223-231.

23. Abshire, T. G., Brown, J. E., and Ezzell, J. W. (2005) Production and validation of the use of gamma phage for identification of *Bacillus anthracis*, *J. Clin. Microbiol. 43*, 4780-4788.

24. Kan, S., Fornelos, N., Schuch, R., and Fischetti, V. A. (2013) Identification of a ligand on the Wip1 bacteriophage highly specific for a receptor on *B. anthracis*, *J. Bacteriol. 195*, 4355-4364.

25. Hoffmaster, A. R., Hill, K. K., Gee, J. E., Marston, C. K., De, B. K., Popovic, T., Sue, D., Wilkins, P. P., Avashia, S. B., Drumgoole, R., Helma, C. H., Ticknor, L. O., Okinaka, R. T., and Jackson, P. J. (2006) Characterization of *Bacillus cereus* isolates associated with fatal pneumonias: strains are closely related to *Bacillus anthracis* and harbor *B. anthracis* virulence genes, *J. Clin. Microbiol. 44*, 3352-3360.

26. Schuch, R., and Fischetti, V. A. (2006) Detailed genomic analysis of the Wbeta and gamma phages infecting *Bacillus anthracis*: implications for evolution of environmental fitness and antibiotic resistance, *J. Bacteriol. 188*, 3037-3051.

27. Marston, C. K., Gee, J. E., Popovic, T., and Hoffmaster, A. R. (2006) Molecular approaches to identify and differentiate *Bacillus anthracis* from phenotypically similar Bacillus species isolates, *BMC Microbiol. 6*, 22.

28. Davison, S., Couture-Tosi, E., Candela, T., Mock, M., and Fouet, A. (2005) Identification of the *Bacillus anthracis* (gamma) phage receptor, *J. Bacteriol. 187*, 6742-6749.

29. Hoffmaster, A. R., Ravel, J., Rasko, D. A., Chapman, G. D., Chute, M. D., Marston, C. K., De, B. K., Sacchi, C. T., Fitzgerald, C., Mayer, L. W., Maiden, M. C., Priest, F. G., Barker, M., Jiang, L., Cer, R. Z., Rilstone, J., Peterson, S. N., Weyant, R. S., Galloway, D. R., Read, T. D., Popovic, T., and Fraser, C. M. (2004) Identification of anthrax toxin genes in a *Bacillus cereus* associated with an illness resembling inhalation anthrax, *Proc. Natl. Acad. Sci. U. S. A. 101*, 8449-8454.

30. Burkett-Cadena, M., Kokalis-Burelle, N., Lawrence, K. S., van Santen, E., and Kloepper, J. W. (2008) Suppressiveness of root-knot nematodes mediated by rhizobacteria, *Biol. Control 47*, 55-59.

31. Liu, Z., Budiharjo, A., Wang, P., Shi, H., Fang, J., Borriss, R., Zhang, K., and Huang, X. (2013) The highly modified microcin peptide plantazolicin is associated with nematicidal activity of *Bacillus amyloliquefaciens* FZB42, *Appl. Microbiol. Biotechnol. 97*, 10081-10090.

32. Ziegler, S., Pries, V., Hedberg, C., and Waldmann, H. (2013) Target identification for small bioactive molecules: finding the needle in the haystack, *Angew. Chem. Int. Ed. 52*, 2744-2792.

33. Li, Z., Hao, P., Li, L., Tan, C. Y., Cheng, X., Chen, G. Y., Sze, S. K., Shen, H. M., and Yao, S. Q. (2013) Design and synthesis of minimalist terminal alkyne-containing diazirine photo-crosslinkers and their incorporation into kinase inhibitors for cell- and tissue-based proteome profiling, *Angew. Chem. Int. Ed. 52*, 8551-8556.

34. Burdine, L., and Kodadek, T. (2004) Target identification in chemical genetics: the (often) missing link, *Chem. Biol. 11*, 593-597.

35. Sahl, H. G., and Brandis, H. (1982) Mode of action of the staphylococcin-like peptide Pep 5 and culture conditions effecting its activity, *Zentralbl. Bakteriol. Mikrobiol. Hyg. A 252*, 166-175.

36. Brazas, M. D., and Hancock, R. E. (2005) Using microarray gene signatures to elucidate mechanisms of antibiotic action and resistance, *Drug Discovery Today 10*, 1245-1252.

37. Shaw, K. J., and Morrow, B. J. (2003) Transcriptional profiling and drug discovery, *Curr. Opin. Pharmacol. 3*, 508-512.

38. Li, X. Z., and Nikaido, H. (2009) Efflux-mediated drug resistance in bacteria: an update, *Drugs 69*, 1555-1623.

39. Jordan, S., Hutchings, M. I., and Mascher, T. (2008) Cell envelope stress response in Gram-positive bacteria, *FEMS Microbiol. Rev. 32*, 107-146.

40. Wecke, T., Zuhlke, D., Mader, U., Jordan, S., Voigt, B., Pelzer, S., Labischinski, H., Homuth, G., Hecker, M., and Mascher, T. (2009) Daptomycin versus Friulimicin B: in-depth profiling of *Bacillus subtilis* cell envelope stress responses, *Antimicrob. Agents Chemother. 53*, 1619-1623.

41. Martin, M., and de Mendoza, D. (2013) Regulation of *Bacillus subtilis* DesK thermosensor by lipids, *Biochem. J. 451*, 269-275.

42. Berenbaum, M. C. (1989) What is synergy?, *Pharmacol. Rev. 41*, 93-141.

43. Ruhr, E., and Sahl, H. G. (1985) Mode of action of the peptide antibiotic nisin and influence on the membrane potential of whole cells and on cytoplasmic and artificial membrane vesicles, *Antimicrob. Agents Chemother. 27*, 841-845.

44. Silverman, J. A., Perlmutter, N. G., and Shapiro, H. M. (2003) Correlation of daptomycin bactericidal activity and membrane depolarization in *Staphylococcus aureus*, *Antimicrob. Agents Chemother. 47*, 2538-2544.

45. Tiyanont, K., Doan, T., Lazarus, M. B., Fang, X., Rudner, D. Z., and Walker, S. (2006) Imaging peptidoglycan biosynthesis in *Bacillus subtilis* with fluorescent antibiotics, *Proc. Natl. Acad. Sci. U. S. A. 103*, 11033-11038.

46. Bindman, N. A., and van der Donk, W. A. (2013) A general method for fluorescent labeling of the N-termini of lanthipeptides and its application to visualize their cellular localization, *J. Am. Chem. Soc. 135*, 10362-10371.

47. Huang, B., Wang, W., Bates, M., and Zhuang, X. (2008) Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy, *Science 319*, 810-813.

48. Nguyen-Mau, S. M., Oh, S. Y., Kern, V. J., Missiakas, D. M., and Schneewind, O. (2012) Secretion genes as determinants of *Bacillus anthracis* chain length, *J. Bacteriol. 194*, 3841-3850.

49. Silver, L. L. (2011) Challenges of antibacterial discovery, *Clin. Microbiol. Rev. 24*, 71-109.

50. Deng, W., Li, C., and Xie, J. (2013) The underling mechanism of bacterial TetR/AcrR family transcriptional repressors, *Cell. Signal. 25*, 1608-1613.

51. Yang, D. C., Peters, N. T., Parzych, K. R., Uehara, T., Markovski, M., and Bernhardt, T. G. (2011) An ATP-binding cassette transporter-like complex governs cell-wall hydrolysis at the bacterial cytokinetic ring, *Proc. Natl. Acad. Sci. U. S. A. 108*, E1052-E1060.

52. Friedman, L., Alder, J. D., and Silverman, J. A. (2006) Genetic changes that correlate with reduced susceptibility to daptomycin in *Staphylococcus aureus*, *Antimicrob. Agents Chemother. 50*, 2137-2145.

53. Lee, J. Y., Janes, B. K., Passalacqua, K. D., Pfleger, B. F., Bergman, N. H., Liu, H., Hakansson, K., Somu, R. V., Aldrich, C. C., Cendrowski, S., Hanna, P. C., and Sherman, D. H. (2007) Biosynthetic analysis of the petrobactin siderophore pathway from *Bacillus anthracis*, *J. Bacteriol. 189*, 1698-1710.

54. Zhang, T., Muraih, J. K., Tishbi, N., Herskowitz, J., Victor, R. L., Silverman, J., Uwumarenogie, S., Taylor, S. D., Palmer, M., and Mintzer, E. (2014) Cardiolipin prevents membrane translocation and permeabilization by daptomycin, *J. Biol. Chem. 289*, 11584-11591.

55. Palmer, K. L., Daniel, A., Hardy, C., Silverman, J., and Gilmore, M. S. (2011) Genetic basis for daptomycin resistance in Enterococci, *Antimicrob. Agents Chemother. 55*, 3345-3356.

56. Romantsov, T., Guan, Z., and Wood, J. M. (2009) Cardiolipin and the osmotic stress responses of bacteria, *Biochim. Biophys. Acta 1788*, 2092-2100.

57. Peters, A. C., Thomas, L., and Wimpenny, J. W. (1991) Effects of salt concentration on bacterial growth on plates with gradients of pH and temperature, *FEMS Microbiol. Lett. 61*, 309-314.

58. Unsay, J. D., Cosentino, K., Subburaj, Y., and Garcia-Saez, A. J. (2013) Cardiolipin effects on membrane structure and dynamics, *Langmuir 29*, 15878-15887.

59. Aguilar, P. S., Cronan, J. E., Jr., and de Mendoza, D. (1998) A *Bacillus subtilis* gene induced by cold shock encodes a membrane phospholipid desaturase, *J. Bacteriol. 180*, 2194-2200.

60. Mileykovskaya, E., and Dowhan, W. (2000) Visualization of phospholipid domains in *Escherichia coli* by using the cardiolipin-specific fluorescent dye 10-N-nonyl acridine orange, *J. Bacteriol. 182*, 1172-1175.

61. Mileykovskaya, E., and Dowhan, W. (2009) Cardiolipin membrane domains in prokaryotes and eukaryotes, *Biochim. Biophys. Acta 1788*, 2084-2091.

62. Strahl, H., Burmann, F., and Hamoen, L. W. (2014) The actin homologue MreB organizes the bacterial cell membrane, *Nat. Commun. 5*, 3442.

63. Kellogg, D. S., Jr., Peacock, W. L., Jr., Deacon, W. E., Brown, L., and Pirkle, D. I. (1963) *Neisseria gonorrhoeae*. I. Virulence genetically linked to clonal variation, *J. Bacteriol. 85*, 1274-1279.

64. (2006) Clinical and Laboratory Standards Institute, In *Methods for Dilution Antimicrobial Susceptibility Tests for Bacteria that Grow Aerobically—Seventh Edition: Approved Standard*, CLSI, Wayne, PA, USA.

65. McClure, R., Balasubramanian, D., Sun, Y., Bobrovskyy, M., Sumby, P., Genco, C. A., Vanderpool, C. K., and Tjaden, B. (2013) Computational analysis of bacterial RNA-Seq data, *Nucleic Acids Res. 41*, e140.

66. Rust, M. J., Bates, M., and Zhuang, X. (2006) Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM), *Nat. Methods 3*, 793-795.

67. Fei, J., Singh, D., Zhang, Q., Park, S., Balasubramanian, D., Golding, I., Vanderpool, C. K., and Ha, T. (2015) RNA biochemistry. Determination of in vivo target search kinetics of regulatory noncoding RNA, *Science 347*, 1371-1374.

68. Bates, M., Huang, B., Dempsey, G. T., and Zhuang, X. (2007) Multicolor super-resolution imaging with photo-switchable fluorescent probes, *Science 317*, 1749-1753.

Appendix C: Crystal structures of nelfinavir analogs

This chapter is taken in part from Maxson et al. and is reproduced with permission from Acta Crystallographica Section E.

Crystal structure and absolute configuration of (3*S*,4aS,8aS)-*N*-(*tert*-butyl)-2-((*S*)-3-(2-chloro- 4-nitro-benzamido)-2-hy-droxy-propyl)-deca-hydro-iso-quinoline-3-carboxamide and (3*S*,4aS,8aS)-*N*-(*tert*-butyl)-2-((*S*)-2-((*S*)-1-(2-chloro-4-nitro-benzoyl)-pyrrolidin-2-yl)-2-hy-droxy-ethyl)-deca-hydro-iso-quinoline-3-carboxamide

Abstract

The crystal structure and absolute configuration of two new nelfinavir analogs, (3*S*,4aS,8aS)-*N*-(*tert*-butyl)-2-((*S*)-3-(2-chloro-4-nitro-benzamido)-2-hy-droxy-propyl)-deca-hydro-iso-quinoline-3-carboxamide (I) and (3*S*,4aS,8aS)-*N*-(*tert*-butyl)-2-((*S*)-2-((*S*)-1-(2-chloro-4-nitro-benzoyl)- pyrrolidin-2-yl)-2-hy-droxy-ethyl)-deca-hydro-iso-quinoline-3-carboxamide (II), have been determined. Each of these molecules exhibits various disordered moieties as well as both intra- and inter-molecular hydrogen bonding. The intra-molecular hydrogen bonding inter-actions in (I) [O—H···O] and in (II) [N—H···O] are between the two chains coming off of the deca-hydro-iso-quinoline group. The inter-molecular hydrogen bonding in (I) [O—H···O; N—H···O] leads to two-dimensional sheets extending along the *ac*-plane. The inter-molecular hydrogen bonding in (II) [O—H···O] leads to one-dimensional chains extending along the *a*-axis.

C.1 Chemical context

Nelfinavir (Viracept) is an FDA approved HIV protease inhibitor identified through structure-based design with a low nanomolar inhibitory concentration against the HIV aspartyl protease (Kaldor *et al.*, 1997). Although nelfinavir is no longer recommended as a first line treatment against HIV due to its inferior efficacy compared to alternative protease inhibitors (Panel on Anti-retroviral Guidelines, 2015), it has been found to have a number of additional biological activities that may have therapeutic utility including anti-viral (against other human viruses) (Yamamoto *et al.*, 2004; Kalu *et al.*, 2014), anti-cancer (Gantt *et al.*, 2013; Koltai, 2015), and anti-virulence activity (Maxson *et al.*, 2015). However, nelfinavir was originally designed with only the HIV protease active site in mind and the structure is likely not optimal for binding to the alternative targets involved in these other activities. We recently reported on

the synthesis of a collection of nelfinavir analogs that may be of inter-est for efforts to repurpose the drug (Maxson *et al.*, 2015).

The syntheses of the title compounds were achieved through a previously reported route that utilizes the configuration of the amino acid starting material to control the stereochemical outcome of the sodium borohydride reduction of the chloro-methyl ketone (Kaldor *et al.*, 1997). However, the reduction of compound (I), derived from achiral glycine, results in a racemic mixture (Figure C.1) while the reduction of compound (II), derived from L-proline, does not benefit from a strong directing influence from the existing chiral center (Figure C.2). The products of the two reductions were carried forward through the remainder of the synthesis to generate the title compounds. The absolute configurations of compounds (I) and (II) as well as the conformations they adopt due to the increased flexibility and rigidity, respectively, relative to nelfinavir was investigated through X-ray diffraction.

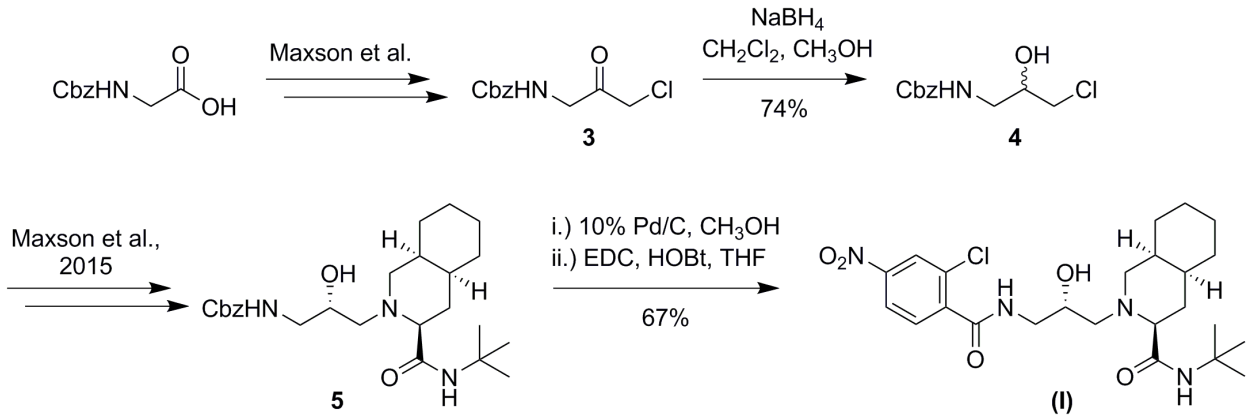


Figure C.1. Scheme showing the synthesis of (I).

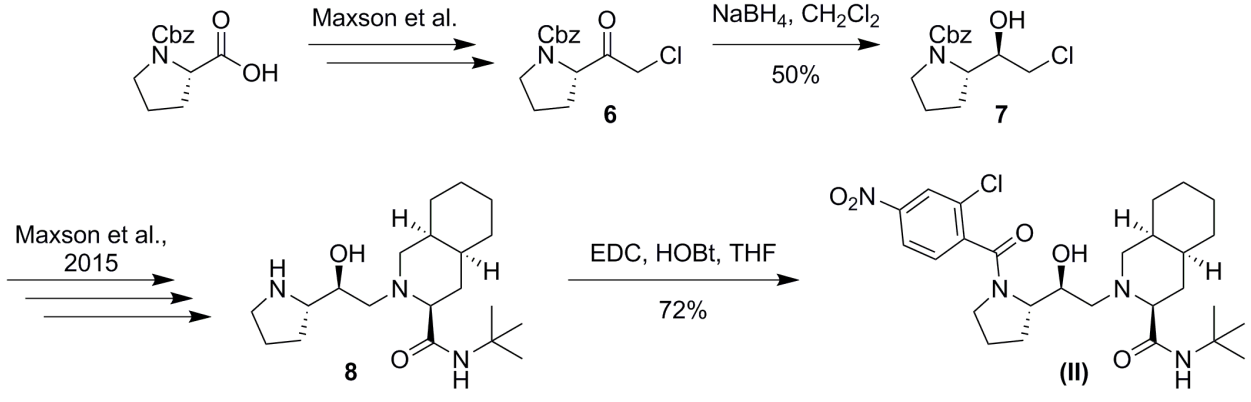


Figure C.2. Scheme showing the synthesis of (II).

C.2 Structural commentary

The core molecular structures of (I) and (II) are comprised of N-(tert-butyl)-2-(2-hy-droxy-alkyl)-deca-hydro-iso-quinoline-3-carboxamide groups. The difference between the two species comes from the substitution at the N position of the deca-hydro-iso-quinoline group. Compound (I) has a (2-chloro-4-nitro-benzamido)-2-hy-droxy-propyl group at the N position of the deca-hydro-iso-quinoline, Figure C.3. Compound (II) has a (2-chloro-4-nitro-benzoyl)-pyrrolidin- 2-yl)-2-hy-droxy-ethyl group at the N position, Figure C.4.

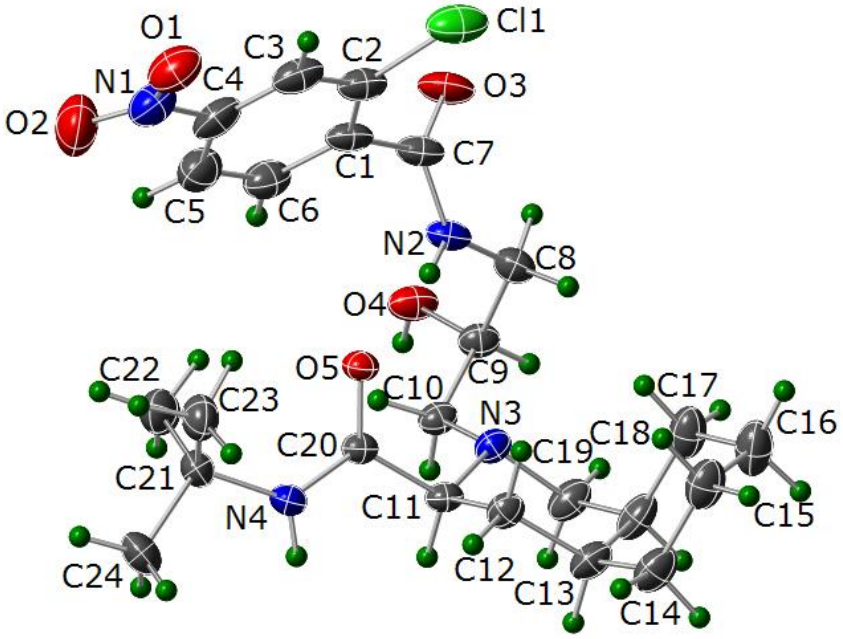


Figure C.3. Plot showing 35% probability ellipsoids for non-H atoms and circles of arbitrary size for H atoms for (I). Only the major component of disordered sites is shown.

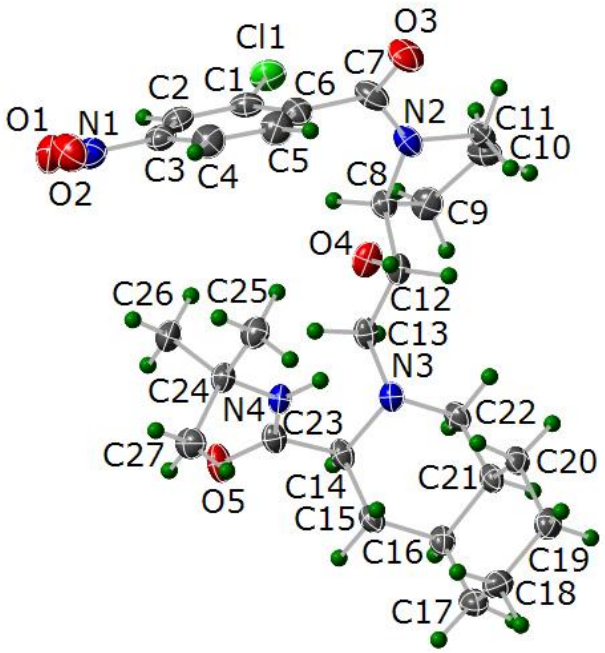


Figure C.4. Plot showing 35% probability ellipsoids for non-H atoms and circles of arbitrary size for H atoms for (II). Only the major component of disordered sites is shown.

There is disorder of the Cl group in (I) over two positions with the site occupancies refining to 0.941 (8):0.059 (8) for Cl1 and Cl1B respectively. The nitro group is disordered over two positions with the site occupancies for each orientation refining to 0.60 (2):0.40 (2). The NO_2 group in one orientation is essentially co-planar with the phenyl ring [O1B—N1B—C4—C3; t = 1(2)°] and in the other orientation is twisted slightly more out of plane [O1—N1—C4—C3; t = -9.0 (13)°]. Both six-membered rings of the deca-hydro-iso-quinoline group in (I) adopt a chair conformation, the dihedral angle between the best fit planes of the cyclo-hexyl and piperidine moieties is 119.9 (15)°. There is one intra-molecular hydrogen bonding inter-action in (I) between the two chains coming off of the deca-hydro-iso-quinoline group (N2—H2···O5), Table C.1. The Flack x parameter of -0.008 (18) and the Hooft y parameter of -0.010 (19) indicate that the absolute configuration of (I) has been assigned correctly.

D—H···*A*	*D*—H	H···*A*	*D*···*A*	*D*—H···*A*
O4—H4*B*···O5^{i}	0.87 (4)	1.94 (4)	2.791 (2)	169 (3)
N2—H2···O5	0.83 (3)	2.11 (3)	2.928 (3)	169 (3)
N4—H4···O3^{ii}	0.84 (3)	2.15 (3)	2.964 (2)	161 (3)

Table C.1. Hydrogen-bond geometry (Å, º) for (I)
Symmetry codes: (i) *x*, *y*, *z*+1; (ii) *x*+1/2, -*y*+3/2, -*z*+2.

There are multiple disordered moieties in (II), the nitro group is disordered over two positions with the site occupancies for the two orientations refining to 0.967 (6):0.033 (8). In both orientations the NO_2 group is twisted out of the plane of the phenyl ring; the major orientation is twisted out of the plane less [O1—N1—C3—C2 t = 10.9 (4)°] than the minor orientation [O1B—N1B—C3—C2 t = -26 (6)°]. The carbonyl group C7—O3 is disordered over two positions with the site occupancies refining to 0.58 (2):0.42 (2). In the minor orientation the CO group is nearly normal to the plane of the phenyl ring [O3B—C7B—C6—C5 t = -89 (3)°], while the major orientation is significantly less out of plane [O3—C7—C6—C5 t = -44 (3)°]. The final two disordered moieties of (II) are a portion of the pyrrolidin-2-yl group and the three methyl groups of the t-butyl. The carbon atoms C10 and C11 of the pyrrolidin-2-yl are disordered over two positions, site occupancies 0.669 (16):0.331 (16). The t-butyl methyl groups also disorder over two positions via a slight rotation around the N4—C24 bond, the site occupancies refine to 0.811 (17):0.189 (17). Similar to (I), both six-membered rings of the deca-hydro-iso-quinoline group in (II) adopt a chair conformation with the dihedral angle between the best fit planes of the cyclo-hexyl and piperidine moieties of 116.3 (17)°. There is one intra-molecular hydrogen bonding inter-action in (II) between the two chains coming off of the deca-hydro-iso-quinoline group (N4—H4C···O4), Table C.2. The Flack x parameter of 0.036 (19) and the Hooft y parameter of 0.03 (2) indicate that the absolute configuration of (II) has been assigned correctly.

D—H···*A*	*D*—H	H···*A*	*D*···*A*	*D*—H···*A*
N4—H4*C*···O4	0.88 (3)	2.60 (3)	3.219 (3)	129 (3)
O4—H4*B*···O5^{i}	0.82 (1)	1.89 (2)	2.709 (2)	170 (4)

Table C.2. Hydrogen-bond geometry (Å, º) for (II)
Symmetry code: (i) *x*-1, *y*, *z*.

C.3 Supra-molecular features

The extended structure of (I) is a two-dimensional sheet of hydrogen bonded molecules extending along the *ac*-plane, Figure C.5a. Each molecule of (I) is inter-molecularly hydrogen bonded to four neighboring molecules *via* O—H···O and N—H···O inter-actions, the details of these inter-actions can be found in Table C.1. The two-dimensional layers stack in an ABAB pattern along the crystallographic *b*-axis, Figure C.5b. The layers are separated by the bulky deca-hydro-iso-quinoline groups which protrude above and below the sheets. The layers alternate between these bulky groups pointing "left" and "right", this along with a slight offset between the A and B layers allows them to inter-digitate.

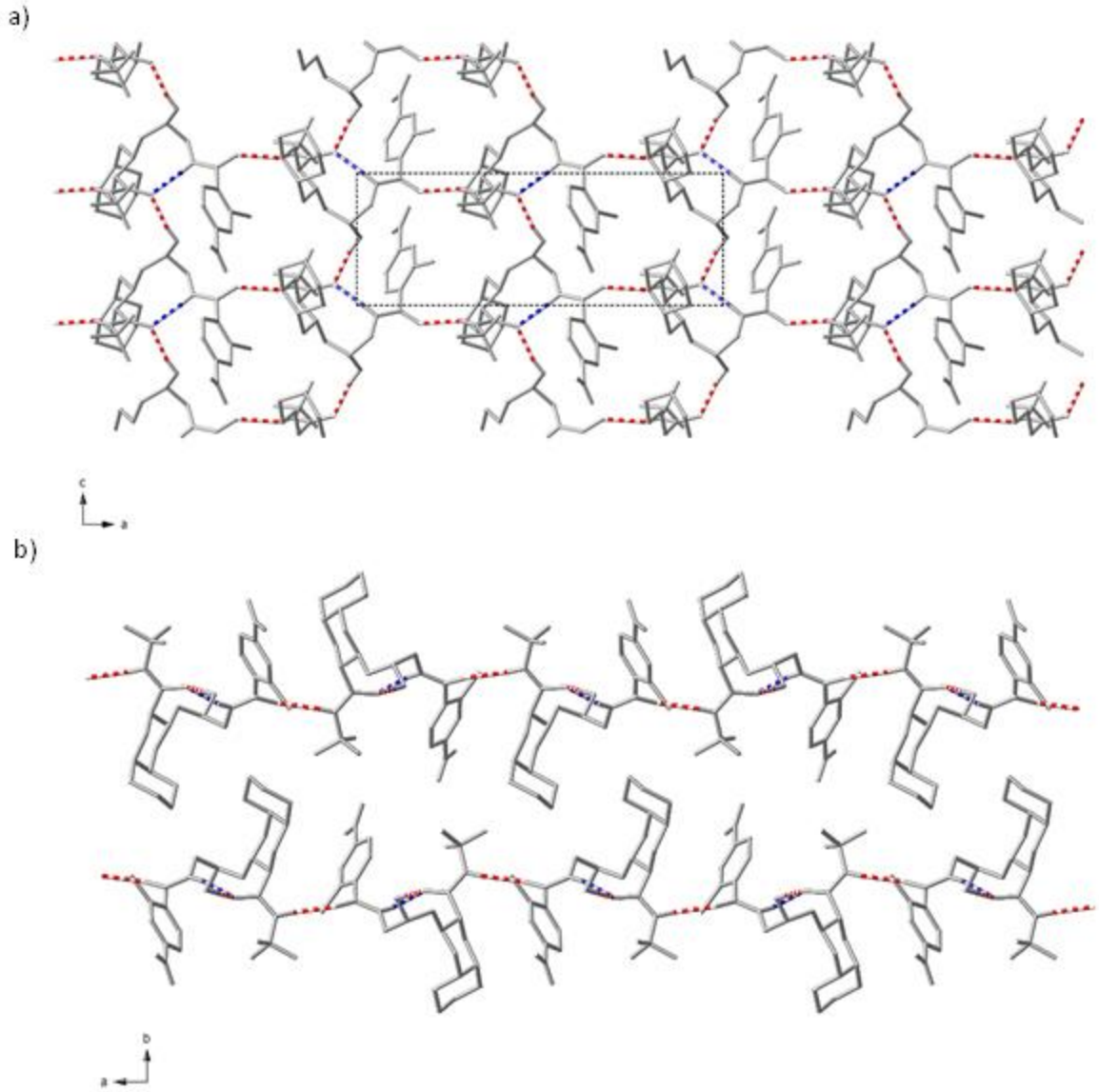


Figure C.5. A plot of the packing of (I) viewed a) along the *b* axis showing a hydrogen bonded two-dimensional sheet overlain with the unit cell; and b) along the *c* axis showing how two layers stack together along the *b* axis. Only the major component of disordered sites is shown. Red dashed lines indicate intermolecular hydrogen bonding; blue dashed lines indicate intramolecular hydrogen bonding.

The extended structure of (II) is a one-dimensional chain of hydrogen bonded molecules extending along the crystallographic *a*-axis, Figure C.6a. Each molecule of (II) is inter-molecularly hydrogen bonded to two neighboring molecules *via* O—H···O inter-actions, the details of these inter-actions can be found in Table C.2. The one-dimensional chains are separated by the bulky deca-hydro-iso-quinoline groups and the *t*-butyl groups which prevent the chains from linking *via* further

hydrogen bonding inter-actions, Figure C.6b.

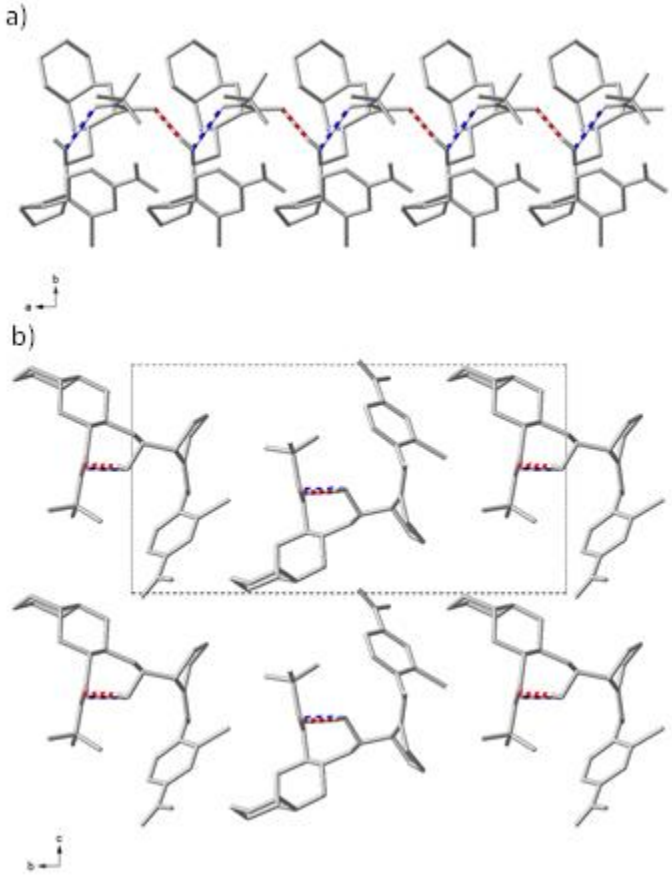


Figure C.6. A plot of the packing of (II) viewed a) along the *c* axis showing a hydrogen bonded one-dimensional chain; and b) along the *a* axis showing how the one-dimensional chains pack together overlain with the unit cell. Only the major component of disordered sites is shown. Red dashed lines indicate intermolecular hydrogen bonding; blue dashed lines indicate intramolecular hydrogen bonding.

C.4 Database survey

A search of the Cambridge Crystallographic Database (Groom & Allen, 2014) returns only three crystal structures with the N-(tert-butyl)-deca-hydro-iso-quinoline-3-carboxamide core. One of the structures is the N-(tert-butyl)-deca-hydro-iso-quinoline-3-carboxamide molecule [CSD code COVYAO] (Zhao *et al.*, 2006). The other two molecules are nelfinavir derivatives like (I) and (II) which were isolated during optimization of the synthesis. The difference between these two molecules comes *via* the substitution at the N position of the deca-hydro-iso-quinoline group.

One compound has a 3-amino-2-hy-droxy-4-(phenyl-thio)-butyl group in this position [CSD code

QONJUY] (Inaba *et al.*, 2000) and the other has a 3-acet-oxy-2-(3-acet-oxy-2- methyl-benzoyl-amino)-4-(phenyl-thio)-butyl group at the N position [CSD code GONKOJ] (Inaba *et al.*, 1998). Each of these molecules has intra-molecular hydrogen bonding between the two chains coming off of the deca-hydro-iso-quinoline group, similar to (I) and (II), involving the N—H of the carboxamide and an oxygen atom on the adjacent chain. The core structure of each of these previously reported materials is similar to (I) and (II) in that both six-membered rings of the deca-hydro-iso-quinoline groups adopt chair conformations. The dihedral angle between the best fit planes of the cyclo-hexyl and piperidine moieties for the 3-amino-2-hy-droxy-4-(phenyl-thio)-butyl substituted molecule is 117.1 (18)°. Similarly, this angle for the 3-acet-oxy-2-(3-acet-oxy-2-methyl-benzoyl-amino)-4-(phenyl-thio)-butyl substituted molecule is 116.8 (14)°.

C.5 Synthesis and crystallization

Compound (I) was synthesized through the inter-mediate chloro-methyl hydroxyl **4** (Figure C.1). Chloro-methyl ketone **3** (571 mg, 2.36 mmol) was dissolved in di-chloro-methane (7 mL) and methanol (4 mL) under nitro-gen. The reaction was cooled to 0 °C and sodium borohydride (63 mg, 1.65 mmol) was added in one portion. The reaction was stirred cold for 1h before being quenched by the slow addition of 2 M HCl (2 mL). The reaction was dried and the solid was dissolved in ethyl acetate. The product was washed twice with water and once with brine, dried over sodium sulfate, and concentrated by rotary evaporation. The product was purified by silica flash column chromatography (gradient of 0-8% EtOAc in DCM) to yield racemic **4** as a colorless oil (423 mg, 75% yield). ^{1}H NMR (500 MHz, $CDCl_3$) d ppm 7.33-7.28 (complex, 5H), 5.63 (t, J = 6 Hz, 1H), 5.06 (s, 2H), 3.88 (s, 2H), 3.48 (m, 2H), 3.39 (m, 1H), 3.22 (m, 1H). ^{13}C NMR (500 MHz, $CDCl_3$) d ppm 157.23, 135.93, 128.36, 128.06, 127.91, 70.52, 66.90, 46.44, 43.96. HRMS (*m/z*): $[M + H]^+$ calc. for $C_{11}H_{15}NO_3Cl$, 244.0740; observed, 244.0741. For the synthesis of compound (I), compound **5** (104 mg, 0.233 mmol) was dissolved in methanol (15 mL) with 10% palladium on carbon (74 mg, 0.070 mmol). The solution was degassed for 30 min before being placed under 1 atm of hydrogen and stirred for 2 h at room temperature. The reaction was filtered through celite, dried to a solid, and taken up in tetra-hydro-furan (5 mL). 2-chloro-4-nitro-benzoic acid (52 mg,

0.256 mmol), 1- ethyl-3-(3-di-methyl-amino-propyl)-carbodiimide hydro-chloride (49 mg, 0.256 mmol), and hydroxy-benzotriazole hydrate (42 mg, 0.256 mmol) were added and the reaction was stirred at room temperature overnight. The reaction was taken up in ethyl acetate, washed once with sodium bicarbonate and once with brine, and dried over sodium sulfate. The product was purified by silica flash column chromatography (gradient of 0-3% MeOH in DCM) to yield (I) as a yellow solid (77 mg, 67%). Crystals suitable for X-ray diffraction were obtained from the vapor diffusion of pentane into a solution of compound (I) in ethyl acetate at room temperature. 1H NMR (500 MHz, CDCl3) d ppm 8.41 (q, J = 4 Hz, 1H), 8.24 (d, J = 2 Hz, 1H), 8.13 (dd, J1 = 2 Hz, J2 = 8.5 Hz, 1H), 7.76 (d, J = 8.5 Hz, 1H), 5.60 (s, 1H), 4.04 (m, 2H), 3.47 (dt, J1 = 4 Hz, J2 = 14 Hz, 1H), 3.35 (br, 1H), 2.71 (dd, J1 = 2 Hz, J2 = 11.5 Hz, 1H), 2.49 (dd, J1 = 3 Hz, J2 = 11.5 Hz, 1H), 2.36 (dd, J1 = 10 Hz, J2 = 12.5 Hz, 1H), 2.22 (dd, J1 = 5 Hz, J2 = 12.5 Hz, 1H), 2.18 (dd, J1 = 3 Hz, J2 = 11.5 Hz, 1H), 1.95 (q, J = 12 Hz, 1H), 1.80-1.08 (complex, 20H). 13C NMR (500 MHz, CDCl3) d ppm 174.16, 167.06, 148.39, 142.00, 132.80, 130.18, 124.96, 121.56, 70.40, 68.29, 59.09, 57.54, 51.27, 43.27, 35.83, 33.55, 31.02, 30.86, 28.39, 26.19, 25.52, 20.18. HRMS (m/z): [M + H]+ calc. for $C_{24}H_{36}N_4O_5Cl$, 495.2374; observed, 495.2376.

Compound (II) was synthesized through the inter-mediate chloro-methyl hydroxyl **7** (Figure C.2). Chloro-methyl ketone **6** (860 mg, 3.05 mmol) was dissolved in di-chloro-methane (7 mL) and methanol (4 mL) under nitro-gen. The reaction was cooled to 0 °C and sodium borohydride (81 mg, 2.14 mmol) was added in one portion. The reaction was stirred cold for 1h before being quenched by the slow addition of 2 M HCl (2 mL). The reaction was dried and the solid was dissolved in ethyl acetate. The product was washed twice with water and once with brine, dried over sodium sulfate, and concentrated by rotary evaporation. TLC analysis showed two diastereomers with the higher Rf compound being the (S,R) product. Both diastereomers were purified by silica flash column chromatography (gradient of 0-10% EtOAc in DCM) to yield the (S,S)-isomer as a white solid (279 mg, 32%) and (S,R)-isomer (7) as a white solid (429 mg, 50%). Characterization of the (S,R)-isomer (7): 1H NMR (500 MHz, CDCl3) d ppm 7.37-7.28 (complex, 5H), 5.13 (dd, J1 = 12.5 Hz, J2 = 16 Hz, 2H), 4.95 (d, J = 2 Hz, 1H), 4.11 (m, 1H), 3.81 (br s, 1H), 3.72 (d, J = 11 Hz, 1H), 3.55 (m, 2H), 3.37 (m, 1H), 2.03 (m, 1H), 1.89 (m, 1H), 1.81 (m, 1H),

1.72 (m, 1H). 13C NMR (500 MHz, CDCl3) d ppm 157.52, 136.04, 128.27, 127.87, 127.65, 74.69, 67.22, 60.57, 47.91, 47.05, 28.12, 23.94. HRMS (m/z): [M + H]+ calc. for $C_{14}H_{19}NO_3Cl$, 284.1053; observed, 284.1055. For the synthesis of compound (II), compound **8** (218 mg, 0.620 mmol) was dissolved in tetra-hydro-furan (6 mL) with 2-chloro-4-nitro-benzoic acid (138 mg, 0.682 mmol), 1-ethyl-3-(3-di-methyl-amino- propyl)-carbodiimide hydro-chloride (131 mg, 0.682 mmol), and hydroxy-benzotriazole hydrate (111 mg, 0.682 mmol). The reaction was stirred at room temperature overnight. The reaction was taken up in ethyl acetate, washed once with sodium bicarbonate and once with brine, and dried over sodium sulfate. The product was purified by silica flash column chromatography (gradient of 0-5% MeOH in DCM) to yield (II) as a yellow solid (248 mg, 72%). Crystals suitable for X-ray diffraction were obtained by layering pentane over a solution of compound (II) in di-chloro-methane at room temperature. 1H NMR (500 MHz, CDCl3) d ppm 8.31 (d, J = 2 Hz, 1H), 8.20 (dd, J1 = 2 Hz, J2 = 8.5 Hz, 1H), 7.54 (d, J = 8.5 Hz, 1H), 6.87 (s, 1H), 5.31 (s, 1H), 4.36 (m, 1H), 3.99 (m, 1H), 3.24 (m, 2H), 2.91 (d, J = 11 Hz, 1H), 2.63 (m, 2H), 2.18-1.13 (complex, 26H). 13C NMR (500 MHz, CDCl3) d ppm 173.83, 172.95, 148.31, 142.45, 128.53, 124.96, 122.35, 121.69, 69.81, 69.73, 60.88, 58.37, 57.98, 50.55, 50.51, 49.05, 35.84, 33.23, 31.07, 30.80, 28.56, 28.20, 26.20, 25.46, 24.53, 20.16. HRMS (m/z): [M + H]+ calc. for $C_{27}H_{40}N_4O_5Cl$, 535.2687; observed, 535.2692.

C.6 Refinement

Crystal data, data collection and structure refinement details for (I) and (II) are summarized in

	(I)	(II)
Crystal data		
Chemical formula	$C_{24}H_{35}ClN_4O_5$	$C_{27}H_{39}ClN_4O_5$
M_r	495.01	535.07
Crystal system, space group	Orthorhombic, $P2_12_12$	Monoclinic, $P2_1$
Temperature (K)	193	168
a, *b*, *c* (Å)	18.8408 (7), 20.2263 (8), 6.7923 (3)	6.4341 (7), 20.280 (2), 11.0377 (12)
a, b, g (°)	90, 90, 90	90, 105.248 (1), 90

Table C.3. Experimental details

V (Å^3)	2588.41 (18)	1389.5 (3)
Z	4	2
Radiation type	Mo Ka	Mo Ka
m (mm^{-1})	0.19	0.18
Crystal size (mm)	0.37 × 0.36 × 0.29	0.86 × 0.65 × 0.15
Data collection		
Diffractometer	Siemens Platform/ApexII CCD diffractometer	Siemens Platform/ApexII CCD diffractometer
Absorption correction	Integration *SHELXTL*/*XPREP* V2005/2 (Bruker, 2014)	Integration *SHELXTL*/*XPREP* V2005/2 (Bruker, 2014)
T_{min}, T_{max}	0.953, 0.960	0.892, 0.980
No. of measured, independent and observed [$I > 2s(I)$] reflections	30342, 5243, 4694	16374, 5627, 5222
R_{int}	0.027	0.024
(sin q/l)$_{max}$ (Å^{-1})	0.623	0.625
Refinement		
$R[F^2 > 2s(F^2)]$, $wR(F^2)$, S	0.035, 0.085, 1.03	0.032, 0.082, 1.04
No. of reflections	5243	5627
No. of parameters	333	377
No. of restraints	53	14
H-atom treatment	H atoms treated by a mixture of independent and constrained refinement	H atoms treated by a mixture of independent and constrained refinement
Dρ$_{max}$, Dρ$_{min}$ (e Å^{-3})	0.34, -0.43	0.24, -0.21
Absolute structure	Flack (1983), 2720 Friedels Hooft, Straver and Spek (2008)	Flack (1983), 2720 Friedels Hooft, Straver and Spek (2008)
Absolute structure parameter	-0.008 (18)	0.036 (19)

Table C.3. (cont.)
Computer programs: *APEX2* V2010.11-3 (Bruker, 2014), *SAINT* V7.68A (Bruker, 2014), *SAINT* V7.68A, *XPREP* V2005/2, *SADABS* V2008/2 (Bruker, 2014), *SHELX2014*-4 (Sheldrick, 2015), *SHELX2014*-6 (Sheldrick, 2015), *SHELXTL* V6.12 (Sheldrick, 2015), *CrystalMaker* v2.1.3 (*CrystalMaker*, 1994), XCIF V6.12 (Bruker, 2014), *publCIF* (Westrip, 2010).

Structural models consisting of the target molecules were developed for (I) and (II). Several disordered sites on each molecule were modelled with disorder. In each case, like distances were restrained to be similar. Since the major and minor components of each disordered site are in such close proximity to each other, the displacement parameters were constrained to be equal. Methyl H atom positions, R—CH_3, were optimized by rotation about R—C bonds with idealized C—H, R–H and H–H distances. All hydroxyl and amine H atoms were found in the difference map in good hydrogen bonding environments (Hamilton & Ibers 1968) and their distances were allowed to refine. The O4—H4B distance in (II) was restrained to be 0.84Å (esd 0.02). Remaining H atoms were included as riding idealized contributors. Methyl, hydroxyl and amine H atom U's were assigned as 1.5 times U_{eq} of the carrier atom; remaining H atom U's were assigned as 1.2 times carrier U_{eq}. On the basis of 2237 unmerged Friedel opposites, the fractional contribution of the racemic twin was negligible (Flack, 1983; Flack & Bernardinelli, 2000) for (I). The absolute structure parameter *y* was calculated using *PLATON* (Spek, 2009). The resulting value was *y*=-0.010 (19) indicating that the absolute structure has been determined correctly (Hooft, Straver and Spek, 2008). On the basis of 2720 unmerged Friedel opposites, the fractional contribution of the racemic twin was negligible (Flack, 1983; Flack & Bernardinelli, 2000) for (II). The absolute structure parameter *y* was calculated using *PLATON* (Spek, 2009). The resulting value was *y*=0.03 (2) indicating that the absolute structure has been determined correctly (Hooft, Straver and Spek, 2008).

C.7 Supporting information

Computing details

For both compounds, data collection: *APEX2* V2010.11-3 (Bruker, 2014); cell refinement: *SAINT* V7.68A (Bruker, 2014); data reduction: *SAINT* V7.68A, *XPREP* V2005/2, *SADABS* V2008/2 (Bruker, 2014); program(s) used to solve structure: *SHELX2014*-4 (Sheldrick, 2015); program(s) used to refine structure: *SHELX2014*-6 (Sheldrick, 2015); molecular graphics: *SHELXTL* V6.12 (Sheldrick, 2015), *CrystalMaker* v2.1.3 (*CrystalMaker*, 1994); software used to prepare material for publication: XCIF

V6.12 (Bruker, 2014), *publCIF* (Westrip, 2010).

(I)

Crystal data

$C_{24}H_{35}ClN_4O_5$	$F(000) = 1056$
$M_r = 495.01$	$D_x = 1.270$ Mg m^{-3}
Orthorhombic, $P2_12_12$	Mo $K\alpha$ radiation, $\lambda = 0.71073$ Å
Hall symbol: P 2 2ab	Cell parameters from 9966 reflections
$a = 18.8408$ (7) Å	$\theta = 2.3$–$24.5°$
$b = 20.2263$ (8) Å	$\mu = 0.19$ mm^{-1}
$c = 6.7923$ (3) Å	$T = 193$ K
$V = 2588.41$ (18) Å^3	Prism, colourless
$Z = 4$	0.37 × 0.36 × 0.29 mm

Data collection

Siemens Platform/ApexII CCD diffractometer	5243 independent reflections
Radiation source: normal-focus sealed tube	4694 reflections with $I > 2\sigma(I)$
Graphite monochromator	$R_{int} = 0.027$
profile data from f and w scans	$\theta_{max} = 26.3°$, $\theta_{min} = 1.5°$
Absorption correction: integration *SHELXTL*/*XPREP* V2005/2 (Bruker, 2014)	$h = -23 \rightarrow 23$
$T_{min} = 0.953$, $T_{max} = 0.960$	$k = -25 \rightarrow 25$
30342 measured reflections	$l = -8 \rightarrow 8$

Refinement

Refinement on F^2	Hydrogen site location: mixed
Least-squares matrix: full	H atoms treated by a mixture of independent and constrained refinement
$R[F^2 > 2\sigma(F^2)] = 0.035$	$w = 1/[\sigma^2(F_o^2) + (0.0355P)^2 + 0.5785P]$ where $P = (F_o^2 + 2F_c^2)/3$
$wR(F^2) = 0.085$	$(\Delta/\sigma)_{max} = 0.001$
$S = 1.03$	$\Delta\rho_{max} = 0.34$ e Å^{-3}
5243 reflections	$\Delta\rho_{min} = -0.43$ e Å^{-3}
333 parameters	Absolute structure: Flack (1983), 2720 Friedels Hooft, Straver and Spek (2008)

53 restraints	Absolute structure parameter: -0.008 (18)

Special details

Experimental. One distinct cell was identified using APEX2 (Bruker, 2014). Four frame series were integrated and filtered for statistical outliers using SAINT (Bruker, 2014) then corrected for absorption by integration using SHELXTL/XPREP V2005/2 (Bruker, 2014) before using SAINT/SADABS (Bruker, 2014) to sort, merge, and scale the combined data No decay correction was applied.
Geometry. All esds (except the esd in the dihedral angle between two l.s. planes) are estimated using the full covariance matrix. The cell esds are taken into account individually in the estimation of esds in distances, angles and torsion angles; correlations between esds in cell parameters are only used when they are defined by crystal symmetry. An approximate (isotropic) treatment of cell esds is used for estimating esds involving l.s. planes.
Refinement. Structure was phased by direct (Sheldrick, 2015). Systematic conditions suggested the ambiguous space group. The space group choice was confirmed by successful convergence of the full-matrix least-squares refinement on F^2. The final map had no significant features. A final analysis of variance between observed and calculated structure factors showed no dependence on amplitude or resolution.

Fractional atomic coordinates and isotropic or equivalent isotropic displacement parameters (Å^2)

	x	*y*	*z*	U_{iso}*/U_{eq}	Occ. (<1)
Cl1	0.29551 (15)	0.69804 (7)	0.67519 (15)	0.0735 (5)	0.941 (8)
Cl1B	0.326 (2)	0.6888 (7)	0.664 (3)	0.0735 (5)	0.059 (8)
O1	0.3558 (8)	0.8988 (8)	0.2519 (16)	0.086 (2)	0.60 (2)
O2	0.4046 (8)	0.9713 (5)	0.445 (2)	0.116 (3)	0.60 (2)
N1	0.3808 (6)	0.9169 (5)	0.4064 (12)	0.0715 (17)	0.60 (2)
O1B	0.3725 (13)	0.9134 (12)	0.271 (3)	0.086 (2)	0.40 (2)
O2B	0.4360 (11)	0.9754 (7)	0.453 (3)	0.116 (3)	0.40 (2)
N1B	0.3991 (9)	0.9260 (7)	0.431 (2)	0.0715 (17)	0.40 (2)
C1	0.38040 (11)	0.78173 (14)	0.8790 (4)	0.0477 (6)	
C2	0.34335 (12)	0.77084 (14)	0.7057 (4)	0.0520 (7)	
C3	0.34488 (14)	0.81498 (16)	0.5532 (4)	0.0585 (7)	
H3A	0.3207	0.8062	0.4333	0.070*	
C4	0.38270 (16)	0.87247 (16)	0.5797 (5)	0.0635 (8)	
C5	0.41913 (17)	0.88668 (15)	0.7505 (5)	0.0675 (8)	
H5A	0.4439	0.9272	0.7658	0.081*	
C6	0.41831 (14)	0.84036 (15)	0.8974 (5)	0.0594 (7)	

H6A	0.4443	0.8485	1.0148	0.071*	
C7	0.38161 (12)	0.73281 (14)	1.0445 (4)	0.0478 (6)	
O3	0.32755 (9)	0.71943 (11)	1.1373 (3)	0.0644 (5)	
O4	0.48956 (9)	0.73984 (10)	1.4849 (3)	0.0534 (5)	
H4B	0.5165 (17)	0.7479 (17)	1.585 (5)	0.080*	
O5	0.56221 (7)	0.75831 (7)	0.8387 (2)	0.0372 (3)	
N2	0.44567 (10)	0.70759 (11)	1.0807 (3)	0.0438 (5)	
H2	0.4786 (16)	0.7171 (15)	1.005 (5)	0.066*	
N3	0.60919 (9)	0.66293 (9)	1.1177 (3)	0.0368 (4)	
N4	0.67223 (10)	0.80363 (9)	0.8528 (3)	0.0417 (4)	
H4	0.7145 (16)	0.7950 (14)	0.884 (4)	0.063*	
C8	0.46111 (12)	0.66114 (13)	1.2385 (4)	0.0459 (6)	
H8A	0.4766	0.6186	1.1805	0.055*	
H8B	0.4172	0.6528	1.3145	0.055*	
C9	0.51828 (11)	0.68604 (12)	1.3772 (3)	0.0420 (5)	
H9A	0.5310	0.6499	1.4712	0.050*	
C10	0.58565 (11)	0.70901 (12)	1.2713 (3)	0.0386 (5)	
H10A	0.5767	0.7528	1.2109	0.046*	
H10B	0.6242	0.7144	1.3691	0.046*	
C11	0.65713 (11)	0.69212 (11)	0.9715 (3)	0.0380 (5)	
H11A	0.7036	0.7026	1.0353	0.046*	
C12	0.66899 (13)	0.64166 (11)	0.8056 (4)	0.0410 (5)	
H12A	0.6230	0.6318	0.7413	0.049*	
H12B	0.7010	0.6610	0.7054	0.049*	
C13	0.70138 (14)	0.57736 (12)	0.8829 (4)	0.0494 (6)	
H13A	0.7500	0.5880	0.9326	0.059*	
C14	0.70957 (16)	0.52528 (13)	0.7203 (4)	0.0590 (7)	
H14A	0.7303	0.5467	0.6025	0.071*	
H14B	0.7434	0.4910	0.7656	0.071*	
C15	0.64047 (16)	0.49205 (14)	0.6624 (4)	0.0611 (7)	
H15A	0.6503	0.4565	0.5661	0.073*	
H15B	0.6089	0.5248	0.5988	0.073*	
C16	0.60342 (19)	0.46287 (13)	0.8419 (5)	0.0696 (9)	
H16A	0.6333	0.4276	0.8999	0.084*	
H16B	0.5576	0.4430	0.8019	0.084*	

C17	0.59030 (17)	0.51683 (13)	0.9945 (4)	0.0610 (8)	
H17A	0.5575	0.5502	0.9387	0.073*	
H17B	0.5671	0.4971	1.1115	0.073*	
C18	0.65836 (16)	0.55060 (13)	1.0574 (4)	0.0550 (7)	
H18A	0.6883	0.5165	1.1243	0.066*	
C19	0.64454 (15)	0.60538 (13)	1.2068 (4)	0.0505 (6)	
H19A	0.6145	0.5878	1.3143	0.061*	
H19B	0.6902	0.6196	1.2650	0.061*	
C20	0.62618 (10)	0.75493 (11)	0.8828 (3)	0.0354 (5)	
C21	0.65543 (13)	0.87079 (12)	0.7767 (4)	0.0480 (6)	
C22	0.60827 (17)	0.90585 (14)	0.9239 (5)	0.0679 (8)	
H22A	0.5644	0.8805	0.9422	0.102*	
H22B	0.5967	0.9501	0.8746	0.102*	
H22C	0.6332	0.9097	1.0500	0.102*	
C23	0.62113 (16)	0.86682 (14)	0.5736 (5)	0.0606 (8)	
H23A	0.5739	0.8468	0.5852	0.091*	
H23B	0.6507	0.8397	0.4866	0.091*	
H23C	0.6167	0.9114	0.5184	0.091*	
C24	0.72639 (15)	0.90696 (15)	0.7599 (6)	0.0682 (9)	
H24A	0.7496	0.9079	0.8890	0.102*	
H24B	0.7182	0.9523	0.7145	0.102*	
H24C	0.7569	0.8838	0.6655	0.102*	

Atomic displacement parameters (Å^2)

	U^{11}	U^{22}	U^{33}	U^{12}	U^{13}	U^{23}
Cl1	0.0607 (12)	0.1010 (6)	0.0588 (5)	-0.0165 (6)	-0.0109 (5)	-0.0190 (5)
Cl1B	0.0607 (12)	0.1010 (6)	0.0588 (5)	-0.0165 (6)	-0.0109 (5)	-0.0190 (5)
O1	0.088 (6)	0.100 (6)	0.068 (2)	0.018 (4)	-0.011 (3)	0.008 (3)
O2	0.141 (8)	0.076 (2)	0.130 (3)	-0.009 (4)	-0.021 (6)	0.021 (2)
N1	0.071 (4)	0.068 (2)	0.075 (3)	0.023 (2)	-0.011 (3)	-0.005 (2)
O1B	0.088 (6)	0.100 (6)	0.068 (2)	0.018 (4)	-0.011 (3)	0.008 (3)
O2B	0.141 (8)	0.076 (2)	0.130 (3)	-0.009 (4)	-0.021 (6)	0.021 (2)
N1B	0.071 (4)	0.068 (2)	0.075 (3)	0.023 (2)	-0.011 (3)	-0.005 (2)
C1	0.0257 (10)	0.0711 (16)	0.0463 (14)	0.0075 (10)	-0.0005 (10)	-0.0111 (13)

C2	0.0330 (12)	0.0761 (18)	0.0471 (14)	0.0110 (12)	-0.0037 (10)	-0.0142 (14)
C3	0.0463 (15)	0.0788 (19)	0.0505 (15)	0.0225 (14)	-0.0057 (12)	-0.0146 (15)
C4	0.0602 (17)	0.0701 (19)	0.0601 (18)	0.0332 (15)	0.0033 (15)	-0.0002 (15)
C5	0.0662 (19)	0.0567 (17)	0.079 (2)	0.0106 (14)	-0.0035 (17)	-0.0068 (16)
C6	0.0464 (14)	0.0732 (19)	0.0586 (17)	0.0057 (14)	-0.0091 (13)	-0.0124 (15)
C7	0.0284 (11)	0.0747 (17)	0.0404 (13)	-0.0032 (11)	-0.0024 (10)	-0.0109 (12)
O3	0.0297 (8)	0.1063 (15)	0.0571 (11)	-0.0049 (9)	0.0044 (8)	0.0005 (12)
O4	0.0413 (9)	0.0750 (13)	0.0440 (9)	0.0114 (9)	-0.0038 (8)	-0.0169 (9)
O5	0.0280 (7)	0.0430 (8)	0.0407 (8)	-0.0008 (6)	-0.0039 (6)	0.0031 (7)
N2	0.0275 (9)	0.0655 (14)	0.0384 (10)	-0.0050 (9)	-0.0011 (8)	-0.0050 (10)
N3	0.0375 (9)	0.0411 (10)	0.0318 (9)	0.0061 (8)	-0.0012 (8)	0.0013 (8)
N4	0.0298 (9)	0.0432 (10)	0.0521 (12)	-0.0032 (8)	-0.0044 (9)	0.0001 (10)
C8	0.0373 (12)	0.0565 (14)	0.0440 (13)	-0.0060 (11)	0.0017 (10)	-0.0014 (12)
C9	0.0383 (12)	0.0525 (13)	0.0352 (11)	0.0052 (10)	-0.0004 (9)	-0.0039 (11)
C10	0.0334 (11)	0.0482 (13)	0.0342 (11)	0.0029 (9)	-0.0056 (9)	-0.0047 (10)
C11	0.0282 (10)	0.0457 (12)	0.0400 (12)	0.0033 (9)	-0.0032 (9)	-0.0016 (10)
C12	0.0393 (12)	0.0460 (12)	0.0378 (12)	0.0059 (10)	0.0050 (10)	0.0019 (10)
C13	0.0482 (13)	0.0531 (14)	0.0470 (14)	0.0192 (12)	0.0032 (12)	0.0006 (12)
C14	0.0673 (17)	0.0557 (16)	0.0540 (16)	0.0224 (14)	0.0116 (14)	0.0021 (13)
C15	0.0818 (19)	0.0466 (14)	0.0549 (16)	0.0096 (13)	0.0141 (15)	-0.0024 (14)
C16	0.097 (2)	0.0436 (14)	0.069 (2)	0.0034 (15)	0.0223 (19)	0.0010 (15)
C17	0.085 (2)	0.0426 (14)	0.0550 (16)	0.0062 (14)	0.0222 (15)	0.0059 (12)
C18	0.0689 (17)	0.0510 (14)	0.0451 (14)	0.0261 (14)	0.0048 (13)	0.0117 (12)
C19	0.0580 (15)	0.0566 (15)	0.0369 (13)	0.0178 (12)	-0.0022 (12)	0.0067 (12)
C20	0.0301 (10)	0.0414 (11)	0.0348 (11)	0.0002 (9)	-0.0006 (9)	-0.0043 (9)
C21	0.0411 (13)	0.0378 (12)	0.0653 (17)	-0.0065	-0.0019	-0.0007

				(10)	(12)	(12)
C22	0.0653 (18)	0.0512 (15)	0.087 (2)	0.0017 (14)	0.0045 (17)	-0.0127 (16)
C23	0.0665 (18)	0.0475 (14)	0.0678 (19)	-0.0086 (13)	-0.0126 (15)	0.0125 (14)
C24	0.0540 (16)	0.0529 (16)	0.098 (2)	-0.0193 (13)	-0.0009 (16)	0.0024 (17)

Geometric parameters (Å, º) for (I)

Cl1—C2	1.739 (3)	C11—C20	1.522 (3)
Cl1B—C2	1.716 (12)	C11—C12	1.536 (3)
O1—N1	1.207 (7)	C11—H11A	1.0000
O2—N1	1.217 (7)	C12—C13	1.529 (3)
N1—C4	1.482 (7)	C12—H12A	0.9900
O1B—N1B	1.226 (9)	C12—H12B	0.9900
O2B—N1B	1.228 (10)	C13—C14	1.534 (4)
N1B—C4	1.511 (9)	C13—C18	1.535 (4)
C1—C2	1.386 (3)	C13—H13A	1.0000
C1—C6	1.390 (4)	C14—C15	1.517 (4)
C1—C7	1.498 (4)	C14—H14A	0.9900
C2—C3	1.368 (4)	C14—H14B	0.9900
C3—C4	1.376 (5)	C15—C16	1.524 (4)
C3—H3A	0.9500	C15—H15A	0.9900
C4—C5	1.379 (5)	C15—H15B	0.9900
C5—C6	1.368 (4)	C16—C17	1.525 (4)
C5—H5A	0.9500	C16—H16A	0.9900
C6—H6A	0.9500	C16—H16B	0.9900
C7—O3	1.228 (3)	C17—C18	1.514 (4)
C7—N2	1.333 (3)	C17—H17A	0.9900
O4—C9	1.418 (3)	C17—H17B	0.9900
O4—H4B	0.87 (4)	C18—C19	1.525 (4)
O5—C20	1.244 (2)	C18—H18A	1.0000
N2—C8	1.455 (3)	C19—H19A	0.9900
N2—H2	0.83 (3)	C19—H19B	0.9900
N3—C11	1.467 (3)	C21—C22	1.514 (4)
N3—C10	1.467 (3)	C21—C23	1.526 (4)

N3—C19	1.471 (3)	C21—C24	1.528 (3)
N4—C20	1.329 (3)	C22—H22A	0.9800
N4—C21	1.487 (3)	C22—H22B	0.9800
N4—H4	0.84 (3)	C22—H22C	0.9800
C8—C9	1.517 (3)	C23—H23A	0.9800
C8—H8A	0.9900	C23—H23B	0.9800
C8—H8B	0.9900	C23—H23C	0.9800
C9—C10	1.531 (3)	C24—H24A	0.9800
C9—H9A	1.0000	C24—H24B	0.9800
C10—H10A	0.9900	C24—H24C	0.9800
C10—H10B	0.9900		
O1—N1—O2	127.3 (9)	H12A—C12—H12B	107.9
O1—N1—C4	121.0 (9)	C12—C13—C14	112.2 (2)
O2—N1—C4	111.6 (9)	C12—C13—C18	110.74 (19)
O1B—N1B—O2B	120.6 (14)	C14—C13—C18	111.5 (2)
O1B—N1B—C4	111.2 (14)	C12—C13—H13A	107.4
O2B—N1B—C4	128.1 (14)	C14—C13—H13A	107.4
C2—C1—C6	118.1 (3)	C18—C13—H13A	107.4
C2—C1—C7	122.7 (2)	C15—C14—C13	113.9 (2)
C6—C1—C7	119.2 (2)	C15—C14—H14A	108.8
C3—C2—C1	121.9 (3)	C13—C14—H14A	108.8
C3—C2—Cl1B	120.7 (7)	C15—C14—H14B	108.8
C1—C2—Cl1B	113.0 (8)	C13—C14—H14B	108.8
C3—C2—Cl1	118.2 (2)	H14A—C14—H14B	107.7
C1—C2—Cl1	119.8 (2)	C14—C15—C16	110.9 (3)
C2—C3—C4	117.6 (3)	C14—C15—H15A	109.5
C2—C3—H3A	121.2	C16—C15—H15A	109.5
C4—C3—H3A	121.2	C14—C15—H15B	109.5
C3—C4—C5	123.0 (3)	C16—C15—H15B	109.5
C3—C4—N1	113.4 (5)	H15A—C15—H15B	108.0
C5—C4—N1	123.7 (5)	C15—C16—C17	109.9 (2)
C3—C4—N1B	128.6 (7)	C15—C16—H16A	109.7
C5—C4—N1B	108.1 (7)	C17—C16—H16A	109.7
C6—C5—C4	117.7 (3)	C15—C16—H16B	109.7

C6—C5—H5A	121.1	C17—C16—H16B	109.7
C4—C5—H5A	121.1	H16A—C16—H16B	108.2
C5—C6—C1	121.6 (3)	C18—C17—C16	112.2 (3)
C5—C6—H6A	119.2	C18—C17—H17A	109.2
C1—C6—H6A	119.2	C16—C17—H17A	109.2
O3—C7—N2	124.9 (3)	C18—C17—H17B	109.2
O3—C7—C1	121.2 (2)	C16—C17—H17B	109.2
N2—C7—C1	113.9 (2)	H17A—C17—H17B	107.9
C9—O4—H4B	109 (2)	C17—C18—C19	111.8 (2)
C7—N2—C8	124.3 (2)	C17—C18—C13	112.8 (2)
C7—N2—H2	118 (2)	C19—C18—C13	110.4 (2)
C8—N2—H2	117 (2)	C17—C18—H18A	107.2
C11—N3—C10	114.32 (17)	C19—C18—H18A	107.2
C11—N3—C19	108.55 (18)	C13—C18—H18A	107.2
C10—N3—C19	110.33 (18)	N3—C19—C18	112.3 (2)
C20—N4—C21	126.27 (19)	N3—C19—H19A	109.2
C20—N4—H4	115 (2)	C18—C19—H19A	109.2
C21—N4—H4	119 (2)	N3—C19—H19B	109.2
N2—C8—C9	112.7 (2)	C18—C19—H19B	109.2
N2—C8—H8A	109.1	H19A—C19—H19B	107.9
C9—C8—H8A	109.1	O5—C20—N4	123.7 (2)
N2—C8—H8B	109.1	O5—C20—C11	120.82 (19)
C9—C8—H8B	109.1	N4—C20—C11	115.44 (18)
H8A—C8—H8B	107.8	N4—C21—C22	108.9 (2)
O4—C9—C8	107.70 (19)	N4—C21—C23	110.9 (2)
O4—C9—C10	109.00 (19)	C22—C21—C23	111.9 (2)
C8—C9—C10	113.42 (19)	N4—C21—C24	106.1 (2)
O4—C9—H9A	108.9	C22—C21—C24	109.8 (2)
C8—C9—H9A	108.9	C23—C21—C24	109.2 (2)
C10—C9—H9A	108.9	C21—C22—H22A	109.5
N3—C10—C9	113.06 (18)	C21—C22—H22B	109.5
N3—C10—H10A	109.0	H22A—C22—H22B	109.5
C9—C10—H10A	109.0	C21—C22—H22C	109.5
N3—C10—H10B	109.0	H22A—C22—H22C	109.5
C9—C10—H10B	109.0	H22B—C22—H22C	109.5

H10A—C10—H10B	107.8	C21—C23—H23A	109.5
N3—C11—C20	111.60 (17)	C21—C23—H23B	109.5
N3—C11—C12	108.59 (18)	H23A—C23—H23B	109.5
C20—C11—C12	108.67 (18)	C21—C23—H23C	109.5
N3—C11—H11A	109.3	H23A—C23—H23C	109.5
C20—C11—H11A	109.3	H23B—C23—H23C	109.5
C12—C11—H11A	109.3	C21—C24—H24A	109.5
C13—C12—C11	111.81 (19)	C21—C24—H24B	109.5
C13—C12—H12A	109.3	H24A—C24—H24B	109.5
C11—C12—H12A	109.3	C21—C24—H24C	109.5
C13—C12—H12B	109.3	H24A—C24—H24C	109.5
C11—C12—H12B	109.3	H24B—C24—H24C	109.5
C6—C1—C2—C3	1.8 (4)	C19—N3—C10—C9	-77.0 (2)
C7—C1—C2—C3	-177.6 (2)	O4—C9—C10—N3	-165.33 (18)
C6—C1—C2—Cl1B	158.4 (14)	C8—C9—C10—N3	-45.4 (3)
C7—C1—C2—Cl1B	-20.9 (14)	C10—N3—C11—C20	-52.4 (2)
C6—C1—C2—Cl1	179.7 (2)	C19—N3—C11—C20	-176.07 (18)
C7—C1—C2—Cl1	0.3 (3)	C10—N3—C11—C12	-172.21 (17)
C1—C2—C3—C4	-2.3 (4)	C19—N3—C11—C12	64.2 (2)
Cl1B—C2—C3—C4	-157.2 (15)	N3—C11—C12—C13	-59.4 (2)
Cl1—C2—C3—C4	179.8 (2)	C20—C11—C12—C13	178.98 (19)
C2—C3—C4—C5	0.7 (4)	C11—C12—C13—C14	176.8 (2)
C2—C3—C4—N1	-179.8 (5)	C11—C12—C13—C18	51.5 (3)
C2—C3—C4—N1B	173.4 (9)	C12—C13—C14—C15	-75.8 (3)
O1—N1—C4—C3	-9.0 (13)	C18—C13—C14—C15	49.0 (3)
O2—N1—C4—C3	168.9 (8)	C13—C14—C15—	-54.1 (3)

		C16	
O1—N1—C4—C5	170.5 (10)	C14—C15—C16—C17	57.2 (3)
O2—N1—C4—C5	-11.6 (11)	C15—C16—C17—C18	-57.7 (3)
O1B—N1B—C4—C3	1 (2)	C16—C17—C18—C19	178.8 (2)
O2B—N1B—C4—C3	-175.8 (15)	C16—C17—C18—C13	53.7 (3)
O1B—N1B—C4—C5	174.8 (17)	C12—C13—C18—C17	77.4 (3)
O2B—N1B—C4—C5	-2 (2)	C14—C13—C18—C17	-48.3 (3)
C3—C4—C5—C6	1.4 (4)	C12—C13—C18—C19	-48.5 (3)
N1—C4—C5—C6	-178.1 (6)	C14—C13—C18—C19	-174.2 (2)
N1B—C4—C5—C6	-172.6 (8)	C11—N3—C19—C18	-64.1 (3)
C4—C5—C6—C1	-2.0 (4)	C10—N3—C19—C18	170.0 (2)
C2—C1—C6—C5	0.5 (4)	C17—C18—C19—N3	-70.8 (3)
C7—C1—C6—C5	179.9 (2)	C13—C18—C19—N3	55.7 (3)
C2—C1—C7—O3	-65.7 (3)	C21—N4—C20—O5	4.6 (4)
C6—C1—C7—O3	115.0 (3)	C21—N4—C20—C11	-177.1 (2)
C2—C1—C7—N2	115.6 (3)	N3—C11—C20—O5	-41.3 (3)
C6—C1—C7—N2	-63.7 (3)	C12—C11—C20—O5	78.4 (3)
O3—C7—N2—C8	-1.1 (4)	N3—C11—C20—N4	140.4 (2)
C1—C7—N2—C8	177.5 (2)	C12—C11—C20—N4	-99.9 (2)
C7—N2—C8—C9	-122.8 (3)	C20—N4—C21—C22	64.5 (3)
N2—C8—C9—O4	69.4 (2)	C20—N4—C21—C23	-59.1 (3)

N2—C8—C9—C10	-51.3 (3)	C20—N4—C21—C24	-177.5 (2)
C11—N3—C10—C9	160.33 (18)		

Hydrogen-bond geometry (Å, º) for (I)

D—H···*A*	*D*—H	H···*A*	*D*···*A*	*D*—H···*A*
O4—H4*B*···O5^{i}	0.87 (4)	1.94 (4)	2.791 (2)	169 (3)
N2—H2···O5	0.83 (3)	2.11 (3)	2.928 (3)	169 (3)
N4—H4···O3^{ii}	0.84 (3)	2.15 (3)	2.964 (2)	161 (3)

Symmetry codes: (i) *x*, *y*, *z*+1; (ii) *x*+1/2, -*y*+3/2, -*z*+2.

(II)

Crystal data

$C_{27}H_{39}ClN_4O_5$	*F*(000) = 572
M_r = 535.07	D_x = 1.279 Mg m^{-3}
Monoclinic, $P2_1$	Mo *K*a radiation, l = 0.71073 Å
Hall symbol: P 2yb	Cell parameters from 7806 reflections
a = 6.4341 (7) Å	q = 2.8–26.3°
b = 20.280 (2) Å	m = 0.18 mm^{-1}
c = 11.0377 (12) Å	*T* = 168 K
b = 105.248 (1)°	Plates, colourless
V = 1389.5 (3) Å^3	0.86 × 0.65 × 0.15 mm
Z = 2	

Data collection

Siemens Platform/ApexII CCD diffractometer	5627 independent reflections
Radiation source: normal-focus sealed tube	5222 reflections with *I* > 2s(*I*)
Graphite monochromator	R_{int} = 0.024
profile data from f and w scans	q_{max} = 26.4°, q_{min} = 1.9°
Absorption correction: integration *SHELXTL*/*XPREP* V2005/2 (Bruker, 2014)	*h* = -8®8
T_{min} = 0.892, T_{max} = 0.980	*k* = -25®25
16374 measured reflections	*l* = -13®13

Refinement

Refinement on F^2	Hydrogen site location: mixed
Least-squares matrix: full	H atoms treated by a mixture of independent and constrained refinement
$R[F^2 > 2s(F^2)] = 0.032$	$w = 1/[s^2(F_o^2) + (0.0447P)^2 + 0.1276P]$ where $P = (F_o^2 + 2F_c^2)/3$
$wR(F^2) = 0.082$	$(D/s)_{max} < 0.001$
$S = 1.04$	$D\rho_{max} = 0.24$ e Å^{-3}
5627 reflections	$D\rho_{min} = -0.21$ e Å^{-3}
377 parameters	Absolute structure: Flack (1983), 2720 Friedels Hooft, Straver and Spek (2008)
14 restraints	Absolute structure parameter: 0.036 (19)

Special details

Experimental. One distinct cell was identified using APEX2 (Bruker, 2014). Four frame series were integrated and filtered for statistical outliers using SAINT (Bruker, 2014) then corrected for absorption by integration using SHELXTL/XPREP V2005/2 (Bruker, 2014) before using SAINT/SADABS (Bruker, 2014) to sort, merge, and scale the combined data. No decay correction was applied.
Geometry. All esds (except the esd in the dihedral angle between two l.s. planes) are estimated using the full covariance matrix. The cell esds are taken into account individually in the estimation of esds in distances, angles and torsion angles; correlations between esds in cell parameters are only used when they are defined by crystal symmetry. An approximate (isotropic) treatment of cell esds is used for estimating esds involving l.s. planes.
Refinement. Structure was phased by direct (Sheldrick, 2015) methods. Systematic conditions suggested the ambiguous space group. The space group choice was confirmed by successful convergence of the full-matrix least-squares refinement on F^2. The final map had no significant features. A final analysis of variance between observed and calculated structure factors showed little dependence on amplitude and resolution.

Fractional atomic coordinates and isotropic or equivalent isotropic displacement parameters (Å^2)

	x	*y*	*z*	U_{iso}*/U_{eq}	Occ. (<1)
Cl1	0.18218 (12)	0.27617 (3)	0.58986 (6)	0.05313 (19)	
O1	0.7185 (5)	0.41190 (18)	0.9393 (3)	0.0657 (8)	0.967 (8)
O2	0.5309 (5)	0.48339 (16)	1.0083 (3)	0.0796 (11)	0.967 (8)
O1B	0.658 (13)	0.391 (2)	0.970 (8)	0.0657 (8)	0.033 (8)
O2B	0.605 (14)	0.4966 (13)	0.961 (9)	0.0796 (11)	0.033 (8)

N1	0.5493 (4)	0.43973 (13)	0.9347 (2)	0.0521 (6)	
O3	-0.3816 (17)	0.3687 (9)	0.5659 (16)	0.064 (2)	0.58 (2)
C7	-0.1974 (15)	0.3696 (13)	0.5438 (6)	0.044 (3)	0.58 (2)
O3B	-0.341 (3)	0.3487 (9)	0.584 (2)	0.064 (2)	0.42 (2)
C7B	-0.209 (2)	0.3832 (18)	0.5437 (8)	0.044 (3)	0.42 (2)
O4	-0.0271 (3)	0.50975 (9)	0.45504 (16)	0.0466 (4)	
H4B	-0.124 (4)	0.5375 (14)	0.441 (3)	0.070*	
O5	0.6776 (3)	0.60681 (10)	0.43924 (17)	0.0497 (5)	
N2	-0.1962 (3)	0.38062 (11)	0.4242 (2)	0.0434 (5)	
N3	0.1541 (3)	0.56188 (9)	0.25105 (17)	0.0349 (4)	
N4	0.3515 (3)	0.61688 (10)	0.48165 (17)	0.0354 (4)	
H4C	0.211 (5)	0.6126 (16)	0.452 (3)	0.053*	
C1	0.1816 (4)	0.34965 (12)	0.6715 (2)	0.0407 (6)	
C2	0.3653 (5)	0.36560 (13)	0.7648 (2)	0.0419 (5)	
H2A	0.4918	0.3393	0.7794	0.050*	
C3	0.3571 (4)	0.42136 (13)	0.8359 (2)	0.0425 (6)	
C4	0.1757 (5)	0.46054 (15)	0.8176 (3)	0.0515 (7)	
H4A	0.1728	0.4977	0.8697	0.062*	
C5	-0.0004 (5)	0.44390 (15)	0.7217 (3)	0.0505 (6)	
H5A	-0.1264	0.4703	0.7071	0.061*	
C6	0.0008 (4)	0.38941 (13)	0.6450 (2)	0.0417 (6)	
C8	-0.0147 (4)	0.40027 (12)	0.3737 (2)	0.0355 (5)	
H8A	0.1254	0.3883	0.4342	0.043*	
C9	-0.0526 (4)	0.35761 (14)	0.2539 (3)	0.0481 (6)	
H9A	-0.0348	0.3849	0.1829	0.058*	
H9B	0.0544	0.3214	0.2674	0.058*	
C10	-0.2747 (10)	0.3296 (4)	0.2237 (6)	0.0504 (15)	0.669 (16)
H10A	-0.2719	0.2816	0.2401	0.060*	0.669 (16)
H10B	-0.3527	0.3379	0.1350	0.060*	0.669 (16)
C11	-0.3755 (11)	0.3670 (4)	0.3125 (7)	0.0423 (15)	0.669 (16)
H11A	-0.4415	0.4086	0.2737	0.051*	0.669 (16)
H11B	-0.4878	0.3400	0.3353	0.051*	0.669 (16)
C10B	-0.2980 (17)	0.3525 (8)	0.2257 (15)	0.0504 (15)	0.331 (16)
H10C	-0.3618	0.3898	0.1705	0.060*	0.331 (16)
H10D	-0.3439	0.3114	0.1777	0.060*	0.331 (16)

C11B	-0.392 (2)	0.3527 (9)	0.3383 (12)	0.0423 (15)	0.331 (16)
H11C	-0.5191	0.3818	0.3268	0.051*	0.331 (16)
H11D	-0.4263	0.3079	0.3630	0.051*	0.331 (16)
C12	-0.0208 (4)	0.47423 (12)	0.3446 (2)	0.0370 (5)	
H12A	-0.1522	0.4846	0.2757	0.044*	
C13	0.1800 (4)	0.49527 (12)	0.3056 (2)	0.0380 (5)	
H13A	0.3061	0.4945	0.3799	0.046*	
H13B	0.2073	0.4636	0.2433	0.046*	
C14	0.3607 (3)	0.59631 (12)	0.2673 (2)	0.0357 (5)	
H14A	0.4542	0.5706	0.2250	0.043*	
C15	0.3232 (4)	0.66580 (12)	0.2091 (2)	0.0369 (5)	
H15A	0.2424	0.6925	0.2564	0.044*	
H15B	0.4642	0.6874	0.2172	0.044*	
C16	0.1982 (4)	0.66434 (13)	0.0704 (2)	0.0380 (5)	
H16A	0.2901	0.6415	0.0230	0.046*	
C17	0.1487 (4)	0.73367 (15)	0.0150 (2)	0.0462 (6)	
H17A	0.2833	0.7598	0.0349	0.055*	
H17B	0.0968	0.7302	-0.0776	0.055*	
C18	-0.0196 (4)	0.76982 (15)	0.0644 (2)	0.0470 (6)	
H18A	0.0388	0.7786	0.1552	0.056*	
H18B	-0.0528	0.8127	0.0209	0.056*	
C19	-0.2244 (4)	0.72936 (14)	0.0435 (2)	0.0451 (6)	
H19A	-0.2910	0.7249	-0.0478	0.054*	
H19B	-0.3279	0.7527	0.0807	0.054*	
C20	-0.1805 (4)	0.66089 (12)	0.1020 (2)	0.0381 (5)	
H20A	-0.3159	0.6351	0.0816	0.046*	
H20B	-0.1306	0.6652	0.1945	0.046*	
C21	-0.0090 (4)	0.62377 (13)	0.0539 (2)	0.0374 (5)	
H21A	-0.0699	0.6159	-0.0380	0.045*	
C22	0.0428 (4)	0.55715 (13)	0.1167 (2)	0.0396 (5)	
H22A	-0.0929	0.5322	0.1070	0.048*	
H22B	0.1344	0.5321	0.0736	0.048*	
C23	0.4782 (3)	0.60558 (11)	0.4055 (2)	0.0359 (5)	
C24	0.4216 (4)	0.62935 (12)	0.6177 (2)	0.0390 (5)	
C25	0.2182 (7)	0.6384 (4)	0.6614 (5)	0.0498 (13)	0.811 (17)

H25A	0.1300	0.5984	0.6430	0.075*	0.811 (17)
H25B	0.1361	0.6760	0.6173	0.075*	0.811 (17)
H25C	0.2573	0.6467	0.7521	0.075*	0.811 (17)
C26	0.5534 (13)	0.5704 (3)	0.6839 (4)	0.0572 (14)	0.811 (17)
H26A	0.4678	0.5300	0.6631	0.086*	0.811 (17)
H26B	0.5913	0.5773	0.7750	0.086*	0.811 (17)
H26C	0.6853	0.5662	0.6559	0.086*	0.811 (17)
C27	0.5599 (11)	0.6921 (3)	0.6388 (7)	0.0569 (15)	0.811 (17)
H27A	0.4787	0.7283	0.5894	0.085*	0.811 (17)
H27B	0.6918	0.6842	0.6124	0.085*	0.811 (17)
H27C	0.5977	0.7037	0.7281	0.085*	0.811 (17)
C25B	0.227 (4)	0.6608 (14)	0.651 (3)	0.0498 (13)	0.189 (17)
H25D	0.0945	0.6390	0.6040	0.075*	0.189 (17)
H25E	0.2201	0.7078	0.6299	0.075*	0.189 (17)
H25F	0.2418	0.6558	0.7415	0.075*	0.189 (17)
C26B	0.461 (5)	0.5602 (8)	0.678 (2)	0.0572 (14)	0.189 (17)
H26D	0.3258	0.5351	0.6569	0.086*	0.189 (17)
H26E	0.5122	0.5647	0.7697	0.086*	0.189 (17)
H26F	0.5692	0.5369	0.6465	0.086*	0.189 (17)
C27B	0.612 (4)	0.6761 (14)	0.659 (3)	0.0569 (15)	0.189 (17)
H27D	0.7446	0.6534	0.6543	0.085*	0.189 (17)
H27E	0.6262	0.6903	0.7457	0.085*	0.189 (17)
H27F	0.5895	0.7148	0.6037	0.085*	0.189 (17)

Atomic displacement parameters (Å^2)

	U^{11}	U^{22}	U^{33}	U^{12}	U^{13}	U^{23}
Cl1	0.0715 (4)	0.0369 (3)	0.0546 (4)	-0.0005 (3)	0.0230 (3)	-0.0055 (3)
O1	0.0580 (14)	0.0857 (18)	0.0498 (14)	0.0073 (14)	0.0076 (10)	-0.0005 (13)
O2	0.0664 (16)	0.097 (2)	0.0719 (18)	-0.0101 (14)	0.0123 (14)	-0.0422 (16)
O1B	0.0580 (14)	0.0857 (18)	0.0498 (14)	0.0073 (14)	0.0076 (10)	-0.0005 (13)
O2B	0.0664 (16)	0.097 (2)	0.0719 (18)	-0.0101 (14)	0.0123 (14)	-0.0422 (16)
N1	0.0564 (14)	0.0602 (15)	0.0405 (12)	-0.0036	0.0143 (10)	0.0005 (11)

				(12)		
O3	0.041 (4)	0.092 (8)	0.068 (4)	-0.008 (3)	0.029 (4)	-0.011 (4)
C7	0.0435 (16)	0.040 (10)	0.0557 (16)	-0.004 (3)	0.0269 (13)	-0.0166 (16)
O3B	0.041 (4)	0.092 (8)	0.068 (4)	-0.008 (3)	0.029 (4)	-0.011 (4)
C7B	0.0435 (16)	0.040 (10)	0.0557 (16)	-0.004 (3)	0.0269 (13)	-0.0166 (16)
O4	0.0528 (11)	0.0454 (10)	0.0399 (10)	0.0172 (8)	0.0090 (8)	-0.0097 (8)
O5	0.0298 (8)	0.0613 (12)	0.0524 (11)	0.0106 (8)	0.0011 (7)	-0.0171 (9)
N2	0.0344 (11)	0.0495 (13)	0.0480 (12)	-0.0039 (9)	0.0136 (9)	-0.0135 (10)
N3	0.0322 (10)	0.0374 (11)	0.0333 (10)	0.0008 (8)	0.0054 (8)	-0.0082 (8)
N4	0.0306 (9)	0.0423 (11)	0.0300 (10)	0.0024 (8)	0.0022 (7)	-0.0053 (8)
C1	0.0582 (15)	0.0342 (12)	0.0379 (13)	-0.0028 (11)	0.0273 (12)	-0.0015 (10)
C2	0.0528 (14)	0.0417 (13)	0.0363 (12)	0.0041 (11)	0.0208 (11)	0.0064 (10)
C3	0.0525 (14)	0.0448 (14)	0.0336 (12)	-0.0042 (11)	0.0174 (10)	0.0019 (10)
C4	0.0611 (17)	0.0552 (16)	0.0424 (15)	0.0031 (13)	0.0213 (13)	-0.0126 (12)
C5	0.0487 (15)	0.0599 (17)	0.0469 (15)	0.0071 (13)	0.0196 (12)	-0.0097 (13)
C6	0.0450 (14)	0.0477 (14)	0.0394 (13)	-0.0040 (11)	0.0233 (11)	-0.0054 (11)
C8	0.0330 (11)	0.0394 (12)	0.0350 (12)	0.0029 (9)	0.0104 (9)	-0.0048 (10)
C9	0.0539 (15)	0.0443 (14)	0.0461 (15)	0.0056 (12)	0.0131 (12)	-0.0130 (12)
C10	0.073 (2)	0.037 (4)	0.0420 (16)	-0.021 (3)	0.0166 (17)	-0.018 (3)
C11	0.0325 (16)	0.035 (3)	0.058 (3)	-0.0071 (18)	0.0089 (19)	-0.004 (2)
C10B	0.073 (2)	0.037 (4)	0.0420 (16)	-0.021 (3)	0.0166 (17)	-0.018 (3)
C11B	0.0325 (16)	0.035 (3)	0.058 (3)	-0.0071 (18)	0.0089 (19)	-0.004 (2)
C12	0.0375 (12)	0.0394 (12)	0.0330 (11)	0.0083 (10)	0.0072 (9)	-0.0068 (9)
C13	0.0360 (12)	0.0345 (12)	0.0410 (12)	0.0050 (9)	0.0059 (9)	-0.0089 (10)
C14	0.0265 (10)	0.0438 (13)	0.0364 (12)	0.0029 (9)	0.0077 (9)	-0.0106 (9)
C15	0.0291 (11)	0.0472 (14)	0.0357 (12)	-0.0043	0.0111 (9)	-0.0062

				(10)		(10)
C16	0.0304 (11)	0.0535 (15)	0.0325 (12)	-0.0009 (10)	0.0128 (9)	-0.0049 (10)
C17	0.0409 (13)	0.0614 (16)	0.0383 (13)	-0.0066 (12)	0.0139 (10)	0.0040 (12)
C18	0.0487 (14)	0.0487 (14)	0.0450 (14)	-0.0002 (12)	0.0148 (11)	0.0057 (12)
C19	0.0378 (13)	0.0583 (16)	0.0398 (13)	0.0056 (12)	0.0111 (10)	0.0010 (12)
C20	0.0272 (10)	0.0520 (14)	0.0355 (12)	-0.0009 (10)	0.0089 (9)	-0.0009 (10)
C21	0.0322 (11)	0.0523 (14)	0.0267 (11)	-0.0045 (10)	0.0061 (8)	-0.0095 (10)
C22	0.0365 (12)	0.0462 (14)	0.0346 (12)	-0.0017 (10)	0.0066 (9)	-0.0137 (10)
C23	0.0315 (11)	0.0353 (12)	0.0387 (12)	0.0057 (9)	0.0052 (9)	-0.0079 (10)
C24	0.0422 (13)	0.0420 (13)	0.0295 (11)	0.0017 (10)	0.0037 (9)	-0.0019 (9)
C25	0.0545 (17)	0.062 (4)	0.0341 (18)	-0.001 (2)	0.0141 (12)	-0.008 (3)
C26	0.054 (3)	0.066 (2)	0.0469 (18)	0.007 (2)	0.005 (2)	0.0227 (17)
C27	0.072 (3)	0.058 (3)	0.039 (3)	-0.015 (3)	0.012 (2)	-0.015 (2)
C25B	0.0545 (17)	0.062 (4)	0.0341 (18)	-0.001 (2)	0.0141 (12)	-0.008 (3)
C26B	0.054 (3)	0.066 (2)	0.0469 (18)	0.007 (2)	0.005 (2)	0.0227 (17)
C27B	0.072 (3)	0.058 (3)	0.039 (3)	-0.015 (3)	0.012 (2)	-0.015 (2)

Geometric parameters (Å, º) for (II)

Cl1—C1	1.742 (2)	C12—H12A	1.0000
O1—N1	1.215 (3)	C13—H13A	0.9900
O2—N1	1.228 (3)	C13—H13B	0.9900
O1B—N1	1.210 (15)	C14—C23	1.525 (3)
O2B—N1	1.219 (15)	C14—C15	1.541 (3)
N1—C3	1.466 (3)	C14—H14A	1.0000
O3—C7B	1.236 (18)	C15—C16	1.531 (3)
O3—C7	1.271 (11)	C15—H15A	0.9900
C7—O3B	1.20 (2)	C15—H15B	0.9900
C7—N2	1.341 (7)	C16—C17	1.533 (4)
C7—C6	1.512 (7)	C16—C21	1.536 (3)
O3B—C7B	1.271 (14)	C16—H16A	1.0000

C7B—N2	1.343 (10)	C17—C18	1.522 (4)
C7B—C6	1.513 (10)	C17—H17A	0.9900
O4—C12	1.425 (3)	C17—H17B	0.9900
O4—H4B	0.824 (14)	C18—C19	1.518 (4)
O5—C23	1.239 (3)	C18—H18A	0.9900
N2—C8	1.475 (3)	C18—H18B	0.9900
N2—C11	1.476 (5)	C19—C20	1.526 (4)
N2—C11B	1.477 (10)	C19—H19A	0.9900
N3—C22	1.470 (3)	C19—H19B	0.9900
N3—C14	1.470 (3)	C20—C21	1.541 (3)
N3—C13	1.471 (3)	C20—H20A	0.9900
N4—C23	1.336 (3)	C20—H20B	0.9900
N4—C24	1.472 (3)	C21—C22	1.515 (4)
N4—H4C	0.88 (3)	C21—H21A	1.0000
C1—C6	1.382 (4)	C22—H22A	0.9900
C1—C2	1.387 (4)	C22—H22B	0.9900
C2—C3	1.385 (4)	C24—C25	1.520 (4)
C2—H2A	0.9500	C24—C27B	1.524 (13)
C3—C4	1.382 (4)	C24—C27	1.535 (5)
C4—C5	1.374 (4)	C24—C26	1.535 (4)
C4—H4A	0.9500	C24—C25B	1.536 (13)
C5—C6	1.393 (4)	C24—C26B	1.545 (12)
C5—H5A	0.9500	C25—H25A	0.9800
C8—C12	1.532 (3)	C25—H25B	0.9800
C8—C9	1.545 (3)	C25—H25C	0.9800
C8—H8A	1.0000	C26—H26A	0.9800
C9—C10	1.492 (6)	C26—H26B	0.9800
C9—C10B	1.531 (11)	C26—H26C	0.9800
C9—H9A	0.9900	C27—H27A	0.9800
C9—H9B	0.9900	C27—H27B	0.9800
C10—C11	1.514 (7)	C27—H27C	0.9800
C10—H10A	0.9900	C25B—H25D	0.9800
C10—H10B	0.9900	C25B—H25E	0.9800
C11—H11A	0.9900	C25B—H25F	0.9800
C11—H11B	0.9900	C26B—H26D	0.9800

C10B—C11B	1.519 (11)	C26B—H26E	0.9800
C10B—H10C	0.9900	C26B—H26F	0.9800
C10B—H10D	0.9900	C27B—H27D	0.9800
C11B—H11C	0.9900	C27B—H27E	0.9800
C11B—H11D	0.9900	C27B—H27F	0.9800
C12—C13	1.526 (3)		
O1B—N1—O2B	126 (3)	N3—C14—C23	112.03 (19)
O1—N1—O2	123.3 (3)	N3—C14—C15	110.24 (17)
O1B—N1—C3	110 (3)	C23—C14—C15	106.64 (18)
O1—N1—C3	119.0 (2)	N3—C14—H14A	109.3
O2B—N1—C3	124 (3)	C23—C14—H14A	109.3
O2—N1—C3	117.8 (3)	C15—C14—H14A	109.3
C7B—O3—C7	13 (3)	C16—C15—C14	112.51 (19)
O3B—C7—O3	22.9 (9)	C16—C15—H15A	109.1
O3B—C7—N2	128.8 (12)	C14—C15—H15A	109.1
O3—C7—N2	116.0 (12)	C16—C15—H15B	109.1
O3B—C7—C6	113.6 (12)	C14—C15—H15B	109.1
O3—C7—C6	120.7 (10)	H15A—C15—H15B	107.8
N2—C7—C6	117.5 (7)	C15—C16—C17	112.4 (2)
C7—O3B—C7B	13 (3)	C15—C16—C21	109.89 (19)
O3—C7B—O3B	22.8 (9)	C17—C16—C21	111.2 (2)
O3—C7B—N2	118.4 (11)	C15—C16—H16A	107.7
O3B—C7B—N2	122.9 (19)	C17—C16—H16A	107.7
O3—C7B—C6	123.1 (11)	C21—C16—H16A	107.7
O3B—C7B—C6	109.5 (15)	C18—C17—C16	113.0 (2)
N2—C7B—C6	117.2 (9)	C18—C17—H17A	109.0
C12—O4—H4B	112 (2)	C16—C17—H17A	109.0
C7—N2—C7B	12 (3)	C18—C17—H17B	109.0
C7—N2—C8	128.8 (4)	C16—C17—H17B	109.0
C7B—N2—C8	128.2 (5)	H17A—C17—H17B	107.8
C7—N2—C11	125.8 (5)	C19—C18—C17	110.9 (2)
C8—N2—C11	105.0 (4)	C19—C18—H18A	109.5
C7—N2—C11B	110.1 (7)	C17—C18—H18A	109.5
C7B—N2—C11B	112.6 (8)	C19—C18—H18B	109.5

C8—N2—C11B	119.2 (6)	C17—C18—H18B	109.5
C22—N3—C14	109.84 (18)	H18A—C18—H18B	108.1
C22—N3—C13	109.01 (18)	C18—C19—C20	111.8 (2)
C14—N3—C13	112.62 (18)	C18—C19—H19A	109.3
C23—N4—C24	126.7 (2)	C20—C19—H19A	109.3
C23—N4—H4C	118.8 (19)	C18—C19—H19B	109.3
C24—N4—H4C	114.2 (19)	C20—C19—H19B	109.3
C6—C1—C2	121.8 (2)	H19A—C19—H19B	107.9
C6—C1—Cl1	120.4 (2)	C19—C20—C21	111.6 (2)
C2—C1—Cl1	117.8 (2)	C19—C20—H20A	109.3
C3—C2—C1	117.3 (2)	C21—C20—H20A	109.3
C3—C2—H2A	121.3	C19—C20—H20B	109.3
C1—C2—H2A	121.3	C21—C20—H20B	109.3
C4—C3—C2	122.9 (2)	H20A—C20—H20B	108.0
C4—C3—N1	118.5 (2)	C22—C21—C16	110.13 (19)
C2—C3—N1	118.6 (2)	C22—C21—C20	111.7 (2)
C5—C4—C3	117.7 (2)	C16—C21—C20	112.0 (2)
C5—C4—H4A	121.2	C22—C21—H21A	107.6
C3—C4—H4A	121.2	C16—C21—H21A	107.6
C4—C5—C6	121.9 (3)	C20—C21—H21A	107.6
C4—C5—H5A	119.1	N3—C22—C21	113.16 (18)
C6—C5—H5A	119.1	N3—C22—H22A	108.9
C1—C6—C5	118.2 (2)	C21—C22—H22A	108.9
C1—C6—C7	120.0 (10)	N3—C22—H22B	108.9
C5—C6—C7	121.4 (9)	C21—C22—H22B	108.9
C1—C6—C7B	130.1 (13)	H22A—C22—H22B	107.8
C5—C6—C7B	111.7 (13)	O5—C23—N4	124.3 (2)
C7—C6—C7B	11 (2)	O5—C23—C14	120.3 (2)
N2—C8—C12	111.36 (19)	N4—C23—C14	115.23 (18)
N2—C8—C9	102.10 (19)	N4—C24—C25	106.6 (3)
C12—C8—C9	112.3 (2)	N4—C24—C27B	114.8 (14)
N2—C8—H8A	110.3	N4—C24—C27	107.9 (3)
C12—C8—H8A	110.3	C25—C24—C27	111.4 (4)
C9—C8—H8A	110.3	N4—C24—C26	109.6 (3)
C10—C9—C8	109.3 (3)	C25—C24—C26	110.9 (3)

C10B—C9—C8	97.7 (6)	C27—C24—C26	110.3 (3)
C10—C9—H9A	109.8	N4—C24—C25B	105.5 (12)
C8—C9—H9A	109.8	C27B—C24—C25B	108.3 (18)
C10—C9—H9B	109.8	N4—C24—C26B	104.9 (9)
C8—C9—H9B	109.8	C27B—C24—C26B	114.1 (13)
H9A—C9—H9B	108.3	C25B—C24—C26B	108.9 (13)
C9—C10—C11	102.3 (6)	C24—C25—H25A	109.5
C9—C10—H10A	111.3	C24—C25—H25B	109.5
C11—C10—H10A	111.3	H25A—C25—H25B	109.5
C9—C10—H10B	111.3	C24—C25—H25C	109.5
C11—C10—H10B	111.3	H25A—C25—H25C	109.5
H10A—C10—H10B	109.2	H25B—C25—H25C	109.5
N2—C11—C10	105.1 (5)	C24—C26—H26A	109.5
N2—C11—H11A	110.7	C24—C26—H26B	109.5
C10—C11—H11A	110.7	H26A—C26—H26B	109.5
N2—C11—H11B	110.7	C24—C26—H26C	109.5
C10—C11—H11B	110.7	H26A—C26—H26C	109.5
H11A—C11—H11B	108.8	H26B—C26—H26C	109.5
C11B—C10B—C9	116.4 (11)	C24—C27—H27A	109.5
C11B—C10B—H10C	108.2	C24—C27—H27B	109.5
C9—C10B—H10C	108.2	H27A—C27—H27B	109.5
C11B—C10B—H10D	108.2	C24—C27—H27C	109.5
C9—C10B—H10D	108.2	H27A—C27—H27C	109.5
H10C—C10B—H10D	107.3	H27B—C27—H27C	109.5
N2—C11B—C10B	94.0 (9)	C24—C25B—H25D	109.5
N2—C11B—H11C	112.9	C24—C25B—H25E	109.5
C10B—C11B—H11C	112.9	H25D—C25B—H25E	109.5
N2—C11B—H11D	112.9	C24—C25B—H25F	109.5
C10B—C11B—H11D	112.9	H25D—C25B—H25F	109.5
H11C—C11B—H11D	110.3	H25E—C25B—H25F	109.5
O4—C12—C13	108.45 (19)	C24—C26B—H26D	109.5

O4—C12—C8	108.7 (2)	C24—C26B—H26E	109.5
C13—C12—C8	110.82 (18)	H26D—C26B—H26E	109.5
O4—C12—H12A	109.6	C24—C26B—H26F	109.5
C13—C12—H12A	109.6	H26D—C26B—H26F	109.5
C8—C12—H12A	109.6	H26E—C26B—H26F	109.5
N3—C13—C12	111.04 (18)	C24—C27B—H27D	109.5
N3—C13—H13A	109.4	C24—C27B—H27E	109.5
C12—C13—H13A	109.4	H27D—C27B—H27E	109.5
N3—C13—H13B	109.4	C24—C27B—H27F	109.5
C12—C13—H13B	109.4	H27D—C27B—H27F	109.5
H13A—C13—H13B	108.0	H27E—C27B—H27F	109.5
C7B—O3—C7—O3B	154 (6)	O3B—C7B—C6—C5	-89 (3)
C7B—O3—C7—N2	-77 (3)	N2—C7B—C6—C5	125 (2)
C7B—O3—C7—C6	76 (3)	O3—C7B—C6—C7	87 (5)
O3—C7—O3B—C7B	-26 (5)	O3B—C7B—C6—C7	67 (4)
N2—C7—O3B—C7B	-88 (4)	N2—C7B—C6—C7	-80 (4)
C6—C7—O3B—C7B	88 (3)	C7—N2—C8—C12	-100.6 (15)
C7—O3—C7B—O3B	-25 (5)	C7B—N2—C8—C12	-85 (2)
C7—O3—C7B—N2	83 (4)	C11—N2—C8—C12	86.4 (4)
C7—O3—C7B—C6	-84 (4)	C11B—N2—C8—C12	96.6 (9)
C7—O3B—C7B—O3	153 (6)	C7—N2—C8—C9	139.3 (15)
C7—O3B—C7B—N2	68 (4)	C7B—N2—C8—C9	155 (2)
C7—O3B—C7B—C6	-76 (3)	C11—N2—C8—C9	-33.6 (4)

O3B—C7—N2—C7B	95 (4)	C11B—N2—C8—C9	-23.5 (9)
O3—C7—N2—C7B	72 (4)	N2—C8—C9—C10	14.1 (5)
C6—C7—N2—C7B	-81 (4)	C12—C8—C9—C10	-105.3 (4)
O3B—C7—N2—C8	-174 (2)	N2—C8—C9—C10B	29.1 (7)
O3—C7—N2—C8	163.8 (14)	C12—C8—C9—C10B	-90.3 (7)
C6—C7—N2—C8	11 (3)	C8—C9—C10—C11	10.2 (8)
O3—C7—N2—C11	-25 (3)	C7—N2—C11—C10	-131.3 (15)
C6—C7—N2—C11	-177.9 (10)	C8—N2—C11—C10	42.0 (7)
O3B—C7—N2—C11B	-10 (4)	C9—C10—C11—N2	-31.3 (9)
O3—C7—N2—C11B	-32 (3)	C8—C9—C10B—C11B	-31.8 (15)
C6—C7—N2—C11B	174.6 (15)	C7—N2—C11B—C10B	-161.4 (15)
O3—C7B—N2—C7	-87 (5)	C7B—N2—C11B—C10B	-174 (2)
O3B—C7B—N2—C7	-61 (4)	C8—N2—C11B—C10B	4.3 (16)
C6—C7B—N2—C7	80 (4)	C9—C10B—C11B—N2	18.3 (18)
O3—C7B—N2—C8	174.9 (19)	N2—C8—C12—O4	55.0 (2)
O3B—C7B—N2—C8	-159.1 (18)	C9—C8—C12—O4	168.76 (19)
C6—C7B—N2—C8	-18 (4)	N2—C8—C12—C13	174.05 (19)
O3—C7B—N2—C11B	-7 (4)	C9—C8—C12—C13	-72.1 (2)
O3B—C7B—N2—C11B	19 (4)	C22—N3—C13—C12	-85.1 (2)
C6—C7B—N2—C11B	161 (2)	C14—N3—C13—C12	152.70 (18)
C6—C1—C2—C3	3.5 (3)	O4—C12—C13—N3	-73.4 (2)
Cl1—C1—C2—C3	-174.92 (18)	C8—C12—C13—	167.38 (18)

		N3	
C1—C2—C3—C4	0.3 (4)	C22—N3—C14—C23	177.20 (18)
C1—C2—C3—N1	-179.6 (2)	C13—N3—C14—C23	-61.1 (2)
O1B—N1—C3—C4	155 (6)	C22—N3—C14—C15	58.6 (2)
O1—N1—C3—C4	-168.9 (3)	C13—N3—C14—C15	-179.68 (18)
O2B—N1—C3—C4	-36 (7)	N3—C14—C15—C16	-56.0 (2)
O2—N1—C3—C4	10.2 (4)	C23—C14—C15—C16	-177.86 (18)
O1B—N1—C3—C2	-26 (6)	C14—C15—C16—C17	176.37 (19)
O1—N1—C3—C2	10.9 (4)	C14—C15—C16—C21	52.0 (2)
O2B—N1—C3—C2	144 (7)	C15—C16—C17—C18	-70.8 (3)
O2—N1—C3—C2	-169.9 (3)	C21—C16—C17—C18	52.9 (3)
C2—C3—C4—C5	-2.2 (4)	C16—C17—C18—C19	-55.0 (3)
N1—C3—C4—C5	177.7 (2)	C17—C18—C19—C20	55.7 (3)
C3—C4—C5—C6	0.4 (4)	C18—C19—C20—C21	-55.3 (3)
C2—C1—C6—C5	-5.2 (4)	C15—C16—C21—C22	-51.3 (2)
Cl1—C1—C6—C5	173.2 (2)	C17—C16—C21—C22	-176.4 (2)
C2—C1—C6—C7	-179.1 (6)	C15—C16—C21—C20	73.6 (2)
Cl1—C1—C6—C7	-0.7 (6)	C17—C16—C21—C20	-51.5 (3)
C2—C1—C6—C7B	176.1 (9)	C19—C20—C21—C22	177.21 (19)
Cl1—C1—C6—C7B	-5.5 (10)	C19—C20—C21—C16	53.1 (3)
C4—C5—C6—C1	3.2 (4)	C14—N3—C22—C21	-61.1 (2)

C4—C5—C6—C7	177.0 (7)	C13—N3—C22—C21	175.09 (19)
C4—C5—C6—C7B	-177.9 (8)	C16—C21—C22—N3	57.2 (2)
O3B—C7—C6—C1	105 (2)	C20—C21—C22—N3	-67.9 (2)
O3—C7—C6—C1	129 (2)	C24—N4—C23—O5	-3.2 (4)
N2—C7—C6—C1	-79 (2)	C24—N4—C23—C14	-178.2 (2)
O3B—C7—C6—C5	-69 (3)	N3—C14—C23—O5	149.4 (2)
O3—C7—C6—C5	-44 (3)	C15—C14—C23—O5	-89.9 (2)
N2—C7—C6—C5	107.5 (16)	N3—C14—C23—N4	-35.4 (3)
O3B—C7—C6—C7B	-95 (4)	C15—C14—C23—N4	85.3 (2)
O3—C7—C6—C7B	-71 (4)	C23—N4—C24—C25	-179.4 (4)
N2—C7—C6—C7B	81 (4)	C23—N4—C24—C27B	43.0 (15)
O3—C7B—C6—C1	110 (3)	C23—N4—C24—C27	60.9 (4)
O3B—C7B—C6—C1	90 (2)	C23—N4—C24—C26	-59.3 (5)
N2—C7B—C6—C1	-57 (3)	C23—N4—C24—C25B	162.1 (12)
O3—C7B—C6—C5	-69 (3)	C23—N4—C24—C26B	-83.0 (13)

Hydrogen-bond geometry (Å, º) for (II)

D—H···*A*	*D*—H	H···*A*	*D*···*A*	*D*—H···*A*
N4—H4*C*···O4	0.88 (3)	2.60 (3)	3.219 (3)	129 (3)
O4—H4*B*···O5^{i}	0.82 (1)	1.89 (2)	2.709 (2)	170 (4)

Symmetry code: (i) *x*-1, *y*, *z*.
Document origin: *publCIF* [Westrip, S. P. (2010). *J. Apply. Cryst.*, **43**, 920-925].

C.8 References

1. Bruker (2014). *APEX2*, *SAINT*, *SHELXTL*, XCIF, *XPREP*, *SADABS*. Bruker AXS, Inc., Madison, Wisconsin, USA.

2. *CrystalMaker* (1994). *CrystalMaker*, a crystal and molecular structures program for Mac and Windows. CrystalMaker Software Ltd., Oxford, England (www.*CrystalMaker*.com).

3. Flack, H. D. (1983). *Acta Cryst.* A**39**, 876–881.

4. Flack, H. D. & Bernardinelli, G. (2000). *J. Appl. Cryst.* **33**, 1143–1148.

5. Gantt, S., Casper, C. & Ambinder, R. F. (2013). *Curr. Opin. Oncol.* **25**, 495–502.

6. Groom, C. R. & Allen, F. H. (2014). *Angew. Chem. Int. Ed.* **53**, 662–671.

7. Hamilton, W. C. & Ibers, J. A. (1968). Hydrogen Bonding in Solids. W.A.Benjamin, Inc., New York, New York, USA.

8. Hooft, R., Straver, L. & Spek, A. (2008). *J. Appl. Cryst.* **41**, 96–103.

9. Inaba, T., Birchler, A. G., Yamada, Y., Sagawa, S., Yokota, K., Ando, K. & Uchida, I. (1998). *J. Org. Chem.* **63**, 7582–7583.

10. Inaba, T., Yamada, Y., Abe, H., Sagawa, S. & Cho, H. (2000). *J. Org. Chem.* **65**, 1623–1628.

11. Kaldor, S. W., Kalish, V. J., Davies, J. F., 2nd, Shetty, B. V., Fritz, J. E., Appelt, K., Burgess, J. A., Campanale, K. M., Chirgadze, N. Y., Clawson, D. K., Dressman, B. A., Hatch, S. D., Khalil, D. A., Kosa, M. B., Lubbehusen, P. P., Muesing, M. A., Patick, A. K., Reich, S. H., Su, K. S. & Tatlock, J. H. (1997). *J. Med. Chem.* **40**, 3979–3985.

12. Kalu, N. N., Desai, P. J., Shirley, C. M., Gibson, W., Dennis, P. A. & Ambinder, R. F. (2014). *J. Virol.* **88**, 5455–5461.

13. Koltai, T. (2015). *F1000Res.* **4**, 1-19.

14. Maxson, T., Deane, C. D., Molloy, E. M., Cox, C. L., Markley, A. L., Lee, S. W. & Mitchell, D. A. (2015). ACS Chem Biol 10, 1217-1226.

15. Panel on Antiretroviral Guidelines (2015). *Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. Department of Health and Human Services. Available at* http://www.aidsinfo.nih.gov/ContentFiles/AdultandAdolescentGL.pdf. Accessed August 29, 2015.

16. Spek, A. L. (2009). *Acta Cryst.* D**65**, 148–155.

17. Sheldrick, G. M. (2015). *Acta Cryst.* C**71**, 3–8.

18. Westrip, S. P. (2010). *J. Appl. Cryst.* **43**, 920–925.

19. Yamamoto, N., Yang, R., Yoshinaka, Y., Amari, S., Nakano, T., Cinatl, J., Rabenau, H., Doerr, H. W.,
20. Hunsmann, G., Otaka, A., Tamamura, H., Fujii, N. & Yamamoto, N. (2004). *Biochem. Biophys. Res. Commun.* **318**, 719–725.

21. Zhao, X. Y., Li, R. M., Ma, C. J. & Zhang, G. Y. (2006). *J. East China University of Sci. and Tech.* **32**, 1449–1453.