STEROID METABOLISM IN THE HUMAN GUT BY MICROBIAL HYDROXYSTEROID
DEHYDROGENASES

BY

HEIDI LEE DODEN

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Animal Sciences
in the Graduate College of the
University of Illinois Urbana-Champaign, 2021

Urbana, Illinois

Doctoral Committee:

Assistant Professor Jason M. Ridlon, Chair
Professor H. Rex Gaskins
Professor Isaac K. O. Cann
Professor Emeritus Steven L. Daniel, Eastern Illinois University

ABSTRACT

Bile acids and cortisol are steroid hormones derived from cholesterol that are important signaling molecules in the human body. Bile acids are essential for nutrient solubilization and absorption in the gut while cortisol is involved in stress response and metabolism throughout the body. The gut microbiome has evolved a network of enzymes that modify steroids produced by the host. One such enzyme activity is oxidoreduction of steroid hydroxyl groups by hydroxysteroid dehydrogenases (HSDHs). Pairs of HSDHs can reversibly epimerize α-hydroxyl groups through an oxo-intermediate to β-hydroxyl groups. These small modifications of steroid structure often greatly influence their physicochemical properties, leading to changes in toxicity, solubility, and their ability to activate or inhibit host receptors.

In this work, two types of bile acid HSDHs were explored: 12α-HSDH, which reversibly converts the C-12 position hydroxyl from the α-orientation to an oxo-group, and 12β-HSDH, which completes the epimerization by converting the C-12 oxo-group to the β-orientation. Three 12α-HSDHs were identified in the human gut microbes *Clostridium scindens* ATCC 35704, *C. hylemonae* DSM 15053, and *Peptacetobacter hiranonis* DSM 13275. After cloning and overexpression of these enzymes in *Escherichia coli*, biochemical analysis revealed a preference for the oxo-intermediate, 12-oxolithocholic (12-oxoLCA), over the 12α-hydroxy substrate, deoxycholic acid (DCA). Phylogenetic analysis suggests 12α-HSDH is widespread across Firmicutes and Actinobacteria. The first gene encoding bile acid 12β-HSDH, converting 12-oxoLCA to epiDCA, was discovered following a screen of 6 candidate enzymes from *C. paraputrificum* ATCC 25780. A phylogenetic analysis of 12β-HSDH led to the identification of

two additional validated bile acid 12β-HSDHs in *Eisenbergiella* sp. OF01-20 and *Olsenella* sp. GAM18. Comparison of the 12α- and 12β-HSDH phylogenies revealed two shared organisms between them, *Collinsella tanakaei* YIT 12063 and *Collinsella stercoris* DSM 13279. Although not yet confirmed in culture, these are the first strains identified with potential bile acid C-12 epimerizing activity.

Additionally, two HSDHs metabolizing cortisol were characterized biochemically and structurally: 20β-HSDH, which interconverts cortisol and 20β-dihydrocortisol, and 20α-HSDH, which reversibly biotransforms cortisol to 20α-dihydrocortisol. A 20β-HSDH from *Bifidobacterium adolescentis* strain L2-32 was cloned, overexpressed and purified. The apo-form without ligands bound and the binary complex with cofactor bound were crystallized. A large, flexible N-terminal domain within the structures was investigated further by gel filtration chromatography and circular dichroism spectroscopy. Deletions of the extended N-terminus caused a loss of activity and structural changes, suggesting the N-terminus is critical for protein stability. Additionally, a 20α-HSDH from *C. scindens* ATCC 35704 was crystallized and its enzymatic mechanism predicted by hybrid quantum mechanics/molecular mechanics (QM/MM) simulations. In order to test the reaction mechanism, substrate-binding and catalytic residues were validated by site-directed mutagenesis and isothermal titration calorimetry.

Microbial HSDHs are predicted regulators of steroid pathways implicated in colorectal cancer, liver cancer, castration-resistant prostate cancer and polycystic ovary syndrome. Due to their regulatory potential, microbial HSDHs may have promise as therapeutics or druggable targets in the future. Thus, the characterization of new bile acid 12α/β-HSDH and cortisol 20α/β-HSDH activities will serve as the foundation for future mechanistic studies on the role of HSDHs in various disease states.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

CHAPTER 1

REVIEW OF LITERATURE[1]

ABSTRACT

Bile acids (BAs) and glucocorticoids are steroid hormones derived from cholesterol that are important signaling molecules in humans and other vertebrates. Hydroxysteroid dehydrogenases (HSDHs) are encoded both by the host and by their resident gut microbiota, and they reversibly convert steroid hydroxyl groups to keto groups. Pairs of HSDHs can reversibly epimerize steroids from α-hydroxy conformations to β-hydroxy, or β-hydroxy to ω-hydroxy in the case of ω-muricholic acid. These reactions often result in products with drastically different physicochemical properties than their precursors, which can result in steroids being activators or inhibitors of host receptors, can affect solubility in fecal water, and can modulate toxicity. Microbial HSDHs modulate sterols associated with diseases such as colorectal cancer, liver cancer, prostate cancer, and polycystic ovary syndrome. Although the role of microbial HSDHs is not yet fully elucidated, they may have therapeutic potential as steroid pool modulators or druggable targets in the future. In this review, we explore metabolism of BAs and glucocorticoids with a focus on biotransformation by microbial HSDHs.

INTRODUCTION

Steroid hormones are signaling molecules derived from cholesterol that include glucocorticoids, mineralocorticoids, androgens, estrogens, progestogens, and bile acids (BAs).[1] Steroid hormones are essential for the regulation of various physiological processes, such as metabolism, salt and water balance, reproduction, inflammation, and stress response.[2] These cholesterol-derived molecules are synthesized in the human adrenal glands, gonads, placenta, and liver.[3,4] All steroids have a cyclopentanoperhydrophenanthrene ring structure, composed of three six-carbon rings denoted A, B, and C along with a five-carbon D ring (**Figure 1.1**), with differing hydroxyl groups and side-chains.[1] Hydroxysteroid dehydrogenases (HSDH) are an important class of enzyme expressed by both host tissues and host-associated microbiota that modify the hydroxyl groups on steroids. These small modifications to steroids greatly impact their physicochemical properties and can change the steroid solubility, toxicity, host receptor affinity, and ability to activate or inhibit host receptors.[5–8] The current review focuses on the importance of gut microbial HSDHs in cholesterol, BA, and glucocorticoid metabolism.

HYDROXYSTEROID DEHYDROGENASES

Hydroxysteroid Dehydrogenase Function

Hydroxysteroid dehydrogenases are nicotinamide adenine dinucleotide (phosphate) (NAD(P)(H))-dependent oxidoreductases that catalyze the reversible conversion of hydroxyl groups to keto groups on steroids.[9] HSDHs are regio- and stereospecific, meaning they are specific for the hydroxyl position on the steroid (C-3 vs. C-7) and for the orientation (α vs. β) of the hydroxyl group, respectively.[5] Pairs of HSDHs can convert steroids from the α-orientation, through an oxo-intermediate, to the epimerized β-orientation and vice versa.

Hydroxysteroid dehydrogenases are found in both host and microbial genomes, although more is known about the physiological function of host hydroxysteroid dehydrogenases, which are typically abbreviated HSDs in literature. In this review, host hydroxysteroid dehydrogenases are denoted "HSD" while bacterial enzymes are denoted "HSDH". Host HSDs are key enzymes in the biosynthesis of steroids in steroidogenic tissues.[10] They also function to activate or inactivate steroids in peripheral tissues, thus regulating local concentrations of steroid hormones.[5] Even though host HSDs catalyze reversible reactions in vitro, they typically function primarily in one direction in vivo on the basis of cofactor balance: either as dehydrogenases or as reductases.[11]

Host HSDs are druggable targets important in the treatment of endocrine-dependent disorders, including cancers.[12] Host-associated microbial HSDHs may also serve as pharmacological targets or, alternatively, may be enriched in the host through engineering and delivering probiotic bacteria with rational sterolbiome phenotypes. One recent example involves identification of a cholesterol 3β-HSDH involved in conversion of cholesterol to coprostanol, the enrichment of which may be important as a probiotic approach to reducing serum cholesterol.[13]

Structural Biology of Hydroxysteroid Dehydrogenases

Hydroxysteroid dehydrogenases belong to one of the following three large and diverse protein superfamilies: short-chain dehydrogenase/reductase (SDR), medium-chain dehydrogenase/reductase (MDR), or aldo-keto reductase (AKR).[5,14] Many SDR and MDR family hydroxysteroid dehydrogenases have been identified in the gut microbiome.[14–17] HSDHs in the AKR superfamily are generally found within mammals[12], although microbial AKR family HSDHs have been reported.[18]

The SDR superfamily is one of the largest, containing proteins spanning all three domains of life.[19] SDR proteins have highly diverse substrate specificities, ranging from sugars to dyes to steroids.[20] Members of this superfamily are non-metalloenzymes and typically 250 amino acids in length.[5] Due to the dependence of dehydrogenase/reductase enzymes on NAD(P)(H) to carry out redox reactions, SDR proteins contain a Rossmann fold domain for binding cofactors. This domain consists of 6–7 β-strands with 3–4 peripheral α-helices on either side.[21,22] Typically, the Rossmann fold domain is located near the N-terminus of SDR proteins, while the C-terminus binds substrates.[20] Most SDR members have a conserved Tyr, Ser, and Lys at the catalytic site. The overall folding pattern is closely conserved across the superfamily, while amino-acid sequence varies greatly.[22] This causes great difficulty in predicting substrate specificities by amino acid homology search alone. HSDHs within the SDR superfamily include but are not limited to host 11β-HSD and 17β-HSD[5], and various microbial BA 12α-HSDHs (see Chapter 2)[23], 12β-HSDH (see Chapter 3)[24], 3α/β-HSDHs[17], and glucocorticoid 20β-HSDH (see Chapter 4)[15].

The MDR family is similar to the SDR family both in number of members and in function, although their structures have marked differences. MDR proteins contain Rossmann fold domains for NAD(P)(H) binding like SDRs, but they are ~350 residues long and many are metal-dependent.[25] They are typically dimeric or tetrameric and many contain a catalytic zinc ion, sometimes along with a structural zinc ion, while others are non-zinc-containing.[26] The zinc-containing MDRs share a strictly conserved Gly, His, and Glu for zinc binding.[27] MDR family HSDHs include host BA 3β-HSD[26] and microbial glucocorticoid 20α-HSDH (see Chapter 5)[14,16].

AKRs are NAD(P)(H)-dependent oxidoreductases acting on carbonyl groups or double bonds and are ~320 amino acids long. They are monomeric with diverse substrate recognition,

including steroids, monosaccharides, and isoflavonoids. An ordered bi–bi kinetic mechanism has been shown for multiple AKR family members, where the cofactor is first to bind and last to leave.[28] Most have a conserved active site with residues Asp, Lys, Tyr, and His. Examples of members of this superfamily involved in steroid metabolism are human 3α-HSD[29], human 20α-HSD[30], and bacterial BA 3β-HSDH[18].

BILE ACID METABOLISM

Host Bile Acid Synthesis and Signaling

Bile acids are amphipathic C_{24} steroids that play an important role in host nutrition[31]. They are essential for solubilization and later absorption of cholesterol, dietary fatty acids, triglycerides, and lipid-soluble vitamins A, D, E, and K. Bile acids assemble into mixed micelles, forming a hydrocarbon interior in order to solubilize these molecules.[31,32]

Bile acid biosynthesis occurs in the liver and begins with the rate-limiting step of cholesterol 7α-hydroxylation by cytochrome P450 7α-hydroxylase (CYP7A1) in hepatocytes (**Figure 1.2**).[31,33] While other carbon positions on cholesterol can be hydroxylated first (C-24, C-25, C-26, C-27), the classical pathway initiates through C-7 hydroxylation catalyzed by CYP7A1.[34,35] The next step alters the ring structure through conversion to 3-oxo-Δ^4 by 3β-hydroxy-Δ^5-C_{27}-steroid oxidoreductase (HSD3B7).[34,36,37] After HSD3B7 action, the intermediate is converted by 12α-hydroxylase (CYP8B1) if the final product contains a 12α-hydroxyl group. Ensuing steps involve additional modification to the ring structure by AKR1D1 and AKR1C1.[37] Then, mitochondrial sterol 27-hydroxylase (CYP27A1) oxidizes the side-chain, followed by removal of three carbon atoms beginning with activation of the sterol by BA coenzyme A (CoA) synthase.[34,38,39] Subsequent reactions are catalyzed by 2-methylacyl-CoA racemase, branched-

chain acyl-CoA oxidase, D-bifunctional protein, and peroxisomal thiolase 2, which cleaves the C-24–C-25 bond.[34,37] The final step in BA biosynthesis is conjugation of the BA-CoA intermediate to either glycine or taurine, catalyzed by BA CoA:amino acid *N*-acyltransferase.[34,40]

Conjugated BAs, called "bile salts" due to their ionized state at physiological pH, have increased solubility and greater amphipathicity. The biosynthetic pathway results in the formation of conjugated cholic acid (CA; 3α,7α,12α-hydroxy) or chenodeoxycholic acid (CDCA; 3α,7α-hydroxy) with their relative proportions determined by levels of 12α-hydroxylase in the liver.[33,34] The ratio of taurine- to glycine-conjugated BAs is dependent on diet in humans. A high-protein diet results in greater taurine conjugation, while vegetarian diets lead to more glycine conjugation.[33] CA and CDCA are the primary BAs produced in humans, whereas other vertebrates produce bile salts that differ in ring hydroxylation pattern, as well as side-chain length and functional groups. The main classes are C_{24} BAs, C_{27} BAs, and C_{27} bile alcohols.[41] C_{24} BAs are common in all vertebrates, but with differing hydroxylation patterns. For example, mice produce CA and convert CDCA to muricholic acids (3,6,7-hydroxy) via hydroxylation and epimerization at C-6. C_{27} bile alcohols are typically synthesized in fish[42] and amphibians, while C_{27} BAs are present in reptiles and birds[41].

Once synthesized, conjugated BAs are actively transported out of hepatocytes into the bile duct. Conjugated BAs are stored in the gallbladder until the gallbladder is emptied into the duodenum in response to a meal.[43] Conjugated bile salts form mixed micelles with cholesterol, lipid-soluble vitamins, and dietary lipids throughout the small intestine. In the ileum, a sodium-dependent transporter (IBAT) takes up BAs into ileocytes.[44] From ileocytes, they are exported by organic solute transporter OSTα/β[45,46] into the portal vein, where they circulate back to the liver in a process known as enterohepatic circulation.[47] However, ~500 mg of BAs each day are not

taken up in the ileum and progress to the colon where they encounter gut microbiota.[37] Microbial metabolites of BAs can be passively absorbed in the colon, travel through the portal vein, and join the recycled host-derived BAs in the liver. Thus, the biliary pool consists of both host- and microbiota-derived BAs that are re-conjugated and, in some species, 7-hydroxylated, as they return to the liver.[48]

In addition to the digestive function of BAs, they are now known to act as hormone signaling molecules. BAs are involved in regulation of their own biosynthesis, as well as energy, glucose, and lipid metabolism.[43] Farnesoid X receptor (FXR, NR1H4) is a BA-activated nuclear receptor expressed in tissues such as liver, intestine, and kidney.[49,50] FXR regulates BA biosynthesis and enterohepatic circulation through many mechanisms. The FXR/SHP (small heterodimer partner) pathway of regulation involves the inhibition of *CYP7A1*, the rate-limiting step in BA formation. FXR induces the nuclear receptor, SHP, which inhibits liver-related homolog-1 (LRH-1) and hepatocyte nuclear factor 4α (HNF4α), both leading to inhibition of *CYP7A1* transcription.[51–53] Another pathway involves FXR, fibroblast growth factor 19 (FGF19), and FGF receptor 4 (FGFR4), which also results in inhibition of *CYP7A1*. Before recirculation back to the liver, BAs stimulate intestinal FXR, which induces FGF19 synthesis in ileocytes.[54] FGF19 is transported to the liver, where it binds FGFR4 and activates the *c-jun* N-terminal kinase (JNK) 1/2 signaling cascade, leading to downregulation of *CYP7A1*.[33,55]

Pregnane X receptor (PXR) and vitamin D receptor (VDR) are both nuclear receptors activated by microbial-derived BAs that also lead to the binding of *CYP7A1* promoter and repression of CYP7A1.[8,56–58] Takeda G-protein receptor 5 (TGR5) is a G-protein-coupled receptor for BAs that is expressed in intestinal and biliary epithelial cells among other cell types.[59,60] TGR5 has widespread effects throughout the body, including regulation of intestinal

motility.[61] Taurine-conjugated BAs activate TGR5 more effectively than unconjugated or glycine-conjugated BAs.[62] TGR5 signaling can activate epidermal growth factor receptor (EGFR).[63] EGFR is also a BA receptor that, once bound, initiates a signaling pathway ending in inhibition of CYP7A1.[43,64] In the gut, primary bile salts can be microbially biotransformed to dozens of metabolites whose concentrations and affinities can impact host physiological response in the intestine.

Microbial Bile Acid Metabolism

Bile acids that enter the colon are metabolized by gut microbiota through a combination of de(re)conjugation, 7α/β-dehydroxylation, and epimerization (**Figure 1.2**). The first step of microbial BA metabolism, known as deconjugation, mainly occurs in the small intestine and involves the hydrolysis of the C-24 *N*-acyl bond linking the conjugated amino acid to the BA. This reaction is catalyzed by bile salt hydrolase (BSH) encoded by diverse microbiota, including *Clostridium*[65,66], *Bacteroides*[67,68], Lactobacillaceae[69], *Bifidobacterium*[70,71], *Enterococcus*[72], and archaea[73]. BSHs have differing substrate specificity and subunit size, but often have conserved active site Cys, Arg, Asp, Asn, and another Arg.[74] BSHs have a pH optimum of 5–6 and are typically intracellular[65,70], although activity has been reported extracellularly in some cases.[66] Interestingly, re-conjugation of BAs by gut microbiota has recently been observed with unique amino acids: Phe, Tyr, and Leu.[75]

There are multiple hypotheses on the evolutionary role of BSH in microbial fitness: interspecies competition, detoxification, and release of an energy source. Deconjugated BAs are more toxic than conjugated bile salts to some bacterial species; thus, deconjugation may serve a competitive function to inhibit other bacteria.[4] However, the reverse may also be true. Some

bacteria are more sensitive to conjugated BAs and, thus, BSH may help them detoxify their environment.[76] Amino acids released from deconjugation could be an important energy source for certain microbiota, such as *Clostridium* that can utilize amino acids through Stickland fermentation.[77]

Deconjugated primary BAs can be 7-dehydroxylated by a select few species within the gut, including *Clostridium scindens*, *C. hylemonae*, and *C. hiranonis* (now reclassified as *Peptacetobacter hiranonis*).[4,78–80] Through this process, the primary BAs CA and CDCA are converted to "secondary" deoxycholic acid (DCA; 3α,12α-hydroxy) and lithocholic acid (LCA; 3α-hydroxy), respectively. Although so few species encode the 7α-dehydroxylation pathway, secondary BAs make up the majority of excreted BAs[74,81,82], meaning these microbiota have extensive dehydroxylation capacity.

The 7-dehydroxylation pathway is encoded by the polycistronic BA-inducible (*baiABCDEFGHI*) operon.[4,83,84] The first step is the import of unconjugated primary BAs by a BA transporter BaiG.[85] Next, ligation of CoA to the unconjugated BA is catalyzed by BA CoA ligase encoded by *baiB*, requiring ATP and Mg^{2+}.[86] Then, the 3α-hydroxyl group is oxidized by BaiA.[87] Three *baiA* genes from *C. scindens* have been reported in *C. scindens* VPI 12708, although completion of the *C. scindens* American Type Culture Collection (ATCC) 35704 genome revealed the presence of only two, with *baiA2* located in the *bai* operon [88–91] These enzymes are NAD(H)-dependent BA 3α-HSDHs that are specific for BA-CoA conjugates.[87] BaiCD is an NADH:flavin-dependent oxidoreductase that creates a C-4=C-5 double bond on 7α-hydroxy BA intermediates, while BaiH has the same function on 7β-hydroxy BAs.[92] CoA is then hydrolyzed by BaiF or BaiK and transferred without requirement of ATP to an incoming primary BA.[93] Subsequent 7α-dehydration is the rate-limiting step in the pathway, catalyzed by the *baiE*

product.[94] 7β-Dehydration is predicted to be carried out by BaiI.[95] Recently, a recombinant flavoprotein encoded by *baiN*, which is not a part of the *bai* operon, was shown to convert 3-dehydro-DCA to a product 4 amu less than the substrate.[96] Further characterization is necessary, but this suggests that *baiN* may catalyze reduction of both Δ^4 and Δ^6-intermediates following 7-dehydration.[96] Alternatively, BaiCD and BaiH were reported to be sufficient for C-4=C-5 and C-6=C-7 metabolism in the oxidative and reductive arms of the pathway.[97] The final step in the pathway, converting the 3-oxo intermediate to a secondary BA, is likely to be carried out by the products of one or both copies of *baiA*.[98] The BA exporter is not yet known.[4] However, two genes co-localized with *baiN* have been proposed, but not yet confirmed, to catalyze the final reaction and BA export, named BaiO and BaiP, respectively.[99] Several additional candidate export proteins were identified through transcriptomic analysis of *C. scindens* ATCC 35704 after BA induction.[91]

The 7α/β-dehydroxylation pathway results in a net two-electron reduction, meaning a net of one NAD^+ is produced when a primary BA is used as an electron acceptor.[74] The 7α/β-dehydroxylation pathway is likely coupled to glucose metabolism, benefitting 7α/β-dehydroxylating bacteria.[91] The pathway may serve another function in producing secondary BAs, which are more hydrophobic and toxic to gut bacteria, to regulate the growth of competing gut microbiota.[7,100] For example, DCA has a minimum inhibitory concentration tenfold lower than CA against many *Lactobacillus* and *Bifidobacterium* species.[100]

Both primary and secondary BAs can be oxidized and epimerized at position C-3, C-7, and/or C-12 reversibly from the α-orientation to an oxo-intermediate and further to the β-orientation by microbial HSDHs. Epimerized BAs have specific nomenclature: those containing 3β-hydroxyl groups are iso-BAs, while 7β- and 12β-BAs are recommended to be denoted epi-

BAs preceded by the hydroxyl position, according to Hofmann et al. (1992).[101] However, 7β-BAs are generally accepted to be named urso-BAs. For simplicity in this review, each prefix refers to only one of the β-hydroxyl positions: iso for 3β-, urso for 7β-, and epi for 12β-hydroxyl (**Figure 1.3**). Similarly to humans, mouse α-muricholic acid (3α,6β,7α-hydroxy) and β-muricholic acid (3α,6β,7β-hydroxy) can be oxidized and epimerized to ω-muricholic acid (3α,6α,7β-hydroxy) via a 6-oxo-intermediate.[102] Numerous microbiota are capable of oxidoreduction of BAs, including *Eggerthella lenta*[103], *C. scindens*[23,87], *P. hiranonis*[23], *C. hylemonae* (see Chapter 2)[23], *Escherichia coli*[104], and *Bacteroides fragilis*[105].

Microbial Bile Acid Hydroxysteroid Dehydrogenases

Microbial HSDHs catalyze the NAD(P)(H)-dependent oxidation and reduction of hydroxyl groups on BAs in the gut (**Figure 1.3**). Human interest in ursodeoxycholic acid (UDCA; 3α,7β-hydroxy) has a long and fascinating history. Asiatic black bear bile has been used in traditional Chinese medicine to treat disease for over 1000 years.[106] In the early 1900s, a BA was isolated from polar bear bile and, later, the same BA was crystallized from the American black bear. This BA was named ursodeoxycholic acid after the Latin name *ursus*.[107] UDCA makes up about 3–4% of the human BA pool but, in contrast to bear bile, is a secondary BA in humans.[108,109] UDCA and other urso-BAs are produced by combined microbial 7α-HSDH and 7β-HSDH activity in the human gut. Both microbial 7α- and 7β-HSDHs are typically NADP(H)-dependent, and they frequently exhibit specificity for dihydroxy-BAs (e.g., CDCA and UDCA) over trihydroxy-BAs (e.g., CA and UCA)[104,105,110–114], although exceptions have been reported.[115,116]

Urso-BAs are more hydrophilic and less toxic both to microbiota and to the host than DCA or LCA.[7] Indeed, DCA and LCA are involved in various diseases, such as cancers of the colon and liver.[117–120] UDCA is currently approved for treatment of biliary disorders[121], is being studied for both chemoprevention and chemotherapy of various cancers[108,122], and is undergoing clinical trials as part of a combination chemotherapy for colorectal cancer (clinicaltrials.gov identifier: NCT00873275). Its mechanism of action likely involves the displacement of more toxic BAs in the BA pool and its choleretic effect of inducing secretion of BAs from the liver.[123] However, UDCA can be 7β-dehydroxylated by certain gut microbiota or isomerized back to 7α-hydroxy prior to 7α-dehydroxylation.[124,125] 7β-Dehydroxylation of UDCA forms LCA, which may explain various toxicities associated with UDCA treatment.[126]

The iso-BA pathway is catalyzed by the paired action of BA 3α- and BA 3β-HSDH. Generally, 3α-HSDHs utilize NAD(H), whereas 3β-HSDHs require NADP(H). They also usually prefer dihydroxy-BAs (derivatives of DCA or CDCA) over trihydroxy-BAs (derivatives of CA).[17,18,112,127] BA 7α-dehydroxylating bacteria express a 3α-HSDH (BaiA) that differs greatly in substrate specificity as it reacts with CoA conjugates, not free BAs.[87] Iso-BAs are present ranging from 0% to about 20% of the total BA pool in the gut.[109] Iso-BAs have greatly decreased detergent nature and are thus less cytotoxic to gut microbiota, as well as the host, than DCA or LCA.[6,17] 3α/β-HSDHs may be of pharmaceutical use with respect to modulating the BA pool in favor of less toxic iso-BAs. Iso-BAs are intrinsically poor detergents and impede nutrient absorption. The liver epimerizes iso-BAs back to the 3α-hydroxyl form via a cytosolic 3β-HSDH.[128] Further studies are needed to determine the viability of developing strategies to favor iso-BAs.

Compared to the iso- and urso-BA pathways, the least is known about the epi-BA pathway. While multiple 12α-HSDHs have been characterized[18,23,103,116,129,130], BA 12β-HSDH was only studied in cell extracts until the discovery of the first gene encoding this activity in the current study (see Chapter 3).[24,131,132] 12-Oxolithocholic acid (12-oxoLCA; 3α-hydroxy,12-oxo), the product of 12α-HSDH oxidation of DCA, is often one of the most abundant oxo-BAs found in human feces, at concentrations of about one half DCA in some studies.[81,133,134] Of note, levels of 12-oxoLCA were increased in rats with high incidence of tumors after being fed a diet high in corn oil or safflower oil.[135] Measurement of epi-BAs is rare in the literature. EpiDCA (3α,12β-hydroxy) was first identified in human feces by Eneroth et al. (1966).[136] Recently, Franco et al. (2019) measured 3-oxo-12β-hydroxy-CDCA in humans, but little is known about concentrations of epiDCA or epiCA (3α,7α,12β-hydroxy) in feces.[81] EpiDCA has also been identified in the biliary bile of angelfish; hence, 12β-HSDH activity is likely present within the microbiome of diverse vertebrates.[41]

Many gut microbial 12α-HSDHs have NADP(H) specificity[18,23,129,130], while others are NAD(H)-specific.[116] 12α-HSDHs generally have higher activity with free and dihydroxy-BAs than conjugated or trihydroxy-BAs.[18,23,129] The only gut microbial BA 12β-HSDH characterized to date, from *Clostridium paraputrificum* ATCC 25780, has affinity for NADP(H) and greater activity with dihydroxy-BAs (see Chapter 3).[24,132] Two additional 12β-HSDHs have been shown to react with 12-oxoLCA and epiDCA with NADP(H) as co-substrate, although their substrate specificities have not been fully characterized (see Chapter 3).[24] Interestingly, 12β-HSDH activity recognizing side-chain cleaved steroids derived from BAs has been observed in multiple environmental microorganisms. This activity is displayed by *Comamonas testosteroni* TA441[137]

and *Pseudomonas* sp. strain Chol1[138] as they convert a 12-oxo-intermediate into 7α,12β-dihydroxy-androsta-1,4-diene-3,17-dione (12β-DHADD) in a cholic acid degradation pathway.

Epi-BAs are understudied compared to urso- and iso-BAs. Thus, their toxicity relative to secondary BAs is untested, although epiDCA and 12-oxoLCA are less hydrophobic than DCA according to LC–MS (see Chapter 3).[24] It is possible that isomerization of primary BAs to iso- or epi-BAs may impede formation of secondary BAs if they cannot be recognized by 7-dehydroxylation pathway enzymes. This could be of therapeutic importance because secondary BAs DCA and LCA are not only toxic to gut microbiota, but also to the human host.

Our knowledge of microbial HSDHs is largely limited to studies in humans and rodents. Notable recent studies extend to black bears in the search for HSDHs capable of forming UDCA [139]. There is a rich diversity of bile salts produced in vertebrates, such as pythocholic acid (16α-hydroxycholic acid; 3α,12α,16α-trihydroxy-5β-cholan-24-oic acid) found in snakes, which is a 16α-hydroxylated derivative of DCA.[106,140] Avicholic acid (3α,7α,16α-trihydroxy-5β-cholan-24-oic acid), found in birds, was identified in a drug screen as a TGR5 agonist.[141] An NAD(P)-dependent 16α-HSD was purified and characterized from rat kidney[142]; however, to our knowledge microbial 16α-HSDH activity has not yet been reported in snake or bird gastrointestinal content.

Physiological Roles of Microbial Bile Acid Hydroxysteroid Dehydrogenases

The physiological function of many microbial BA HSDHs remains unclear, although species and strain context seem likely to be important. In all cases, these redox reactions affect NAD(P)/NAD(P)H ratios and BA oxo-groups provide substrates for disposal of excess reducing equivalents or acquisition of hydrides in order to detoxify molecular oxygen close to the gut

mucosa. Oxidation and epimerization of BA α-hydroxyl groups to β-hydroxyl groups is also

thought to function in detoxification by converting hydrophobic BAs to hydrophilic BAs that are

less damaging to biological membranes.[7,17] For example, isoDCA has a minimum inhibitory

concentration of more than double that of DCA against various Gram-negative *Bacteroides* and

Gram-positive species.[17] In contrast, some HSDHs seem to favorably produce DCA from oxo-

derivatives, suggesting they may function to maintain high concentrations of DCA in the

environment (see Chapter 2).[23]

Culture-based studies indicate that the oxidation and epimerization of primary BAs

affects the extent of BA 7α-dehydroxylation.[143] There are several hypotheses that could explain

this observation. First, there is currently a paucity of knowledge relating to substrate specificity

of the BA transporter, BaiG, and whether oxo- and iso-BAs are efficiently imported. Our recent

study indicates that 3,7-dioxoLCA is converted to CDCA and low levels of LCA by *C.

scindens*[143], albeit to lower levels than CDCA addition, suggesting import is occurring. Second,

BA 7α-dehydroxylating bacteria appear to lack significant 3β-HSDH activity and, as a result,

iso-primary BAs (3β-hydroxy) are not substrates for the BA 7α-dehydroxylation pathway.[143] As

noted above, the first oxidation step and the last reductive step in the BA 7α-dehydroxylation

pathway are catalyzed by 3α-HSDH (BaiA). A 3β-hydroxyl group, thus, prevents key oxidation

steps that lead to 7α-dehydration. Indeed, LCA was not observed in cultures of *C. scindens* VPI

12708 induced with CA (resulting in upregulation of Bai enzymes) and then incubated with

isoCDCA.[143] While trace levels of isoLCA (<1%) have been reported in vitro during BA

metabolism by *C. scindens* ATCC 35704[144], this may be due to the minor promiscuity known for

some bacterial HSDHs.[96] Iso-secondary BAs (e.g., isoDCA and isoLCA) are second only to

DCA and LCA in abundance in stool[109] and are less toxic than LCA and DCA to intestinal

bacteria.[17] Iso-BA epimerizing HSDHs also show substrate specificity preference toward secondary BAs.[18] It is, therefore, hypothesized that isoLCA and isoDCA are generated from LCA and DCA, respectively, in the gastrointestinal tract. A third point is that enrichment of primary oxo- and β-hydroxy-BAs comes at the expense of primary BAs such as CA and CDCA, which induce expression of the *bai* operon.[145] Indeed, culture-based studies indicate that *C. scindens* VPI 12708 is capable of converting 3,7-dioxocholanoic acid and 7-oxoLCA to LCA only if the cells were preincubated with CA.[143]

Numerous gut bacteria, including *Bacteroides* spp. and *E. coli*, encode 7α-HSDH and produce 7-oxo-BAs that are released into the lumen.[105,146,147] The formation of 7-oxo-primary BAs precludes 7α/β-dehydration by the *bai* pathway and must be reduced to proceed. It is, therefore, not surprising that BA 7α-dehydroxylating bacteria express NADP-dependent 7α-HSDH.[114] The BA 7α-HSDH is predicted to be important both in regulating the NAD(H)-dependent BA 7α-dehydroxylating pathway intracellularly and in reducing 7-oxo-BAs imported from the environment.

BA 7α-dehydroxylating bacteria also encode BA 12α-HSDH (see Chapter 2).[23,148] The formation of 12-oxo-BAs reduces toxicity of BAs toward gut bacteria[7], which is likely why a wide diversity of gut bacteria encode 12α-HSDH.[18,23,143,149] However, substrate specificity of 12α-HSDHs in 7α-dehydroxylating bacteria favors the reductive direction, converting 12-oxoLCA to DCA (see Chapter 2).[23] We, therefore, hypothesize that BA 7α-dehydroxylating bacteria express BA 12α-HSDH principally to "retoxify" 12-oxoLCA that was generated by bacteria less resistant to DCA.

We recently demonstrated extensive oxidation of BAs by *Eggerthella lenta*.[143] Indeed, *E. lenta* strains C592 and DSM 2243 encode 3α-, 3β-, 7α-, and 12α-HSDHs capable of converting

CA to trioxo-cholanoic acid under a nitrogen or carbon dioxide atmosphere. However, BA oxidation was inhibited under a hydrogen gas atmosphere (**Figure 1.4**). Genomic analysis revealed genes encoding energy conserving hydrogenase (*echABCDEF*) and Rnf complex (*rnfABCDEG*), as well as a complete Wood–Ljungdahl pathway, suggesting that *E. lenta* is an acetogen.[143,150] The classical acetogen fixes CO_2 or CO in the presence of H_2[151]; however, acetogens are known to utilize a wide range of electron donors. Under this scheme, *E. lenta* HSDH enzymes are hypothesized to generate NADH by oxidizing BAs, which provides reducing equivalents to fix CO_2. In the presence of H_2, *E. lenta* hydrogenases reduce NAD^+ via molecular hydrogen, and BA oxidation is prevented. Additional studies will be needed to confirm this hypothesis linking BA metabolism and H_2 partial pressure in a novel acetogen.

The role of microbial BA HSDHs in host physiology is also relatively unclear. While the involvement of oxo- and β-BAs in host signaling pathways has not been fully explored, there is evidence that products in the iso-BA pathway activate various host receptors. For example, along with LCA, 3-oxoLCA has been shown to activate the BA receptors FXR, VDR, and PXR.[8,56] In contrast, 12-oxoLCA, 7-oxoLCA, and UDCA did not efficiently activate either FXR or VDR.[8,50] Recently, 3-oxoLCA and a planar iso-BA, isoalloLCA, were shown to be regulators of interleukin (IL)-17a expressing T helper cells (T_H17) and regulatory T cells (T_{reg}) in mice.[152] Determining the full spectrum of both primary and secondary oxo- and β-derivatives against BA-responsive nuclear and G protein-coupled receptors will be important future work.

GLUCOCORTICOID METABOLISM

Host Glucocorticoid Synthesis

Glucocorticoids are involved in diverse essential physiological processes throughout the body.[153] Cortisol and corticosterone are the primary C_{21} glucocorticoids present in humans. However, cortisol concentrations are about 10 times greater than corticosterone.[154] Cortisol plays a major role in the stress response and maintenance of blood glucose concentration, as well as in inhibition of protein synthesis in muscle, of lipogenesis in fat cells, and of the immune system.[155]

Cortisol is synthesized in the adrenal gland from cholesterol and involves the action of both cytochrome P450 enzymes and hydroxysteroid dehydrogenases, much like BA biosynthesis (**Figure 1.2**). The first step is catalyzed by CYP11A1, which side-chain cleaves cholesterol and results in pregnenolone.[9] This is the rate-limiting step and precursor to many other steroid hormones, including progesterone, corticosterone, aldosterone, testosterone, and estradiol.[156] 17α-Hydroxyprogesterone is then produced by CYP17A1 (17-hydroxylase/17,20 lyase) and HSD3B2 (3β-HSD/$\Delta^{5/4}$-isomerase type 2). CYP21A2 converts 17α-hydroxyprogesterone to 11-deoxycortisol. The last reaction results in the formation of cortisol through the action of CYP11B1.[9,10] Cortisol circulates in serum at concentrations between 100 and 600 nM.[9] Cortisol then acts in peripheral tissues by binding to the nuclear glucocorticoid receptor, resulting in regulation of numerous genes, including those involved in inflammation, immune function, and gluconeogenesis. Cortisol can also bind to mineralocorticoid receptor, which regulates electrolyte balance.[157,158] Cortisol concentrations are tightly regulated by 11β-HSD isoforms 1 and 2. 11β-HSD1/2 interconvert cortisol (C-11 hydroxyl) to its inactive form, cortisone (C-11 ketone), which cannot bind the glucocorticoid receptor or mineralocorticoid receptor. 11β-HSD1 functions primarily as a reductase to activate cortisol in the liver, muscle, and bone. In contrast,

11β-HSD2 acts as a dehydrogenase, inactivating cortisol to cortisone in the kidney, colon, and salivary glands.[9]

Human tissues metabolize cortisol in various ways, leading to its excretion primarily in urine. However, low levels of cortisol and its derivatives are secreted in bile and enter the gut.[159] Cortisol undergoes 5α- or 5β-reduction in the liver, while cortisone is only 5β-reduced.[160] After 3α-reduction, 5α/β-tetrahydrocortisol and tetrahydrocortisone are produced, which are the main metabolites of cortisol and cortisone in urine, respectively.[9] Cortisol can also be metabolized by 20α- and 20β-HSDs, yielding either 20α- or 20β-dihydrocortisol.[161] Carbonyl reductase-1 (CBR1) has 20β-HSD activity producing 20β-dihydrocortisol, while a host 20α-HSD has been observed with specificity for progesterone, but not cortisol.[9,162] 20α/β-Reduction of tetrahydrocortisol and tetrahydrocortisone results in α/β-cortols or α/β-cortolones.[163]

Host Androgen Synthesis

Androgens are important for metabolic homeostasis and reproductive function in men, as well as women. Androgens are C_{19} steroids that are synthesized in the Leydig cells of the testes or adrenal glands.[164] The primary active androgens in circulation are testosterone and dihydrotestosterone, although, in the adrenal glands, the major products are the androgen precursors dehydroepiandrosterone (and its sulfate ester), androstenedione, and 11β-hydroxyandrostenedione (11β-OHAD).[165]

Androgen biosynthesis in the adrenal cortex begins with side-chain cleavage of cholesterol to pregnenolone by CYP11A1. Then, CYP17A1 hydroxylase and 17,20-lyase activities produce dehydroepiandrosterone (DHEA). HSD3B2 (3β-HSD/$\Delta^{5/4}$-isomerase type 2) converts DHEA to androstenedione. Alternatively, AKR1C3 (17β-HSD) can produce

androstenediol from DHEA, and HSD3B2 then yields testosterone. Androstenedione can be further converted to 11β-OHAD by adrenal-specific CYP11B1 (11β-hydroxylase).[166]

Even though 11β-OHAD makes up a large proportion of adrenal steroidogenesis, it has historically largely been ignored (except in fishes) due to its low androgenic activity.[167] Storbeck et al. (2013) reported that 11β-OHAD leads to the formation of 11-ketotestosterone (11KT)[168], a potent 11-oxygenated C_{19} androgen involved in castration-resistant prostate cancer[169,170] and polycystic ovary syndrome.[170,171] This is important because, although 11β-OHAD is primarily produced in the adrenal glands by CYP11B1, peripheral side-chain cleavage of cortisol to 11β-OHAD also occurs.[172] Peripheral 11β-OHAD is not formed by CYP17A1.[173] Thus, the enzyme responsible for cortisol-derived 11β-OHAD may be an unknown host enzyme and/or of microbial origin. Intriguingly, 11β-OHAD has been shown to be produced from side-chain cleavage of cortisol by human gut microbiota.[14,174–176]

Androgens signal throughout the body by binding to androgen receptor (AR) expressed in various cell types, including B cells, T cells, neutrophils, and macrophages[177], as well as colon cancer cell lines[178]. Nuclear AR is a ligand-dependent transcription factor that, when activated by an androgen, regulates expression of cell growth, differentiation, and even carcinogenesis in some cases.[179] Intestinal cells express both nuclear AR and membrane AR.[178–180] Importantly, the gut microbiome has evolved enzymes that catalyze many of the same reactions described for host glucocorticoid and androgen metabolism. This indicates that the host endocrine system has interkingdom components in need of further exploration.

Microbial Cortisol Metabolism

The earliest evidence of microbial biotransformation of cortisol was observed when rectal infusion of cortisol in ulcerative colitis patients led to an increase in urinary excretion of 17-ketosteroids.[181] This increase in urinary steroids was not detected when cortisol treatment coincided with oral neomycin[182], suggesting microbial biotransformation of cortisol. Thereafter, side-chain cleavage of cortisol or steroid-17,20-desmolase activity was observed when human fecal samples produced C_{19} steroids after incubation with cortisol.[176]

In 1984, a bacterium was isolated from human fecal material exhibiting steroid-17,20-desmolase activity producing 11β-OHAD from cortisol (**Figure 1.2**).[174,175] This organism was named *Clostridium scindens*, formerly *Clostridium* strain 19, which also has BA 7α-dehydroxylation activity.[4] Additional organisms with steroid-17,20-desmolase activity were then isolated: *Butyricicoccus desmolans* ATCC 43058 (formerly *Eubacterium desmolans*), *C. cadaveris* AGR2141[183], and the urinary microbe *Propionimicrobium lymphophilum* ACS-093-V-SCH5.[184,185] The operon encoding this activity (*desABCD*) has since been identified by performing RNA-Seq after inducing *C. scindens* ATCC 35704 with cortisol.[14] The inducible *desABCD* operon consists of steroid-17,20-desmolase (DesAB) encoded by *desAB*, a 20α-HSDH (DesC), and a putative transporter (DesD) (**Figure 1.5**).[14,186] *C. scindens* ATCC 35704 DesAB was determined to be a heterotetramer and recognized both cortisol and 11-deoxycortisol, which only differs from cortisol in the absence of an 11β-hydroxyl group.[186]

C-20 reduced metabolites of cortisol have been observed in human urine, likely attributable to host enzymes that produce 20α- or 20β-dihydrocortisol and their derivatives [163,187]. However, Winter et al. (1982) showed that gut microbiota can reduce cortisol to 20β-dihydrocortisol, exhibiting 20β-HSDH (DesE) activity.[188] *B. desmolans* and *C cadaveris* express

20β-HSDH [183], along with *Bifidobacterium adolescentis.*[188] Additionally, the gut microbe *Clostridium scindens* ATCC 35704 can convert cortisol to 20α-dihydrocortisol.[174] Thus, gut microbiota encode 20α- and 20β-HSDHs that biotransform cortisol (**Figure 1.5**).

Human gut microbiota are also capable of 21-dehydroxylation of corticosteroids. 21-Dehydroxylase activity was first detected in *Eggerthella lenta* (formerly *Eubacterium lentum*).[189,190] *E. lenta* 21-dehydroxylase has substrate specificity for 11-deoxycorticosterone, deoxycortisol, dehydrocorticosterone, and corticosterone.[191,192] The enzyme requires NAD(P)H and flavin or only reduced flavin mononucleotide for activity.[192] Although this enzyme seems to be specific for corticosterone, 21-dehydroxylation of cortisol to 21-deoxycortisol also occurs.[176] Interestingly, 21-deoxycortisol is a substrate for 11β-HSD2[193] while the 21-dehydroxylation product of corticosterone is a potent inhibitor.[194]

Microbial Cortisol Hydroxysteroid Dehydrogenases

Host hydroxysteroid dehydrogenases have been established as important for biosynthesis and modulation of steroid hormones such as androgens, estrogens, and glucocorticoids for years.[5] Since the discovery of steroid hormone-converting HSDHs in the human gut microbiome, gut bacteria have been proposed to play an important role beyond that of the host in modification of steroids.[14] Within the steroid-17,20-desmolase pathway, two HSDHs have been identified that convert cortisol to 20α- or 20β-dihydrocortisol and may act as enzymatic switches to control formation of 11β-OHAD (**Figure 1.5**).

20β-Dihydrocortisol is excreted in urine at rates comparable to that of free cortisol in healthy individuals.[161,187] Urinary excretion of 20α-dihydrocortisol occurs at rates of about 1.5 times the excretion of cortisol.[161,187] Although the physiologic role of 20α- and 20β-

dihydrocortisol is not extensively studied, they are elevated in patients with Cushing's syndrome[187], as well as in patients with hypertension.[195]

One of the first organisms studied expressing 20β-HSDH activity was the soil microbe *Streptomyces hydrogenans*.[196] This enzyme reacted with not only cortisol, but also cortisone, cortexolone (lacks C-11 oxygen group), and their 21-aldehydes.[196] More recently, the genes encoding 20β-HSDH in *B. desmolans* and *C. cadaveris*, organisms that were previously shown to have this activity in culture, have been identified.[183,184] The gene is denoted *desE* due to its involvement in the DesAB pathway and because it forms an operon with the *desAB* genes.[14,184] Both *B. desmolans* and *C. cadaveris* are capable of cortisol side-chain cleavage, as well as 20β-oxidoreduction.[183,184] 20β-HSDH has been characterized in detail from *B. desmolans* ATCC 43058, which exhibits specificity for cortisol as substrate and is NAD(H)-dependent.[184] *Bifidobacterium scardovii* ATCC BAA-773 and the urinary tract microbe *Propionimicrobium lymphophilum* ACS-093-V-SCH5 also express 20β-HSDH according to HPLC[184], and *P. lymphophilum* has also been shown to encode *desAB*.[184,185] Additionally, the SDR family NAD(H)-dependent 20β-HSDH product of *desE* in *B. adolescentis* strain L2-32 has been characterized in this study. It is specific for cortisol and was crystallized in both the apo-form without any binding and the binary form with NADH bound at 2.2 and 2.0 Å, respectively (see Chapter 4).[15]

Thus far, 20α-HSDH activity seems to be significantly less widespread than 20β-HSDH, with only one organism shown to exhibit the activity.[14,197] Reduction of cortisol at the C-20 position to 20α-dihydrocortisol was observed in pure cultures of *C. scindens* along with steroid-17,20-desmolase activity.[175] 20α-HSDH from *C. scindens* ATCC 35704 was initially characterized from cell extracts and shown to be NAD(H)-dependent.[198] The gene for 20α-

HSDH was identified in 2013 after RNA-Seq analysis revealed a cortisol-inducible operon including *desAB* and *desC*, encoding steroid-17,20-desmolase and 20α-HSDH, respectively.[14] In this study, the *C. scindens* ATCC 35704 20α-HSDH was crystallized for further characterization of the enzymatic mechanism. Hybrid quantum mechanical molecular modeling simulations revealed a reaction mechanism involving a multistep proton relay, which was validated by site-directed mutagenesis experiments of active site and substrate binding residues (see Chapter 5).[16] An amino-acid homology search based on *C. scindens* ATCC 35704 20α-HSDH within the National Center for Biotechnology Information (NCBI) database uncovered two additional organisms, *Denitratisoma oestradiolicum* DSM 16959 and *Intestinibacillus* sp. Marseille-P4005, which may express 20α-HSDH, although activity has not yet been confirmed (see Chapter 3).[24]

Microbial 20α- and 20β-HSDH may be important regulators of the steroid-17,20-desmolase/DesAB pathway. By competing for cortisol as substrate with DesAB, they would decrease the potential for 11β-OHAD formation. Microbial steroid-17,20-desmolase activity may be one of the important missing enzymes contributing to peripheral 11β-OHAD production in the body.[199] Recent work showed that *Clostridium scindens* ATCC 35704 and the urinary microbe *Propionimicrobium lymphophilum* ACS-093-V-SCH5 can side-chain cleave both cortisol and glucocorticoid drugs[185], suggesting microbial production of 11β-OHAD may occur in both the gut and urinary tract. As mentioned above, 11β-OHAD can be further converted to highly androgenic 11KT.[168] This has compelling implications for androgen-dependent diseases, such as castration-resistant prostate cancer, or diseases defined by androgen excess, such as polycystic ovary syndrome.[170] Further studies are necessary to assess the efficacy of utilizing 20α- and/or 20β-HSDH to mediate 11β-OHAD formation in vivo.

CONCLUSIONS

Overall, both host and microbial HSDHs play pivotal roles in BA and glucocorticoid metabolism. Research on the importance of HSDH-derived BAs on host physiology is in its infancy. However, the immense diversity of these BA metabolites, due to combinations of HSDH activity, means that the gut harbors a multitude of potential candidates for host receptor signaling. Gut microbial cortisol HSDHs are likely important regulators of steroid-17,20-desmolase activity, although additional research is needed to ascertain the physiological significance of 20α- and 20β-HSDH products. New microbial HSDHs are continually being discovered and characterized, which will allow mechanistic study of their impacts in disease models.

Microbial HSDHs may have potential as therapeutic modulators in diseases such as colorectal cancer, liver cancer, castration-resistant prostate cancer, and polycystic ovary syndrome. However, to work toward therapeutics, we must first connect HSDH function to host phenotypes through mechanistic experiments, such as gnotobiotic animal studies.[200,201] Such avenues include developing genetic knockouts of HSDHs in microbes naturally encoding them or, when genetic systems are unavailable, engineering genetically tractable microbes to encode HSDHs. Furthermore, crystal structures of microbial HSDHs will aid in any necessary mutagenesis to rationally design substrate specificity for these enzymes. Integrating functional studies, genetic manipulation, structural biology, and gnotobiotic animal experiments will be imperative to reach a clearer picture of microbial steroid metabolism in the future.

FUNDING

ACKNOWLEDGMENTS
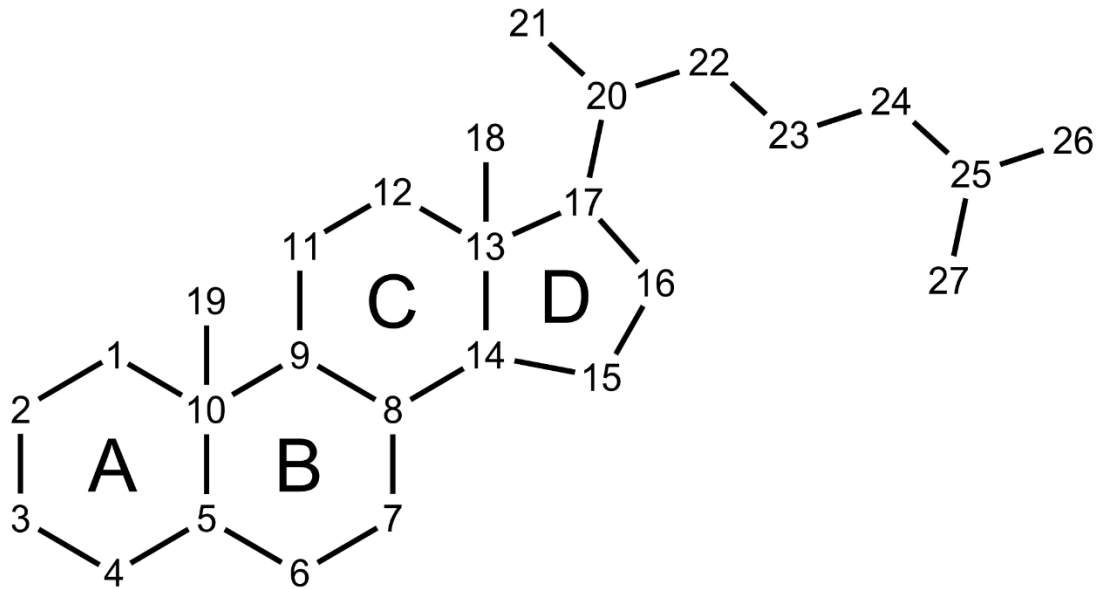
Figure 1.1. Steroid structure. Steroids have a cyclopentanoperhydrophenanthrene ring structure. Cholesterol, the precursor to human steroid hormones, contains 27 carbons, while the major classes of steroid hormones contain the following: C_{24} bile acids, C_{19} androgens, C_{18} estrogens, and C_{21} glucocorticoids, mineralocorticoids, and progestogens.
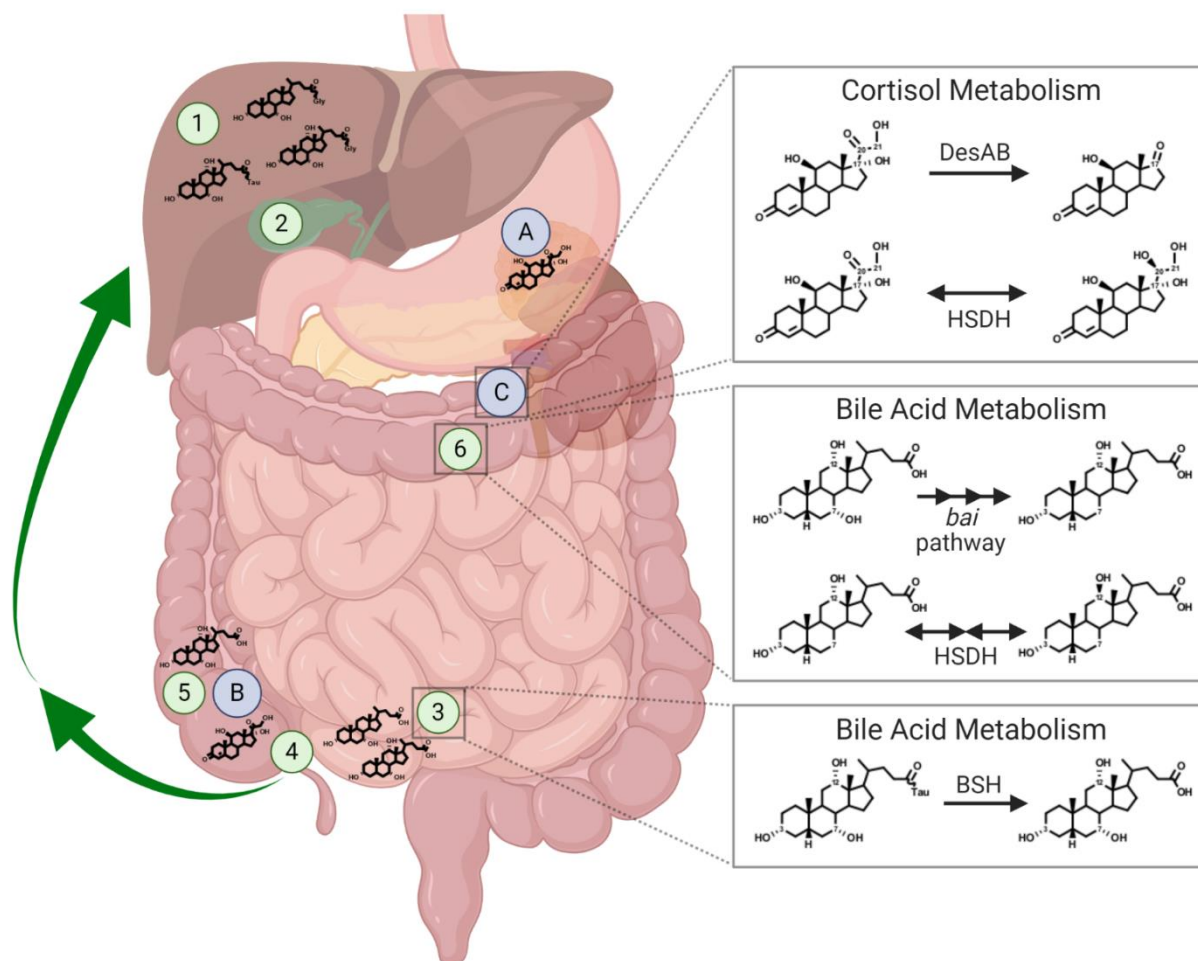
Figure 1.2. Synthesis and microbial metabolism of bile acids and cortisol. (1) The bile acids (BAs) cholic acid (CA) and chenodeoxycholic acid (CDCA) are synthesized and conjugated to glycine (Gly) or taurine (Tau) in the liver. (2) They are then stored in the gallbladder until they are released in response to a meal. (3) Microbial deconjugation of amino acids, catalyzed by bile salt hydrolase (BSH), primarily occurs in the small intestine. (4) BAs are taken up in the terminal ileum and undergo enterohepatic circulation back to the liver indicated by green arrows. (5) About 5% of BAs are not recycled and proceed to the colon. (6) Gut microbiota residing in the colon can 7α-dehydroxylate CA or CDCA to secondary BAs in a pathway encoded by the BA-inducible (*bai*) operon. Microbial hydroxysteroid dehydrogenases (HSDHs) interconvert BA

Figure 1.2 (cont.)

hydroxyl groups between the α- and β-conformations through an oxo-intermediate. (A) Cortisol

is synthesized in the adrenal glands. (B) Cortisol and its derivatives are principally excreted in

urine; however, low levels are secreted in bile and enter the gut. (C) In the gut, cortisol can be

side-chain cleaved by microbiota encoding steroid-17,20-desmolase (DesAB) or reduced to 20α-

or 20β-dihydrocortisol by HSDHs.
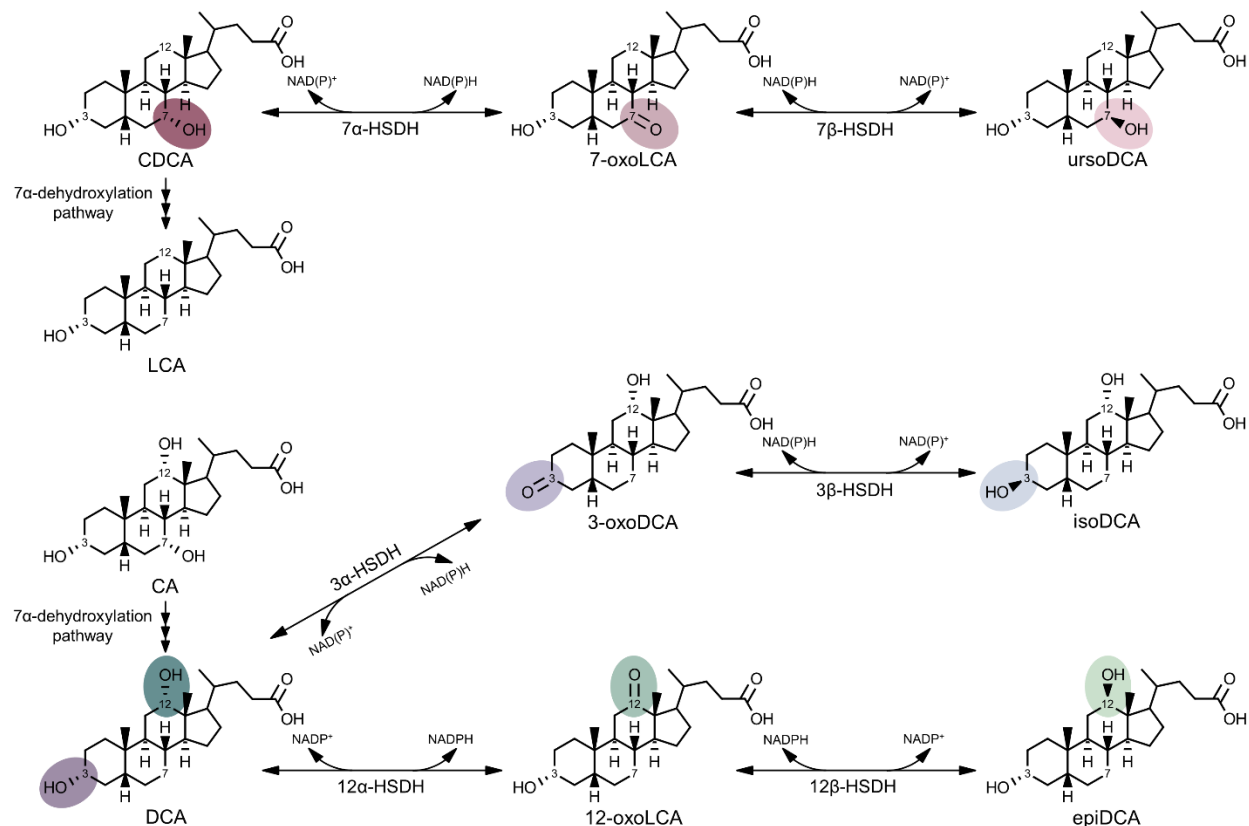
Figure 1.3. Microbial bile acid hydroxysteroid dehydrogenase metabolism. After

deconjugation by bile salt hydrolase, the primary bile acids (BAs) chenodeoxycholic acid

(CDCA) and cholic acid (CA) can be 7α-dehydroxylated or reversibly biotransformed by

NAD(P)(H)-dependent hydroxysteroid dehydrogenases (HSDHs). CDCA is converted to the

oxo-intermediate, 7-oxolithocholic acid (7-oxoLCA), and further to ursoDCA (UDCA) in the

urso-BA pathway catalyzed by 7α- and 7β-HSDH. The secondary BAs lithocholic acid (LCA)

and deoxycholic acid (DCA) are produced through the multi-step 7α-dehydroxylation of CDCA

and CA, respectively. 3α-HSDH biotransforms DCA into 3-oxoDCA, and 3β-HSDH converts 3-

oxoDCA to isoDCA in the iso-BA pathway. DCA is converted to 12-oxoLCA by 12α-HSDH

and from 12-oxoLCA to epiDCA by 12β-HSDH. HSDHs can recognize other BAs with the

correct hydroxyl group position and orientation beyond those depicted.
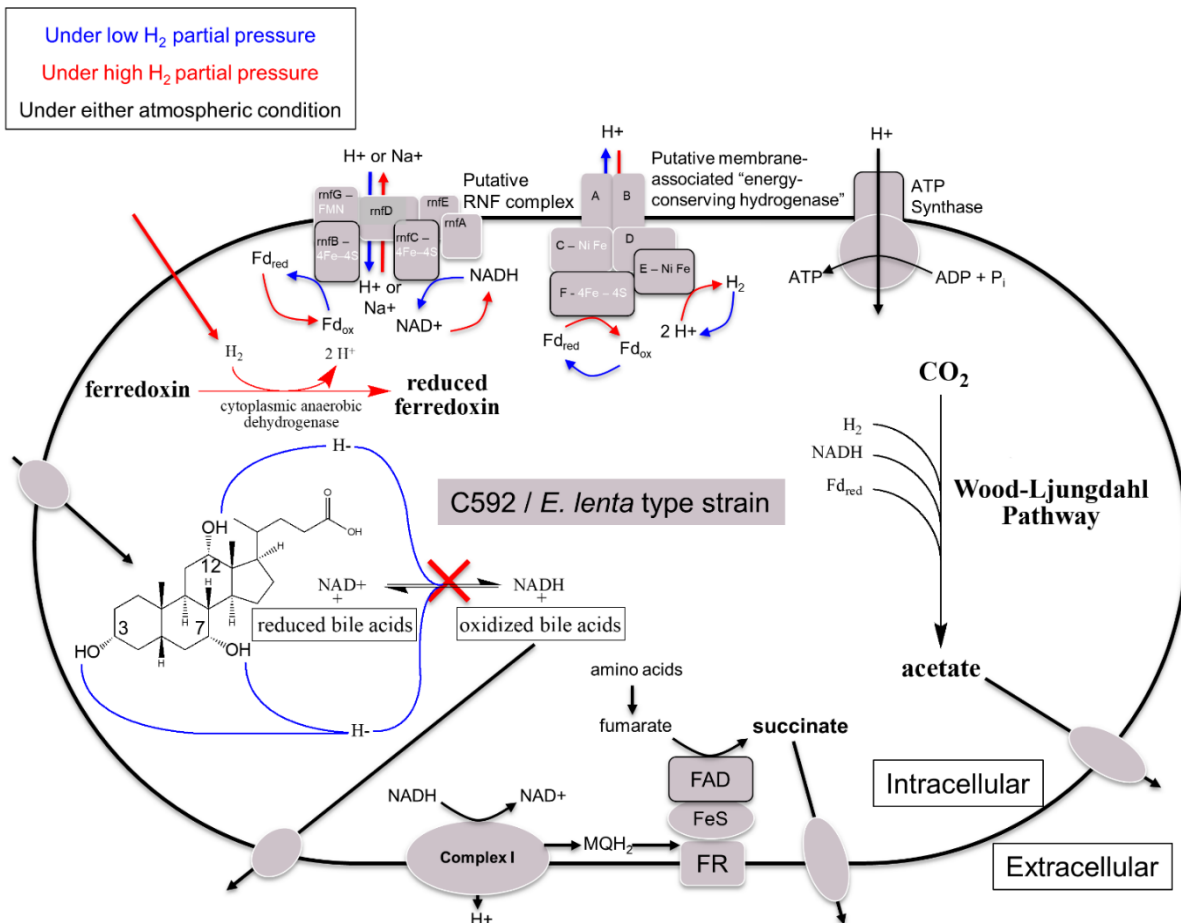
Figure 1.4. Proposed model for the role of *Eggerthella lenta* hydroxysteroid dehydrogenases: bile acid oxidation provides reductant for the Wood–Ljungdahl Pathway (WLP) of acetogenesis. This model is based on biochemical and genomic data demonstrating that *E. lenta* strains contain complete WLP genes, and that bile acid oxidation is inhibited by a hydrogen gas atmosphere.
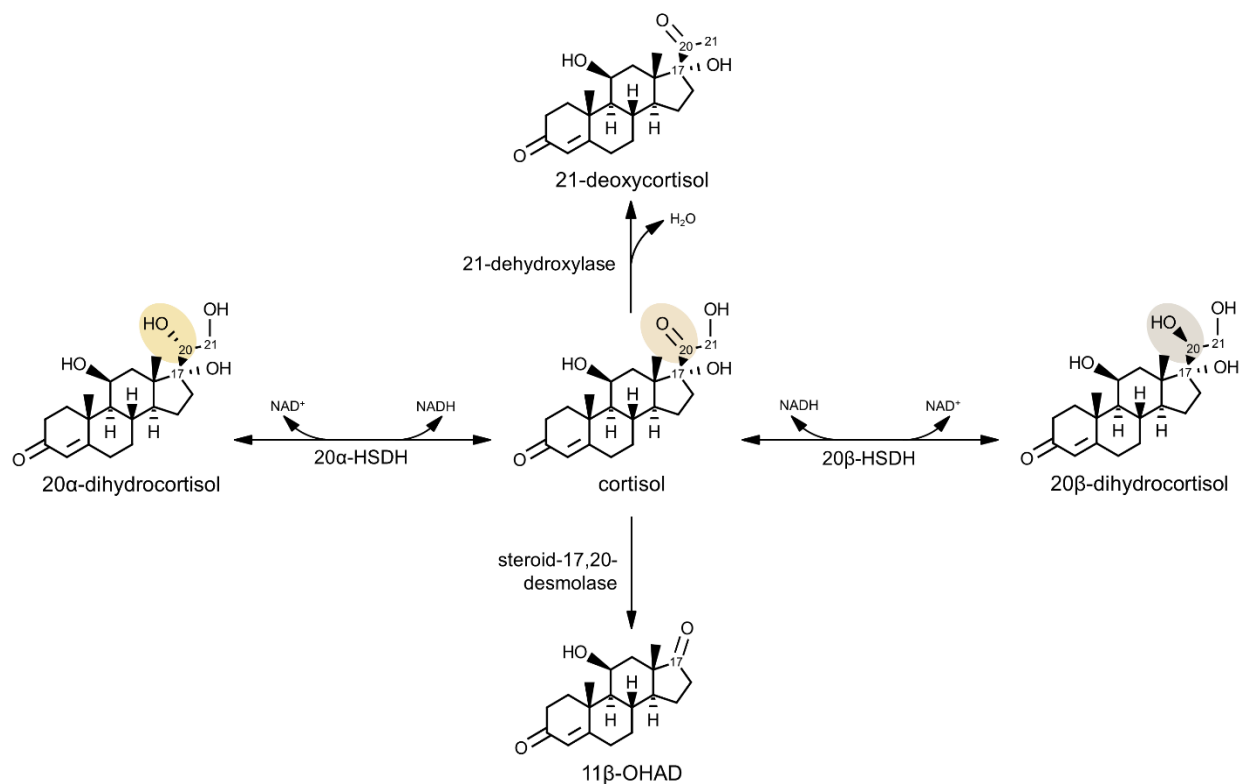
Figure 1.5. Microbial cortisol hydroxysteroid dehydrogenase metabolism. Cortisol can be reversibly biotransformed by 20β-hydroxysteroid dehydrogenase (20β-HSDH; DesE) to 20β-dihydrocortisol, or by 20α-HSDH (DesC) to 20α-dihydrocortisol. Steroid-17,20-desmolase (DesAB) converts cortisol to 11β-hydroxyandrostenedione (11β-OHAD). 21-Dehydroxylase catalyzes conversion of cortisol to 21-deoxycortisol.

REFERENCES

1. Norman AW, Henry HL. Steroid Hormones: Chemistry, Biosynthesis, and Metabolism. In: Norman AW, Henry HL, editors. Hormones. London: Academic Press; 2015. page 27–53.
2. Litwack G. Steroid hormones. In: Litwack G, editor. Human Biochemistry. London: Academic Press; 2018. page 467–506.
3. Payne AH, Hales DB. Overview of steroidogenic enzymes in the pathway from cholesterol to active steroid hormones. Endocr Rev 2004; 25:947–70.
4. Ridlon JM, Harris SC, Bhowmik S, Kang DJ, Hylemon PB. Consequences of bile salt biotransformations by intestinal bacteria. Gut Microbes 2016; 7:22–39.
5. Penning TM. Molecular endocrinology of hydroxysteroid dehydrogenases. Endocr Rev 1997; 18:281–305.
6. Hofmann AF, Roda A. Physicochemical properties of bile acids and their relationship to biological properties: An overview of the problem. J Lipid Res 1984; 25:1477–89.
7. Watanabe M, Fukiya S, Yokota A. Comprehensive evaluation of the bactericidal activities of free bile acids in the large intestine of humans and rodents. J Lipid Res 2017; 58:1143–52.
8. Makishima M, Lu TT, Xie W, Whitfield GK, Domoto H, Evans RM, Haussler MR, Mangelsdorf DJ. Vitamin D receptor as an intestinal bile acid sensor. Science 2002; 296:1313–7.
9. Schiffer L, Barnard L, Baranowski ES, Gilligan LC, Taylor AE, Arlt W, Shackleton CHL, Storbeck KH. Human steroid biosynthesis, metabolism and excretion are differentially reflected by serum and urine steroid metabolomes: A comprehensive review. J Steroid Biochem Mol Biol 2019; 194:105439.
10. Miller WL, Auchus RJ. The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. Endocr Rev 2011; 32:81–151.
11. Agarwal AK, Auchus RJ. Minireview: Cellular redox state regulates hydroxysteroid dehydrogenase activity and intracellular hormone potency. Endocrinology 2005; 146:2531–8.
12. Penning TM, Wangtrakuldee P, Auchus RJ. Structural and functional biology of aldo-keto reductase steroid-transforming enzymes. Endocr Rev 2019; 40:447–75.
13. Kenny DJ, Plichta DR, Shungin D, Koppel N, Hall AB, Fu B, Vasan RS, Shaw SY, Vlamakis H, Balskus EP, et al. Cholesterol metabolism by uncultured human gut bacteria influences host cholesterol level. Cell Host Microbe 2020; 28:245–57.
14. Ridlon JM, Ikegawa S, Alves JMP, Zhou B, Kobayashi A, Iida T, Mitamura K, Tanabe G, Serrano M, De Guzman A, et al. *Clostridium scindens*: a human gut microbe with a high potential to convert glucocorticoids into androgens. J Lipid Res 2013; 54:2437–49.
15. Doden HL, Pollet RM, Mythen SM, Wawrzak Z, Devendran S, Cann I, Koropatkin NM, Ridlon JM. Structural and biochemical characterization of 20β-hydroxysteroid dehydrogenase from *Bifidobacterium adolescentis* strain L2-32. J Biol Chem 2019; 294:12040–53.
16. Bernardi R, Doden H, Melo M, Devendran S, Pollet R, Mythen S, Bhowmik S, Lesley S, Cann I, Luthey-Schulten Z, et al. Bacteria on steroids: the enzymatic mechanism of an NADH-dependent dehydrogenase that regulates the conversion of cortisol to androgen in the gut microbiome. 2020; bioRxiv 2020.06.12.149468.

17.	Devlin AS, Fischbach MA. A biosynthetic pathway for a prominent class of microbiota-derived bile acids. Nat Chem Biol 2015; 11:685–90.

18.	Mythen SM, Devendran S, Méndez-García C, Cann I, Ridlon JM. Targeted synthesis and characterization of a gene cluster encoding NAD(P)H-dependent 3α-, 3β-, and 12α-hydroxysteroid dehydrogenases from *Eggerthella* CAG:298, a gut metagenomic sequence. Appl Environ Microbiol 2018; 84:e02475-17.

19.	Filling C, Berndt KD, Benach J, Knapp S, Prozorovski T, Nordling E, Ladenstein R, Jörnvall H, Oppermann U. Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. J Biol Chem 2002; 277:25677–84.

20.	Kallberg Y, Oppermann U, Jörnvall H, Persson B. Short-chain dehydrogenases/reductases (SDRs). Coenzyme-based functional assignments in completed genomes. Eur J Biochem 2002; 269:4409–17.

21.	Rossmann MG, Moras D, Olsen KW. Chemical and biological evolution of a nucleotide-binding protein. Nature 1974; 250:194–9.

22.	Kavanagh KL, Jörnvall H, Persson B, Oppermann U. The SDR superfamily: Functional and structural diversity within a family of metabolic and regulatory enzymes. Cell Mol Life Sci 2008; 65:3895–906.

23.	Doden H, Sallam LA, Devendran S, Ly L, Doden G, Daniel SL, Alves JMP, Ridlon JM. Metabolism of oxo-bile acids and characterization of recombinant 12α-hydroxysteroid dehydrogenases from bile acid 7α-dehydroxylating human gut bacteria. Appl Environ Microbiol 2018; 84:e00235-18.

24.	Doden HL, Wolf PG, Gaskins HR, Anantharaman K, Alves JMP, Ridlon JM. Completion of the gut microbial epi-bile acid pathway. bioRxiv 2021.

25.	Persson B, Hedlund J, Jornvall H. The MDR superfamily. Cell Mol Life Sci 2008; 65:3879–94.

26.	Nordling E, Jornvall H, Persson B. Medium-chain dehydrogenases/reductases (MDR): Family characterizations including genome comparisons and active site modelling. Eur J Biochem 2002; 269:4267–76.

27.	Knoll M, Pleiss J. The Medium-Chain Dehydrogenase/Reductase Engineering Database: A systematic analysis of a diverse protein family to understand sequence-structure-function relationship. Protein Sci 2008; 17:1689–97.

28.	Jez JM, Bennett MJ, Schlegel BP, Lewis M, Penning TM. Comparative anatomy of the aldo-keto reductase superfamily. Biochem J 1997; 326:625–36.

29.	Khanna M, Qin KN, Wang RW, Cheng KC. Substrate specificity, gene structure, and tissue-specific distribution of multiple human 3α-hydroxysteroid dehydrogenases. J. Biol. Chem.1995; 270:20162–8.

30.	Zhang Y, Dufort I, Rheault P, Luu-The V. Characterization of a human 20α-hydroxysteroid dehydrogenase. J Mol Endocrinol 2000; 25:221–8.

31.	Vlahcevic ZR, Heuman DM, Hylemon PB. Physiology and pathophysiology of enterohepatic circulation of bile acids. In: Zakim D, Boyer T, editors. Hepatology: A Textbook of Liver Disease. Philadelphia: Saunders; 1996. page 376–417.

32.	Hofmann AF, Hagey LR. Key discoveries in bile acid chemistry and biology and their clinical applications: History of the last eight decades. J Lipid Res 2014; 55:1553–95.

33.	Hylemon PB, Zhou H, Pandak WM, Ren S, Gil G, Dent P. Bile acids as regulatory molecules. J Lipid Res 2009; 50:1509–20.

34. Russell DW. The enzymes, regulation, and genetics of bile acid synthesis. Annu Rev Biochem 2003; 72:137–74.
35. Pandak WM, Kakiyama G. The acidic pathway of bile acid synthesis: Not just an alternative pathway. Liver Res 2019; 3:88–98.
36. Schwarz M, Björkhem I, Wright AC, Davis DL, Nazer H, Björkhem I, Russell DW. The bile acid synthetic gene 3β-hydroxy-Δ5-C27-steroid oxidoreductase is mutated in progressive intrahepatic cholestasis. J Clin Invest 2000; 106:1175–84.
37. Chiang JYL. Bile acid metabolism and signaling. Compr Physiol 2013; 3:1191–212.
38. Steinberg SJ, Wang SJ, Kim DG, Mihalik SJ, Watkins PA. Human very-long-chain acyl-CoA synthetase: Cloning, topography, and relevance to branched-chain fatty acid metabolism. Biochem Biophys Res Commun 1999; 257:615–21.
39. Steinberg SJ, Wang SJ, McGuinness MC, Watkins PA. Human liver-specific very-long-chain acyl-Coenzyme A synthetase: cDNA cloning and characterization of a second enzymatically active protein. Mol Genet Metab 1999; 68:32–42.
40. Falany CN, Johnson MR, Barnes S, Diasio RB. Glycine and taurine conjugation of bile acids by a single enzyme: Molecular cloning and expression of human liver bile acid CoA:amino acid N- acyltransferase. J Biol Chem 1994; 269:19375–9.
41. Hofmann AF, Hagey LR, Krasowski MD. Bile salts of vertebrates: Structural variation and possible evolutionary significance. J Lipid Res 2010; 51:226–46.
42. Goto T, Holzinger F, Hagey LR, Cerrè C, Ton-Nu HT, Schteingart CD, Steinbach JH, Shneider BL, Hofmann AF. Physicochemical and physiological properties of 5α-cyprinol sulfate, the toxic bile salt of cyprinid fish. J Lipid Res 2003; 44:1643–51.
43. Chiang JYL. Bile acids: Regulation of synthesis. J Lipid Res 2009; 50:1955–66.
44. Craddock AL, Love MW, Daniel RW, Kirby LC, Walters HC, Wong MH, Dawson PA. Expression and transport properties of the human ileal and renal sodium- dependent bile acid transporter. Am J Physiol 1998; 274:157–69.
45. Dawson PA, Hubbert M, Haywood J, Craddock AL, Zerangue N, Christian W V., Ballatori N. The heteromeric organic solute transporter α-β, Ostα-Ostβ, is an ileal basolateral bile acid transporter. J Biol Chem 2005; 280:6960–8.
46. Ballatori N, Christian W V., Lee JY, Dawson PA, Soroka CJ, Boyer JL, Madejczyk MS, Li N. OSTα-OSTβ: A major basolateral bile acid and steroid transporter in human intestinal, renal, and biliary epithelia. Hepatology 2005; 42:1270–9.
47. Kullak-Ublick GA, Stieger B, Meier PJ. Enterohepatic bile salt transporters in normal physiology and liver disease. Gastroenterology 2004; 126:322–42.
48. Rembacz KP, Woudenberg J, Hoekstra M, Jonkers EZ, Van Den Heuvel FAJ, Buist-Homan M, Woudenberg-Vrenken TE, Rohacova J, Marin ML, Miranda MA, et al. Unconjugated bile salts shuttle through hepatocyte peroxisomes for taurine conjugation. Hepatology 2010; 52:2167–76.
49. Makishima M, Okamoto AY, Repa JJ, Tu H, Learned RM, Luk A, Hull M V., Lustig KD, Mangelsdorf DJ, Shan B. Identification of a nuclear receptor for bile acids. Science 1999; 284:1362–5.
50. Parks DJ, Blanchard SG, Bledsoe RK, Chandra G, Consler TG, Kliewer SA, Stimmel JB, Willson TM, Zavacki AM, Moore DD, et al. Bile acids: Natural ligands for an orphan nuclear receptor. Science 1999; 284:1365–8.

51. Goodwin B, Jones SA, Price RR, Watson MA, McKee DD, Moore LB, Galardi C, Wilson JG, Lewis MC, Roth ME, et al. A regulatory cascade of the nuclear receptors FXR, SHP-1, and LRH-1 represses bile acid biosynthesis. Mol Cell 2000; 6:517–26.

52. Zhang M, Chiang JYL. Transcriptional regulation of the human sterol 12α-hydroxylase gene (CYP8B1): Roles of hepatocyte nuclear factor 4α in mediating bile acid repression. J Biol Chem 2001; 276:41690–9.

53. Del Castillo-Olivares A, Campos JA, Pandak WM, Gil G. The role of α1-fetoprotein transcription factor/LRH-1 in bile acid biosynthesis: A known nuclear receptor activator that can act as a suppressor of bile acid biosynthesis. J Biol Chem 2004; 279:16813–21.

54. Holt JA, Luo G, Billin AN, Bisi J, McNeill YY, Kozarsky KF, Donahee M, Wang DY, Mansfield TA, Kliewer SA, et al. Definition of a novel growth factor-dependent signal cascade for the suppression of bile acid biosynthesis. Genes Dev 2003; 17:1581–91.

55. Song KH, Li T, Owsley E, Strom S, Chiang JYL. Bile acids activate fibroblast growth factor 19 signaling in human hepatocytes to inhibit cholesterol 7α-hydroxylase gene expression. Hepatology 2009; 49:297–305.

56. Staudinger JL, Goodwin B, Jones SA, Hawkins-Brown D, MacKenzie KI, LaTour A, Liu Y, Klaassen CD, Brown KK, Reinhard J, et al. The nuclear receptor PXR is a lithocholic acid sensor that protects against liver toxicity. Proc Natl Acad Sci U S A 2001; 98:3369–74.

57. Li T, Chiang JYL. Mechanism of rifampicin and pregnane X receptor inhibition of human cholesterol 7α-hydroxylase gene transcription. Am J Physiol - Gastrointest Liver Physiol 2005; 288:G74–84.

58. Han S, Chiang JYL. Mechanism of vitamin D receptor inhibition of cholesterol 7α-hydroxylase gene transcription in human hepatocytes. Drug Metab Dispos 2009; 37:469–78.

59. Keitel V, Cupisti K, Ullmer C, Knoefel WT, Kubitz R, Häussinger D. The membrane-bound bile acid receptor TGR5 is localized in the epithelium of human gallbladders. Hepatology 2009; 50:861–70.

60. Ward JBJ, Mroz MS, Keely SJ. The bile acid receptor, TGR5, regulates basal and cholinergic-induced secretory responses in rat colon. Neurogastroenterol Motil 2013; 25:708–11.

61. Alemi F, Poole DP, Chiu J, Schoonjans K, Cattaruzza F, Grider JR, Bunnett NW, Corvera CU. The receptor TGR5 mediates the prokinetic actions of intestinal bile acids and is required for normal defecation in mice. Gastroenterology 2013; 144:145–54.

62. Hegyi P, Maléth J, Walters JR, Hofmann AF, Keely SJ. Guts and gall: Bile acids in regulation of intestinal epithelial function in health and disease. Physiol Rev 2018; 98:1983–2023.

63. Yasuda H, Hirata S, Inoue K, Mashima H, Ohnishi H, Yoshiba M. Involvement of membrane-type bile acid receptor M-BAR/TGR5 in bile acid-induced activation of epidermal growth factor receptor and mitogen-activated protein kinases in gastric carcinoma cells. Biochem Biophys Res Commun 2007; 354:154–9.

64. Qiao L, Studer E, Leach K, McKinstry R, Gupta S, Decker R, Kukreja R, Valerie K, Nagarkatti P, El Deiry W, et al. Deoxycholic acid (DCA) causes ligand-independent activation of epidermal growth factor receptor (EGFR) and FAS receptor in primary hepatocytes: Inhibition of EGFR/mitogen-activated protein kinase-signaling module enhances DCA-induced apoptosis. Mol Biol Cell 2001; 12:2629–45.

65. Coleman JP, Hudson LL. Cloning and characterization of a conjugated bile acid hydrolase gene from *Clostridium perfringens*. Appl Environ Microbiol 1995; 61:2514–20.

66. Kishinaka M, Umeda A, Kuroki S. High concentrations of conjugated bile acids inhibit bacterial growth of *Clostridium perfringens* and induce its extracellular cholylglycine hydrolase. Steroids 1994; 59:485–9.

67. Stellwag EJ, Hylemon PB. Purification and characterization of bile salt hydrolase from *Bacteroides fragilis* subsp. *fragilis*. Biochim Biophys Acta 1976; 452:165–76.

68. Yao L, Seaton SC, Ndousse-Fetter S, Adhikari AA, Dibenedetto N, Mina AI, Banks AS, Bry L, Devlin AS. A selective gut bacterial bile salt hydrolase alters host metabolism. Elife 2018; 7:e37182.

69. Elkins CA, Moser SA, Savage DC. Genes encoding bile salt hydrolases and conjugated bile salt transporters in *Lactobacillus johnsonii* 100-100 and other *Lactobacillus* species. Microbiology 2001; 147:3403–12.

70. Tanaka H, Hashiba H, Kok J, Mierau I. Bile salt hydrolase of *Bifidobacterium longum*-Biochemical and genetic characterization. Appl Environ Microbiol 2000; 66:2502–12.

71. Kim GB, Miyamoto CM, Meighen EA, Lee BH. Cloning and characterization of the bile salt hydrolase genes (*bsh*) from *Bifidobacterium bifidum* strains. Appl Environ Microbiol 2004; 70:5603–12.

72. Wijaya A, Hermann A, Abriouel H, Specht I, Yousif NMK, Holzapfel WH, Franz CMAP. Cloning of the bile salt hydrolase (*bsh*) gene from *Enterococcus faecium* FAIR-E 345 and chromosomal location of *bsh* genes in food Enterococci. J Food Prot 2004; 67:2772–8.

73. Jones B V., Begley M, Hill C, Gahan CGM, Marchesi JR. Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. Proc Natl Acad Sci U S A 2008; 105:13580–5.

74. Ridlon JM, Kang D-J, Hylemon PB. Bile salt biotransformations by human intestinal bacteria. J Lipid Res 2006; 47:241–59.

75. Quinn RA, Melnik A V., Vrbanac A, Fu T, Patras KA, Christy MP, Bodai Z, Belda-Ferre P, Tripathi A, Chung LK, et al. Global chemical effects of the microbiome include new bile-acid conjugations. Nature 2020; 579:123–9.

76. Grill JP, Cayuela C, Antoine JM, Schneider F. Isolation and characterization of a *Lactobacillus amylovorus* mutant depleted in conjugated bile salt hydrolase activity: Relation between activity and bile salt resistance. J Appl Microbiol 2000; 89:553–63.

77. Mead GC. The amino acid-fermenting Clostridia. J Gen Microbiol 1971; 67:47–56.

78. Kitahara M, Takamine F, Imamura T, Benno Y. Assignment of *Eubacterium* sp. VPI 12708 and related strains with high bile acid 7α-dehydroxylating activity to *Clostridium scindens* and proposal of *Clostridium hylemonae* sp. nov., isolated from human faeces. Int J Syst Evol Microbiol 2000; 50:971–8.

79. Kitahara M, Takamine F, Imamura T, Benno Y. *Clostridium hiranonis* sp. nov., a human intestinal bacterium with bile acid 7α-dehydroxylating activity. Int J Syst Evol Microbiol 2001; 51:39–44.

80. Chen XJ, Wang ZQ, Zhou ZY, Zeng NY, Huang QF, Wang ZW, Tang WL, Zhou HW. Characterization of *Peptacetobacter hominis* gen. nov., sp. nov., isolated from human faeces, and proposal for the reclassification of *Clostridium hiranonis* within the genus *Peptacetobacter*. Int J Syst Evol Microbiol 2020; 70:2988–97.

81. Franco P, Porru E, Fiori J, Gioiello A, Cerra B, Roda G, Caliceti C, Simoni P, Roda A. Identification and quantification of oxo-bile acids in human faeces with liquid chromatography–mass spectrometry: A potent tool for human gut acidic sterolbiome studies. J Chromatogr A 2019; 1585:70–81.

82. Kakiyama G, Muto A, Takei H, Nittono H, Murai T, Kurosawa T, Hofmann AF, Pandak WM, Bajaj JS. A simple and accurate HPLC method for fecal bile acid profile in healthy and cirrhotic subjects: Validation by GC-MS and LC-MS. J Lipid Res 2014; 55:978–90.

83. Hylemon PB, Melone PD, Franklund C V, Lund E, Bjorkhemt I. Mechanism of intestinal 7α-dehydroxylation of cholic acid: evidence that allo-deoxycholic acid is an inducible side-product. J Lipid Res 1991; 32:89–96.

84. Ridlon JM, Kang DJ, Hylemon PB. Isolation and characterization of a bile acid inducible 7α-dehydroxylating operon in *Clostridium hylemonae* TN271. Anaerobe 2010; 16:137–46.

85. Mallonee DH, Hylemon PB. Sequencing and expression of a gene encoding a bile acid transporter from *Eubacterium* sp. strain VPI 12708. J Bacteriol 1996; 178:7053–8.

86. Mallonee DH, Adams JL, Hylemon PB. The bile acid-inducible *baiB* gene from *Eubacterium* sp. strain VPI 12708 encodes a bile acid-coenzyme A ligase. J Bacteriol 1992; 174:2065–71.

87. Mallonee DH, Lijewski MA, Hylemon PB. Expression in *Escherichia coli* and characterization of a bile acid-inducible 3α-hydroxysteroid dehydrogenase from *Eubacterium* sp. strain VPI 12708. Curr Microbiol 1995; 30:259–63.

88. Coleman JP, White WB, Lijewski M, Hylemon PB. Nucleotide sequence and regulation of a gene involved in bile acid 7-dehydroxylation by *Eubacterium* sp. strain VPI 12708. J Bacteriol 1988; 170:2070–7.

89. Gopal-Srivastava R, Mallonee DH, White WB, Hylemon PB. Multiple copies of a bile acid-inducible gene in *Eubacterium* sp. strain VPI 12708. J Bacteriol 1990; 172:4420–6.

90. White WB, Franklund C V., Coleman JP, Hylemon PB. Evidence for a multigene family involved in bile acid 7-dehydroxylation in *Eubacterium* sp. strain VPI 12708. J Bacteriol 1988; 170:4555–61.

91. Devendran S, Shrestha R, Alves JMP, Wolf PG, Ly L, Hernandez AG, Fields CJ, Daniel SL, Ridlon M. *Clostridium scindens* ATCC 35704: Integration of nutritional requirements, the complete genome sequence, and global transcriptional responses to bile acids. Appl Environ Microbiol 2019; 85:e00052-19.

92. Kang DJ, Ridlon JM, Moore DR, Barnes S, Hylemon PB. *Clostridium scindens baiCD* and *baiH* genes encode stereo-specific 7α/7β-hydroxy-3-oxo-Δ4-cholenoic acid oxidoreductases. Biochim Biophys Acta 2008; 1781:16–25.

93. Ridlon JM, Hylemon PB. Identification and characterization of two bile acid coenzyme A transferases from *Clostridium scindens*, a bile acid 7α-dehydroxylating intestinal bacterium. J Lipid Res 2012; 53:66–76.

94. Dawson JA, Mallonee DH, Bjorkhem I, B P. Expression and characterization of a C24 bile acid 7α-dehydratase from *Eubacterium* sp. strain VPI 12708 in *Escherichia coli*. J Lipid Res 1996; 37:1258–67.

95. Ridlon JM, Bajaj JS. The human gut sterolbiome: Bile acid-microbiome endocrine aspects and therapeutics. Acta Pharm Sin B 2015; 5:99–105.

96. Harris SC, Devendran S, Alves JMP, Mythen SM, Hylemon PB, Ridlon JM. Identification of a gene encoding a flavoprotein involved in bile acid metabolism by the human gut bacterium *Clostridium scindens* ATCC 35704. Biochim Biophys Acta- Mol Cell Biol Lipids 2018; 1863:276–83.

97. Funabashi M, Grove TL, Wang M, Varma Y, McFadden ME, Brown LC, Guo C, Higginbottom S, Almo SC, Fischbach MA. A metabolic pathway for bile acid dehydroxylation by the gut microbiome. Nature 2020; 582:566–70.

98. Bhowmik S, Jones DH, Chiu HP, Park IH, Chiu HJ, Axelrod HL, Farr CL, Tien HJ, Agarwalla S, Lesley SA. Structural and functional characterization of BaiA, an enzyme involved in secondary bile acid synthesis in human gut microbe. Proteins Struct Funct Bioinforma 2014; 82:216–29.

99. Heinken A, Ravcheev DA, Baldini F, Heirendt L, Fleming RMT, Thiele I. Systematic assessment of secondary bile acid metabolism in gut microbes reveals distinct metabolic capabilities in inflammatory bowel disease. Microbiome 2019; 7:1–18.

100. Kurdi P, Kawanishi K, Mizutani K, Yokota A. Mechanism of growth inhibition by free bile acids in Lactobacilli and Bifidobacteria. J Bacteriol 2006; 188:1979–86.

101. Hofmann AF, Sjovall J, Kurz G, Radominska A, Schteingart CD, Tint GS, Vlahcevic ZR, Setchell KDR. A proposed nomenclature for bile acids. J Lipid Res 1992; 33:599–604.

102. Eyssen H, De Pauw G, Stragier J, Verhulst A. Cooperative formation of ω-muricholic acid by intestinal microorganisms. Appl Environ Microbiol 1983; 45:141–7.

103. Macdonald IA, Jellett JF, Mahony DE, Holdeman L V. Bile salt 3α- and 12α-hydroxysteroid dehydrogenases from *Eubacterium lentum* and related organisms. Appl Environ Microbiol 1979; 37:992–1000.

104. Yoshimoto T, Higashi H, Kanatani A, Lin XS, Nagai H, Oyama H, Kurazono K, Tsuru D. Cloning and sequencing of the 7α-hydroxysteroid dehydrogenase gene from *Escherichia coli* HB101 and characterization of the expressed enzyme. J Bacteriol 1991; 173:2173–9.

105. Bennett MJ, McKnight SL, Coleman JP. Cloning and characterization of the NAD-dependent 7α-hydroxysteroid dehydrogenase from *Bacteroides fragilis*. Curr Microbiol 2003; 47:475–84.

106. Wang DQH, Carey MC. Therapeutic uses of animal biles in traditional Chinese medicine: An ethnopharmacological, biophysical chemical and medicinal review. World J Gastroenterol 2014; 20:9952–75.

107. Hagey LR, Crombie DL, Espinosa E, Carey MC, Igimi H, Hofmann AF. Ursodeoxycholic acid in the Ursidae: Biliary bile acids of bears, pandas, and related carnivores. J Lipid Res 1993; 34:1911–7.

108. Goossens JF, Bailly C. Ursodeoxycholic acid and cancer: From chemoprevention to chemotherapy. Pharmacol Ther 2019; 203:107396.

109. Hamilton JP, Xie G, Raufman JP, Hogan S, Griffin TL, Packard CA, Chatfield DA, Hagey LR, Steinbach JH, Hofmann AF. Human cecal bile acids: Concentration and spectrum. Am J Physiol Gastrointest Liver Physiol 2007; 293:G256–63.

110. Coleman JP, Hudson LL, Adams MJ. Characterization and regulation of the NADP-linked 7α-hydroxysteroid dehydrogenase gene from *Clostridium sordellii*. J Bacteriol 1994; 176:4865–74.

111. Macdonald LA, Rochon YP, Hutchison DM, Holdeman L V. Formation of ursodeoxycholic acid from chenodeoxycholic acid by a 7β-hydroxysteroid dehydrogenase-elaborating *Eubacterium aerofaciens* strain cocultured with 7α-hydroxysteroid dehydrogenase-elaborating organisms. Appl Environ Microbiol 1982; 44:1187–95.

112. Edenharder R, Pfützner A, Hammann R. Characterization of NAD-dependent 3α- and 3β-hydroxysteroid dehydrogenase and of NADP-dependent 7β-hydroxysteroid dehydrogenase from *Peptostreptococcus productus*. Biochim Biophys Acta 1989; 1004:230–8.

113. Liu L, Aigner A, Schmid RD. Identification, cloning, heterologous expression, and characterization of a NADPH-dependent 7β-hydroxysteroid dehydrogenase from *Collinsella aerofaciens*. Appl Microbiol Biotechnol 2011; 90:127–35.

114. Baron SF, Franklund C V., Hylemon PB. Cloning, sequencing, and expression of the gene coding for bile acid 7α-hydroxysteroid dehydrogenase from *Eubacterium* sp. strain VPI 12708. J Bacteriol 1991; 173:4558–69.

115. Sutherland JD, Williams CN. Bile acid induction of 7α- and 7β-hydroxysteroid dehydrogenases in *Clostridium limosum*. J Lipid Res 1985; 26:344–50.

116. Macdonald IA, Meier EC, Mahony DE, Costain GA. 3α-, 7α- And 12α-hydroxysteroid dehydrogenase activities from *Clostridium perfringens*. Biochim Biophys Acta 1976; 450:142–53.

117. Bernstein C, Holubec H, Bhattacharyya AK, Nguyen H, Payne CM, Zaitlin B, Bernstein H. Carcinogenicity of deoxycholate, a secondary bile acid. Arch Toxicol 2011; 85:863–71.

118. Cao H, Xu M, Dong W, Deng B, Wang S, Zhang Y, Wang S, Luo S, Wang W, Qi Y, et al. Secondary bile acid-induced dysbiosis promotes intestinal carcinogenesis. Int J Cancer 2017; 140:2545–56.

119. Yoshimoto S, Loo TM, Atarashi K, Kanda H, Sato S, Oyadomari S, Iwakura Y, Oshima K, Morita H, Hattori M, et al. Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome. Nature 2013; 499:97–101.

120. Ma C, Han M, Heinrich B, Fu Q, Zhang Q, Sandhu M, Agdashian D, Terabe M, Berzofsky JA, Fako V, et al. Gut microbiome–mediated bile acid metabolism regulates liver cancer via NKT cells. Science 2018; 360:eaan5931.

121. Magouliotis DE, Tasiopoulou VS, Svokos AA, Svokos KA, Chatedaki C, Sioka E, Zacharoulis D. Ursodeoxycholic acid in the prevention of gallstone formation after bariatric surgery: an updated systematic review and meta-analysis. Obes Surg 2017; 27:3021–30.

122. Kim EK, Cho JH, Kim EJ, Kim YJ. Ursodeoxycholic acid inhibits the proliferation of colon cancer cells by regulating oxidative stress and cancer stem-like cell growth. PLoS One 2017; 12:e0181183.

123. Lazaridis KN, Gores GJ, Lindor KD. Ursodeoxycholic acid "mechanisms of action and clinical use in hepatobiliary disorders." J Hepatol 2001; 35:134–46.

124. White BA, Fricke RJ, Hylemon PB. 7β-Dehydroxylation of ursodeoxycholic acid by whole cells and cell extracts of the intestinal anaerobic bacterium, *Eubacterium* species V.P.I 12708. J Lipid Res 1982; 23:145–53.

125. Macdonald IA, Hutchison DM. Epimerization versus dehydroxylation of the 7α-hydroxyl-group of primary bile acids: Competitive studies with *Clostridium absonum* and 7α-dehydroxylating bacteria (*Eubacterium* sp.). J Steroid Biochem 1982; 17:295–303.

126. Kotb MA. Molecular mechanisms of ursodeoxycholic acid toxicity & side effects: Ursodeoxycholic acid freezes regeneration & induces hibernation mode. Int J Mol Sci 2012; 13:8882–914.

127. Edenharder R, Pfützner M, Hammann R. NADP-dependent 3β-, 7α- and 7β-hydroxysteroid dehydrogenase activities from a lecithinase-lipase-negative *Clostridium* species 25.11.c. Biochim Biophys Acta 1989; 1002:37–44.

128. Marschall HU, Oppermann UCT, Svensson S, Nordling E, Persson B, Höög JO, Jörnvall H. Human liver class I alcohol dehydrogenase isozyme: The sole cytosolic 3β-hydroxysteroid dehydrogenase of iso bile acids. Hepatology 2000; 31:990–6.

129. Macdonald IA, Jellett JF, Mahony DE. 12alpha-Hydroxysteroid dehydrogenase from *Clostridium* group P strain C48-50 ATCC #29733: partial purification and characterization. J Lipid Res 1979; 20:234–9.

130. Harris JN, Hylemon PB. Partial purification and characterization of NADP-dependent 12α-hydroxysteroid dehydrogenase from *Clostridium leptum*. Biochim Biophys Acta 1978; 528:148–57.

131. Edenharder R, Schneider J. 12β-Dehydrogenation of bile acids by *Clostridium paraputrificum*, *C. tertium*, and *C. difficile* and epimerization at carbon-12 of deoxycholic acid by cocultivation with 12α-dehydrogenating *Eubacterium lentum*. Appl Environ Microbiol 1985; 49:964–8.

132. Edenharder R, Pfützner A. Characterization of NADP-dependent 12β-hydroxysteroid dehydrogenase from *Clostridium paraputrificum*. Biochim Biophys Acta 1988; 962:362–70.

133. Fischer S, Ben F, Paumgartner G. Unchanged levels of keto bile acids in bile after cholecystectomy. Digestion 1991; 48:202–9.

134. Eneroth P, Hellstrom K, Sjovall J. A method for quantitative determination of bile acid in human feces. Acta Chem. Scand.1968; 22:1720–44.

135. Reddy BS, Maeura Y. Tumor promotion by dietary fat in azoxymethane-induced colon carcinogenesis in female F344 rats: Influence of amount and source of dietary fat. J Natl Cancer Inst 1984; 72:745–50.

136. Eneroth P, Gordon B, Ryhage R, Sjövall J. Identification of mono- and dihydroxy bile acids in human feces by gas-liquid chromatography and mass spectrometry. J Lipid Res 1966; 7:511–23.

137. Horinouchi M, Hayashi T, Koshino H, Malon M, Yamamoto T, Kudo T. Identification of genes involved in inversion of stereochemistry of a C-12 hydroxyl group in the catabolism of cholic acid by *Comamonas testosteroni* TA441. J Bacteriol 2008; 190:5545–54.

138. Holert J, Kulić Ž, Yücel O, Suvekbala V, Suter MJF, Möller HM, Philipp B. Degradation of the acyl side chain of the steroid compound cholate in *Pseudomonas* sp. strain Chol1 proceeds via an aldehyde intermediate. J Bacteriol 2013; 195:585–95.

139. Song C, Wang B, Tan J, Zhu L, Lou D. Discovery of tauroursodeoxycholic acid biotransformation enzymes from the gut microbiome of black bears using metagenomics. Sci Rep 2017; 7:1–8.

140. Haselwood GAD, Wootton VM. Comparative studies of "bile salts". Pythocholic acid. Biochem J 1951; 49:67–71.

141. Pellicciari R, Gioiello A, Sabbatini P, Venturoni F, Nuti R, Colliva C, Rizzo G, Adorini L, Pruzanski M, Roda A, et al. Avicholic acid: A lead compound from birds on the route to potent TGR5 modulators. ACS Med Chem Lett 2012; 3:273–7.

142. Meigs RA, Ryan KJ. 16α-hydroxysteroid dehydrogenase of rat kidney. J Biol Chem 1966; 241:4011–5.

143. Harris SC, Devendran S, Méndez- García C, Mythen SM, Wright CL, Fields CJ, Hernandez AG, Cann I, Hylemon PB, Ridlon JM. Bile acid oxidation by *Eggerthella lenta* strains C592 and DSM 2243. Gut Microbes 2018; 9:523–39.

144. Marion S, Studer N, Desharnais L, Menin L, Escrig S, Meibom A, Hapfelmeier S, Bernier-Latmani R. In vitro and in vivo characterization of *Clostridium scindens* bile acid transformations. Gut Microbes [Internet] 2019; 10:481–503. Available from: https://doi.org/10.1080/19490976.2018.1549420

145. White BA, Lipsky RL, Fricke RJ, Hylemon PB. Bile acid induction specificity of 7α-dehydroxylase activity in an intestinal *Eubacterium* species. Steroids 1980; 35.

146. Fukiya S, Arata M, Kawashima H, Yoshida D, Kaneko M, Minamida K, Watanabe J, Ogura Y, Uchida K, Itoh K, et al. Conversion of cholic acid and chenodeoxycholic acid into their 7-oxo derivatives by *Bacteroides intestinalis* AM-1 isolated from human feces. FEMS Microbiol Lett 2009; 293:263–70.

147. Tanaka N, Nonaka T, Tanabe T, Yoshimoto T, Tsuru D, Mitsui Y. Crystal structures of the binary and ternary complexes of 7α-hydroxysteroid dehydrogenase from *Escherichia coli*. Biochemistry 1996; 35:7715–30.

148. Masuda N, Oda H. 7α-Dehydroxylation of bile acids by resting cells of an unidentified, Gram-positive, nonsporeforming anaerobic bacterium. Appl Environ Microbiol 1983; 45:456–62.

149. Kisiela M, Skarka A, Ebert B, Maser E. Hydroxysteroid dehydrogenases (HSDs) in bacteria - A bioinformatic perspective. J Steroid Biochem Mol Biol [Internet] 2012; 129:31–46. Available from: http://dx.doi.org/10.1016/j.jsbmb.2011.08.002

150. Hylemon PB, Harris SC, Ridlon JM. Metabolism of hydrogen gases and bile acids in the gut microbiome. FEBS Lett 2018; 592:2070–82.

151. Ragsdale SW, Pierce E. Acetogenesis and the Wood–Ljungdahl pathway of CO_2 fixation. Biochim Biophys Acta 2008; 1784:1873–98.

152. Hang S, Paik D, Yao L, Kim E, Jamma T, Lu J, Ha S, Nelson BN, Kelly SP, Wu L, et al. Bile acid metabolites control T_H17 and T_{reg} cell differentiation. Nature 2019; 576:143–8.

153. Cain DW, Cidlowski JA. Immune regulation by glucocorticoids. Nat Rev Immunol 2017; 17:233–47.

154. Morris DJ. Why do humans have two glucocorticoids: A question of intestinal fortitude. Steroids 2015; 102:32–8.

155. Norman AW, Henry HL. Adrenal Corticoids. In: Norman AW, Henry HL, editors. Hormones. London: Academic Press; 2015. page 223–38.

156. Miller WL. Steroidogenesis: Unanswered questions. Trends Endocrinol Metab 2017; 28:771–93.

157. Wang, J.-C., & Harris C. Glucocorticoid signaling: From molecules to mice to man. London: Springer; 2015.

158. Arriza JL, Weinberger C, Cerelli G, Glaser TM, Handelin BL, Housman DE, Evans RM. Cloning of human mineralocorticoid receptor complementary DNA: Structural and functional kinship with the glucocorticoid receptor. Science 1987; 237:268–75.

159. Peterson RE, Wyngaarden JB, Guerra SL, Brodie BB, Bunim JJ. The physiological disposition and metabolic fate of hydrocortisone in man. J Clin Investig Investig 1955; 34:1779–94.

160. Gold BNI, Smith LL, Moore FD. Cortisol metabolism in man: observations of pathways, pool sizes of metabolites and rates of formation of metabolites. J Clin Invest 1959; 38:2238–52.
161. Eisenschmid B, Heilmann P, Oelkers W, Rejaibi R, Schöneshöfer M. 20-Dihydroisomers of cortisol and cortisone in human urine: excretion rates under different physiological conditions. J Clin Chem Clin Biochem 1987; 25:345–50.
162. Morgan RA, Beck KR, Nixon M, Homer NZM, Crawford AA, Melchers D, Houtman R, Meijer OC, Stomby A, Anderson AJ, et al. Carbonyl reductase 1 catalyzes 20β-reduction of glucocorticoids, modulating receptor activation and metabolic complications of obesity. Sci Rep 2017; 7:1–11.
163. Shackleton CHL, Roitman E, Monder C, Bradlow HL. Gas chromatographic and mass spectrometric analysis of urinary acidic metabolites of cortisol. Steroids 1980; 36:289–98.
164. Schiffer L, Arlt W, Storbeck KH. Intracrine androgen biosynthesis, metabolism and action revisited. Mol Cell Endocrinol 2018; 465:4–26.
165. Rege J, Nakamura Y, Satoh F, Morimoto R, Kennedy MR, Layman LC, Honma S, Sasano H, Rainey WE. Liquid chromatography-tandem mass spectrometry analysis of human adrenal vein 19-carbon steroids before and after ACTH stimulation. J Clin Endocrinol Metab 2013; 98:1182–8.
166. Turcu A, Smith JM, Auchus R, Rainey WE. Adrenal androgens and androgen precursors-Definition, synthesis, regulation and physiologic actions. Compr Physiol 2014; 4:1369–81.
167. Pretorius E, Arlt W, Storbeck K-H. A new dawn for androgens: Novel lessons from 11-oxygenated C19 steroids. Mol Cell Endocrinol 2017; 441:76–85.
168. Storbeck K, Bloem LM, Africander D, Schloms L, Swart P, Swart AC. 11β-Hydroxydihydrotestosterone and 11-ketodihydrotestosterone, novel C19 steroids with androgenic activity: A putative role in castration resistant prostate cancer? Mol Cell Endocrinol 2013; 377:135–46.
169. Pretorius E, Africander DJ, Vlok M, Perkins MS, Quanson J, Storbeck K-H. 11-Ketodihydrotestosterone in castration resistant prostate cancer: Potent androgens which can no longer be ignored. PLoS One 2016; 11:e0159867.
170. Turcu AF, Auchus RJ. Clinical significance of 11-oxygenated androgens. Curr Opin Endocrinol Diabetes Obes 2017; 24:252–9.
171. O'Reilly MW, Kempegowda P, Jenkinson C, Taylor AE, Quanson JL, Storbeck KH, Arlt W. 11-oxygenated C19 steroids are the predominant androgens in polycystic ovary syndrome. J Clin Endocrinol Metab 2017; 102:840–8.
172. Swart AC, Storbeck KH. 11β-hydroxyandrostenedione: Downstream metabolism by 11βHSD, 17βHSD and SRD5A produces novel substrates in familiar pathways. Mol Cell Endocrinol 2015; 408:114–23.
173. Shackleton CHL, Neres MS, Hughes BA, Stewart PM, Kater CE. 17-Hydroxylase/C17,20-lyase (CYP17) is not the enzyme responsible for side-chain cleavage of cortisol and its metabolites. Steroids 2008; 73:652–6.
174. Winter J, Morris GN, O'Rourke-Locascio S, Bokkenheuser VD, Mosbach EH, Cohen BI, Hylemon PB. Mode of action of steroid desmolase and reductases synthesized by *Clostridium "scindens"* (formerly *Clostridium* strain 19). J Lipid Res 1984; 25:1124–31.

175. Bokkenheuser VD, Morris GN, Ritchie AE, Holdeman L V, Winter J. Biosynthesis of androgen from cortisol by a species of *Clostridium* recovered from human fecal flora. J Infec Dis 1984; 149:489–94.

176. Cerone-McLernon AM, Winter J, Mosbach EH, Bokkenheuser VD. Side-chain cleavage of cortisol by fecal flora. Biochim Biophys Acta 1981; 666:341–7.

177. Lai JJ, Lai KP, Zeng W, Chuang KH, Altuwaijri S, Chang C. Androgen receptor influences on body defense system via modulation of innate and adaptive immune systems. Am J Pathol 2012; 181:1504–12.

178. Gu S, Papadopoulou N, Nasir O, Föller M, Alevizopoulos K, Lang F, Stournaras C. Activation of membrane androgen receptors in colon cancer inhibits the prosurvival signals Akt/Bad in vitro and in vivo and blocks migration via vinculin/actin signaling. Mol Med 2011; 17:48–58.

179. D'Errico I, Moschetta A. Nuclear receptors, intestinal architecture and colon cancer: an intriguing link. Cell Mol Life Sci 2008; 65:1523–43.

180. Catalano MG, Pfeffer U, Raineri M, Ferro P, Curto A, Capuzzi P, Corno F, Berta L, Fortunati N. Altered expression of androgen-receptor isoforms in human colon-cancer tissues. Int J Cancer 2000; 89:325–30.

181. Nabarro JDN, Moxham A, Walker G, Slater JDH. Rectal Hydrocortisone. Br Med J 1957; 2:272–4.

182. Wade AP, Slater JD, Kellie AE, Holliday ME. Urinary excretion of 17-ketosteroids following rectal infusion of cortisol. J Clin Endocrinol Metab 1959; 19:444–53.

183. Bokkenheuser VD, Winter J, Morris GN, Locascio S. Steroid desmolase synthesis by *Eubacterium desmolans* and *Clostridium cadavaris*. Appl Environ Microbiol 1986; 52:1153–6.

184. Devendran S, Méndez-García C, Ridlon JM. Identification and characterization of a 20β-HSDH from the anaerobic gut bacterium *Butyricicoccus desmolans* ATCC 43058. J Lipid Res 2017; 58:916–25.

185. Ly LK, Rowles JL, Paul HM, Alves JMP, Yemm C, Wolf PM, Devendran S, Hudson ME, Morris DJ, Erdman JW, et al. Bacterial steroid-17,20-desmolase is a taxonomically rare enzymatic pathway that converts prednisone to 1,4-androstanediene-3,11,17-trione, a metabolite that causes proliferation of prostate cancer cells. J Steroid Biochem Mol Biol 2020; 199:105567.

186. Devendran S, Mythen SM, Ridlon JM. The desA and desB genes from *Clostridium scindens* ATCC 35704 encode steroid-17,20-desmolase. J Lipid Res 2018; 59:1005–14.

187. Schoneshofer M, Weber B, Nigam S. Increased urinary excretion of free 20α- and 20β-dihydrocortisol in a hypercortisolemic but hypocortisoluric patient with Cushing's disease. Clin Chem 1983; 29:385–9.

188. Winter J, Cerone-McLernon A, O'Rourke S, Ponticorvo L, Bokkenheuser VD. Formation of 20β-dihydrosteroids by anaerobic bacteria. J Steroid Biochem 1982; 17:661–7.

189. Bokkenheuser VD, Winter J, Dehazya P, Kelly WG. Isolation and characterization of human fecal bacteria capable of 21-dehydroxylating corticoids. Appl Environ Microbiol 1977; 34:571–5.

190. Winter J, Bokkenheuser VD. 21-Dehydroxylation of corticoids by anaerobic bacteria isolated from human fecal flora. J Steroid Biochem 1978; 9:379–84.

191. Feighner SD, Hylemon PB. Characterization of a corticosteroid 21-dehydroxylase from the intestinal anaerobic bacterium, Eubacterium lentum. J Lipid Res 1980; 21:585–93.

192. Feighner SD, Bokkenheuser VD, Winter J, Hylemon PB. Characterization of a C21 neutral steroid hormone transforming enzyme, 21-dehydroxylase, in crude cell extracts of *Eubacterium lentum*. Biochim Biophys Acta 1979; 574:154–63.

193. Barnard L, Gent R, van Rooyen D, Swart AC. Adrenal C11-oxy C21 steroids contribute to the C11-oxy C19 steroid pool via the backdoor pathway in the biosynthesis and metabolism of 21-deoxycortisol and 21-deoxycortisone. J Steroid Biochem Mol Biol 2017; 174:86–95.

194. Latif SA, Pardo HA, Hardy MP, Morris DJ. Endogenous selective inhibitors of 11β-hydroxysteroid dehydrogenase isoforms 1 and 2 of adrenal origin. Mol Cell Endocrinol 2005; 243:43–50.

195. Kornel L, Miyabo S, Saito Z, Cha R-W, Wu F-T. Corticosteroids in human blood. VIII. Cortisol metabolites in plasma of normotensive subjects and patients with essential hypertension. J Clin Endocrinol Metab 1975; 40:949–58.

196. Szymanski ES, Furfine CS. 20β-Hydroxysteroid oxidoreductase: Kinetics and binding of corticosteroids and corticosteroid-21-aldehydes. J Biol Chem 1977; 252:205–11.

197. Javdan B, Lopez JG, Chankhamjon P, Lee YCJ, Hull R, Wu Q, Wang X, Chatterjee S, Donia MS. Personalized mapping of drug metabolism by the human gut microbiome. Cell 2020; 181:1661–79.

198. Krafft AE, Hylemon PB. Purification and characterization of a novel form of 20α-hydroxysteroid dehydrogenase from *Clostridium scindens*. J Bacteriol 1989; 171:2925–32.

199. Ly LK, Doden HL, Ridlon JM. Gut feelings about bacterial steroid-17,20-desmolase. Mol Cell Endocrinol 2021; 525:111174.

200. Koppel N, Balskus EP. Exploring and understanding the biochemical diversity of the human microbiota. Cell Chem Biol 2016; 23:18–30.

201. Fischbach MA. Microbiome: Focus on causation and mechanism. Cell 2018; 174:785–90.

CHAPTER 2

METABOLISM OF OXO-BILE ACIDS AND CHARACTERIZATION OF RECOMBINANT 12α-HYDROXYSTEROID DEHYDROGENASES FROM BILE ACID 7α-DEHYDROXYLATING HUMAN GUT BACTERIA[1]

ABSTRACT

Bile acids are important cholesterol-derived nutrient signaling hormones, synthesized in the liver, that act as detergents to solubilize dietary lipids. Bile acid 7α-dehydroxylating gut bacteria generate the toxic bile acids deoxycholic acid and lithocholic acid from host bile acids. The ability of these bacteria to remove the 7-hydroxyl group is partially dependent on 7α-hydroxysteroid dehydrogenase (HSDH) activity, which reduces 7-oxo-bile acids generated by other gut bacteria. 3α-HSDH has an important enzymatic activity in the bile acid 7α-dehydroxylation pathway. 12α-HSDH activity has been reported for the low-activity bile acid 7α-dehydroxylating bacterium *Clostridium leptum*; however, this activity has not been reported for high-activity bile acid 7α-dehydroxylating bacteria, such as *Clostridium scindens*, *Clostridium hylemonae*, and *Peptacetobacter hiranonis*. Here, we demonstrate that these strains express bile acid 12α-HSDH. The recombinant enzymes were characterized from each species

---

and shown to preferentially reduce 12-oxolithocholic acid to deoxycholic acid, with low activity against 12-oxochenodeoxycholic acid and reduced activity when bile acids were conjugated to taurine or glycine. Phylogenetic analysis suggests that 12α-HSDH is widespread among Firmicutes, Actinobacteria in the Coriobacteriaceae family, and human gut Archaea.

IMPORTANCE

12α-HSDH activity has been established in the medically important bile acid 7α-dehydroxylating bacteria *C. scindens*, *P. hiranonis*, and *C. hylemonae*. Experiments with recombinant 12α-HSDHs from these strains are consistent with culture-based experiments that show a robust preference for 12-oxolithocholic acid over 12-oxochenodeoxycholic acid. Phylogenetic analysis identified novel members of the gut microbiome encoding 12α-HSDH. Future reengineering of 12α-HSDH enzymes to preferentially oxidize cholic acid may provide a means to industrially produce the therapeutic bile acid ursodeoxycholic acid. In addition, a cholic acid-specific 12α-HSDH expressed in the gut may be useful for the reduction in deoxycholic acid concentration, a bile acid implicated in cancers of the gastrointestinal (GI) tract.

INTRODUCTION

Bile acids are synthesized from cholesterol in the liver, conjugated to the amino acid taurine or glycine, and secreted and stored in the gallbladder during the interdigestive period. During a meal, the gallbladder is hormonally stimulated to contract, releasing bile into the duodenum. In the small bowel, conjugated bile acids measure in the low-millimolar range and function to solubilize cholesterol, dietary lipids, and lipid-soluble vitamins.[1] Conjugated bile acids are also directly antimicrobial, reducing the microbial load at the site of host nutrient

absorption. In the terminal ileum, ~95% of bile acids are returned to the liver after high-affinity transport by apical bile salt transporters into enterocytes and entrance into portal circulation after basolateral transport.[2] However, roughly 400 to 800 mg per day of bile salts enters the large intestine, where anaerobic bacteria perform several important biotransformations.

The "gateway reaction" is hydrolysis of the amino acid conjugate by bacterial bile salt hydrolases (BSH) which are found in all major bacterial phyla and in the two main species of human gut archaea.[3] Gut microbes are also capable of removing the 7α-hydroxyl group from host cholic acid (CA; 5β-cholanic acid-3α,7α,12α-triol), forming deoxycholic acid (DCA; 5β-cholanic acid-3α,12α-diol), and from chenodeoxycholic acid (CDCA; 5β-cholanic acid-3α,7α-diol), forming lithocholic acid (LCA; 5β-cholanic acid-3α-ol).[4,5] DCA and LCA are the main components of the bile acid profile of stool in healthy humans[6] and have been associated causally with diseases of the gastrointestinal (GI) tract, including cancers of the colon[7], liver[8], and esophagus[9], as well as cholesterol gallstone disease in a subset of patients[10]. 7α-Hydroxy bile acids, such as ursodeoxycholic acid (UDCA; 5β-cholanic acid-3α,7α-diol), are important in the therapy of the aforementioned GI diseases and function mainly to dilute and counteract the influence of microbe-derived DCA and LCA.[11–13] UDCA synthesis from CA requires enzymatic oxidation of the 12α-hydroxyl group, followed by Wolff-Kishner reduction to remove the C-12 carbonyl group.[14] Therefore, understanding the role of bacterial 12α-dehydrogenation in microbial bile acid biotransformation and identifying genes encoding 12α-hydroxysteroid dehydrogenases (HSDH) in gut microbes are important in understanding host disease processes and potentially in aiding in the biotechnological generation of therapeutic agents in disease treatment.

Strains of bile acid 7α-dehydroxylating bacteria (BA7) that have been isolated to date from human feces, or inferred by sequence, are within *Clostridium* clusters XIVa, IV, and XI.[5] Those strains in *Clostridium* cluster XIVa that express high BA7 activity and encode a polycistronic bile acid-inducible (*bai*) operon include strains of *C. scindens*, *Peptacetobacter hiranonis* (formerly *Clostridium hiranonis*), and *C. hylemonae*.[5] HSDH enzymes play an important role in secondary bile acid formation by these gut bacterial species. Bile acids are pumped inside the cell by a proton-dependent transport protein (BaiG)[15], ligated to coenzyme A (CoA) by the ATP-dependent CoA-ligase (BaiB)[16] or ATP-independent CoA-transferases (BaiF and BaiK)[17,18]. Diverse gut microbiota express 7α-HSDHs, which convert host primary bile acids to 7-oxo-bile acids.[4] An NADP-dependent 7α-HSDH gene has been identified and characterized in *C. scindens* VPI 12708[19] and is inferred by sequence similarity in *C. scindens* ATCC 35704, *C. hylemonae* DSM 15053, and *P. hiranonis* DSM 13275[5]. Additionally, after CoA ligation, two oxidation steps precede the rate-limiting 7α-dehydration step (BaiE)[20,21], one of which, the *baiA* gene, encodes a 3α-HSDH[22,23]. The other, encoded by the *baiCD* and *baiH* genes, introduces a C-4=C-5 bond in 7α-hydroxyl and 7β-hydroxyl bile acid substrates, respectively.[24] After 7α/β-dehydration, a series of reductions (catalyzed by the *baiN* gene product) occur to convert 5β-chol-4,6-dienoic acid-12α-ol-3-one to DCA or LCA.[24,25]

Roughly half of the host primary bile acids synthesized in the liver contain a 12α-hydroxyl group.[1] Members of the gastrointestinal microbiota have evolved 12α-HSDHs capable of oxidizing and epimerizing the 12α-hydroxyl group from host CA and its microbial metabolites, including DCA.[26] *C. leptum*, reported to express low BA7 activity[27], has also been reported to express a 12α-HSDH[28]; however, the gene(s) encoding 12α-HSDH have not been identified. A previous bioinformatics analysis of bacterial HSDH enzymes identified a gene

encoding a putative 12α-HSDH (accession no. WP_006441568.1) in *C. hylemonae*[29] based on a single deduced amino acid sequence from the only known gene encoding 12α-HSDH identified from *Clostridium* sp. strain ATCC 29733.[30,31] This phylogeny also predicted that BA7 strains *C. scindens* ATCC 35704, *C. hylemonae* DSM 15053, and *P. hiranonis* DSM 13275 harbor 12α-HSDHs; however, the genes predicted to encode this activity have yet to be reported. A major clue to the expression of 12α-HSDH in *P. hiranonis* (formerly strain HD-17) came from a study where resting cells of *P. hiranonis* produced significant levels of 12-oxoLCA when incubated aerobically with CA.[32] We therefore sought to evaluate the potential of *C. scindens* ATCC 35704, *C. hylemonae* DSM 15053, and *P. hiranonis* DSM 13275 to metabolize oxo-bile acids, and specifically to detect 12α-HSDH activity. Further, we sought to characterize purified recombinant 12α-HSDHs from these medically important gut microbes.

RESULTS

Transformation of 12-oxoLCA and other oxo-bile acids by C. scindens ATCC 35704, C. hylemonae DSM 15053, and P. hiranonis DSM 13275 in BHI cultures

Studies to date have yet to examine the metabolism of 12-oxo-bile acid substrates in anaerobic cultures by BA7 strains. We therefore limited our screening of substrates to oxo-bile acids possessing a 12α-hydroxyl or 12-oxo group (i.e., CA derivatives) (**Table 2.1**). 12-oxoLCA was converted quantitatively to DCA in 24-h cultures of each strain (**Table 2.1** and **Figure 2.1**). In contrast, all three converted 12-oxoCDCA to 7-oxoDCA rather than to DCA. We did not detect the formation of 12-oxoLCA when CA or its C-3– or C-7–oxo derivative was the substrate, nor did we detect oxidation of the 12α-hydroxyl group when DCA was the substrate.

All three strains reduced 3-oxoCA to CA, but only P. hiranonis DSM 13275 generated DCA

*(**Table 2.1**)*

7-oxoDCA was converted to DCA by *C. scindens* ATCC 35704 and *P. hiranonis* DSM

13275 but not by *C. hylemonae* DSM 15053. DCA and 3-oxoDCA were not metabolized by

these strains. Dioxo- and trioxo-bile acids were also evaluated for metabolism by BA7 bacteria.

When 3,12-dioxoLCA was the substrate, all three strains reduced the 12-oxo-group and formed

3-oxoDCA. Substrates 7,12-dioxoLCA and 3,7,12- trioxoLCA were converted to a mixture of

CA, DCA, and 3-oxoDCA. Collectively, these results indicated that *C. scindens* ATCC 35704,

C. hylemonae DSM 15053, and *P. hiranonis* DSM 13275 reduce oxo-bile acids, and each strain

possesses 12α-HSDH activity.

Cloning, overexpression, purification, and characterization of 12α-HSDH from C. scindens

ATCC 35704, C. hylemonae DSM 15053, and P. hiranonis DSM 13275

Recently, genes encoding 12α-HSDH were reported in *Eggerthella lenta* DSM 2243

(Elen_2515)[33] and *Eggerthella* CAG:298 (accession no. CDD59475)[34]. We aligned the amino

acid sequences of the 12α-HSDHs from *Eggerthella* strains as well as 12α-HSDH reported in

Clostridium sp. ATCC 29733 (accession no. ERJ00208.1), with putative 12α-HSDH from *C.

hylemonae* DSM 15053 (CLOHYLEM_04236), *C. scindens* ATCC 35704 (CLOSCI_02455),

and *P. hiranonis* DSM 13275 (CLOHIR_01081). Multiplesequence alignment is depicted in

Figure 2.2. Each protein in the alignment is a member of the 3-ketoacyl-(acyl-carrier-protein)

short-chain dehydrogenase/reductase (SDR) family, defined by a conserved N-terminal

Rossmann-fold responsible for NAD(P)-binding (GGGX5GXG) and catalytic triad (SYK)

(**Figure 2.2**). The genomic context of the genes encoding putative 12α-HSDH in each 7α-dehydroxylating bacterium is shown in **Figure 2.3**.

To determine whether the genes CLOHYLEM_04236 (259 amino acids), CLOSCI_02455 (266 amino acids), and CLOHIR_01081 (266 amino acids) encode novel 12α-HSDHs, we amplified the genes by PCR and cloned each gene into pET51(b)+ for overexpression in *Escherichia coli*. Each recombinant protein was overexpressed as an N-terminal streptavidin fusion protein, which was purified by affinity chromatography using Strep-Tactin resin and resolved by SDS-PAGE analysis (**Figure 2.3**). Recombinant CLOHYLEM_04236 (rCHYL) had a theoretical subunit mass of 28.0 kDa and yielded an observed subunit mass of 28.5 ± 0.3 kDa on SDS-PAGE (three independent protein gels). rCLOSCI_02455 (rCSCI) from *C. scindens* ATCC 35704 had a deduced subunit molecular mass of 28.2 kDa, with an observed subunit mass of 27.3 ± 0.9 kDa, and rCLOHIR_01081 (rCHIR) (theoretical subunit mass, 28.5 kDa) yielded an observed subunit mass of 28.5 ± 0.1 kDa.

TLC and mass spectrometry of enzyme reaction products

Next, we confirmed functional 12α-HSDH activity of purified rCSCI, rCHYL, and rCHIR in the presence of 12-oxoLCA substrate and reduced or oxidized pyridine nucleotide cofactors (**Figure 2.4**). DCA standard yielded a major mass ion of 391.2858 *m/z* in negative-ion mode, consistent with the molecular mass of DCA at 392.57 atomic mass units (amu). rCSCI was NADPH dependent, according to thin-layer chromatography (TLC) separation of bile acid metabolites. The single reaction product was eluted from TLC silica and yielded a major mass ion of 391.2856 *m/z*. Similarly, rCHYL showed specificity for NADPH, with no significant formation of product with NADH after 24 h, as determined by TLC. A product, whose migration

was identical to DCA, yielded a major mass ion in negative mode of 391.2856 *m/z*. When

incubated with 12-oxoLCA in the presence of NADPH (but not NADH), purified rCHIR yielded

a single product on TLC that comigrated with DCA, with a major mass ion of 391.2858 *m/z*.

pH optimization of r12α-HSDHs

In order to optimize the enzyme-catalyzed C-12 oxidoreduction, we monitored the

conversion of pyridine nucleotides at 340 nm in various buffer systems adjusted to pH

increments from pH 4.5 to 10.0 (**Figure 2.5**). The optimum pH for rCSCI in the oxidative

direction with DCA as the substrate and $NADP^+$ as the cosubstrate was 7.5, and the optimum in

the reductive direction with 12-oxoLCA as the substrate and NADPH as the cosubstrate was 7.0.

In the oxidative direction, the optimum pH for rCHYL with $NADP^+$ as cofactor and DCA as the

substrate was 8.0, while the optimum pH with NADPH and 12-oxoLCA was 7.0. The optimum

pH for rCHIR in the oxidative direction was found to be 7.0 and was between 6.5 and 7.0 in the

reductive direction.

Kinetic analysis and substrate specificity of r12α-HSDHs

Kinetic analysis was performed at the optimum pH for each enzyme and the particular

reaction direction. In the oxidative direction, rCHYL, rCSCI, and rCHIR displayed comparable

Km values for DCA at ~150 µM, a concentration approximating levels of DCA in fecal water

obtained from Western individuals (**Table 2.2**; see also **Supplementary Figure 2.1** in the

supplemental material).[35] The *Km* values for $NADP^+$ were on the same order as those for DCA.

The K_m value for 12-oxoLCA in the reductive direction was an order of magnitude lower for

rCHYL and ~4.5- and 5-fold lower for rCSCI and rCHIR, respectively (**Table 2.2**). V_{max} and k_{cat}

values were between 1.6- and 2.9-fold higher in the oxidative than the reductive direction. However, catalytic efficiencies (K_m/k_{cat}) were 2.8- to 3.5-fold greater in the reductive than oxidative direction.

In the reductive direction, recombinant NADPH-dependent 12α-HSDH had between 9.55 and 3.65% relative activity toward 12-oxoCDCA compared to 12-oxoLCA, suggesting hindrance by the 7α-hydroxyl group (**Table 2.3**). NADH was not a cosubstrate, as determined by spectrophotometric assay and overnight reaction mixtures separated by TLC (**Figure 2.3**). In the oxidative direction, the relative activity toward CA was between 54.44 and 82.32% that of the activity toward DCA. rCSCI had 16.5-fold lower activity toward taurine-conjugated DCA (TDCA) than did DCA, while rCHIR and rCHYL showed 2.4-fold and 2.2-fold lower activity toward TDCA than DCA, respectively. Glycine conjugation resulted in a relative activity similar to taurine conjugation with rCSCI, but rCHYL and rCHIR showed 4.1-fold and 3.9-fold decreases in activity relative to DCA, respectively (**Table 2.3**).

Phylogenetic analysis of bacterial 12α-HSDH

Next, we determined the phylogenetic positions of CSCI, CHIR, and CHYL relative to nearly 10,000 similar amino acid sequences in the NCBI NR database, by inferring a maximum likelihood tree, from which the subtree containing the sequences of interest was extracted for presentation (**Figure 2.6**). CHYL clustered with other *Clostridiales* sequences, including the first characterized 12α-HSDH from *Clostridium* sp. ATCC 29733.[29] CHIR is a lone sequence which shares a common node with a larger set of clusters composed principally of *Firmicutes* and *Actinobacteria*. Within this larger cluster is CSCI, which is closely related to two metagenomic sequences, one sequence from *Eubacterium* sp. strain CAG:192 and one from *Eggerthella* sp.

strain CAG:298 (accession no. CDD59475.1). We recently reported that the protein with accession number CDD59475.1 has 12α-HSDH activity.[34]

The *Actinobacteria* represented in the phylogeny, from the family *Coriobacteriaceae*, are of particular interest due to previous reports of 12α-HSDH activity. Wegner et al. demonstrated that mouse fecal isolate *Enterorhabdus mucosicola* DSM 19490T expresses bile salt hydrolase, as well as 3α-HSDH and 12α-HSDH activities.[36] Our phylogenetic analysis suggests that the NAD(P)-dependent oxidoreductase (accession no. WP_028027349.1) is the 12α-HSDH encoded by *E. mucosicola*, and that closely related *Enterorhabdus caecimuris* protein (accession no. WP_016309125.1) is also a 12α-HSDH. Within this cluster are proteins from other *Coriobacteriaceae* members, such as an equol-producing strain of *Adlercreutzia equolifaciens*[37] and an asaccharolytic strain of *Senegalemassilia anaerobia*. Sharing a node (support value of 0.946) with this cluster are additional *Coriobacteriaceae* family members in the genus *Collinsella*, including *Collinsella aerofaciens* (accession no. WP_006235414.1), which was recently reported to express 12α-HSDH activity.[36] Another genus within the *Coriobacteriaceae*, *Eggerthella*, is represented at various points within the tree and deserves mention for two reasons. First, early reports establish *E. lenta* strains to be capable of oxidation and epimerization of bile acids, with clear 12α-HSDH activity.[38–40] Second, in addition to the protein with accession number CDD59475.1[34], we recently characterized Elen_2515 from *E. lenta* DSM 2243T, which was found to have bile acid 12α-HSDH activity[33].

While 12α-HSDH is well represented among Gram-positive members of the *Actinobacteria* and *Firmicutes*, the only taxa representing Gram-negative bacteria within this portion of the tree were strains of *Bacteroides pectinophilus* (**Figure 2.6**). There is a large cluster of *Bacteroides* sp. proteins, including several from *Bacteroides fragilis*, which were further

removed from the portion of the tree represented in **Figure 2.6** (see also **Supplementary Figure 2.2**). The *B. fragilis* proteins are members of the SDR family; however, they share only 31% amino acid identity with CSCI. Suspecting that the *B. fragilis* cluster represented in **Supplementary Figure 2.2** might contain the 7α-HSDH reported for *B. fragilis* ATCC 25285, we compared the amino acid sequence of a 3-oxoacyl-[acyl-carrier-protein] reductase (accession no. WP_005799187.1) to that of 7α-HSDH (accession no. WP_005792012.1).[41] However, the *B. fragilis* 7α-HSDH shared only 34% amino acid identity with the protein with accession number WP_005799187.1, despite both proteins being members of the FabG family of proteins. The sequence identity between WP_005799187.1 and the 12α-HSDH from *C. scindens* (CSCI) was only 31%. Multiple forms of 7α-HSDH were reported among strains of *B. fragilis*[42], and further research will be needed to determine the functions of the proteins represented in this portion of the phylogeny.

Archaea were also identified in an isolated cluster (**Figure 2.6**) populated by human gut isolates and metagenomic sequences of *Methanobrevibacter smithii* and *Methanosphaera stadtmanae* DSM 3091, consistent with a previous report.[29] Our phylogeny suggests that 12α-HSDH, like bile salt hydrolase[3], was horizontally transferred from the *Firmicutes* to these *Archaea*.

The low-activity BA7 bacteria *C. leptum*, *Paraclostridium bifermentans*, and *Paeniclostridium sordellii* were also represented in the larger phylogeny (**Supplementary Figure 2.2**). *C. leptum* was previously reported to express 12α-HSDH[28]; however, the gene encoding this enzyme has not been identified. *C. leptum* did not appear in a previous phylogeny based on the 12α-HSDH from *Clostridium* sp. ATCC 29733.[30,31] However, we located a *C. leptum* protein (accession no. WP_003532012.1) that shares 61% identity with a deduced protein

from *P. sordellii* (accession no. WP_057547571.1), as well as *C. perfringens* (accession no. WP_096515955.1), an organism that has previously been shown to express 3α-, 7α-, and 12α-HSDH activities.[43] When the protein with accession number WP_057547571.1 was aligned with the amino acid sequence of a 7α-HSDH previously characterized from *P. sordellii*, it was determined that they shared only 36.29% identity. *Paraclostridium bifermentans* encodes a gene with deduced amino acid sequence (accession no. WP_021433828.1) sharing 86% identity with WP_057547571.1 from *P. sordellii*. Further work will be required to determine if these proteins identified in the phylogeny have 12α-HSDH activity.

DISCUSSION

It was shown in previous studies that *C. scindens* VPI 12708 generates a complex mixture of metabolites from the bile acid CA during anaerobic growth in culture medium (**Figure 2.7**).[4,24] These studies identified 3α-HSDH and 7α-HSDH activities, and the genes encoding these enzymes have been reported in strains of *C. scindens*[19,22], as well as *P. hiranonis*[44] and *C. hylemonae*[45]. However, the formation of 12-oxointermediates has not been reported in anaerobic batch culture-based studies of these strains, but it was reported under aerobic conditions by resting cells of *P. hiranonis*.[32] Oxidative conditions would be expected to funnel reductant toward detoxification of molecular oxygen, resulting in bile acid oxidation. *C. leptum*, which expresses BA7 activity but appears to lack several genes in the pathway harbored by *C. scindens*, *P. hiranonis*, and *C. hylemonae*[5], also expresses 12α-HSDH[28]. Intriguingly, our extensive phylogenetic analysis may have identified the gene encoding 12α-HSDH in *C. leptum*; however, further work will be needed for confirmation.

The limitations of previous studies that cultured BA7 bacteria under low redox potential relate to their use of host-derived primary bile acids (CA, CDCA, and UDCA), rather than microbe-derived oxidized bile acid metabolites. Consistent with previous reports[24,44,45], 12α-HSDH activity was not observed when CA, 3-dehydro-CA, 7-oxoDCA, or DCA was added to anaerobic cultures of *C. scindens*, *C. hylemonae*, or *P. hiranonis* (**Table 2.1**). However, when we introduced 12-oxoLCA, 3,12-dioxoLCA, 7,12- dioxoLCA, and 3,7,12-trioxocholanic acid, we observed reduction of the 12-oxo-group, providing evidence for 12α-HSDH activity. Surprisingly, 12-oxoCDCA was not converted to DCA in batch cultures of *C. scindens*, *P. hiranonis*, and *C. hylemonae*. Instead, 12-oxoCDCA was converted quantitatively to 7-oxoDCA, whose formation could follow two metabolic routes, both depicted in **Figure 2.7**. Both reduction of bile acid oxo-groups and 7α-dehydroxylation result in a net 2-electron reduction.[4] We therefore speculate that a reduction of oxo-bile acids fulfills the redox balance needs of the organisms in batch culture. It could also be that these oxo-derivatives do not induce the bile acid 7α-dehydroxylation pathway. Future work will need to focus on understanding reductant flow and induction conditions in these gut bacteria when substrates are oxo-bile acids versus primary bile acids.

The purified recombinant 12α-HSDH enzymes from these strains displayed an order of magnitude lower activity toward 12-oxoCDCA than 12-oxoLCA (**Table 2.2**). This is consistent with kinetic analysis of 12α-HSDH enzymes from gut bacteria that lack BA7 activity and the *bai* operon.[30,33,34] Unlike *C. scindens*, *P. hiranonis*, and *C. hylemonae*, non-BA7 gut bacteria appear to oxidize the C-12 of DCA.[26,33,44,45]

Our phylogenetic analysis updates and extends previous work on 12α-HSDH enzymes in gut bacteria.[29] Indeed, we have recently verified that *Eggerthella* strains harbor distinct 12α-

HSDH enzymes, both of which (Elen_2515 and the protein with accession no. CDD59475) are represented in our phylogeny (**Figure 2.6**). Furthermore, newly reported members of the gut microbiota recently obtained through culturomics approaches are represented in our phylogeny, expanding the potential for exploring bile acid metabolism by diverse gut microbes. The widespread distribution of 12α-HSDH among *Firmicutes*, *Coriobacteriaceae* family members of the *Actinobacteria*, and gut *Archaea* may represent an important detoxification mechanism to protect against the antimicrobial nature of DCA. DCA is roughly 10-fold more toxic to many gut bacteria than CA.[45] 12-oxoLCA is a commonly detected metabolite in human fecal samples; thus, the formation of this metabolite may have measurable effects on the microbial community and host physiology.[46–48] Given their potential as redox-active centers, further research on oxo-bile acids is warranted.

The 12-oxo group of 12-oxoLCA would be expected to be an electron sink for bile acid 7α-dehydroxylating gut bacteria, and reduction of this group would be favorable in an anaerobic environment. However, this may be a proximate cause, and a different evolutionary pressure may have evolved to respond to C-12 oxidation by other members of the gut microbiota. Oxidation of bile acid hydroxyl groups reduces the detergent nature of the bile acid and its toxicity to gut bacteria.[49–52] Therefore, 12α-HSDH activity in bile acid 7α-dehydroxylating gut bacteria may have arisen to maintain toxic levels of DCA in fecal water, perhaps to reduce competitors for key nutrients. Furthermore, the observations of Masuda and Oda[32] suggest that redox potential affects the direction of this reaction, and that oxidation of C-12 is more probable closer to the mucosa where oxygen partial pressures are increased.[53]

At present, there are no structures for gut bacterial 12α-HSDHs. Such information would provide important insight into the key residues that might be altered to engineer an enzyme with

increased activity toward CA, a step toward improving Wolff-Kishner reduction in the synthesis of the therapeutic bile acid UDCA. Further, our data suggest that shifting CA metabolism toward the formation of 12-oxoCDCA may reduce the rate of formation of DCA, a bile acid implicated in cancers of the liver[8] and colon[7], as well as cholesterol gallstone formation[10], and roles in cardiovascular disease are becoming apparent[54]. Theoretical studies also suggest that oxo-bile acids, including 12-oxo-derivatives, are potential enzyme inhibitors, effectors of host nuclear receptors, and substrates for particular P450 monooxygenases.[55] Further studies are needed to determine the role of microbial bile acid oxidation on host physiology.

MATERIALS AND METHODS

Bacterial strains and culture conditions.

     C. scindens ATCC 35704, *C. hylemonae* DSM 15053, and *P. hiranonis* DSM 13275 were obtained from -80°C glycerol stocks from culture collections at the University of Illinois at Urbana-Champaign (UIUC) and Eastern Illinois University (EIU). Each strain was grown anaerobically at 37°C in butyl rubber-stoppered crimp-sealed culture tubes (18 by 150 mm, series 2048; Bellco Glass, Inc., Vineland, NJ, USA; 27.2 ml approximate stoppered volume at 1 atm [101.29 kPa]) containing brain heart infusion (BHI) broth consisting of the following (grams per liter): BHI broth (BD Difco), 37; NaHCO3, 7.5; glucose, 2.0; yeast extract, 5.0; and resazurin, 0.001. The culture medium was prepared anaerobically by boiling and cooling the medium under 100% CO2, adding cysteine・HCI・H2O (0.5 g/liter) to the medium, and dispensing the medium under 100% CO2 into culture tubes (10 ml per tube). Tubes were subsequently crimp sealed and autoclaved. After autoclaving, the pH of the medium approximated 6.8. Stock solutions (10.6 mM) of bile acids were prepared in 100% methanol,

added to tubes, sealed with stoppers, and degassed by sparging and then flushing the headspace gas with argon. Bile acid solutions were added (0.1 ml) via sterile needles and syringes to tubes of sterile BHI broth to achieve an initial concentration of 0.1 mM after inoculation. In all experiments, growth was initiated by injecting 0.5 ml of inoculum per culture. The final concentration of methanol in cultures approximated 0.9% and did not influence the growth of the strains being tested; growth was also not observed following the addition of methanolic bile acid solutions to tubes of sterile BHI broth in the absence of inoculum.

Escherichia coli DH5α (Turbo) competent cells were from New England BioLabs (NEB) (Ipswich, MA, USA) and *E. coli* BL21-CodonPlus(DE3)-RIPL cells were purchased from Stratagene (La Jolla, CA, USA) and grown in Luria-Bertani (LB) broth at 37°C.

Chemicals

5β-Cholanic acid-3α, 7α, 12α-triol (CA), 5β-cholanic acid-3α, 12α-diol (DCA), 5β-cholanic acid-3α, 7α-diol (CDCA), 5β-cholanic acid-3α, 12α-diol *N*-(2-sulfoethyl)-amide (tauroDCA), glycoDCA, 5β-cholanic acid-3α, 7α, 12α-triol *N*-(2-sulfoethyl)-amide (tauroCA), 5β-cholanic acid-3α, 7α, 12α-triol *N*-(carboxymethyl)-amide (glycoCA), 5β-cholanic acid-3α, 7α-diol *N*-(2-sulfoethyl)-amide (tauroCDCA), and 5β-cholanic acid-3α, 7α-diol *N*-(carboxymethyl)-amide (glycoCDCA) were purchased from Sigma- Aldrich (St. Louis, MO, USA). 5β-Cholanic acid-3α, 7α-diol-12-one (12-oxoCDCA), 5β-cholanic acid-3α-ol-12-one (12-oxoLCA), 5β-cholanic acid-12α-ol-3-one (3-oxoDCA), 5β-cholanic acid-3α, 12α-diol-7-one (7-oxoDCA), 5β-cholanic acid-3,7,12-trione (3,7,12-trioxoLCA), 5β-cholanic acid-3α-ol-7,12-dione (7,12-dioxoLCA), and 5β-cholanic acid-3,12-dione (3,12-dioxoLCA) were purchased from Steraloids, Inc. (Newport, RI, USA).[56] Isopropyl β-D-1-thiogalactopyranoside (IPTG) was

purchased from Gold Biotechnology (St. Louis, MO, USA). Strep-Tactin resin was purchased from IBA GmbH (Gottingen, Germany). All other reagents were of the highest possible purity and were purchased from Fisher Scientific (Pittsburgh, PA, USA).

Detection of bile acids in cultures

TLC was used for the detection of bile acids in BHI cultures and in standards prepared from sterile BHI broth supplemented with a bile acid.[24,57,58] Briefly, a 1-ml sample of culture or sterile broth standard was transferred to duplicate 2-ml microcentrifuge tubes (0.5 ml of sample per tube). To each sample, 100 µl of 3 N HCl and 500 µl of ethyl acetate were added. The microcentrifuge tubes were capped, mixed by vortexing, and spun in a microcentrifuge for 1 min at 14,000 rpm. The organic phase (top layer) from both samples was removed and transferred to a 20-ml glass scintillation vial. The extraction step was repeated for each sample, and the combined ethyl acetate extracts (~2 ml) were dried at room temperature under a stream of nitrogen. Methanol (100 µl) was added to each vial, and the resuspended extracts were immediately spotted (50 µl) onto silica plates (Whatman AL SIL G, 250 µm thick, non-UV; Fisher Scientific). Individual and mixed methanolic standards (0.1 and 1.0 mM) were also prepared by dissolving pure bile acids directly in methanol and spotting (50 µl) onto silica plates. Unless indicated otherwise, TLC plates were developed in a tank containing 50 ml of solvent (cyclohexane, ethyl acetate, and glacial acetic acid, 12:12:1 [vol/vol/vol]). Once the solvent reached approximately 25 mm from the top of the plate, the plate was removed from the tank, sprayed with charring agent solution consisting of methanol (150 ml), water (150 ml), concentrated sulfuric acid (10 ml), and $MnCl_2 \cdot 4H_2O$ (1 g), heated to 115°C for 15 min, and

visualized with long-UV light (365 nm). Bile acids were identified by comparing *Rf* values of bile acid standards to those of bile acids detected in cultures.

Isolation of genomic DNA

Genomic DNA for PCR and molecular cloning applications was extracted from *P. hiranonis* DSM 13275, *C. scindens* ATCC 35704, and *C. hylemonae* DSM 15053 using the Fast DNA isolation kit from Mo Bio (Carlsbad, CA, USA), according to the manufacturer's protocol.

Heterologous expression of 12α-HSDH enzymes in E. coli and protein purification

The pET- 51b(+) vector was obtained from Novagen (San Diego, CA, USA). Restriction enzymes were purchased from NEB (Ipswich, MA). Inserts were generated by PCR amplification using cloning primers obtained from Integrated DNA Technologies, Inc. (Coralville, IA, USA). The cloning primers used in this study are listed in **Table 2.4**. Inserts for directional cloning were amplified using the Phusion high-fidelity polymerase (Stratagene, La Jolla, CA, USA) and cloned into pET-51b(+) after the insert and vector were doubly digested with the appropriate restriction endonuclease and treated with DNA ligase. Recombinant plasmid was transformed into chemically competent *E. coli* DH5α cells via a heat shock method, plated, and grown for 16 h at 37°C on lysogeny broth (LB) agar plates supplemented with ampicillin (100 µg/ml). A single colony from each transformation was inoculated into LB broth (5 ml) containing ampicillin (100 µg/ml) and grown to saturation. The cells were subsequently centrifuged (3,220 x *g*, 15 min, 4°C), and plasmids were extracted from the resulting cell pellet using the QIAprep Spin miniprep kit (Qiagen, Valencia, CA, USA). The sequence of the insert

was determined by Sanger sequencing (W. M. Keck Center for Comparative and Functional Genomics at the University of Illinois at Urbana-Champaign).

For protein expression, the correct recombinant plasmids extracted from the *E. coli* DH5α cells were transformed into *E. coli* BL21 CodonPlus (DE3)-RIPL chemically competent cells by heat shock method and cultured for 24 h at 37°C on LB agar plates supplemented with ampicillin (100 µg/ml) and chloramphenicol (50 µg/ml). Selected colonies were inoculated into LB broth (10 ml) supplemented with antibiotics and grown at 37°C for 6 h with vigorous aeration. The precultures were then added to fresh LB broth (1 liter), supplemented with ampicillin (100 µg/ml), and aerated at 37°C until reaching an optical density at 600 nm (OD600) of 0.3. To induce recombinant protein expression, IPTG was added to each culture at a final concentration of 0.1 mM, and the temperature was decreased to 16°C. Following 16 h of incubation, cells were pelleted by centrifugation (4,000 x *g*, 30 min, 4°C) and resuspended in 30 ml of binding buffer (20 mM Tris-HCl, 150 mM NaCl, 10% glycerol, 10 mM 2-mercaptoethanol [pH 7.9]). The cell suspension was subjected to four passages through an EmulsiFlex C-3 cell homogenizer (Avestin, Ottawa, Canada), and the cell lysate was clarified by centrifugation at 20,000 x *g* for 30 min at 4°C.

The recombinant 12α-HSDHs were purified using Strep-Tactin resin (IBA GmbH), as per the manufacturing protocol. The protein was eluted with an elution buffer composed of 20 mM Tris-HCl, 150 mM NaCl, 10% glycerol, 10 mM 2-mercaptoethanol (pH 7.9), and 2.5 mM D-desthiobiotin. The purity of the proteins was assessed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and the protein bands were visualized by Coomassie brilliant blue G-250 staining. The concentrations of the proteins were calculated based on their extinction coefficients and molecular weights. The observed subunit weight for each was calculated by

migration distance of purified protein to standard proteins in ImageJ

(https://imagej.nih.gov/ij/index.html).

Enzyme assays

The buffers used to investigate the optimal pHs of the recombinant 12α-HSDH enzymes (rCHYL, rCSCI, and rCHIR) contained 150 mM NaCl and varied contents of the following buffering agents: 50 mM sodium acetate (pH 4 to 6), 50 mM sodium phosphate (pH 6.5 to 7.5), and 50 mM Tris-Cl (pH 8 to 9). The linearity of the 12α-HSDH activity over time and the enzyme concentration were determined by monitoring the oxidation and reduction of NADP(H) aerobically at 340 nm ($\varepsilon = 6{,}220$ M^{-1} · cm^{-1}) continuously for 1.5 min in a reaction mixture with 12α-HSDH and bile acid substrates. The reaction mixtures for rCHYL in the oxidative direction were composed of sodium phosphate buffer at pH 8.0, 18 nM enzyme with either 700 μM NADP$^+$ and various concentrations of DCA or 900 μM DCA and differing NADP$^+$ concentrations. The reductive direction consisted of pH 7.0 sodium phosphate buffer and 150 μM NADPH and different 12-oxoLCA concentrations or 100 μM 12-oxoLCA and various NADPH concentrations, with 12.5 nM enzyme. rCSCI oxidative reaction mixtures included 10 nM enzyme sodium phosphate buffer (pH 7.5), and 400 μM NADP$^+$ with various concentrations of DCA or 1,500 μM DCA with different NADP$^+$ concentrations. The reductive mixtures were sodium phosphate buffer at pH 7.0, 8 nM enzyme, and 150 μM NADPH with changing 12-oxoLCA concentrations or 50 μM 12-oxoLCA with differing NADPH concentrations. The oxidative reaction mixtures for rCHIR were sodium phosphate buffer (pH 7), and 10 nM enzyme with either 900 μM NADP$^+$ and various DCA concentrations or 1,600 μM DCA and different NADPμ concentrations. The reductive mixtures were composed of sodium phosphate buffer (pH

7), 200 µM NADPH with various 12-oxoLCA concentrations or 150 µM 12-oxoLCA with differing NADPH concentrations, and 8 nM enzyme. The kinetic parameters were estimated by fitting the data to the Michaelis-Menten equation by the nonlinear regression method using the enzyme kinetics module in GraphPad Prism (GraphPad Software, La Jolla, CA, USA). The substrate specificity analysis was performed according to the previously stated reaction mixtures, using saturating concentrations of substrate and pyridine nucleotide.

TLC and mass spectrometry of bile acid metabolites

Reaction mixtures were set up with 50 µM 12-oxoLCA substrate and NAD(P)(H) cofactor, either 12.5 nM rCHYL, 8 nM rCSCI, or 8 nM rCHIR, and the optimal pH 7.0 buffer up to 1 ml. The reaction mixtures were incubated overnight at room temperature and stopped with 150 µl of 1 N HCl. The mixtures were extracted through vortexing with two volumes of ethyl acetate for 1 min and then two more volumes and another 1 min of vortexing. The organic phase was recovered and evaporated under nitrogen gas. The residue was dissolved in 30 µl ethyl acetate and spotted on the TLC plate (silica gel IB2-F flexible TLC sheet, 20 by 20 cm, with 250-µm analytical layer; J.T. Baker, Avantor Performance Materials, LLC, PA, USA) with 8 µl of 10 mM DCA and 12-oxoLCA standards. The mobile phase consisted of 70:20:2 toluene–1,4-dioxane–acetic acid.

The bile acid metabolites separated on the TLC were visualized using a 10% phosphomolybdic acid in ethanol spray and heating for 15 min at 100°C. The bile acid metabolites were then extracted from TLC plates and sent for mass spectrometry (MS) analysis at the UIUC Mass Spectrometry Laboratory. The mass spectrometer (Thermo Scientific LTQ Orbitrap XL Hybrid Ion Trap-Orbitrap) was operated with an electrospray ionization (ESI)

66

source in negative-ion mode. The sample was introduced through syringe pump injection at a flow rate of 10 µl・min^{-1}. The capillary voltage was -28 V, with a source temperature of 235°C. The mass spectrogram data were processed with the Thermo Xcalibur software.

Phylogenetic analysis of bacterial 12α-HSDHs

Sequences from the NCBI NR protein database were identified by BLASTP (limited to a maximum of 10,000 sequences, with an E value threshold of 1E^{-10}) using as query protein with accession no. ERJ00208.1 (270 residues long) from *Clostridium* sp. ATCC 29733. Sequences shorter than 200 or longer than 350 amino acids were discarded to minimize fragments and other artifacts, as well as false positives (e.g., long sequences that have some similarity due to shared domains but are not likely to be homologues). The resulting 9,968 sequences were aligned using Muscle version 3.8.31.[59]

Maximum likelihood phylogenetic analysis was performed using FastTree version 2.1.8[60], with gamma-distributed heterogeneity rates and employing the WAG substitution model. The resulting tree was drawn and processed in Dendroscope version 3.5.9[61], and cosmetic adjustments were performed in Inkscape (http://inkscape.org).
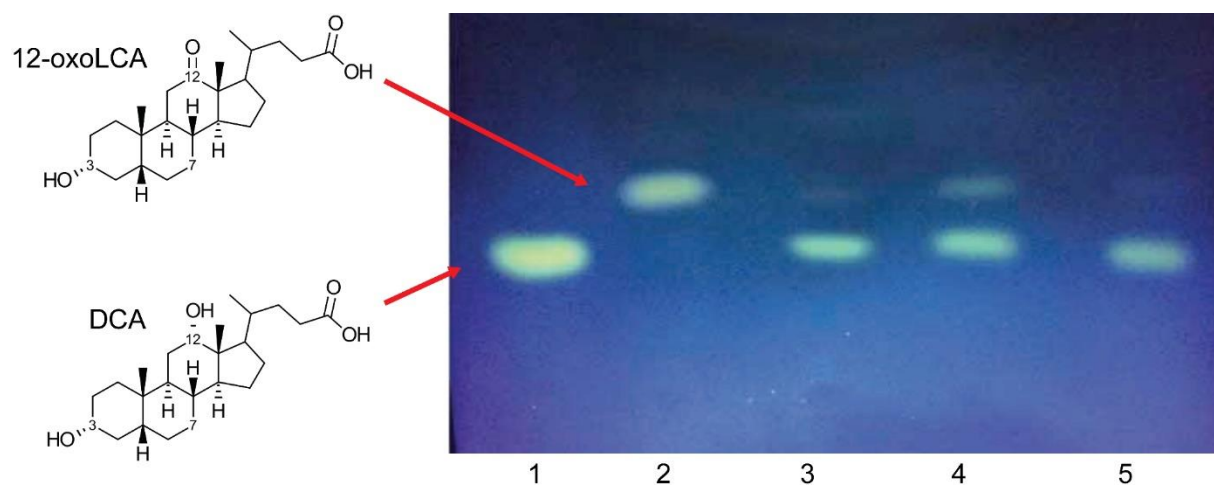
Figure 2.1. Thin-layer chromatography (TLC) demonstrating the conversion of 12-oxolithocholic acid (12-oxoLCA) to deoxycholic acid (DCA) by anaerobic cultures of *C. scindens* ATCC 35704, *C. hylemonae* DSM 15053, and *P. hiranonis* DSM 13275. DCA and 12-oxoLCA standards were spotted in lanes 1 and 2, respectively. 12-oxoLCA was added to cultures of *C. scindens* ATCC 35704 (lane 3), *C. hylemonae* DSM 15053 (lane 4), and *P. hiranonis* DSM 13275 (lane 5).

```
Elen_2515       1 ----------MDMGL-KDRVVLITGGG---------GGIARGIERAFATEGAKPILTDLFP
CLOHYLEM_04236  1 ----------MGIFDGKTAIITGGGKA-------RSIGYGIAVAYAKEGANLALTGRNE
ClostATCC29733  1 MDFIDFKEMGRMGIFDGKVAIITGGGKA------KSIGYGIAVAYAKEGANLVLTGRNE
CLOHIR_01081    1 ----------MGFLNGKTVIVTGGGRSVLSDGRCGSIGYGIVTAPAKEGANLVITGRNV
CLOSCI_02455    1 ----------MGFLTGKTAIITGGGRATLSDGSCGSIGYGIATAYAKEGANLGLTGRNL
CDD59475.1      1 ----------MGFLEGKTAIITGGGRAVLKDGSCGSIGYGIATAYAKEGANLVITGRNV
consensus       1           MGfldgKtaiiTGGGra l dg cgsIgyGIatAyAkEGAnlvlTgrnv
                                            ***            * *

Elen_2515      42 GGLBAAKEELERDFGSEVFTILANGSVEEEVRAAV----EAGABHFGGRIDVLINNAQAS
CLOHYLEM_04236 43 QKLLDAKEELERLYGIKVLPLQADVTPDEKSEEVVKETVQRVVD-TFGRIDVLINNAQAS
ClostATCC29733 54 QKLLDAKEELERLYGIKVLPLAVDVTPSDESEDRVKEAVQRVIA-EFGRIDVLINNAQAS
CLOHIR_01081   50 KKLEDAKEEIERLYGVKVLPVRADVSAGGDNKAVVDEVIKQTID-TFGRIDVLUNNAQAS
CLOSCI_02455   50 KKLEDAKEELERLYGIKVLAVQADVSAGADNKAVVEQVIKQTVE-EFGRIDVLINNAQAS
CDD59475.1     50 QKLEDAKEELERLYGIKVLPIQADVSAGNDNAATVQNVIDKTIE-BFGRIDVLINNAQAS
consensus      61 qkLedAKEElERlyGikVlpiqadvsagednravV evi ktie efGRIDVLiNNAQAS

Elen_2515      98 ASGLTLVQNSEEDFDLAVRSGLYATFPYMKHAYPYLKETAGSVINFASGAGIGGNPGQSS
CLOHYLEM_04236 102 ASGIPLSMLMKDHFDLGIYSGLYAVPYYMRACYPYLKETQGSVINFASGAGLFGNAGQSS
ClostATCC29733 113 ASGIPLSMQKDHFDLGIYSGLYATFYYMRLCYPYLKETQGSVINFASGAGLFGNVGQCS
CLOHIR_01081   109 ASGVTLEDHTTEQFDLAIYSGLYATFYYMQACLPYLKETKGSVINFASGAGLFGNYGQCA
CLOSCI_02455   109 ASGVSIADHTTEQFDLAIYSGLYAAYYYMQACYPYLREAKGSVINFASGAGLFGHYGQCS
CDD59475.1     109 ASGVSLAEHTTDQFDLAIYSGLYAAFYYMQACYPHLKETKGTVINFASGAGLFGNVGQCS
consensus      121 ASGvtl dhtteqFDLaiySGLYAtfyYMqacyPyLkEtkGsVINFASGAGlfGn GQcs
                                                                       ◆

Elen_2515     158 YAAAKEGIRGMSRVAASEWGPDNINVNIVCPIVMTKALEEWRBREPEMYEKNVNAIPLGR
CLOHYLEM_04236 162 YAAAKEGIRGISRVAATEWGKDNINVNVVCPLAMTAQLENFKBAYPEAYEKNLXAVPMGR
ClostATCC29733 173 YAAAKEGIRGLSRVAATEWGKDNINVNVVCPLAMTAQLENFKLSYPEAYEKNLRGVPMGR
CLOHIR_01081  169 YAAAKEGVRGLTRVAATEWGQFGINVNIICPLAMTAQLENFEKAYPEAFKENVKMPPAGH
CLOSCI_02455  169 YAAAKEGIRGLTRVAAEWGKDGINVNVVCPLAWTVQLENFEKAYPDAFKANVKMPPAGH
CDD59475.1    169 YAAAKEGIRGLTRVAANEWGADDINVNVICPLAWTAQLENFAEAYPDAFETNVHMPPMGH
consensus     181 YAAAKEGiRGlsRVAAtEWGkdnINVNvvCPla TaqLEnfreayPeayekNvkmpPmGr
                   ◆          ◆

Elen_2515     218 PGDAEKDVGRVCVPLASPDASFVTGDTIMVQGGSQMKP
CLOHYLEM_04236 222 PGDPEKDIGRVCVHLGSPDLKYMSGETLTLEGGMGQRP
ClostATCC29733 233 PGDPELDIGRVCVQLASPDFKYMSGETLTLEGGMGQRP
CLOHIR_01081  229 YGDAEKEIGRVCVQLASPDFKYMSGETITLEGGMGLRP
CLOSCI_02455  229 YGDVEKEIGRVCVQLASPDFKFMSGETITLEGGMGLRP
CDD59475.1    229 YGNVETEIGRPCVQLASPDFRPMSGETITLEGGMGLRP
consensus     241 fGd EkdiGRvCVqLaSPDfkfmsGeTitleGGmGlrP
```

Figure 2.2. Boxshade representation of multiple-sequence alignment of 12α-hydroxysteroid dehydrogenases from *Eggerthella lenta* strains and *Clostridium* species. Depicted in the alignment are 12α-HSDH from *Eggerthella lenta* DSM 2243 (Elen_2515) (33), *Eggerthella* CAG:298 (accession no. CDD59475) (34), *Clostridium* sp. strain ATCC 29733 (ClostATCC29733), *Clostridium scindens* ATCC 35704 (CSCI), *Clostridium hylemonae* DSM 15053 (CHYL), and *Peptacetobacter hiranonis* DSM 13275 (CHIR). Asterisks mark amino acids predicted to bind pyridine nucleotide cofactor, and diamonds indicate conserved active-site residues in the SDR family.
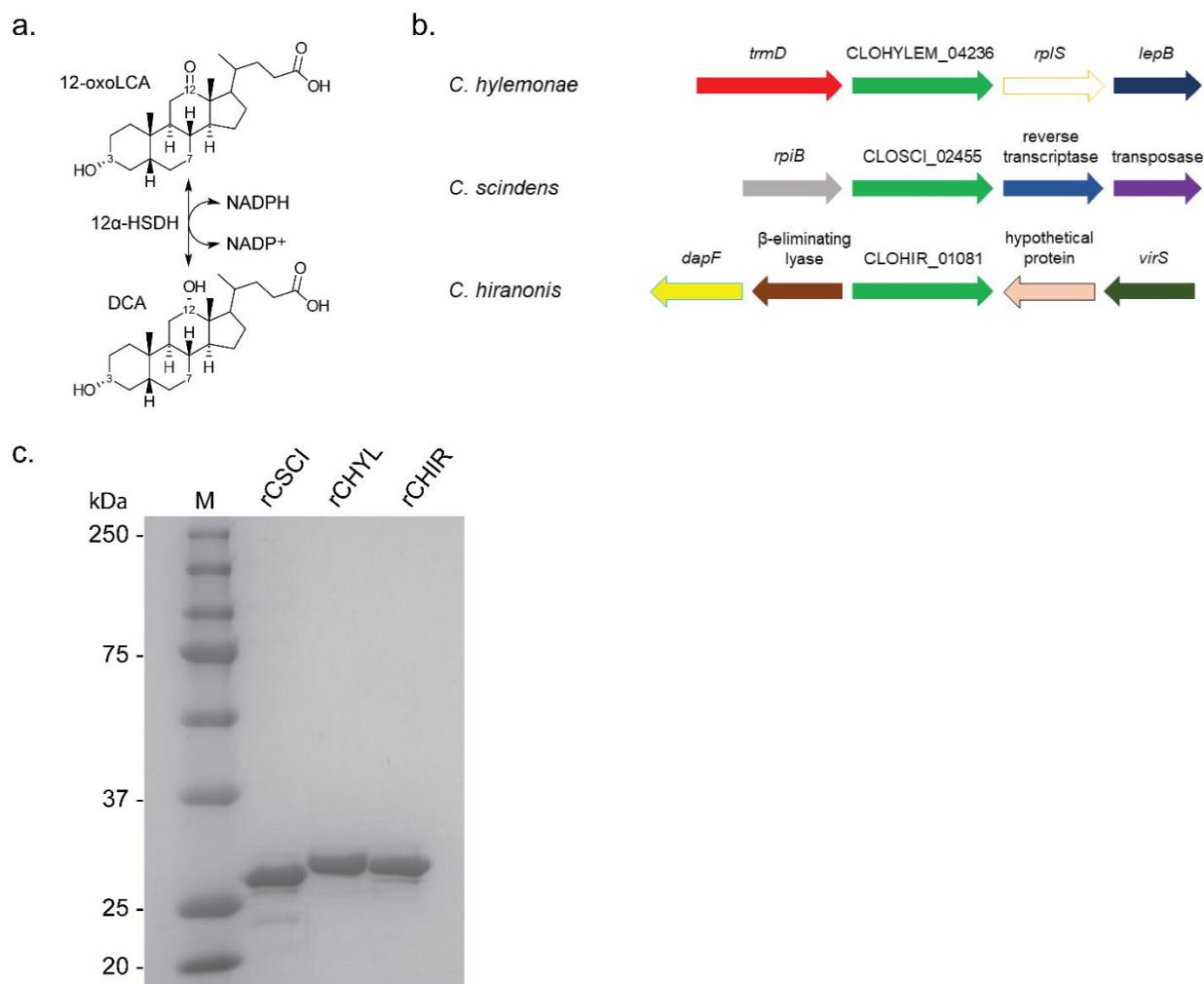
Figure 2.3. Overexpression and purification of recombinant 12α-hydroxysteroid dehydrogenases from bile acid 7α-dehydroxylating bacteria. (a) Reaction catalyzed by 12α-HSDH. (b) Genomic context of genes predicted to encode 12α-HSDH in BA7 bacteria. (c) SDS-PAGE of 2 µg purified recombinant 12α-HSDHs overexpressed in *E. coli* and affinity purified on Strep-Tactin resin. Lane M, molecular mass markers (values are in kilodaltons).
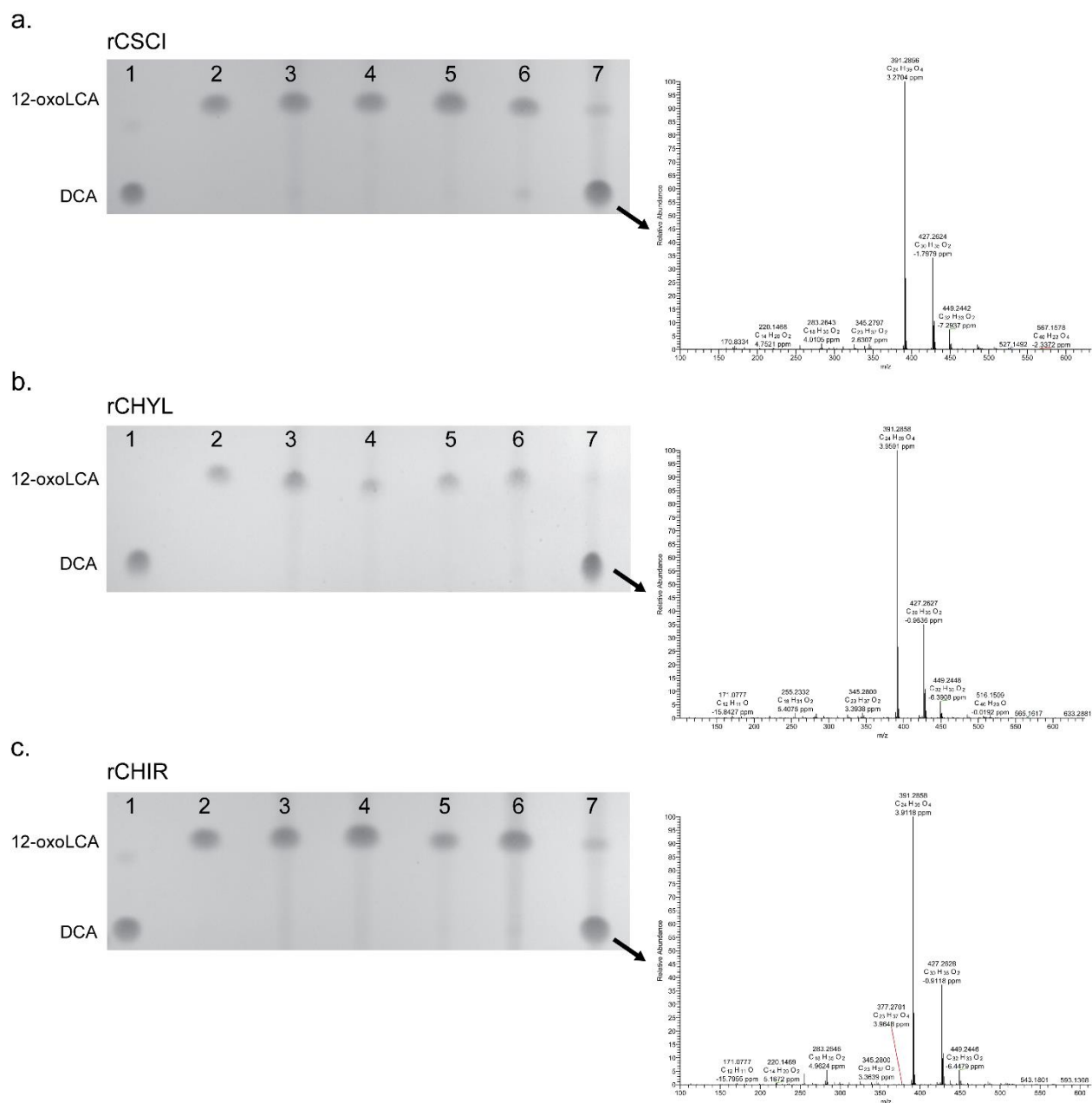
Figure 2.4. TLC and mass spectrometry of 12α-hydroxysteroid dehydrogenase overnight reaction products. In lanes 1 and 2, DCA and 12-oxoLCA were spotted as standards, respectively. Purified recombinant protein (12.5 nM rCHYL, 8 nM rCSCI, 8 nM rCHIR) was incubated overnight under the following conditions: lane 3, 12-oxoLCA plus NADPH, no enzyme; lane 4, 12-oxoLCA plus NAD$^+$ plus enzyme; lane 5, 12-oxoLCA plus NADP$^+$ plus enzyme; lane 6, 12-oxoLCA plus NADH plus enzyme; lane 7, 12-oxoLCA plus NADPH plus

Figure 2.4. (cont.)

enzyme. Arrow depicts spot analyzed by electrospray ionization mass spectrometry in negative

mode, and compared to authentic DCA standard. Formula weight for DCA is 392.58. (a to c)

TLC of CSCI (a), CHYL (b), and CHIR (c) reaction products.
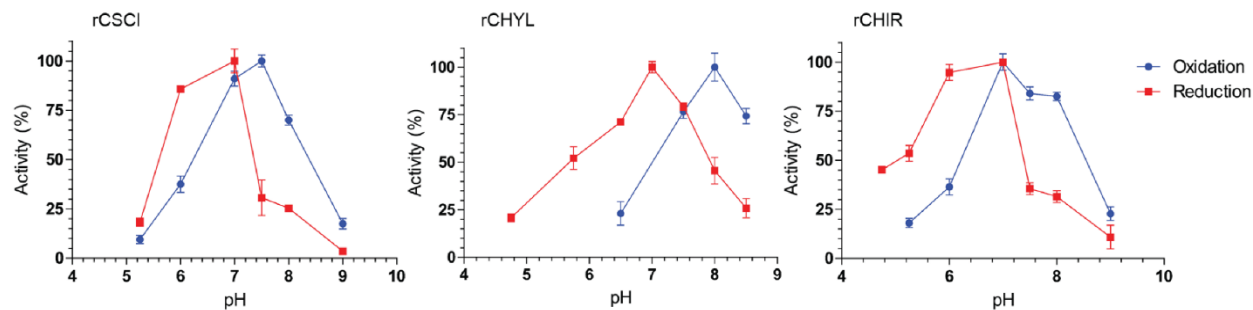
Figure 2.5. Optimal pHs for purified recombinant 12α-hydroxysteroid dehydrogenases.
Substrates were 12-oxoLCA with NADPH as cofactor (reductive direction, red) or DCA with
$NADP^+$ as cofactor (oxidative direction, blue). See Materials and Methods for buffer
compositions. Experiments were repeated three or more times; values represent means ±
standard deviation (SD).

Firmicutes
Actinobacteria
Archaea
Bacteroidetes

[Clostridium] scindens CLOSCI_02455

[Clostridium] hiranonis CLOHIR_01081

[Clostridium] hylemonae CLOHYLEM_0423

Figure 2.6 (cont.)

Figure 2.6. Subtree of phylogenetic analysis of 12α-hydroxysteroid dehydrogenases. Branch colors reflect bacterial phylum and the domain *Archaea*. Descriptors for 12α-HSDHs characterized in the present study are boxed and in red font. Accession numbers are shown to the right of the organism names.
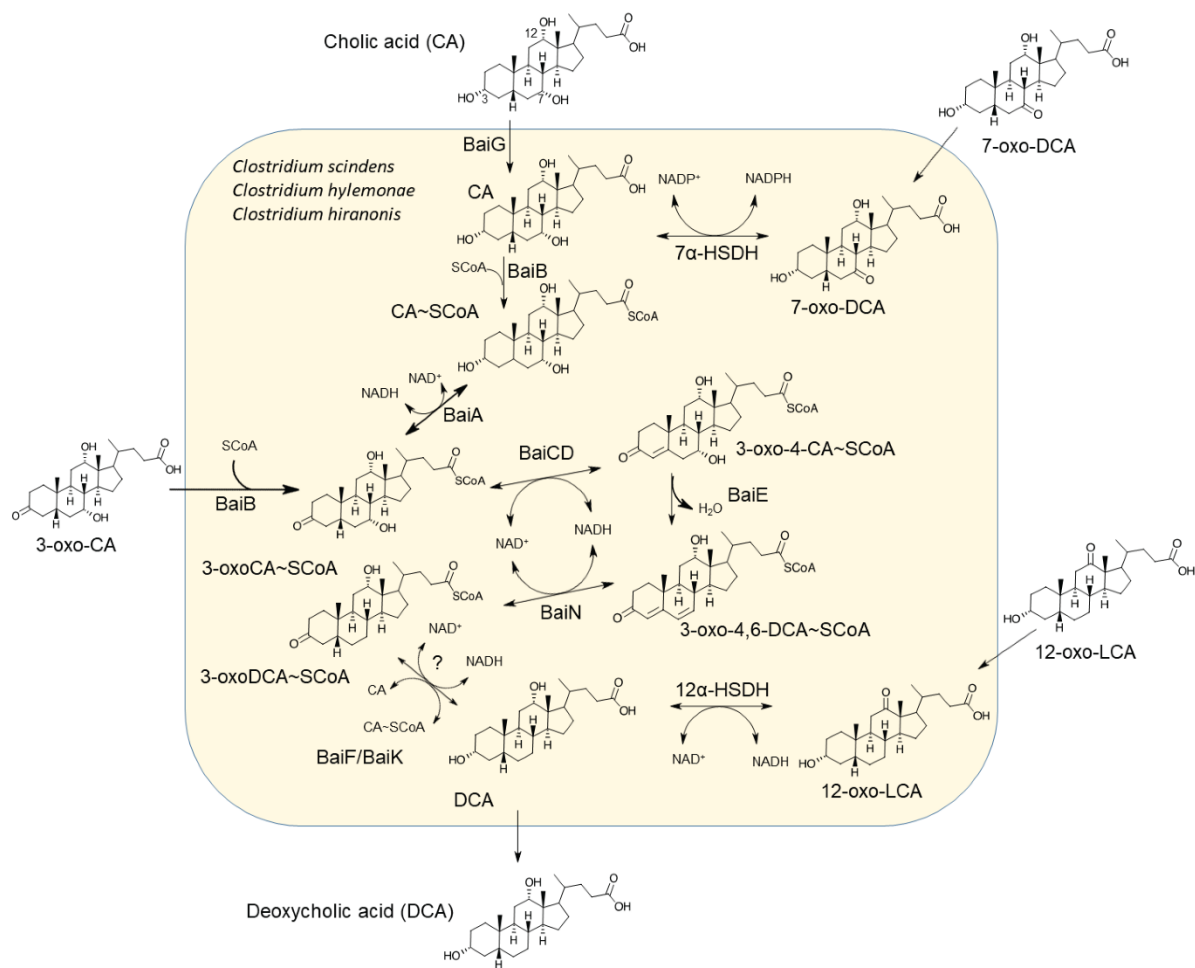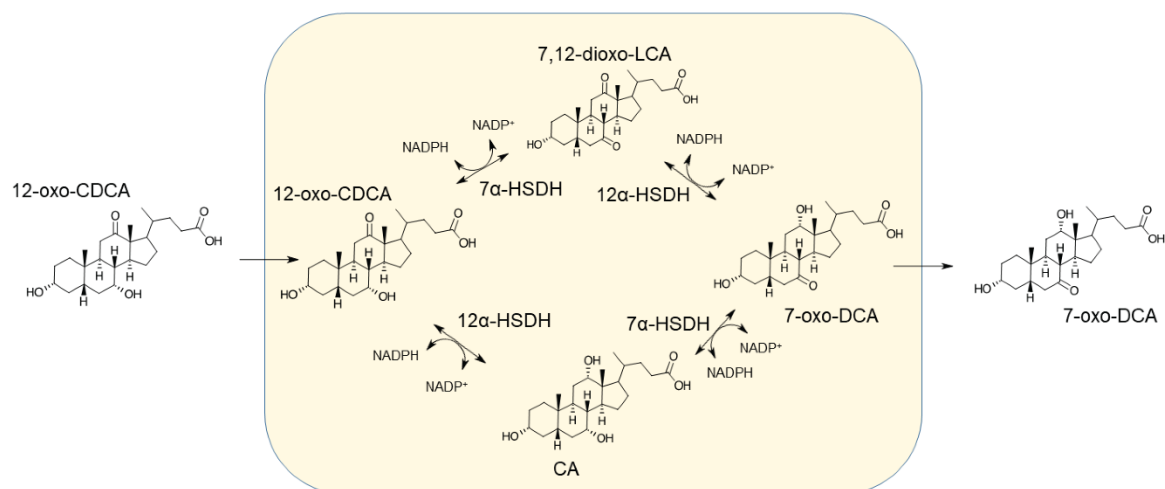
Figure 2.7. Schematic model of bile acid metabolism by bile acid 7α-dehydroxylating bacteria. (a) See introduction for information regarding the core bile acid 7α-dehydroxylation
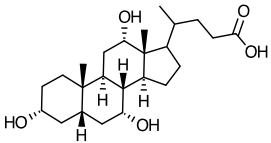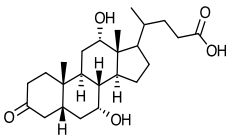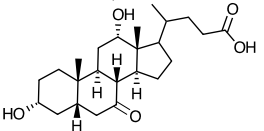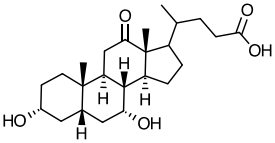
Figure 2.7 (cont.)

pathway catalyzed by bile acid-inducible (Bai) enzymes. Transport of oxo-bile acid derivatives

hypothesized to import via sodium dependent-bile acid transporter (BaiG). See **Table 2.1** for

anaerobic culture-based bile acid conversion studies. (b) Alternative routes for conversion of 12-

oxochenodeoxycholic acid (12-oxoCDCA) to 7-oxodeoxycholic acid (7-oxoDCA), observed in

bile acid conversion experiments (**Table 2.1**).
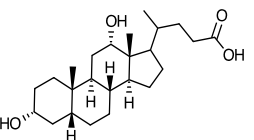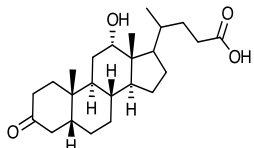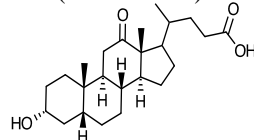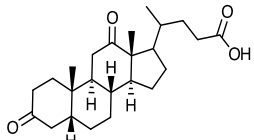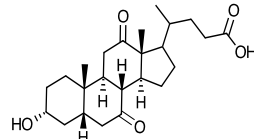
TABLES

Table 2.1. Effect of structure on bile acid transformation by *C. scindens* ATCC 35704, *C. hylemonae* DSM 15053, and *P. hiranonis* DSM 13275.
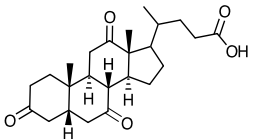
Bile acid added to culture[a]	R_f[b] value	Bile acid detected after growth (24 h)									
		CA	3-oxoCA	7-oxoDCA	12-oxoCDCA	DCA	3-oxoDCA	12-oxoLCA	3,12-dioxoLCA	7,12-dioxoLCA	3,7,12-trioxoLCA
Cholic acid (CA)	0.04 ± 0.02	*C. hylemonae*	-[c]	-	-	*C. scindens* *C. hylemonae* *P. hiranonis*	*P. hiranonis*	-	-	-	-
3-oxocholic acid (3-oxoCA)	0.21 ± 0.03	*C. scindens* *C. hylemonae* *P. hiranonis*	*C. scindens* *C. hylemonae*	*P. hiranonis*	-	*C. scindens* *C. hylemonae* *P. hiranonis*	*P. hiranonis*	-	-	-	-
7-oxodeoxycholic acid (7-oxoDCA)	0.14± 0.03	*C. scindens*	-	*C. hylemonae* *P. hiranonis*	-	*C. scindens* *P. hiranonis*	*P. hiranonis*	-	-	-	-
12-oxochenodeoxycholic acid (12-oxoCDCA)	0.27 ± 0.06	-	-	*C. scindens* *C. hylemonae* *P. hiranonis*	*P. hiranonis*	-	-	-	-	-	-

Table 2.1 (cont.)

Deoxycholic acid (DCA)

| | 0.38 ± 0.14 | - | - | - | - | C. scindens C. hylemonae P. hiranonis | - | - | - | - | - |

3-oxodeoxycholic acid (3-oxoDCA)

| | 0.56 ± 0.11 | - | - | - | - | - | C. scindens C. hylemonae P. hiranonis | - | - | - | - |

12-oxolithocholic acid (12-oxoLCA)

| | 0.51 ± 0.14 | - | - | - | - | C. scindens C. hylemonae P. hiranonis | - | - | - | - | - |

3,12-dioxolithocholic acid (3,12-dioxoLCA)

| | 0.55 ± 0.01[d] | - | - | - | - | - | C. scindens C. hylemonae P. hiranonis | - | - | - | - |

7,12-dioxolithocholic acid (7,12-dioxoLCA)

| | 0.22 ± 0.02[d] | C. scindens | C. hylemonae | - | - | C. scindens P. hiranonis | - | - | - | - | - |

Table 2.1 (cont.)

3,7,12-trioxolithocholic
acid (3,7,12-trioxoLCA)

| | 0.46 ± 0.03^{d} | *C. scindens* *P. hiranonis* | *C. scindens* *P. hiranonis* | - | - | *C. hylemonae* | - | - | - | - |

[a] Final concentration of bile acid added to BHI culture was 0.1 mM.

[b] Values determined from TLC analysis represent the means ± standard deviation (SD) of three or more replications. TLC solvent (cyclohexane, ethyl acetate, and glacial acetic acid; 12:12:1 [v/v/v]) was used for detection. Detection limits approximated ≤10 µM.

[c] -, not detected.

[d] The presence or absence of the added bile acid following growth in BHI broth was determined with an additional TLC solvent (toluene, 1-4 dioxane, and glacial acetic acid; 70:20:2 [v/v/v]). Detection limits approximated ≤10 µM.

Table 2.2. Steady-state kinetic parameters of 12α-HSDH homologs.

Homolog	Kinetic Parameters	Substrate or coenzyme[a]			
		12-oxoLCA	NADPH	DCA	NADP+
rCSCI	K_m (µM)	17.52 (0.32)	34.01 (0.32)	178.61 (0.66)	79.88 (0.37)
	K_{cat} (sec^{-1})	31.13 (1.56)	26.71 (1.22)	90.93 (4.46)	90.74 (4.35)
	V_{max} (µmole•min^{-1}•mg^{-1})	66.15 (3.31)	56.74 (2.59)	193.19 (9.47)	192.79 (9.24)
	K_{cat}/K_m (µM^{-1}•sec^{-1})	1.77 (0.109)	0.79 (0.042)	0.51 (0.029)	1.14 (0.063)
rCHYL	K_m (µM)	32.70 (0.55)	48.31 (0.37)	147.80 (0.38)	110.53 (0.38)
	K_{cat} (sec^{-1})	45.09 (2.19)	44.52 (2.30)	72.52 (3.55)	66.72 (3.18)
	V_{max} (µmole•min^{-1}•mg^{-1})	96.76 (4.70)	95.53 (4.94)	155.63 (7.64)	143.18 (6.82)
	K_{cat}/K_m (µM^{-1}•sec^{-1})	1.38 (0.08)	0.92 (0.056)	0.49 (0.028)	0.60 (0.033)
rCHIR	K_m (µM)	34.29 (0.34)	40.37 (0.38)	175.44 (0.50)	127.94 (0.42)
	K_{cat} (sec^{-1})	72.09 (3.50)	76.24 (3.88)	130.62 (7.88)	129.75 (6.68)
	V_{max} (µmole•min^{-1}•mg^{-1})	151.82 (7.37)	160.57 (8.16)	275.08 (16.60)	273.26 (14.07)
	K_{cat}/K_m (µM^{-1}•sec^{-1})	2.10 (0.12)	1.89 (0.11)	0.74 (0.052)	1.01 (0.060)

[a]The concentrations used in the assays were based on K_m values (refer to Materials and Methods). Values represent the mean ± SD based on three or more replications.

Table 2.3. Substrate specificity of purified 12α-HSDHs.

Substrate[a]	Cofactor	rCSCI[b] Activity	rCSCI[b] Relative activity (%)	rCHYL[b] Activity	rCHYL[b] Relative activity (%)	rCHIR[b] Activity	rCHIR[b] Relative activity (%)
12-oxoLCA	NADPH	38.59 ± 2.38	100	79.42 ± 3.58	100	133.56 ± 4.91	100
12-oxoLCA	NADH	-[c]	-	-	-	-	-
12-oxoCDCA	NADPH	3.69 ± 0.12	9.55	3.31 ± 0.11	4.16	4.88 ± 0.16	3.65
DCA	$NADP^+$	153.42 ± 6.2	100	127.68 ± 3.50	100	239.16 ± 2.89	100
CA	$NADP^+$	83.55 ± 1.55	54.44	102.79 ± 2.85	80.51	196.87 ± 4.36	82.32
TDCA	$NADP^+$	9.28 ± 0.57	6.05	57.56 ± 1.78	45.08	96.45 ± 6.48	40.33
GDCA	$NADP^+$	9.76 ± 1.56	6.36	30.86 ± 0.67	24.17	61.81 ± 3.61	25.85
CDCA	$NADP^+$	-	-	-	-	-	-
DCA	NAD^+	-	-	-	-	-	-

[a] 12-oxolithocholic acid (12-oxoLCA), 12-oxochenodeoxycholic acid (12-oxoCDCA), deoxycholic acid (DCA), cholic acid (CA), taurodeoxycholic acid (TDCA), glycodeoxycholic acid (GDCA), chenodeoxycholic acid (CDCA). For both the reductive and oxidative directions, the highest activity was set to 100%. The concentrations used in the assays were based on Km values (refer to Materials and Methods).
[b] Values represent the means $\pm$ SD based on three or more replications.
[c] -, no activity detected.

Table 2.4. Sequence of oligonucleotide cloning primers.

Gene ID	Primer pair	MW (kDa)[a]	Extinction coefficient (M$^{-1} \cdot$cm^{-1})
CLOSCI_02455	ATATAGGATCCTATGGGATTTTTAACAGGTAAGACAGCC (5'-Forward Primer-3')	28.24	32,110
	ATATAAAGCTTTTATGGCCTAAGCCCCATTCC (5'-Reverse Primer-3')		
CLOHYLEM_04236	ATATATGGTACCGATGGGTATATTTGACGGAAAAACAGCTA	27.96	24,995
	ATATATAAGCTTTTATGGGCGCTGTCCCATGC		
CLOHIR_01081	ATATAGGATCCTATGGGATTTTTAAATGGAAAAACAG	28.49	29,130
	ATATAGAGCTCTTAAGGTCTTAATCCCATTCCA		

[a] Deduced recombinant 12α-HSDH molecular weight

SUPPLEMENTARY FIGURES

a. CLOHYLEM_04236

b. CLOSCI_02455
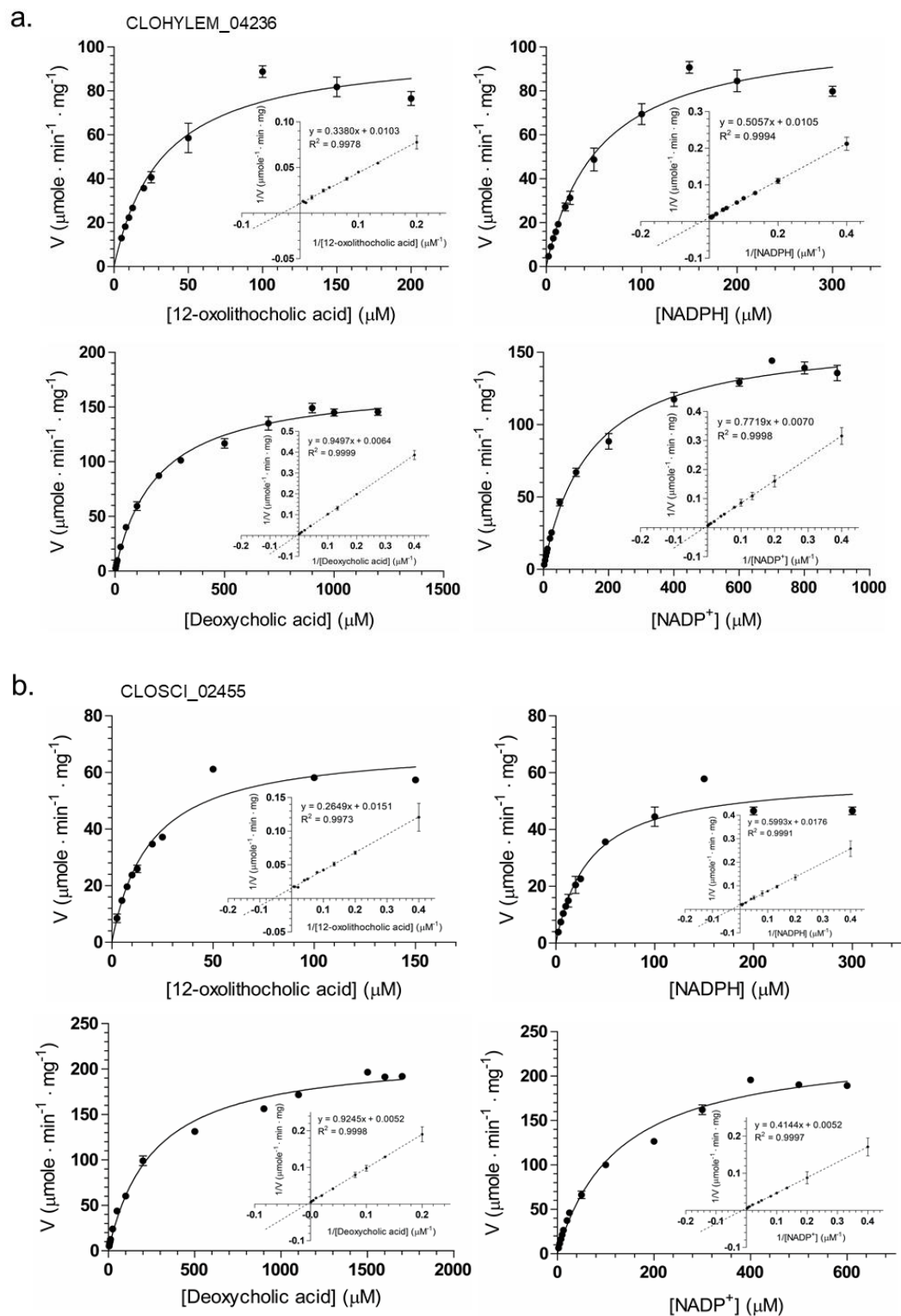
Supplementary Figure 2.1 (cont.)
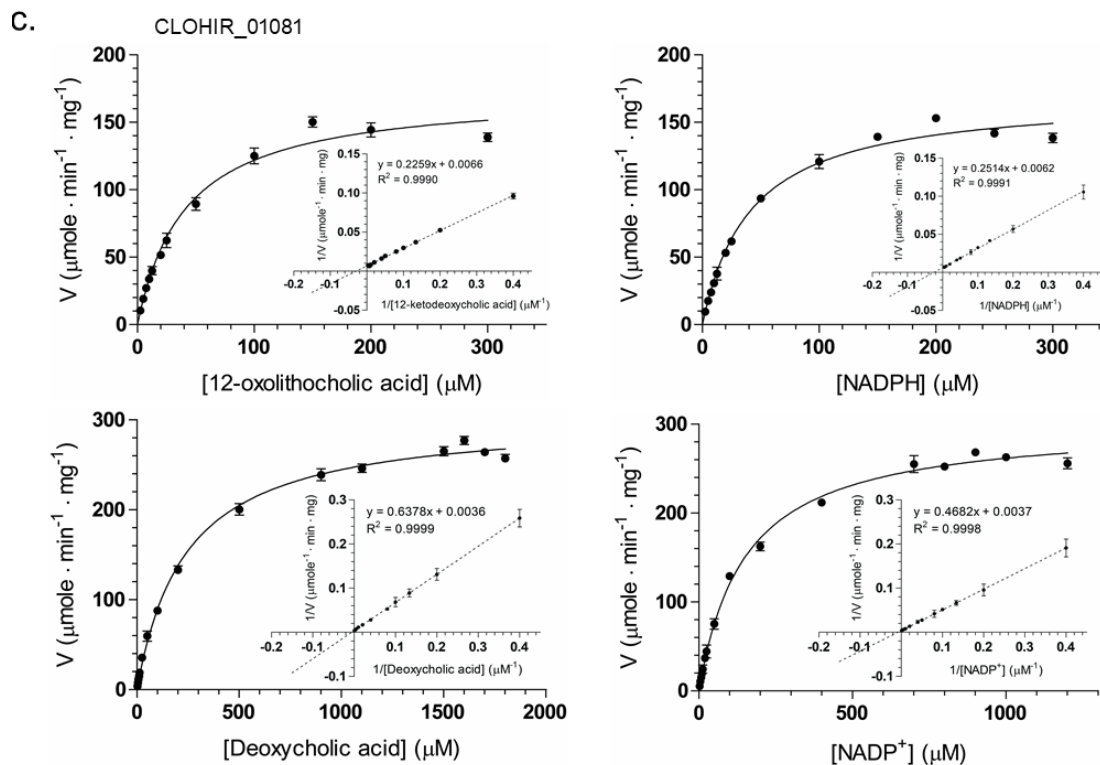
Supplementary Figure 2.1. Kinetic Analysis of rCHYL, rCSCI, rCHIR. Michaelis-Menten

and Lineweaver-Burk plot of enzyme initial velocity with varying concentration of substrate (left

panel) and co-substrate (right panel) in both the reductive (top panels) and oxidative (bottom

panels). Enzyme reaction conditions are described in Materials and Methods. The

oxidation/reduction of NAD(H) was measured by continuous spectrophotometry at 340 nm for 5

min and the initial velocities were used calculated to calculate kinetic constants (**Table 2.1**).

Each data-point represents ± S.E.M from four independent experiments.
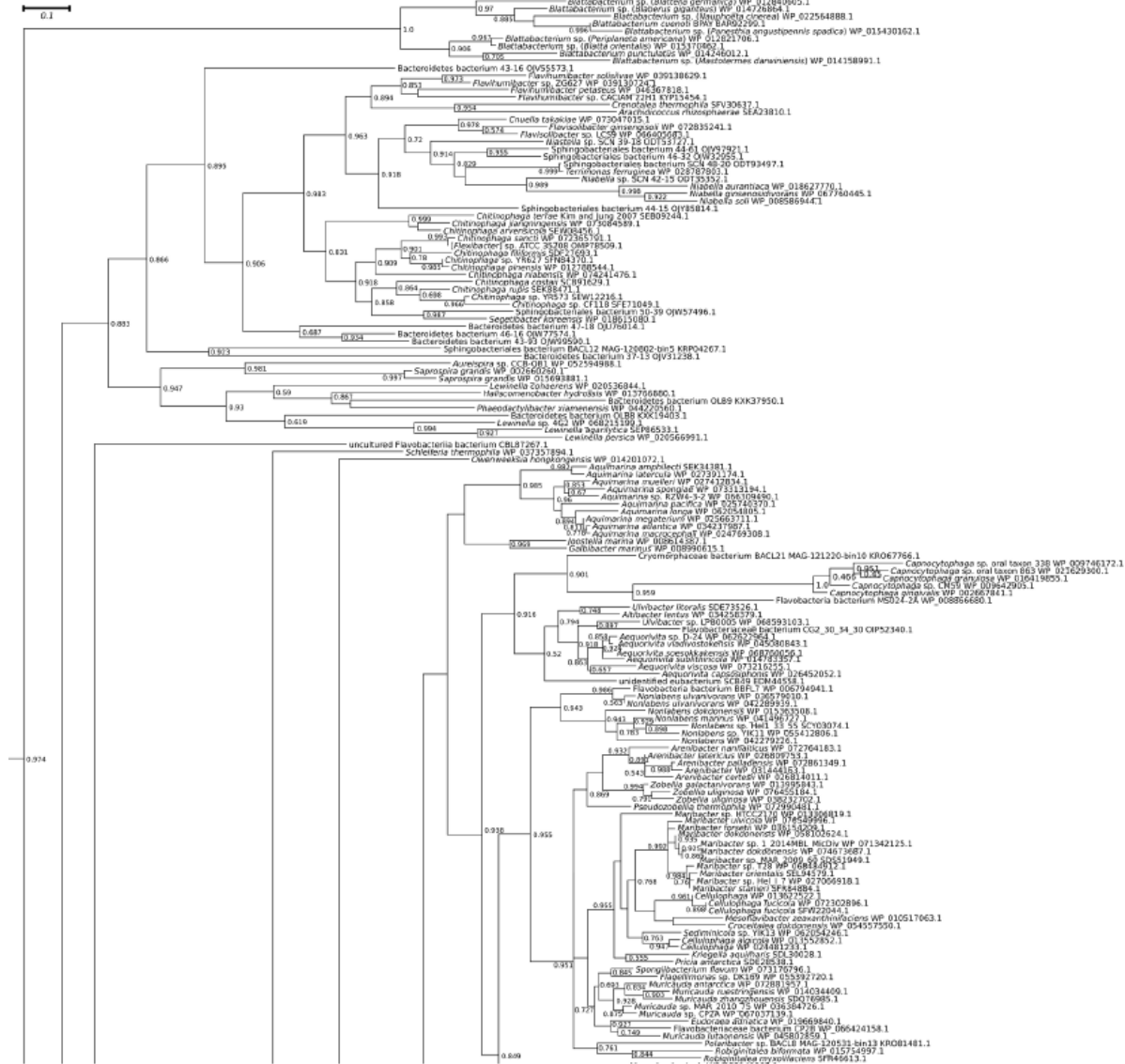
Supplementary Figure 2.2 (cont.)
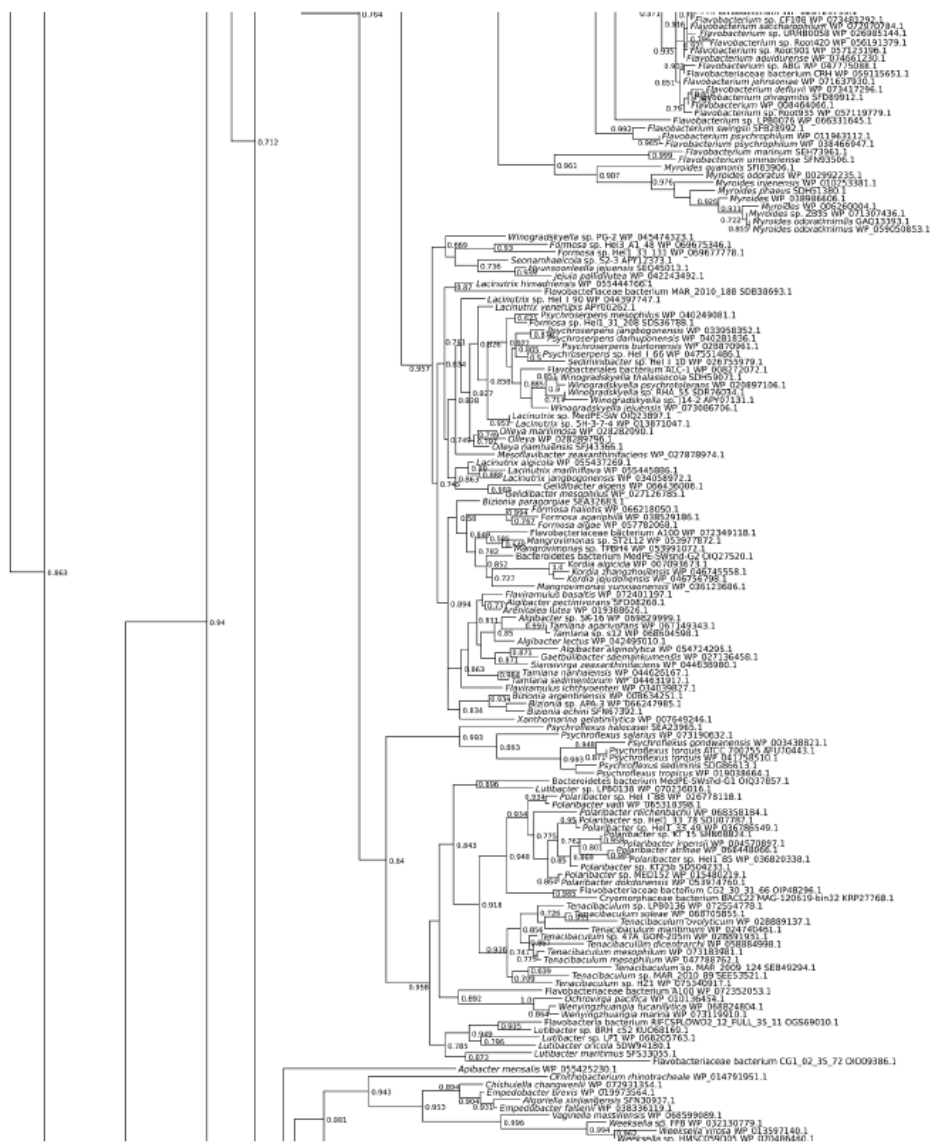
Supplementary Figure 2.2 (cont.)

Supplementary Figure 2.2 (cont.)

Supplementary Figure 2.2 (cont.)

Supplementary Figure 2.2 (cont.)
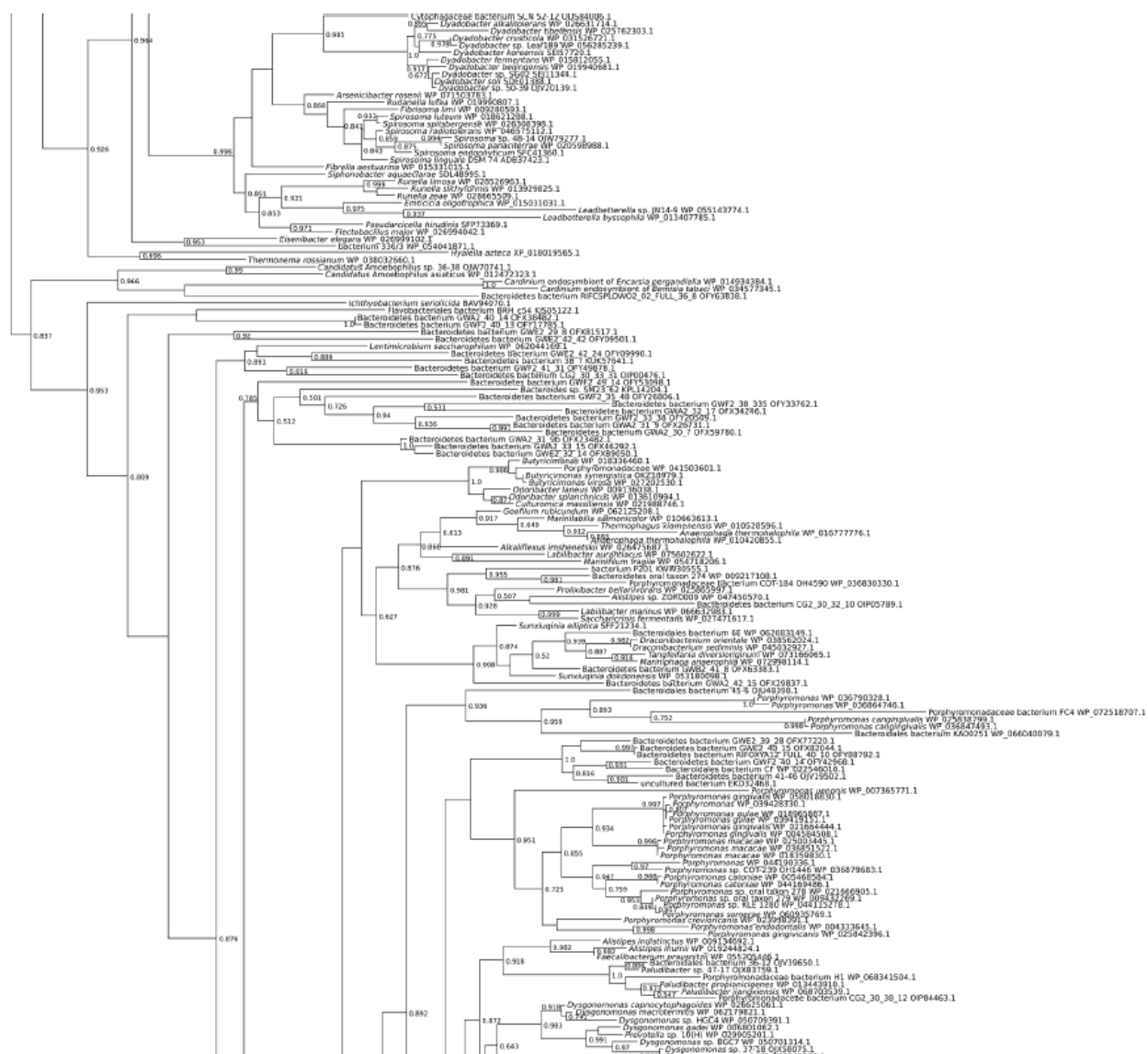
Supplementary Figure 2.2 (cont.)
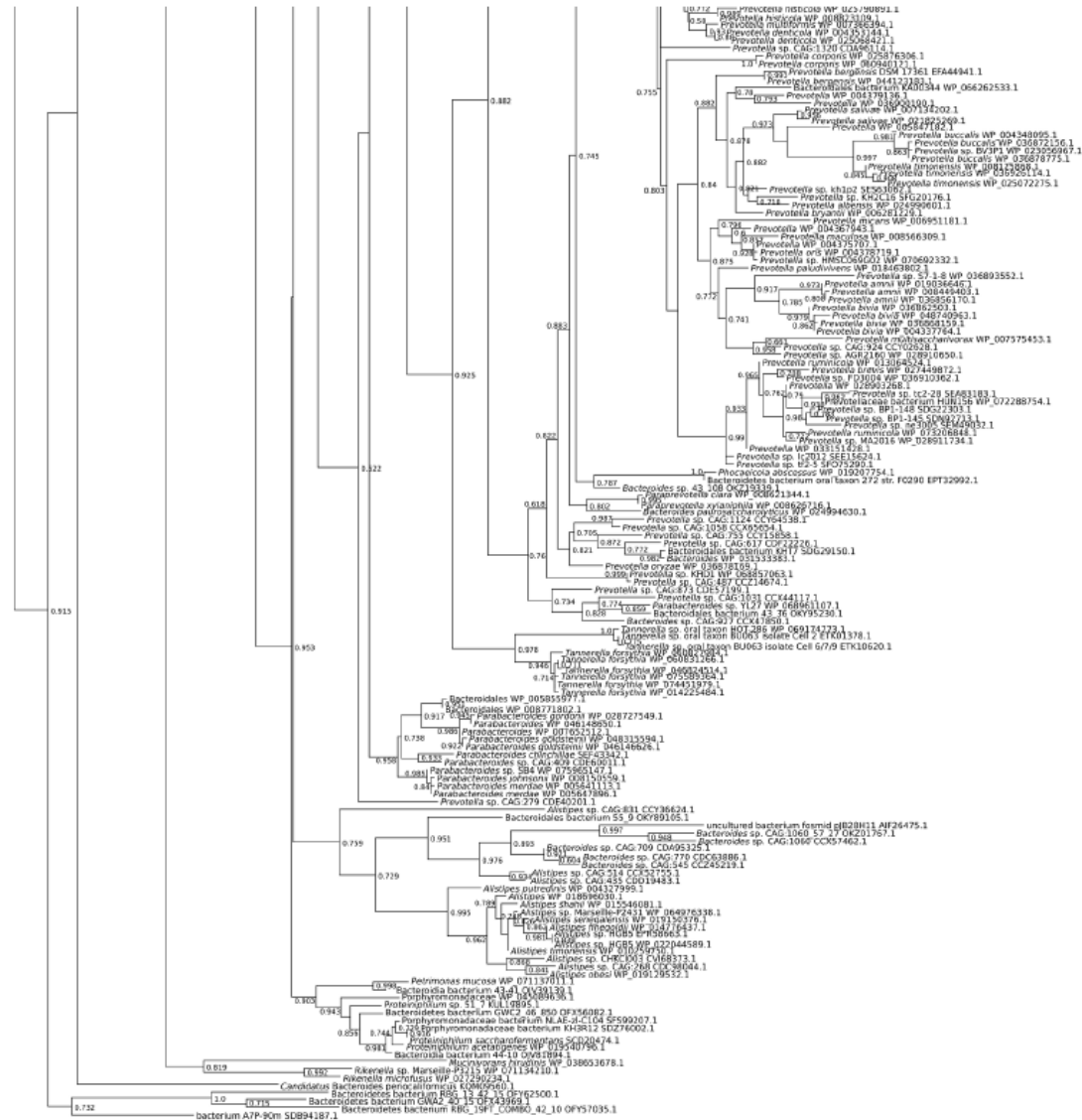
Supplementary Figure 2.2 (cont.)

Supplementary Figure 2.2. Sub-tree of phylogenetic analysis, displaying a large cluster

dominated by Bacteroidetes. Numbers on nodes represent clade support, from zero to one, as

calculated by FastTree (only values above 0.5 are shown).

REFERENCES

1. Vlahcevic ZR, Heuman DM, Hylemon PB. 1996. Physiology and pathophysiology of enterohepatic circulation of bile acids, p 376–417. *In* Zakim D, Boyer T (ed), Hepatology: a textbook of liver disease, 3rd ed, vol 1. Saunders, Philadelphia, PA.

2. Dawson PA, Karpen SJ. Intestinal transport and metabolism of bile acids. J Lipid Res 2015; 56:1085–1099. https://doi.org/10.1194/jlr.R054114.

3. Jones BV, Begley M, Hill C, Gahan CG, Marchesi JR. Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. Proc Natl Acad Sci U S A 2008; 105:13580–13585. https://doi.org/10.1073/pnas.0804437105.

4. Ridlon JM, Kang DJ, Hylemon PB. Bile salt biotransformations by human intestinal bacteria. J Lipid Res 2006; 47:241–259. https://doi.org/10.1194/jlr.R500013-JLR200.

5. Ridlon JM, Harris SC, Bhowmik S, Kang DJ, Hylemon PB. Consequences of bile salt biotransformations by intestinal bacteria. Gut Microbes 2016; 7:22–39. https://doi.org/10.1080/19490976.2015.1127483.

6. Kakiyama G, Muto A, Takei H, Nittono H, Murai T, Kurosawa T, Hofmann AF, Pandak WM, Bajaj JS. A simple and accurate HPLC method for fecal bile acid profile in healthy and cirrhotic subjects: validation by GC-MS and LC-MS. J Lipid Res 2014; 55:978–990. https://doi.org/10.1194/jlr.D047506.

7. Bernstein C, Holubec H, Bhattacharyya AK, Nguyen H, Payne CM, Zaitlin B, Bernstein H. Carcinogenicity of deoxycholic acid, a secondary bile acid. Arch Toxicol 2011; 85:863– 871. https://doi.org/10.1007/s00204-011-0648-7.

8. Yoshimoto S, Loo TM, Atarashi K, Kanda H, Sato S, Oyadomari S, Iwakura Y, Oshima K, Morita H, Hattori M, Honda K, Ishikawa Y, Hara E, Ohtani N. Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome. Nature 2013; 499:97–101. https://doi.org/10.1038/nature12347.

9. Wu J, Gong J, Geng J, Song Y. Deoxycholic acid induces the overexpression of intestinal mucin, MUC2, via NF-αB signaling pathway in human esophageal adenocarcinoma cells. BMC Cancer 2008; 8:333. https://doi.org/10.1186/1471-2407-8-333.

10. Berr F, Kullak-Ublick GA, Paumgartner G, Munzing W, Hylemon PB. 7α-Dehydroxylating bacteria enhance deoxycholic acid input and cholesterol saturation of bile in patients with gallstones. Gastroenterology 1996; 111:1611–1620. https://doi.org/10.1016/S0016-5085(96)70024-0.

11. Magouliotis DE, Tasiopoulou VS, Svokos AA, Chatedaki C, Sioka E, Zacharoulis D. Ursodeoxycholic acid in the prevention of gallstone formation after bariatric surgery: an updated systematic review and meta-analysis. Obes Surg 2017; 27:3021–3030. https://doi.org/10.1007/s11695-017-2924-y.

12. Peng S, Huo X, Rezaei D, Zhang Q, Zhang X, Yu C, Asanuma K, Cheng E, Pham TH, Wang DH, Chen M, Souza RF, Spechler SJ. In Barrett's esophagus patients and Barrett's cell lines, ursodeoxycholic acid increases antioxidant expression and prevents DNA damage by bile acids. Am J Physiol Gastrointest Liver Physiol 2014; 307:G129–G139. https://doi.org/10.1152/ajpgi.00085.2014.

13. Kim EK, Cho JH, Kim E, Kim YJ. Ursodeoxycholic acid inhibits the proliferation of colon cancer cells by regulating oxidative stress and cancer stem-like cell growth. PLoS One 2017; 12:e0181183. https://doi.org/10.1371/journal.pone.0181183.

14. Sutherland JD, MacDonald IA, Forrest TP. The enzymatic and chemical synthesis of ursodeoxycholic and chenodeoxycholic acid from cholic acid. Prep Biochem 1982; 12:307–321.

15. Mallonee DH, Hylemon PB. Sequencing and expression of a gene encoding a bile acid transporter from *Eubacterium* sp. strain VPI 12708. J Bacteriol 1996; 178:7053–7058. https://doi.org/10.1128/jb.178.24.7053-7058.1996.

16. Mallonee DH, Adams JL, Hylemon PB. The bile acid-inducible *baiB* gene from *Eubacterium* sp. strain VPI 12708 encodes a bile acidcoenzyme A ligase. J Bacteriol 1992; 174:2065–2071. https://doi.org/10.1128/jb.174.7.2065-2071.1992.

17. Ye HQ, Mallonee DH, Wells JE, Bjorkhem I, Hylemon PB. The bile acid-inducible *baiF* gene from *Eubacterium* sp. strain VPI 12708 encodes a bile acid-coenzyme A hydrolase. J Lipid Res 1999; 40:17–23.

18. Ridlon JM, Hylemon PB. Identification and characterization of two bile acid coenzyme A transferases from *Clostridium scindens*, a bile acid 7α-dehydroxylating intestinal bacterium. J Lipid Res 2012; 53:66–76. https://doi.org/10.1194/jlr.M020313.

19. Baron SF, Franklund CV, Hylemon PB. Cloning, sequencing, and expression of the gene coding for bile acid 7α-hydroxysteroid dehydrogenase from *Eubacterium* sp. strain VPI 12708. J Bacteriol 1991; 173: 4558–4569. https://doi.org/10.1128/jb.173.15.4558-4569.1991.

20. Dawson JA, Mallonee DH, Bjorkhem I, Hylemon PB. Expression and characterization of a C24 bile acid 7α-dehydratase from *Eubacterium* sp. strain VPI 12708 in *Escherichia coli*. J Lipid Res 1996; 37:1258–1267.

21. Bhowmik S, Chiu HP, Jones DH, Chiu HJ, Miller MD, Xu Q, Farr CL, Ridlon JM, Wells JE, Elsliger MA, Wilson IA, Hylemon PB, Lesley SA. Structure and functional characterization of a bile acid 7α dehydratase BaiE in secondary bile acid synthesis. Proteins 2016; 84:316–331. https://doi.org/10.1002/prot.24971.

22. Mallonee DH, Lijewski MA, Hylemon PB. Expression in *Escherichia coli* and characterization of a bile acid-inducible 3α-hydroxysteroid dehydrogenase from *Eubacterium* sp. strain VPI 12708. Curr Microbiol 1995; 30:259–263. https://doi.org/10.1007/BF00295498.

23. Bhowmik S, Jones DH, Chiu HP, Park IH, Chiu HJ, Axelrod HL, Farr CL, Tien HJ, Agarwalla S, Lesley SA. Structural and functional characterization of BaiA, an enzyme involved in secondary bile acid synthesis in human gut microbe. Proteins 2014; 82:216–229. https://doi.org/10.1002/prot.24353.

24. Hylemon PB, Melone PD, Franklund CV, Lund E, Bjorkhem I. Mechanism of intestinal 7α-dehydroxylation of cholic acid: evidence that allo-deoxycholic acid is an inducible side-product. J Lipid Res 1991; 32:89–96.

25. Harris SC, Devendran S, Alves JMP, Mythen SM, Hylemon PB, Ridlon JM. Identification of a gene encoding a flavoprotein involved in bile acid metabolism by the human gut bacterium *Clostridium scindens* ATCC 35704. Biochim Biophys Acta 2017; 1863:276–283. https://doi.org/10.1016/j.bbalip.2017.12.001.

26. Edenharder R, Schneider J. 12α-Dehydrogenation of bile acids by *Clostridium paraputrificum*, *C. tertium*, and *C. difficile* and epimerization at carbon-12 of deoxycholic acid by co-cultivation with 12α-dehydrogenating *Eubacterium lentum*. Appl Environ Microbiol 1985; 49:964–968.

27. Doerner KC, Takamine F, LaVoie CP, Mallonee DH, Hylemon PB. Assessment of fecal bacteria with bile acid 7α-dehydroxylating activity for the presence of *bai*-like genes. Appl Environ Microbiol 1997; 63:1185–1188.

28. Harris JN, Hylemon PB. Partial purification and characterization of NADP-dependent 12α-hydroxysteroid dehydrogenase from Clostridium leptum. Biochim Biophys Acta 1978; 528:148–157. https://doi.org/10.1016/0005-2760(78)90060-7.

29. Kisiela M, Skarka A, Ebert B, Maser E. Hydroxysteroid dehydrogenases (HSDs) in bacteria: a bioinformatic perspective. J Steroid Biochem Mol Biol 2012; 129:31– 46. https://doi.org/10.1016/j.jsbmb.2011.08.002.

30. MacDonald IA, Jellett JF, Mahony DE. 12α-Hydroxysteroid dehydrogenase from *Clostridium* group P strain C48-50 ATCC no. 29733: partial purification and characterization. J Lipid Res 1979; 20:234–239.

31. Aigner A, Gross R, Schmid R, Braun M, Mauer S. April 2011. Novel 12α-hydroxysteroid dehydrogenases, production and use thereof. US patent 20110091921A1.

32. Masuda N, Oda H. 7α-Dehydroxylation of bile acids by resting cells of an unidentified, Gram-positive, non-spore forming anaerobic bacterium. Appl Environ Microbiol 1983; 45:456–462.

33. Harris SC, Devendran S, Mendez-Garcia C, Mythen SM, Wright CL, Fields CJ, Hernandez AG, Cann I, Hylemon PB, Ridlon JM. Bile acid oxidation by *Eggerthella lenta* strains C592 and DSM 2243T. Gut Microbes 2018; 9:523-539. https://doi.org/10.1080/19490976.2018.1458180.

34. Mythen SM, Devendran S, Mendez-Garcia C, Cann I, Ridlon JM. Targeted synthesis and characterization of a gene cluster encoding NAD(P)H-dependent 3α-, 3α-, and 12α-hydroxysteroid dehydrogenases from *Eggerthella* CAG:298, a gut metagenomic sequence. Appl Environ Microbiol 2018; 84:e02475-17. https://doi.org/10.1128/AEM.02475-17.

35. Ditscheid B, Keller S, Jahreis G. Faecal steroid excretion in humans is affected by calcium supplementation and shows genderspecific differences. Eur J Nutr 2009; 48:22–30. https://doi.org/10.1007/s00394-008-0755-2.

36. Wegner K, Just S, Gau L, Mueller H, Gerard P, Lepage P, Clavel T, Rohn S. Rapid analysis of bile acids in different biological matrices using LC-ESI-MS/MS for the investigation of bile acid transformation by mammalian gut bacteria. Anal Bioanal Chem 2017; 409:1231–1245. https://doi.org/10.1007/s00216-016-0048-1.

37. Maruo T, Sakamoto M, Ito C, Toda T, Benno Y. *Adlercreutzia equolifaciens* gen. nov., sp. nov., an equol-producing bacterium isolated from human feces, and emended description of the genus *Eggerthella*. Int J Syst Evol Microbiol 2008; 58:1221–1227. https://doi.org/10.1099/ijs.0.65404-0.

38. MacDonald IA, Jellett JF, Mahony DE, Holdeman LV. Bile salt 3α- and 12α-hydroxysteroid dehydrogenases from *Eubacterium lentum* and related organisms. Appl Environ Microbiol 1979; 37:992–1000.

39. Hirano S, Masuda N. Transformation of bile acids by. Eubacterium lentum. Appl Environ Microbiol 1981; 42:912–915.

40. Edenharder R, Mielek K. Epimerization, oxidation and reduction of bile acids by *Eubacterium lentum*. Syst Appl Microbiol 1984; 5:287–298. https://doi.org/10.1016/S0723-2020(84)80031-4.

41. Bennett MJ, McKnight SL, Coleman JP. Cloning and characterization of the NAD-dependent 7α-hydroxysteroid dehydrogenase from *Bacteroides fragilis*. Curr Microbiol 2003; 47:475–484. https://doi.org/10.1007/s00284-003-4079-4.

42. Hylemon PB, Sherrod JA. Multiple forms of 7α-hydroxysteroid dehydrogenase in selected strains of *Bacteroides fragilis*. J Bacteriol 1975; 122:418–424.

43. MacDonald IA, Meier MC, Mahony DE, Costain GA. 3α-, 7α-, and 12α-hydroxysteroid dehydrogenase activities from *Clostridium perfringens*. Biochim Biophys Acta 1976; 450:142–153. https://doi.org/10.1016/0005-2760(76)90086-2.

44. Wells JE, Hylemon PB. Identification and characterization of a bile acid 7α-dehydroxylation operon in *Clostridium* sp. strain TO-931, a highly active 7α-dehydroxylating strain isolated from human feces. Appl Environ Microbiol 2000; 66:1107–1113. https://doi.org/10.1128/AEM.66.3.1107-1113.2000.

45. Ridlon JM, Kang DJ, Hylemon PB. Isolation and characterization of a bile acid 7α-dehydroxylating operon in *Clostridium hylemonae* TN271. Anaerobe 2010; 16:137–146. https://doi.org/10.1016/j.anaerobe.2009.05.004.

46. Setchell KDR, Lawson AM, Tanida N, Sjovall J. General methods for the analysis of metabolic profiles of bile acids and related compounds in feces. J Lipid Res 1983; 24:1085–1100.

47. Reddy BS, Sharma C, Simi B, Engle A, Laakso K, Puska P, Korpela R. Metabolic epidemiology of colon cancer: effect of dietary fiber on fecal mutagens and bile acids in healthy subjects. Cancer Res 1987; 47:644–648.

48. Mai V, Katki HA, Harmsen H, Gallaher D, Schatzkin A, Baer DJ, Clevidence B. Effects of a controlled diet and black tea drinking on the fecal microflora composition and the fecal bile acid profile of human volunteers in a double-blinded randomized feeding study. J Nutr 2004; 134: 473–478. https://doi.org/10.1093/jn/134.2.473.

49. Begley M, Gahan CG, Hill C. The interaction between bacteria and bile. FEMS Microbiol Rev 2005; 29:625– 651. https://doi.org/10.1016/j.femsre.2004.09.003.

50. Hofmann AF, Roda A. Physicochemical properties of bile acids and their relationship to biological properties: an overview of the problem. J Lipid Res 1984; 25:1477–1489.

51. Devlin AS, Fischbach MA. A biosynthetic pathway for a prominent class of microbiota-derived bile acids. Nat Chem Biol 2015; 11:685– 690. https://doi.org/10.1038/nchembio.1864.

52. Watanabe M, Fukiya S, Yokota A. Comprehensive evaluation of the bactericidal activities of free bile acids in the large intestine of humans and rodents. J Lipid Res 2017; 58:1143–1152. https://doi.org/10.1194/jlr.M075143.

53. Albenberg L, Esipova TV, Judge CP, Bittinger K, Chen J, Laughlin A, Grunberg S, Baldassano RN, Lewis JD, Li H, Thom SR, Bushman FD, Vinogradov SA, Wu GD. Correlation between intraluminal oxygen gradient and radial partitioning of intestinal microbiota. Gastroenterology 2014; 147:1055.e8–1063.e8. https://doi.org/10.1053/j.gastro.2014.07.020.

54. Brown JM, Hazen SL. Microbial modulation of cardiovascular disease. Nat Rev Microbiol 2018; 16:171–181. https://doi.org/10.1038/nrmicro.2017.149.

55. Trifunovic J, Borcic V, Mikov M. Bile acids and their oxo derivatives: potential inhibitors of carbonic anhydrase I and II, androgen receptor antagonists and CYP3A4 substrates. Biomed Chromatogr 2017; 31:e3870. https://doi.org/10.1002/bmc.3870.

56. Hofmann AF, Sjovall J, Kurz G, Radominska A, Schteingart CD, Tint GS, Vlahcevic ZR, Setchell KDR. A proposed nomenclature for bile acids. J Lipid Res 1992; 33:599–604.

57. Eneroth P. Thin-layer chromatography of bile acids. J Lipid Res 1963; 4:11–16.

58. Wall PE. Thin-layer chromatography: a modern practical approach. Royal Society of Chemistry, Cambridge, United Kingdom. 2007.

59. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 2004; 5:113. https://doi.org/10.1186/1471-2105-5-113.

60. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximumlikelihood trees for large alignments. PLoS One 2010; 5:e9490. https://doi.org/10.1371/journal.pone.0009490.

61. Huson DH, Scornavacca C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst Biol 2012; 61:1061–1067. https://doi.org/10.1093/sysbio/sys062.

CHAPTER 3

COMPLETION OF THE GUT MICROBIAL EPI-BILE ACID PATHWAY[1]

ABSTRACT

Bile acids are detergent molecules that solubilize dietary lipids and lipid-soluble vitamins. Humans synthesize bile acids with α-orientation hydroxyl groups which can be biotransformed by gut microbiota to toxic, hydrophobic bile acids, such as deoxycholic acid (DCA). Gut microbiota can also convert hydroxyl groups from the α-orientation through an oxo-intermediate to the β-orientation, resulting in more hydrophilic, less toxic bile acids. This interconversion is catalyzed by regio- (C-3 vs. C-7) and stereospecific (α vs. β) hydroxysteroid dehydrogenases (HSDHs). So far, genes encoding the urso- (7α-HSDH & 7β-HSDH) and iso-(3α-HSDH & 3β-HSDH) bile acid pathways have been described. Recently, multiple human gut clostridia were reported to encode 12α-HSDH, which interconverts DCA and 12-oxolithocholic acid (12-oxoLCA). 12β-HSDH completes the epi-bile acid pathway by converting 12-oxoLCA to the 12β-bile acid denoted epiDCA; however, gene(s) encoding this enzyme have yet to be identified. We confirmed 12β-HSDH activity in cultures of *Clostridium paraputrificum* ATCC 25780. From six candidate *C. paraputrificum* ATCC 25780 oxidoreductase genes, we discovered the first gene (DR024_RS09610) encoding bile acid 12β-HSDH. Phylogenetic analysis revealed unforeseen diversity for 12β-HSDH, leading to validation of two additional bile acid 12β-HSDHs through a synthetic biology approach. By comparison to a previous phylogenetic

---

analysis of 12α-HSDH, we identified the first potential C-12 epimerizing strains: *Collinsella tanakaei* YIT 12063 and *Collinsella stercoris* DSM 13279. A Hidden Markov Model search against human gut metagenomes located putative 12β-HSDH genes in about 30% of subjects within the cohorts analyzed, indicating this gene is relevant in the human gut microbiome.

INTRODUCTION

The human liver produces all 14 enzymes necessary to convert cholesterol into the dihydroxy bile acid chenodeoxycholic acid (3α,7α-dihydroxy-5β-cholan-24-oic acid; CDCA) and the trihydroxy bile acid cholic acid (3α,7α,12α-trihydroxy-5β-cholan-24-oic acid; CA).[1] These bile acids are conjugated to taurine or glycine in the liver helping to lower the *pK$_a$* and maintain solubility, impermeability to cell membranes, and lower the critical micellar concentration, allowing for efficient emulsification of dietary lipids and lipid-soluble vitamins.[2] Bile acids are effective detergents owing to the α-orientation of the hydroxyl groups which produce a hydrophilic-face above the plane of the cyclopentanephenanthrene steroid nucleus, and a hydrophobic-face below the plane of the hydrocarbon rings.[1] Conjugated bile acids emulsify lipids throughout the duodenum, jejunum, and ileum. Once bile acids reach the terminal ileum, high affinity transporters (intestinal bile acid transporter, IBAT) actively transport both conjugated and unconjugated bile acids from the intestinal lumen into ileocytes where they are bound to ileal bile acid binding protein (IBABP) and exported across the basolateral membrane into portal circulation and returned to the liver.[3] This process of recycling of bile acids is known as the enterohepatic circulation (EHC), responsible for recirculating the ~2 g bile acid pool 8-10 times daily. While ~95% efficient, roughly 600-800 mg bile acids escape active transport and enter the large intestine.[2]

Anaerobic bacteria adapted to inhabiting the large intestine have evolved enzymes to modify the structure of host bile acids.[2] Conjugated bile acids are hydrolyzed, releasing the amino acids, by bile salt hydrolases (BSH) in diverse gut bacteria representing the major phyla, including Bacteroidetes, Firmicutes, and Actinobacteria as well as the domain Archaea.[4] By contrast, the unconjugated primary bile acids CA and CDCA are 7α-dehydroxylated by a select few species of gram-positive Firmicutes mostly in the genus *Clostridium*, forming deoxycholic acid (3α,12α-dihydroxy-5β-cholan-24-oic acid; DCA) and lithocholic acid (3α-hydroxy-5β-cholan-24-oic acid; LCA), respectively.[1,5] The secondary bile acids DCA and LCA have increased hydrophobicity relative to their primary counterparts, which is associated with elevated toxicity.[6] DCA and LCA have been causally linked to cancers of the colon[7], liver[8], and esophagus[9]. Importantly, gut microbiota can produce less toxic oxo-bile acids and β-hydroxy bile acids as well.[6]

Bile acid 3α-, 7α-, and 12α-hydroxyl groups can be reversibly oxidized and epimerized to the β-orientation by pyridine nucleotide-dependent hydroxysteroid dehydrogenases (HSDHs) distributed across the major phyla including Firmicutes, Bacteroidetes, Actinobacteria, Proteobacteria, as well as methanogenic archaea.[1,10] HSDH enzymes that recognize bile acids are regio- (C-3 vs. C-7) and stereospecific (α vs. β) for hydroxyl groups decorating the steroid nucleus. Thus, bile acid 12α-HSDH reversibly converts the C-12 position of bile acids from the α-orientation, such as on DCA, to 12-oxo bile acids, such as 12-oxolithocholic acid (12-oxoLCA).[11–14] Bile acid 12β-HSDH completes the epimerization by interconverting 12-oxo bile acids to the 12β-configuration, forming epi-bile acids. We recently identified and characterized NAD(H)- and NADP(H)-dependent 12α-HSDHs from *Eggerthella* sp. CAG:298[15], *Clostridium scindens, C. hylemonae*, and *Peptacetobacter hiranonis* (formerly *Clostridium hiranonis*)[10]. In

addition to these recently reported 12α-HSDHs, multiple genes encoding enzymes in the urso-
(7α- & 7β-HSDH) and iso- (3α- & 3β-HSDH) bile acid pathways have been described to date
(**Figure 3.1**).[5] However, a gene encoding 12β-HSDH to complete the epi-bile acid pathway has
not yet been reported.

The first indication that gut bacteria may encode 12β-HSDH was suggested by the
detection of 12β-hydroxy bile acids in human feces.[16–18] Edenharder and Schneider (1985)
reported 12β-dehydrogenation of bile acids by *Clostridium paraputrificum*, and epimerization of
DCA by co-culture with *E. lenta* and *C. paraputrificum*.[19] Thereafter, Edenharder and Pfützner
(1988) characterized crude NADP(H)-dependent 12β-HSDH activity from *C. paraputrificum* D
762-06.[20] However, little is known about the potential diversity of gut bacteria capable of
forming 12β-hydroxy bile acids that molecular analysis is predicted to yield. Here, we report the
identification of a gene encoding NADP(H)-dependent 12β-HSDH from *C. paraputrificum*
ATCC 25780 and characterization of the recombinant gene products purified after heterologous
expression in *E. coli* from *C. paraputrificum.* We also identify novel taxa encoding bile acid 12β-
HSDH by phylogenetic analysis, confirmed by a synthetic biology approach.

RESULTS

C. paraputrificum ATCC 25780 possesses bile acid 12β-HSDH activity

We first investigated the bile acid metabolizing capability of *C. paraputrificum* ATCC
25780 because previous studies reported bile acid 12β-HSDH activity in other *C. paraputrificum*
strains, but did not identify the gene(s) responsible.[20] The epi-bile acid pathway of DCA involves
the reversible conversion of DCA (3α,12α) to 12-oxoLCA (3α,12-oxo) through the action of
12α-HSDH, and 12-oxoLCA to epiDCA (3α,12β) by 12β-HSDH (**Figure 3.1**). *C. paraputrificum*

ATCC 25780 was incubated with two potential substrates of 12β-HSDH, 12-oxoLCA and

epiDCA, along with DCA as a control. In order to contrast the product formed by bile acid 12β-

HSDH with that formed by bile acid 12α-HSDH activity, *Clostridium scindens* ATCC 35704,

which is known to express 12α-HSDH, was incubated with the same substrates. When 12-

oxoLCA was incubated in cultures of *C. paraputrificum* ATCC 25780, the primary product

eluted at 13.97 min with 391.28 m/z in negative ion mode (**Figure 3.2**). This is consistent with

the elution time of epiDCA standard and its 392.57 amu formula weight. With epiDCA as

substrate, the culture produced a major peak of 391.28 at 13.96 min and a minor peak of 389.27

m/z at 14.34 min, which suggests epiDCA was not converted in large quantities to 12-oxoLCA

(390.56 amu). *C. paraputrificum* incubation with DCA did not result in detectable formation of

12-oxoLCA or epiDCA products. Taken together, these data demonstrate *C. paraputrificum*

ATCC 25780 expresses bile acid 12β-HSDH activity, but not bile acid 12α-HSDH. *C. scindens*

ATCC 35704 incubation with 12-oxoLCA resulted in a main product (15.57 min and 391.28

m/z) consistent with DCA (392.57 amu), demonstrating bile acid 12α-HSDH activity. In

addition, reaction with DCA resulted in a peak at 15.57 min and 391.28 m/z along with a peak

agreeing with 12-oxoLCA at 14.34 min and 389.27 m/z. When epiDCA was incubated with

cultures of *C. scindens* ATCC 35704, we did not observe formation of 12-oxoLCA.

Identification of a gene encoding bile acid 12β-HSDH

After bile acid 12β-HSDH activity was confirmed in *C. paraputrificum* ATCC 25780, its

genome was searched for genes encoding proteins annotated as oxidoreductases within the NCBI

database. HSDHs are NAD(P)-dependent and often members of the large and diverse SDR

(short-chain dehydrogenase/reductase) family.[21] Five SDR family oxidoreductase proteins and

one aldo/keto reductase were identified as 12β-HSDH candidates in the *C. paraputrificum* ATCC 25780 genome and pursued for further study. These six genes were amplified from genomic DNA of *C. paraputrificum* ATCC 25780, cloned into the pET-28a(+) vector, and overexpressed in *E. coli* (**Supplementary Table 3.1**). The N-terminal His$_6$-tagged recombinant proteins were purified by metal-affinity chromatography and resolved by SDS-PAGE (**Figure 3.3A**).

Two out of the six recombinant proteins (WP_027096909.1, WP_027099631.1) were not soluble and bands at the expected molecular masses were apparent in the membrane fraction by SDS-PAGE. These proteins were not explored further. The other four 12β-HSDH candidates (WP_027099077.1, WP_027098355.1, WP_027097937.1, WP_027098604.1) were soluble and visualized by SDS-PAGE. The four soluble recombinant proteins were then screened for pyridine nucleotide-dependent bile acid 12β-HSDH activity by TLC and spectrophotometric assay. Screening reactions were prepared with 12-oxoLCA and NADPH, or epiDCA and NADP$^+$ in pH 7.0 phosphate buffer.

Only WP_027099077.1 exhibited 12β-HSDH activity by TLC and spectrophotometric assay, which was also confirmed by LC-MS (**Figure 3.3B**). Reaction products of WP_027099077.1 with 12-oxoLCA and NADPH, epiDCA and NADP$^+$, DCA and NADP$^+$, and no substrate control were subjected to LC-MS. In the presence of purified recombinant WP_027099077.1 and NADPH, 12-oxoLCA was reduced quantitatively (2 hydrogen addition) to a product that eluted at 13.12 min with 391.28 m/z in negative ion mode. This is consistent with the 392.57 amu formula weight and elution time for epiDCA based on the substrate standard. Additionally, epiDCA was oxidized to a product with an elution time of 13.40 min at 389.27 m/z, agreeing with the retention time and formula weight of 390.56 amu for authentic 12-oxoLCA. DCA (392.57 amu) was not converted by WP_027099077.1 as the sole peak observed

matched DCA standard at 14.60 min and 391.29 m/z. The interconversion of 12-oxoLCA and epiDCA, but no activity with DCA, indicates stereospecificity for the 12β-hydroxy position. Thus, DR024_RS09610 has been identified as the first gene reported that encodes bile acid 12β-HSDH (WP_027099077.1).

Recombinant *C. paraputrificum* WP_027099077.1, hereafter referred to as Cp12β-HSDH, had a theoretical subunit molecular mass of 27.4 kDa. The observed subunit molecular mass was 26.4 ± 0.5 kDa by SDS-PAGE, calculated from three independent protein gels. WP_027099077.1 is predicted to be a cytosolic protein that is not membrane-associated by TMHMM v. 2.0.[22]

Biochemical characterization of recombinant Cp12β-HSDH

The approximate native molecular mass of Cp12β-HSDH was determined by size-exclusion chromatography. Cp12β-HSDH exhibited an elution volume of 15.04 ± 0.02 mL, corresponding to a 54.67 ± 0.79 kDa molecular mass relative to protein standards (**Figure 3.4A**). The size-exclusion data along with the theoretical subunit molecular mass of 27.4 kDa suggests Cp12β-HSDH assembles a homodimeric quaternary structure in solution. In order to optimize the enzymatic activity of Cp12β-HSDH, the conversion of pyridine nucleotides at 340 nm was measured in buffers between pH 6.0 and 8.0 by spectrophotometric assay (**Figure 3.4B**). The optimum pH for Cp12β-HSDH in the oxidative direction with epiDCA as the substrate and NADP$^+$ as co-substrate was pH 7.5. In the reductive direction where 12-oxoLCA was the substrate and NADPH the co-substrate, the optimum pH was 7.0.

Michaelis-Menten kinetics were performed at the pH optimum for each direction. In the reductive direction, Cp12β-HSDH displayed a K_m value for 12-oxoLCA at 18.76 ± 0.40 µM

which was similar to that of NADPH (**Table 3.1**; **Supplementary Figure 3.1**). The K_m value in the oxidative direction with epiDCA as substrate was about twice the K_m determined for 12-oxoLCA. The K_m for NADP$^+$ was 36.84 ± 0.55 µM. The V_{max} and k_{cat} were greater in the oxidative than the reductive direction. However, the catalytic efficiency (k_{cat}/K_m) of 12-oxoLCA as substrate was greater than the oxidative direction with epiDCA as substrate.

Pyridine nucleotide cofactor and bile acid substrate-specificity of Cp12β-HSDH were determined by relative activity compared to either 12-oxoLCA or epiDCA through spectrophotometric assay (**Table 3.2**). NAD$^+$ and NADH were not co-substrates for Cp12β-HSDH. DCA (3α,12α) as well as CA (3α,7α,12α) were not substrates, which is expected because they are 12α-hydroxy bile acids not 12β-hydroxy bile acids. CDCA (chenodeoxycholic acid; 3α,7α) lacks a 12-hydroxyl group, and as expected was not a substrate. The CA derivatives 12-oxoCDCA (3α,7α,12-oxo) and epiCA (3α,7α,12β) had ~12% and 27% activity, respectively, relative to bile acids lacking a 7α-hydroxyl group. The activity of 3,12-dioxoLCA was ~19% compared to 12-oxoLCA. Altogether, the results suggest Cp12β-HSDH is specific for NADP(H) and favors 12-oxoLCA and epiDCA over their 7α-hydroxy counterparts.

Phylogenetic analysis of Cp12β-HSDH

The Cp12β-HSDH sequence from *C. paraputrificum* ATCC 25780 (WP_027099077.1) was used in a BLASTP search against the NCBI non-redundant protein database in order to determine its prevalence across bacteria. A maximum likelihood phylogeny of 5,000 sequences was constructed, revealing that many sequences most similar to Cp12β-HSDH are found in Firmicutes and Actinobacteria (**Supplementary Figure 3.2**). Within the 5,000-member phylogeny, a subtree (highlighted grey) of the most closely related proteins to Cp12β-HSDH was

selected for closer inspection (**Figure 3.5**). Cp12β-HSDH clustered most closely with other *C. paraputrificum* sequences (WP_099327725, WP_049179624, WP_111937163). These sequences are encoded by *C. paraputrificum s*trains isolated from preterm infants, namely strain LH025, LH141, and LH058[23], or isolated from feces (Gcol.A11).[24]

Firmicutes harbor the majority of sequences within the 12β-HSDH subtree, spanning genera including *Eisenbergiella*, *Ruminococcus*, and *Coprococcus*. To determine if other organisms within the tree have bile acid 12β-HSDH activity, the gene encoding WP_118677302.1 from *Eisenbergiella* sp. OF01-20 was synthesized by IDT in the codon-usage of *E. coli* (**Supplementary Table 3.1**), cloned into pET-28a(+), overexpressed in *E. coli*, and purified by affinity chromatography (**Figure 3.5**). Recombinant WP_118677302.1 was screened by spectrophotometric assay with NAD(P)(H) against 12-oxoLCA, epiDCA, and DCA. Relative to Cp12β-HSDH, WP_118677302.1 exhibited 88% activity with 12-oxoLCA, and 83% activity with epiDCA. WP_118677302.1 did not show conversion of DCA, confirming that this enzyme has bile acid 12β-HSDH activity (**Table 3.2**).

The subtree also contains many sequences from Actinobacteria, the genera *Collinsella* and *Olsenella* among them. C*ollinsella* species are of interest because *C. aerofaciens* expresses BSH and various HSDHs recognizing sterols[25], including bile acid 12α-HSDH.[26] To determine if a member of the Actinobacteria encodes a bile acid 12β-HSDH, a sequence more distantly related to Cp12β-HSDH, *Olsenella* sp. GAM18 WP_120179297.1, was chosen for gene synthesis and protein overexpression because it had not been shown previously to metabolize bile acids (**Figure 3.5**). 12-oxoLCA, epiDCA and DCA were tested as substrates and conversion was measured by spectrophotometric assay. Recombinant WP_120179297.1 displayed activity with 12-oxoLCA at 128% relative to Cp12β-HSDH, 69% relative activity with epiDCA, and

showed no reaction with DCA (**Table 3.2**). These data confirm that the more distantly related WP_120179297.1 has bile acid 12β-HSDH activity.

Within the extended subtree are various *Novosphingobium s*pecies. These Alphaproteobacteria deserve mention due to their ability to biodegrade aromatic compounds, such as phenanthrene[27] and estrogen.[28] To test if this cluster has bile acid 12β-HSDH activity, WP_007678535.1 from *Novosphingobium* sp. AP12 was synthesized, cloned, overexpressed, and purified (**Figure 3.5**). The potential 12β-HSDH activity of WP_007678535.1 was screened using 12-oxoLCA, epiDCA, and DCA as substrates. WP_007678535.1 exhibited no activity with these bile acid substrates (**Table 3.2**). Because *Novosphingobium s*trains are frequently plant-associated or isolated from aquatic environments[29], this enzyme may be specific for other substrates.

The genomic context of 12β-HSDH genes from *C. paraputrificum* ATCC 25780, *Eisenbergiella* sp. OF01-20, and *Olsenella* sp. GAM18 was explored (**Supplementary Figure 3.3**). The three 12β-HSDH genes did not appear to be organized within an operon nor was the genomic context conserved across these organisms.

Two organisms present in the 12β-HSDH subtree, *Collinsella tanakaei* and *Collinsella stercoris* (**Figure 3.5**, asterisks), were also found in a phylogenetic analysis of putative 12α-HSDHs (see Chapter 2).[10] Due to strain variation within species, we inspected the sequences further and determined that the pairs of HSDHs are encoded by the same strain within each species. *Collinsella tanakaei* YIT 12063 12α-HSDH (WP_009141301.1) and 12β-HSDH (WP_009140706.1) are encoded by the genes HMPREF9452_RS06335 and HMPREF9452_RS03390, respectively. *Collinsella stercoris* DSM 13279 also contains both putative 12α-HSDH (WP_040360544.1; COLSTE_RS02900) and 12β-HSDH

(WP_006720039.1; COLSTE_RS01465).[10] While the paired activity of 12α/12β-HSDH has not been tested in culture, these organisms may be novel epi-epimerizing strains that convert bile acid 12α-hydroxyl groups to the epi-configuration. To our knowledge, these are the first strains identified with C-12 epimerizing ability.

Hidden Markov Model search of putative 12β-HSDH genes in human gut metagenomes

To understand the distribution of potential 12β-HSDH genes in the human colonic microbiome, a Hidden Markov Model (HMM) search was performed against metagenome assembled genomes (MAGs) from four publicly available cohorts[30–33] using reference sequences from the 12β-HSDHs characterized in this paper (**Figure 3.5**). Putative 12β-HSDH genes inferred by HMM search were found in ~30% of the subjects (198/666) (**Figure 3.6A**). Twenty-two subjects exhibited two different organisms containing the gene. This gene was found in healthy subjects as well as subjects with the following disease states: colorectal cancer, colorectal adenoma, fatty liver, hypertension, and type 2 diabetes.

Two hundred twenty microbial genomes contained putative 12β-HSDH genes among 16,936 total available genomes. Putative 12β-HSDH genes were most often identified in the phylum Firmicutes, which was dominated by genes in *Lachnospira eligens* (formerly *Eubacterium eligens*) (**Figure 3.6B**). The gene from *L. eligens* was widespread across subjects in each of the four cohorts. This large proportion of hits from *L. eligens* may reflect its higher relative abundance allowing it to be assembled better into genomes. Sequences from this organism also appeared multiple times in the 12β-HSDH subtree (**Figure 3.5**). *Lachnospira eligens* is a pectin degrader capable of promoting anti-inflammatory cytokine IL-10 production *in vitro*[34] and has been proposed as a probiotic for atherosclerosis[35]. The gene was also present in

C. paraputrificum along with other unidentified *Clostridium* sp. and *Eubacterium* sp. Actinobacteria had few members with the gene, represented by *Collinsella intestinalis*, *Collinsella tanakaei*, and *Olsenella* sp. *Phocaeicola coprocola* (formerly *Bacteroides coprocola*) was the only member of phylum Bacteroidetes with the gene.

Phylogenetic analysis of regio- and stereospecific HSDHs

Next, the phylogenetic relationship between Cp12β-HSDH (WP_027099077.1) and other regio- and stereospecific HSDHs was explored. To accomplish this, we updated the HSDH phylogeny presented by Mythen et al. (2018) by including additional bacterial or archaeal HSDH sequences of known or putative function along with representative eukaryotic sequences (**Figure 3.7**; **Supplementary Table 3.2**).[15] The sequences included span the known HSDH functional capacities, with some recognizing bile acids and others recognizing steroids like cortisol or progesterone. Most members of each HSDH class are clustered together, which is apparent by each highlight color encompassing more than one HSDH of the same known function. Furthermore, most bacterial HSDHs grouped separately from their eukaryotic counterparts.

Prokaryotic sequences were interspersed among the eukaryotic with some exceptions in grouping by HSDH function. Cp12β-HSDH, the two other confirmed bile acid 12β-HSDHs (WP-118677302.1, WP_120179297.1), and additional similar sequences from across our bile acid 12β-HSDH subtree formed their own cluster. These sequences shared a branch with bacterial bile acid 12α-HSDHs as well as eukaryotic 3β-HSD/$\Delta^5 \rightarrow \Delta^4$/isomerases. Bile acid 12α-HSDH sequences included various clostridia (EDS06338.1, EEG75500.1, EEA85268.1, ERJ00208.1)[10,36] and *Eggerthella* (CDD59475.1).[15] *Collinsella aerofaciens* (EBA39192.1), which has been reported to express bile acid 12α-HSDH activity[26], grouped with the known bile

acid 12α-HSDHs along with two human gut archaeal sequences from *Methanosphaera stadtmanae* and *Methanobrevibacter smithii*.

Clostridial (gram-positive) bile acid 7α-HSDHs (AAB61151.1 etc.)[37] clustered separately from those expressed by *E. coli* (BAA01384.1)[38] and those predicted in *Bacteroides* (gram-negative), similarly to the Mythen et. al. (2018) phylogeny. Bile acid 7β-HSDHs did not closely cluster with other classes of bacterial HSDHs. Instead, the nearest neighbours to the three known bile acid 7β-HSDHs (WP_006236005.1, WP_004843516.1, AET80684.1)[25,39,40] included in this tree were eukaryotic steroid 11β- and 17β-HSDs.

Bacterial 3α-HSDHs clustered together, excluding one outlier from *Eggerthella lenta* (ACV54671.1).[41] Within the bile acid 3α-HSDH group, four enzymes predicted with this function formed a separate branch apart from the confirmed bile acid 3α-HSDHs. Likewise, three known bile acid 3β-HSDHs grouped together (ACV55294.1, ACV54192.1, EDN78833.1)[41], while one *Eggerthella* sequence was most closely related to putative bile acid 3β-HSDHs from *Lactobacillus* spp. identified by BLAST search in Mythen et. al (2018).

Bacterial steroid 20β-HSDHs convert the glucocorticoid cortisol to 20β-dihydrocortisol. Two experimentally confirmed 20β-HSDHs (WP_003810233.1, WP_051643274.1) (see Chapter 4)[42,43] grouped with putative sequences from both gut and urinary tract isolates. To date, only one steroid 20α-HSDH sequence, which interconverts cortisol and 20α-dihydrocortisol, has been reported (*C. scindens* EDS07887.1) (see Chapter 5).[44,45] Therefore, we performed a BLASTP search and found two sequences with high similarity, WP_145772308.1 from *Denitratisoma oestradiolicum* DSM 16959 and WP_107631222.1 from *Intestinibacillus* sp. Marseille-P4005. *D. oestradiolicum* DSM 16959 was isolated from sludge in a municipal wastewater treatment plant and can use 17β-estradiol as a sole carbon and energy source.[46] *Intestinibacillus*

massiliensis strain Marseille-P3216, a close relative to *Intestinibacillus* sp. Marseille-P4005 found in our tree, was isolated from the human colon and is most closely related to the species *Butyricicoccus desmolans* (formerly *Eubacterium desmolans*) by 16S rRNA gene sequence.[47] Interestingly, *B. desmolans* ATCC 43058 encodes a 20β-HSDH (WP_051643274.1).[43]

Eukaryotic HSDH sequences, typically denoted HSD, were spread throughout the phylogeny, but generally grouped with like sequences. The 17β- and 11β-HSD sequences did not form a group, instead clustering by type. For example, *Homo sapiens* 11β-HSD type 1 (NP_861420.1) was closely related to *Rattus norvegicus* 11β-HSD type 1 (NP_058776.2) and more distantly related to *Homo sapiens* 11β-HSD type 2 (NP_000187.3). 11β-HSD type 1 and type 2 both interconvert steroids between active and inactive forms, such as cortisol and cortisone.[48] However, 11β-HSD type 1 primarily acts as a reductase in many tissues while 11β-HSD type 2 functions as a dehydrogenase.[48]

DISCUSSION

Microbial bile acid HSDHs have been studied since the early 1970s, with much of the original work focusing on 3α- and 7α-HSDHs.[14,49] Thereafter, 3β-, 7β- and 12α-HSDH activity was observed in cultures of various microbiota[1], including *Eggerthella lenta* (formerly *Eubacterium lentum*) which is capable of oxidizing CA and DCA at C-12 and epimerizing bile acids at C-3.[50] In the mid-1980s, *C. paraputrificum*, *C. tertium*, and *Clostridioides difficile* each in binary cultures with *E. lenta* were shown to epimerize DCA via a 12-oxo-intermediate to epiDCA.[19] Since then, HSDH genes encoding the iso- and urso-bile acid pathways and 12α-HSDH were identified, but not 12β-HSDH.[5] In this work, we identified the first bile acid 12β-HSDH gene, completing the microbial epi-bile acid pathway.

Edenharder & Pfützner (1988) initially characterized NADP(H)-dependent 12β-HSDH from crude extracts of the fecal isolate *C. paraputrificum* strain D 762-06, with differing results from our findings.[20] Gel filtration analysis of crude extract from *C. paraputrificum* strain D 762-06 suggested a molecular mass of 126 kDa; whereas our current work with Cp12β-HSDH from ATCC 25780 is estimated at 54.6 KDa by gel filtration chromatography. The strain used in this study, *C. paraputrificum* ATCC 25780, was also isolated from feces.[51] It is possible that these are the same NADP(H)-dependent enzymes by sequence from two different strains of *C. paraputrificum* and that the recombinant protein quaternary structure is unstable, resulting in a dimeric form in our study. Alternatively, these bacterial strains may have distinct versions of 12β-HSDH with different amino acid sequences, as we have shown previously for 12α-HSDH from *Eggerthella lenta*.[15,52] Indeed, the 12β-HSDH from *C. paraputrificum* strain D 762-06 was reported to be partially membrane associated, whereas hydropathy prediction by TMHMM v. 2.0 found no evidence of transmembrane domains in Cp12β-HSDH. In addition, pH optima for the conversion of 12-oxoLCA between Cp12β-HSDH (7.0) and the native 12β-HSDH (10.0) from strain D 762-06 differed. Oxidation of epiDCA was optimal at pH 7.5 for Cp12β-HSDH, and reported as pH 7.8 for the crude native enzyme from strain D 762-06.[20] Further work will be needed to determine if distinct bile acid 12β-HSDHs are present in *C. paraputrificum* strains.

Cp12β-HSDH exhibited a dimeric quaternary structure by size-exclusion chromatography under our experimental conditions. Although future crystallization of Cp12β-HSDH would better illustrate its true polymeric state, HSDHs are often either tetrameric[42,53] or dimeric.[54,55] Cp12β-HSDH was more specific for bile acids lacking a position 7-hydroxyl group: epiDCA and 12-oxoLCA, over epiCA and 12-oxoCDCA. Cp12β-HSDH also had lower activity with 3,12-dioxoLCA versus 12-oxoLCA. This indicates that both the 7-hydroxyl and 3-oxo groups hinder

the ability of Cp12β-HSDH to convert the substrate. An x-ray crystal structure of Cp12β-HSDH may shed light on why this apparent steric hindrance occurs.

Phylogenetic analysis of Cp12β-HSDH coupled with synthetic biological "sampling" and validation at different points along the branches revealed shared 12β-HSDH function among *Eisenbergiella* sp. OF01-20 and *Olsenella* sp. GAM18, lending functional credibility to sequences throughout the subtree (**Figure 3.5**; **Table 3.2**). *Eisenbergiella* sp. OF01-20 was originally sequenced from a human gut microbiota cultivation project (Integrated Microbial Genomes [IMG] Genome ID: 2840324701). *Eisenbergiella* spp. are often present at relative abundances of less than 0.1% in human fecal samples.[56,57] *Olsenella* sp. GAM18 was initially isolated from humans (IMG Genome ID: 2841219092). The relative abundance of *Olsenella* was shown to be about 2% within the gut microbiome of some individuals.[58] Our subtree includes more abundant gut taxa such as *Ruminococcus* (relative abundance ~5%)[59,60] and *Collinsella* (relative abundance ~8%)[59], as well. Due to limitations in 16S rDNA sequencing depth, it is difficult to conclude if the species in our subtree are found at relevant levels in the human gut or if 12β-HSDH genes are present. Therefore, we performed a HMM search to assess the relative prevalence of 12β-HSDH genes. About 30% of subjects had putative 12β-HSDH genes, indicating the relevance of this gene in the human gut microbiome. The HMM search revealed that 220 microbial genomes out of 16,936 total contained putative 12β-HSDH genes. While concrete prevalence is difficult to establish, putative 12β-HSDH genes are less widespread than the ubiquitous bile-acid metabolizing gene, bile salt hydrolase[4], which was present in 2,456/16,936 total genomes in these cohorts. These data expand the limited metagenomic work that has focused on bile acid HSDH genes in the human gut.[61]

Two organisms from our 12β-HSDH subtree were also identified in a 12α-HSDH phylogeny from Doden et al. (2018) (see Chapter 2). Putative proteins WP_009140706.1 (12β-HSDH) and WP_009141301.1 (12α-HSDH) are both present in *Collinsella tanakaei* YIT 12063 and are encoded by the genes HMPREF9452_RS03390 and HMPREF9452_RS06335, respectively. Similarly, *Collinsella stercoris* DSM 13279 encodes both putative 12β-HSDH (WP_006720039.1; COLSTE_RS01465) and 12α-HSDH (WP_040360544.1; COLSTE_RS02900).[10] Although the dual 12α/12β-HSDH activity is untested in culture, we predict these strains are novel C-12 epimerizers. Epimerizing strains have been identified for the C-3[19,41] and C-7 hydroxyl[1,62] positions, however, this is the first indication of bacteria capable of C-12 epimerization.

The sequence WP_007678535.1 from *Novosphingobium* sp. AP12, whose recombinant enzyme product did not exhibit bile acid 12β-HSDH activity with the substrates tested, may be specific for aerobic bile acid degradation products. Environmental microorganisms, such as *Comamonas testosteroni* TA441 and *Pseudomonas* sp. strain Chol1, encode a CA degradation pathway involving conversion of a 12-oxo-intermediate to 7α,12β-dihydroxy-androsta-1,4-diene-3,17-dione (12β-DHADD).[63,64] Thus, sequences in the extension of the subtree may have 12β-HSDH activity, but with specificity for side-chain cleaved steroids rather than bile acids.

Indeed, this function joins the vast repertoire of HSDHs already studied in many Firmicutes and Actinobacteria.[1] Bile acid 12α-HSDH activity has been detected in *Eggerthella* species[15,52] in the phylum Actinobacteria and various clostridia[10–12,36] in the phylum Firmicutes. Similarly, 3α- and 3β-HSDH are widespread among Firmicutes[1,65], and 3α-HSDH has also been reported in *Eggerthella* species.[13,15,41] 7α- and 7β-HSDH were shown in numerous Firmicutes[14,37,65] and the Actinobacteria *Collinsella aerofaciens*.[25] Along with these bile acid-

specific HSDHs, the glucocorticoid 20α- and 20β-HSDHs are evident in both Firmicutes[43,45] and Actinobacteria such as *Bifidobacterium adolescentis*[42]. Until this study, there were no reports of genes encoding 12β-HSDH and the activity had only been shown in *C. paraputrificum*, *C. tertium* and *C. difficile.*[19] Thus, our phylogenetic analysis revealed hitherto unknown diversity for bile acid 12β-HSDHs within the Firmicutes and Actinobacteria. Bacteroidetes sequences were notably absent within our 12β-HSDH subtree and only one sequence was identified in our HMM search, although Bacteroidetes have been shown to encode multiple other HSDHs.[1,49] Interestingly, *C. tertium* and *C. difficile* enzymes were also not present in our phylogenetic analysis even though this activity has been reported for strains of these clostridia[19], indicating that genes encoding other forms of bile acid 12β-HSDH are present in the gut microbiome.

The distribution pattern of microbial HSDHs is becoming increasingly clear (**Figures 3.5 & 3.7**), although in many cases the evolutionary pressures on gut microbes for encoding particular regio- and stereospecific HSDH enzymes is not clear. As observed with BSH enzymes, the functional importance of HSDHs may be strain-dependent. In some strains, the mere ability to acquire or dispose of reducing equivalents may be important, and the class of enzyme unimportant. Bile acid hydroxylation patterns affect the binding and activation/inhibition of host nuclear receptors.[67] HSDH enzymes may thus act in interkingdom-signaling, a hypothesis that has recent support based on the effect of oxidized and epimerized bile acids on the function of regulatory T cells.[68,69]

The concerted action of pairs of HSDHs result in bile acid products with reduced toxicity for microbes expressing the HSDH(s) or for an important inter-species partner, which was likely a factor in the evolution of these enzymes. Examples of strains of species capable of epimerizing bile acid hydroxyl groups are found in the literature, and the physicochemical properties and

reduced toxicity of β-hydroxy bile acids are known, providing hypotheses for physiological function. *Clostridium limosum* (now *Hathewaya limosa*) expresses both bile acid-inducible NADP-dependent 7α- and 7β-HSDH capable of converting CDCA to UDCA.[70] CDCA is more hydrophilic and more toxic to bacteria than UDCA.[6,71] Indeed, treatment with UDCA increases the hydrophilicity of the biliary pool, reducing cellular toxicity and improving biliary disorders.[72] Similarly, strains of *Eggerthella lenta*[15,41] and *Ruminococcus gnavus*[41] express both NADPH-dependent 3α- and 3β-HSDHs capable of forming 3β-bile acids (iso-bile acids). Iso-bile acids are also more hydrophilic and less toxic to bacteria than the α-hydroxy isomers.[41] At least some strains of *R. gnavus* also express NADPH-dependent 7β-HSDH, contributing to the epimerization of CDCA to UDCA.[39] It may be speculated that *R. gnavus* HSDHs function in detoxification of hydrophobic bile acids such as CDCA and DCA; however, further work is needed. Analogous to *E. lenta* and *R. gnavus, C. paraputrificum* is another example of a strain encoding multiple HSDHs that favor formation of β-hydroxy bile acids.[19] *C. paraputrificum* strains encode the iso-bile acid pathway as well as NADPH-dependent 12β-HSDH.[19,20] While little is known about the biological effects of 12β-bile acids (epi-bile acids), the physicochemical properties relative to 12α-hydroxy bile acids should approximate that of iso- and urso-derivatives.[6,41,71] An important question emerging from these observations is whether one particular epimeric product rather than another has important consequences on the fitness of the bacterium generating them, or if the increased hydrophilicity and reduced toxicity are the key factors.

Since the initial detection of epi-bile acids by Eneroth et. al. and Ali et. al.[16–18], the measurement of bile acid metabolomes in clinical samples has become commonplace[73], yet few studies measure or report epi-bile acids. Recently, 12β-hydroxy and 12-oxo-bile acids have been

quantified in human feces by Franco et. al. (2019). 12-oxoLCA was the most abundant oxo-bile acid in feces at concentrations of about one half that of DCA in stool. While epiDCA itself was not measured, 3-oxo-12β-hydroxy-CDCA was shown at 12 ± 4 µg/g wet feces.[74] Additionally, epiDCA has been reported in biliary bile of angelfish, likely produced from bacterial origin, so the 12β-HSDH gene may be widespread among resident microbiota of diverse vertebrate taxa.[75] A critical limitation to the study of epi-bile acids is the absence of commercially available standards, although there are methods available for chemical synthesis.[76,77] The newly identified bile acid 12β-HSDHs could be employed for the enzymatic production of epi-bile acid standards from oxo-intermediates.

The physiological effects of epi-bile acids are poorly characterized, particularly in the GI tract. Borgström and colleagues compared infusion of CA, ursoCA, and epiCA on bile flow, lipid secretion, bile acid synthesis, and bile micellar formation. In contrast to ursoCA and CA, epiCA was secreted into bile in an unconjugated form. The 12β-hydroxyl group may hinder the enzyme responsible for conjugation. Additionally, epiCA infusion increased the rate of secretion of newly synthesized bile salts.[78] Another study reported increased 12-oxoLCA levels in rats with high tumor incidence when they were fed a high safflower oil or corn oil diet.(Reddy 1984) While the toxicity of epi-bile acids has not yet been tested relative to the secondary bile acids DCA or LCA, both 12-oxoLCA and epiDCA are less hydrophobic than DCA by LC-MS (**Figures 3.2 & 3.3**). Due to the involvement of DCA in cancers of the liver and colon[7,8], bile acid 12β-HSDH may be of therapeutic importance in modulating the bile acid pool in favor of epiDCA over toxic DCA. Future studies with animal models will be imperative to determine the effects of epi-bile acids on host physiology.

MATERIALS AND METHODS

Bacterial strains and chemicals

Clostridium paraputrificum ATCC 25780 and *Clostridium scindens* ATCC 35704 were obtained from 80°C glycerol stocks from culture collections at the University of Illinois Urbana-Champaign (UIUC). *E. coli* DH5α (Turbo) competent cells from New England Biolabs (Ipswich, MA) and NovaBlue GigaSingles™ Competent cells from Novagen (San Diego, CA, USA) were used for cloning, and *E. coli* BL21-Codon-Plus (DE3) RIPL was purchased from Stratagene (La Jolla, CA, USA) and used for protein overexpression. 5β-Cholanic acid-3α, 7α, 12α-triol (CA), 5β-cholanic acid-3α,12α-diol (DCA), and 5β-cholanic acid-3α,7α-diol (CDCA) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Authentic 5β-cholanic acid-3α,12β-diol (epiDCA) and 5β-cholanic acid-3α,7α,12β-diol (epiCA) were generously obtained from Lee R. Hagey (University of California, San Diego). Other bile acids were purchased from Steraloids (Newport, RI, USA). All other reagents were of the highest possible purity and purchased from Fisher Scientific (Pittsburgh, PA, USA).

Whole cell bile acid conversion assay

C. paraputrificum ATCC 25780 and *C. scindens* ATCC 35704 were cultivated in anaerobic brain heart infusion (BHI) broth for 24 hrs. Two mL anaerobic BHI was inoculated with 1:10 dilution of either organism along with 50 μM bile acid substrate and incubated at 37°C for 24 hours. The bacterial cultures were centrifuged at $10,000 \times g$ for 5 min to remove bacterial cells and the conditioned medium was adjusted to pH 3.0. Solid phase extraction was used to extract bile acid products from bacterial culture. Waters tC18 vacuum cartridges (3 cc) (Milford, MA, USA) were preconditioned with 6 mL 100% hexanes, 3 mL 100% acetone, 6 mL 100%

methanol, and 6 mL water (pH 3.0). The conditioned medium was added to the cartridges and vacuum was applied to pull media through dropwise. Cartridges were washed with 6 mL water (pH 3.0) and 40% methanol. Bile acid products were eluted with 3 mL 100% methanol. Eluates were then evaporated under nitrogen gas and the residues dissolved in 200 μL 100% methanol for LC-MS analysis.

Liquid chromatography-mass spectrometry

LC-MS analysis for all samples was performed using a Waters Acquity UPLC system coupled to a Waters SYNAPT G2-Si ESI mass spectrometer (Milford, MA, USA). LC was performed with a Waters Acquity UPLC HSS T3 C18 column (1.8 μm particle size, 2.1 mm x 100 mm) at a column temperature of 40°C. Samples were injected at 1 μL. Mobile phase A was water and B was acetonitrile. The mobile phase gradient was as follows: 0 min 100% mobile phase A, 0.5 min 100% A, 25 min 0% A, 25.1 min 100% A, 28 min 100% A. The flow rate was 0.5 mL/min. MS was carried out in negative ion mode with a desolvation temperature of 300°C and desolvation gas flow of 700 L/hr. The capillary voltage was 3,000 V. Source temperature was 100°C and cone voltage was 30 V. Chromatographs and mass spectrometry data were analyzed using Waters MassLynx software (Milford, MA, USA).

Isolation of genomic DNA

Genomic DNA was extracted from *C. paraputrificum* ATCC 25780 using the Fast DNA isolation kit from Mo-Bio (Carlsbad, CA, USA) according to the manufacturer's protocol for polymerase chain reaction and molecular cloning applications.

Heterologous expression of potential 12β-HSDH proteins

The pET-28a(+) and pET-46 Ek/LIC vectors were obtained from Novagen (San Diego, CA, USA). Restriction enzymes were purchased from NEB (Ipswich, MA, USA). Inserts were generated by PCR amplification with cloning primers from Integrative DNA Technologies (Coralville, IA, USA) of *C. paraputrificum* ATCC 25780 genomic DNA or genes synthesized in *E. coli* K12 codon usage (IDT, Coralville, IA, USA). Cloning primers and genes created by gene synthesis are listed in **Supplementary Table 3.1**. Inserts were amplified using the Phusion High Fidelity Polymerase (Stratagene, La Jolla, CA, USA) and cloned into pET-28a(+) after insert and vector were double digested with the appropriate restriction endonuclease and treated with DNA Ligase, or annealed into pET-46 Ek/LIC after treatment with T4 DNA Polymerase. Recombinant plasmids were transformed via heat shock method, plated, and grown overnight at 37°C on lysogeny broth (LB) agar plates supplemented with antibiotic (50 µg/mL kanamycin or 100 µg/mL ampicillin). Vectors were either transformed into chemically competent *E. coli* DH5α cells and grown with kanamycin (pET-28a(+)) or transformed into NovaBlue GigaSingles™ Competent cells and grown with ampicillin (pET-46 Ek/LIC). A single colony from each transformation was inoculated into LB medium (5 mL) containing the corresponding antibiotic and grown to saturation. Recombinant plasmids were extracted from cell pellets using the QIAprep Spin Miniprep kit (Qiagen, Valencia, CA, USA). The sequence of the inserts was confirmed by Sanger sequencing (W. M. Keck Center for Comparative and Functional Genomics at the University of Illinois at Urbana-Champaign).

For protein expression, the extracted recombinant plasmids were transformed into *E. coli* BL-21 CodonPlus (DE3) RIPL chemically competent cells by heat shock method and cultured overnight at 37°C on LB agar plates supplemented with ampicillin or kanamycin (100 µg/ml; 50

µg/mL) and chloramphenicol (50 µg/ml). Selected colonies were inoculated into 10 mL of LB medium supplemented with antibiotics and grown at 37°C for 6 hours with vigorous aeration. The pre-cultures were added to fresh LB medium (1 L), supplemented with antibiotics, and aerated at 37°C until reaching an OD_{600nm} of 0.3. IPTG was added to each culture at a final concentration of 0.1 mM to induce and the temperature was decreased to 16°C for a 16-hour incubation. Cells were pelleted and resuspended in binding buffer (20 mM Tris-HCl, 300 mM NaCl, 10 mM 2-mercaptoethanol, pH 7.9). The cells were subjected to five passages through an EmulsiFlex C-3 cell homogenizer (Avestin, Ottawa, Canada), and the cell debris was separated by centrifugation.

The recombinant protein in the soluble fraction was then purified using TALON® Metal Affinity Resin (Clontech Laboratories, Mountain View, CA, USA) per the manufacturer's protocol. The recombinant protein was eluted using an elution buffer composed of 20 mM Tris-HCl, 300 mM NaCl, 10 mM 2-mercaptoethanol, and 250 mM imidazole at pH 7.9. The resulting purified protein was analyzed using sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). The observed subunit mass for each was calculated by migration distance of purified protein to standard proteins in ImageJ (https://imagej.nih.gov/ij/docs/faqs.html). TMHMM v. 2.0 was used to predict transmembrane helices.[22]

Enzyme Assays

Pure recombinant 12β-HSDH reaction mixtures were made using 50 µM substrate, 150 µM cofactor and 10 nM enzyme in 150 mM NaCl, 50 mM sodium phosphate buffer at the pH optima 7.0 or 7.5. Reactions were monitored by spectrophotometric assay measuring the oxidation or reduction of NADP(H) aerobically at 340 nm (6,220 $M^{-1}.cm^{-1}$) continuously for 1.5

min on a NanoDrop 2000c UV-Vis spectrophotometer (Fisher Scientific, Pittsburgh, PA, USA) using a 10 mm quartz cuvette (Starna Cells, Atascadera, CA, USA). Additional reactions were incubated overnight at room temperature and extracted by vortexing with two volumes ethyl acetate twice. The organic layer was recovered and evaporated under nitrogen gas. The products were dissolved in 50 µL methanol and LC-MS was performed as described above or used for thin layer chromatography.

The buffers for investigation of the optimal pH of recombinant 12β-HSDH contained 150 mM NaCl and one of the following buffering agents: 50 mM sodium acetate (pH 6.0), 50 mM sodium phosphate (pH 6.5 to 7.5), and 50 mM Tris-Cl (pH 8.0). Substrate specificity was performed according to the above reaction conditions at the optimal pH.

The reaction mixtures for kinetic analysis were 10 nM enzyme, sodium phosphate buffer (pH 7.0), and 150 µM NADPH for varying concentrations of 12-oxoLCA or 80 µM 12-oxoLCA for varying NADPH concentrations in the reductive direction. The oxidative reaction mixture contained 10 nM enzyme, sodium phosphate buffer (pH 7.5), and 300 µM $NADP^+$ when epiDCA concentrations were changed or 100 µM epiDCA when $NADP^+$ was varied. Kinetic parameters were estimated with GraphPad Prism (GraphPad Prism, La Jolla, CA, USA) to fit the data using nonlinear regression to the Michaelis-Menten equation.

Thin layer chromatography

Reaction mixtures were made using 50 µM substrate, 150 µM cofactor and 10 nM enzyme in 150 mM NaCl, 50 mM sodium phosphate buffer at pH 7.0. Reactions were incubated overnight at room temperature and extracted by vortexing with two volumes ethyl acetate twice. The organic layer was recovered and evaporated under nitrogen gas. The products were

dissolved in 50 μL methanol and spotted on a TLC plate (silica gel IB2-F flexible TLC sheet, 20 x 20 cm, 250 μm analytical layer; J. T. Baker, Avantor Performance Materials, LLC, PA, USA). The steroids were separated with a 70:20:2 toluene–1,4-dioxane–acetic acid mobile phase and visualized using a 10% phosphomolybdic acid in ethanol spray and heating for 15 min at 100°C.[79]

Native molecular weight determination

Size-exclusion chromatography was performed using a Superose 6 10/300 GL analytical column (GE Healthcare, Piscataway, NJ, USA) connected to an ÄKTAxpress chromatography system (GE Healthcare, Piscataway, NJ, USA) at 4°C. The column was equilibrated with 50 mM Tris-Cl and 150 mM NaCl at a pH of 7.5. The purified protein was loaded onto the analytical column at a concentration of 10 mg/mL and eluted at a flow rate of 0.3 ml/min. The native molecular mass of 12β-HSDH was determined by comparing its elution volume to that of Gel Filtration Standard proteins (Bio-Rad, Hercules, CA, USA): thyroglobulin, γ-globulin, ovalbumin, myoglobin, vitamin B_{12}.

Phylogenetic Analysis

The sequence of the *C. paraputrificum* 12β-HSDH protein (accession number WP_027099077.1) was used as query for a similarity search against the NCBI non-redundant protein database by BLASTP[80], with a maximum E-value threshold of 1e-10 and a limit of 5,000 results. Retrieved sequences were aligned with Muscle v. 3.8.1551[81] and analyzed by maximum likelihood with RAxML v. 8.2.11.[82] Selection of the best-fitting amino acid substitution model and number of bootstrap pseudoreplicates were performed automatically, and substitution rate

heterogeneity was modeled with gamma distributed rate categories. The resulting phylogenetic tree was formatted by Dendroscope v. 3.5.10[83] and further cosmetic modifications were performed with the vector editor Inkscape, v. 0.92.4 (https://inkscape.org).

For closer analysis of the phylogenetic affiliation of *C. paraputrificum* ATCC 25780 12β-HSDH, sequences from the well-supported subtree where this sequence is located in the 5,000-sequence tree, plus an outgroup, were reanalyzed for confirming the relative placement of all sequences nearest to Cp12β-HSDH. The methods used were the same as described above for the full tree.

A maximum-likelihood tree of representative HSDH sequences was inferred by selecting sequences from each HSDH subfamily, based on the tree from Mythen et al. (2018)[15], with the addition of eukaryotic, archaeal, and other bacterial sequences deposited in the public databases. Phylogenetic inference methods were the same as described above.

Hidden Markov Model Search

A Hidden Markov Model (HMM) search was performed using a custom HMM profile against a concatenated file of metagenome assembled genomes (MAGs) from four publicly available cohorts.[30–33] MAGs were filtered for genome completeness, quality, and contamination as described.[84] For generation of the custom 12β-HSDH profile, reference sequences from the 12β-HSDHs characterized in this paper were aligned with MAFFT, manually trimmed, and constructed using hmmscan.[85] The MAG database was searched using HMMSearch version 3.3.0[85], using an individually identified cut-off of 350.00. Resulting hits were then filtered to remove results less than 70% completeness and closest matched species were recorded. The

HMM search raw data file is publicly available at the following link:

https://github.com/AnantharamanLab/doden_et_al_2021.

ACKNOWLEDGMENTS
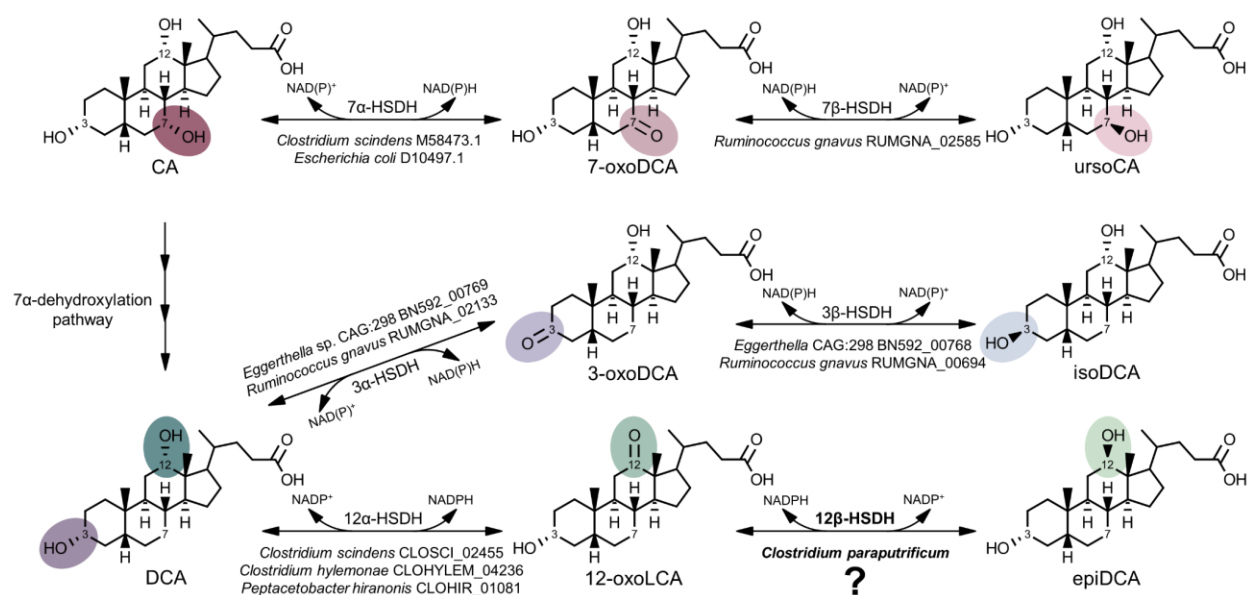
FIGURES

Figure 3.1. A gene encoding 12β-HSDH completes the gut microbial epi-bile acid pathway.

Cholic acid (CA) is converted to the oxo-intermediate, 7-oxodeoxycholic acid (7-oxoDCA), and further to ursoCA through the urso-bile acid pathway catalyzed by NAD(P)-dependent 7α- and 7β-HSDH. The secondary bile acid deoxycholic acid (DCA) is formed through the multi-step 7α-dehydroxylation of CA. DCA is biotransformed to 3-oxoDCA by 3α-HSDH and to isoDCA by 3β-HSDH in the iso-bile acid pathway. DCA can be converted to 12-oxolithocholic acid (12-oxoLCA) by 12α-HSDH and from 12-oxoLCA to epiDCA by 12β-HSDH. Examples of bacteria expressing each HSDH are shown below the reaction followed by corresponding gene annotations. Prior to this study, a gene encoding 12β-HSDH had not been identified.
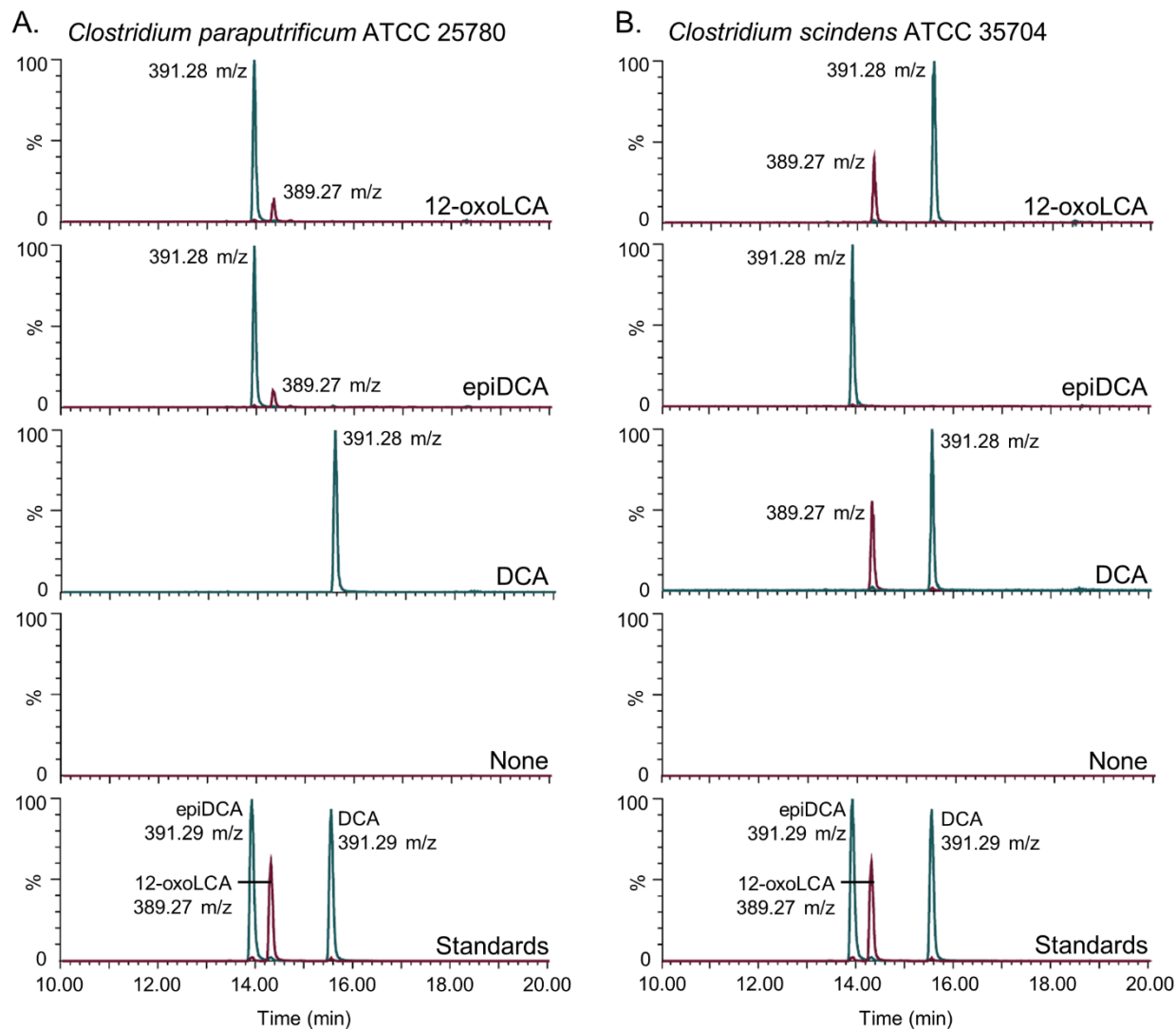
Figure 3.2. *Clostridium paraputrificum* **ATCC 25780 expresses 12β-HSDH while** *C. scindens*
ATCC 35704 expresses 12α-HSDH by whole-cell LC-MS. (A) Representative negative ion
mode LC-MS chromatograms in single ion monitoring mode overlaid with linked vertical axes
of *C. paraputrificum* reaction products from 50 μM substrate compared to deoxycholic acid
(DCA), 12-oxolithocholic acid (12-oxoLCA) and epiDCA standards. (B) As a control,
representative negative ion mode LC-MS chromatograms in single ion monitoring mode overlaid
with linked vertical axes of *C. scindens* products from 50 μM substrate was used to demonstrate
12α-HSDH activity. Standards are shown in A and B for ease of comparison to products.

Figure 3.2 (cont.)

Formula weight for DCA is 392.57 atomic mass units (amu), 12-oxoLCA is 390.56 amu,

epiDCA is 392.57 amu.
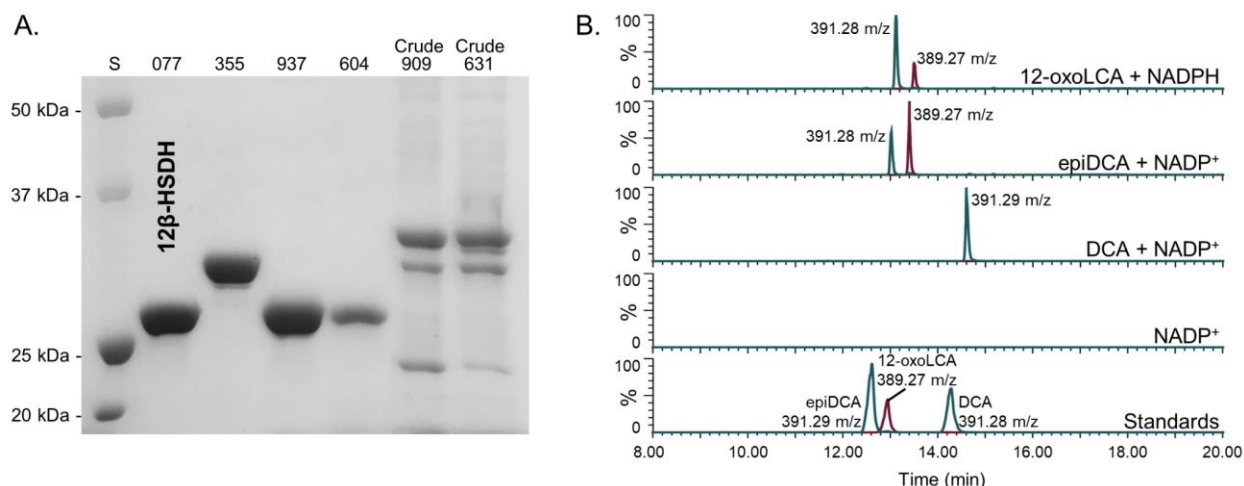
Figure 3.3. Identification of a gene encoding 12β-HSDH. (A) SDS-PAGE of candidate

Clostridium paraputrificum 12β-HSDH proteins that were heterologously expressed in *E. coli*

and purified with TALON® metal affinity resin. Lanes are as follows: S, molecular weight

protein standard; 077, WP_027099077.1; 355, WP_027098355.1; 937, WP_027097937.1; 604,

WP_027098604.1; 909, WP_027096909.1; 631, WP_027099631.1. (B) Representative negative

ion mode LC-MS chromatograms in single ion monitoring mode overlaid with linked vertical

axes of WP_027099077.1 reaction products compared to deoxycholic acid (DCA), 12-

oxolithocholic acid (12-oxoLCA) and epiDCA standards. Standards were run on a separate day

and show a slight offset in elution time. Reactions consisted of 10 nM WP_027099077.1 with 50

μM (or no) substrate, 150 μM pyridine nucleotide in 50 mM sodium phosphate, 150 mM sodium

chloride buffer at pH 7.0. Formula weight for DCA is 392.57 atomic mass units (amu), 12-

oxoLCA is 390.56 amu, epiDCA is 392.57 amu.
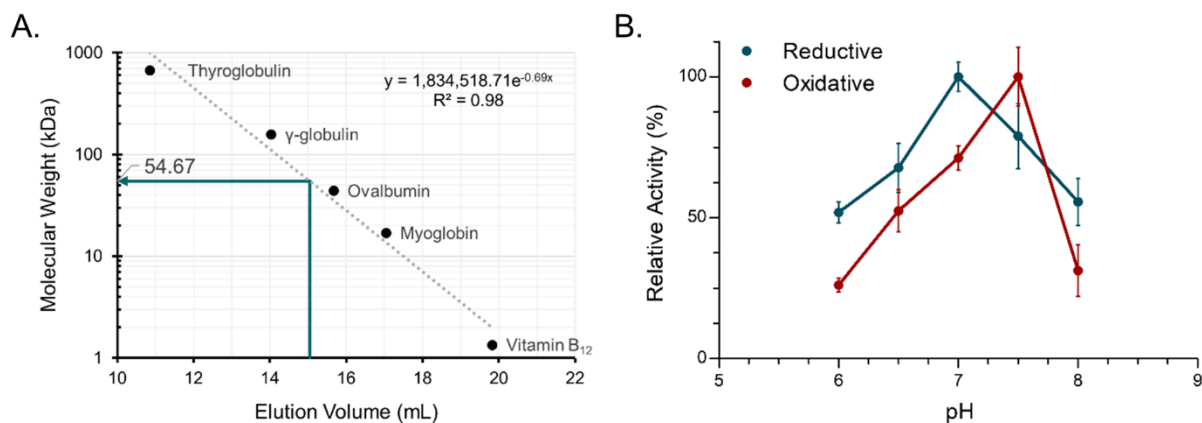
Figure 3.4. Biochemical characterization of recombinant *C. paraputrificum* 12β-HSDH. (A) Native molecular size analysis of 10 mg/mL purified 12β-HSDH via size-exclusion chromatography. (B) Effect of pH on 12β-HSDH activity. The reaction in the reductive direction (blue) consisted of 12-oxoLCA as substrate with NADPH as cofactor. The oxidative reaction (red) was epiDCA with $NADP^+$. See Materials and Methods for buffer compositions.
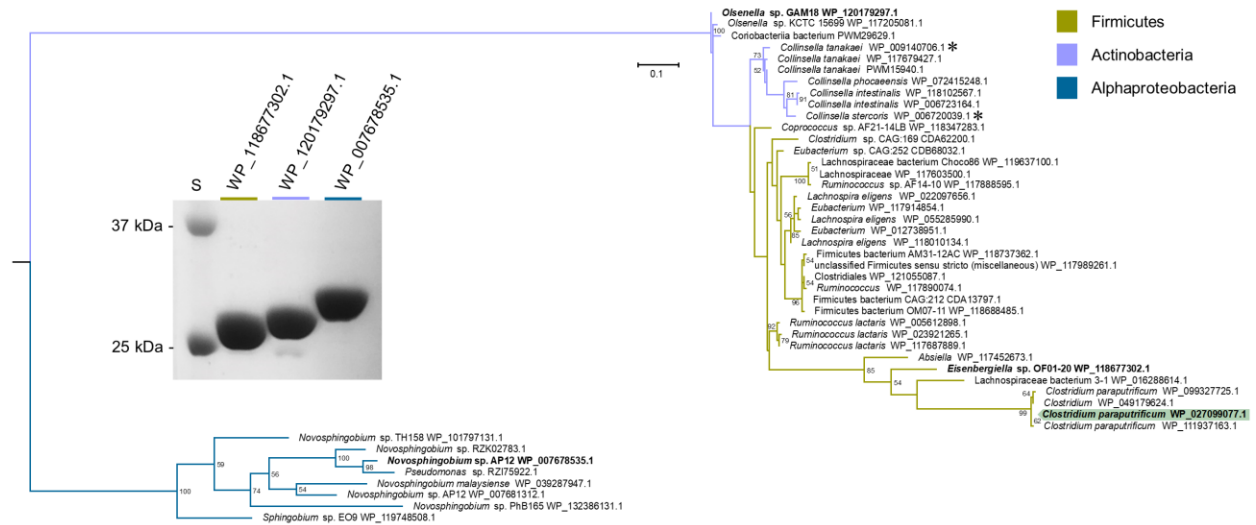
Figure 3.5. Maximum-likelihood tree based on a subset of the taxa present in the full phylogenetic analysis of 12β-HSDH and SDS-PAGE of proteins explored further. Sequences selected for this analysis were those nearest to the *C. paraputrificum* 12β-HSDH (highlighted), plus an outgroup. For the full tree with about 5,000 sequences, see **Supplementary Figure 3.2**. Taxonomic affiliations are indicated by branch colors as specified in the legend. Bolded sequences were chosen for further study. Asterisks indicate novel C-12 epimerizing organisms. (Inset) SDS-PAGE of purified recombinant *Eisenbergiella* WP_118677302.1, *Olsenella* WP_120179297.1, and *Novosphingobium* WP_007678535.1 heterologously expressed in *E. coli* and purified with TALON® metal affinity resin. S, molecular weight protein standard.
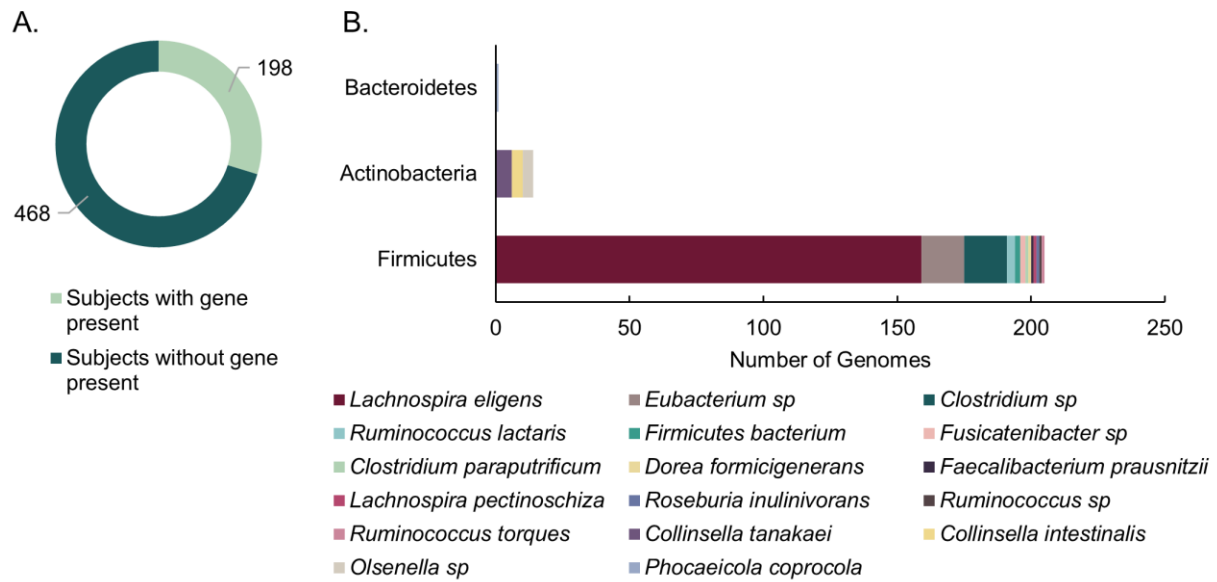
Figure 3.6. 12β-HSDH Hidden-Markov Model search. (A) Number of subjects identified with putative 12β-HSDH genes present in their gut metagenomes. The metagenomes analyzed were from 4 distinct cohorts. (B) Distribution of microbial genomes with putative 12β-HSDH genes present across the 4 metagenomic studies.
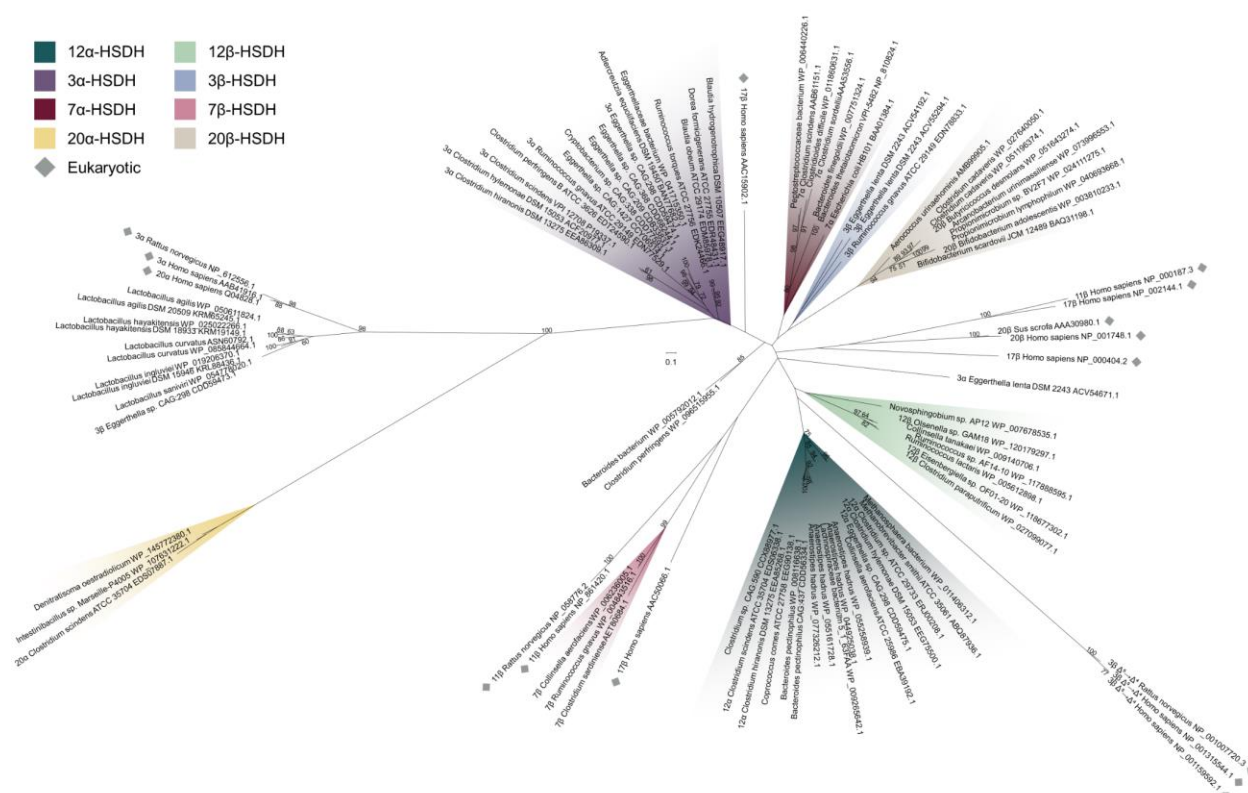
Figure 3.7. Maximum-likelihood phylogenetic analysis of regio- and stereospecific HSDHs.
Clusters are shaded by function or marked as eukaryotic, as displayed in the legend. Sequences
with experimentally determined activities are labeled with their function followed by organism
and accession number. See **Supplementary Table 3.2** for sequence information.

TABLES

Table 3.1. Steady-state kinetic parameters of purified recombinant 12β-HSDH.

Enzyme	Kinetic parameter	Substrate or cofactor[a]			
		12-oxoLCA[b]	NADPH	epiDCA	NADP[+]
Cp12β-HSDH	K_m (μM)	18.76 ± 0.40[c]	29.16 ± 0.42	44.43 ± 1.13	36.84 ± 0.55
	k_{cat} (s[-1])	26.62 ± 1.62	28.15 ± 1.74	59.32 ± 5.47	44.61 ± 3.77
	V_{max} (μmol • min[-1] • mg[-1])	58.41 ± 3.56	61.75 ± 3.82	130.13 ± 12.01	97.87 ± 8.27
	k_{cat}/K_m (μM[-1] • s[-1])	1.42 ± 0.11	0.97 ± 0.07	1.34 ± 0.15	1.21 ± 0.12

[a]Assays were performed at saturating concentrations of components not being tested (refer to Materials and Methods).

[b]12-oxolithocholic acid (12-oxoLCA), epideoxycholic acid (epiDCA).

[c]Values represent the mean ± SD of three or more replicates.

Table 3.2. Substrate and pyridine nucleotide specificity of purified recombinant *C. paraputrificum* 12β-HSDH, *Eisenbergiella* WP_118677302.1, *Olsenella* WP_120179297.1, and *Novosphingobium* WP_007678535.1.

Enzyme	Substrate[a,b]	Cofactor	Activity ($\mu mol \cdot min^{-1} \cdot mg^{-1}$)	Relative activity (%)
Cp12β-HSDH	12-oxoLCA	NADPH	18.26 ± 1.01[c]	100
	12-oxoLCA	NADH	-[d]	-
	12-oxoCDCA	NADPH	2.21 ± 0.82	12.08
	3,12-dioxoLCA	NADPH	3.49 ± 0.34	19.09
WP_118677302.1	12-oxoLCA	NADPH	16.04 ± 1.23	87.85
WP_120179297.1	12-oxoLCA	NADPH	23.29 ± 2.57	127.57
WP_007678535.1	12-oxoLCA	NADPH	-	-
Cp12β-HSDH	epiDCA	NADP$^+$	33.42 ± 0.81	100
	epiDCA	NAD$^+$	-	-
	epiCA	NADP$^+$	8.99 ± 0.90	26.88
	DCA	NADP$^+$	-	-
	CA	NADP$^+$	-	-
	CDCA	NADP$^+$	-	-
WP_118677302.1	epiDCA	NADP$^+$	27.85 ± 1.12	83.32
WP_120179297.1	epiDCA	NADP$^+$	23.02 ± 2.57	68.86
WP_007678535.1	epiDCA	NADP$^+$	-	-

[a]12-oxolithocholic acid (12-oxoLCA), 12-oxochenodeoxycholic acid (12-oxoCDCA), deoxycholic acid (DCA), cholic acid (CA).

[b]Assays were performed with 10 nM enzyme, 50 uM substrate and 150 uM cofactor at optimum pH.

[c]Values represent the mean ± SD of three or more replicates.

[d]-, no activity detected.
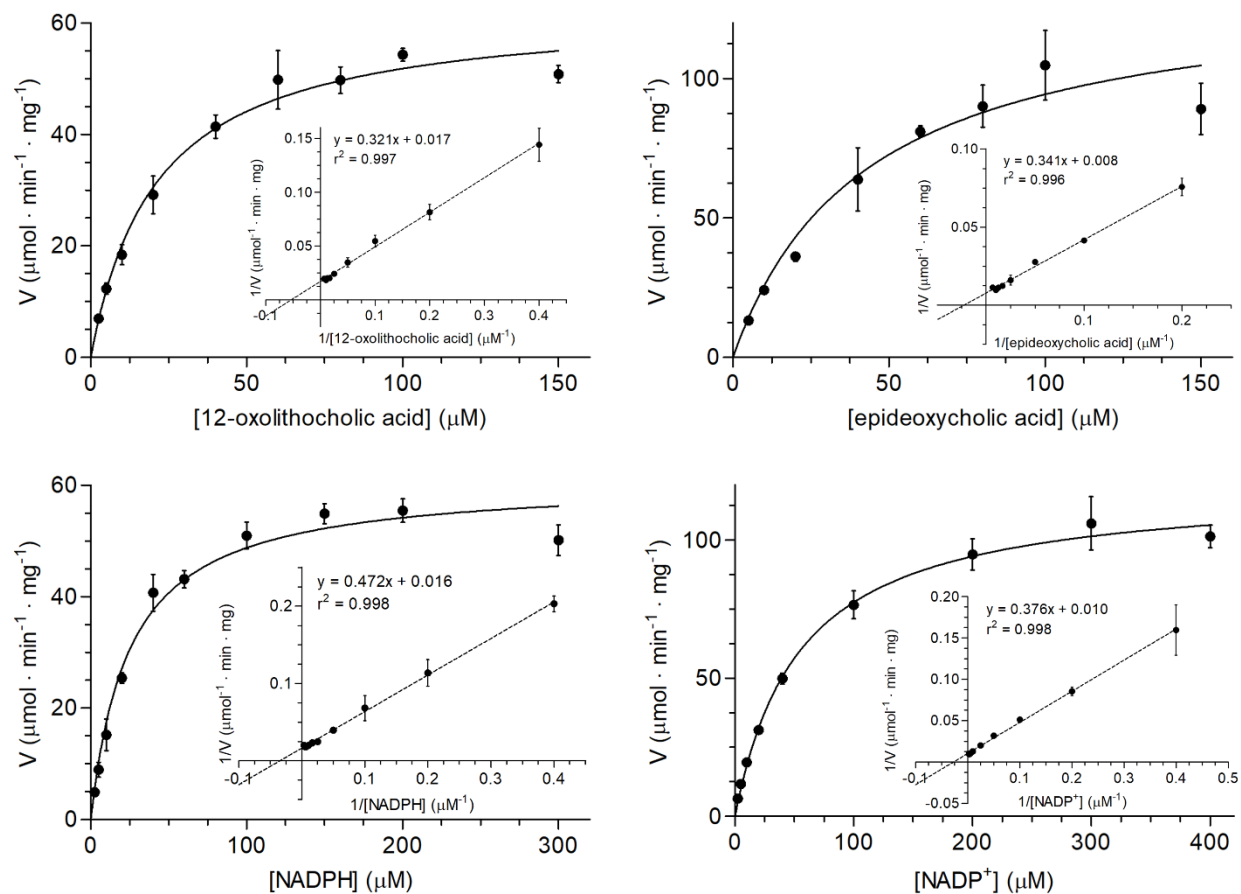
Supplementary Figure 3.1. Kinetic analysis of Cp12β-HSDH activity. Michaelis-Menten and Lineweaver-Burk plots of initial enzyme velocity with varying concentration of substrate (top panels) and cofactor (bottom panels) in both the reductive (left panels) and oxidative (right panels) directions. See Materials and Methods for enzyme reactions. Oxidation/reduction of NAD(H) was measured by continuous spectrophotometry at 340 nm for 1.5 min and the initial velocities were used to calculate kinetic constants (Table 1). Each point represents the mean ± SD of three or more replicates.
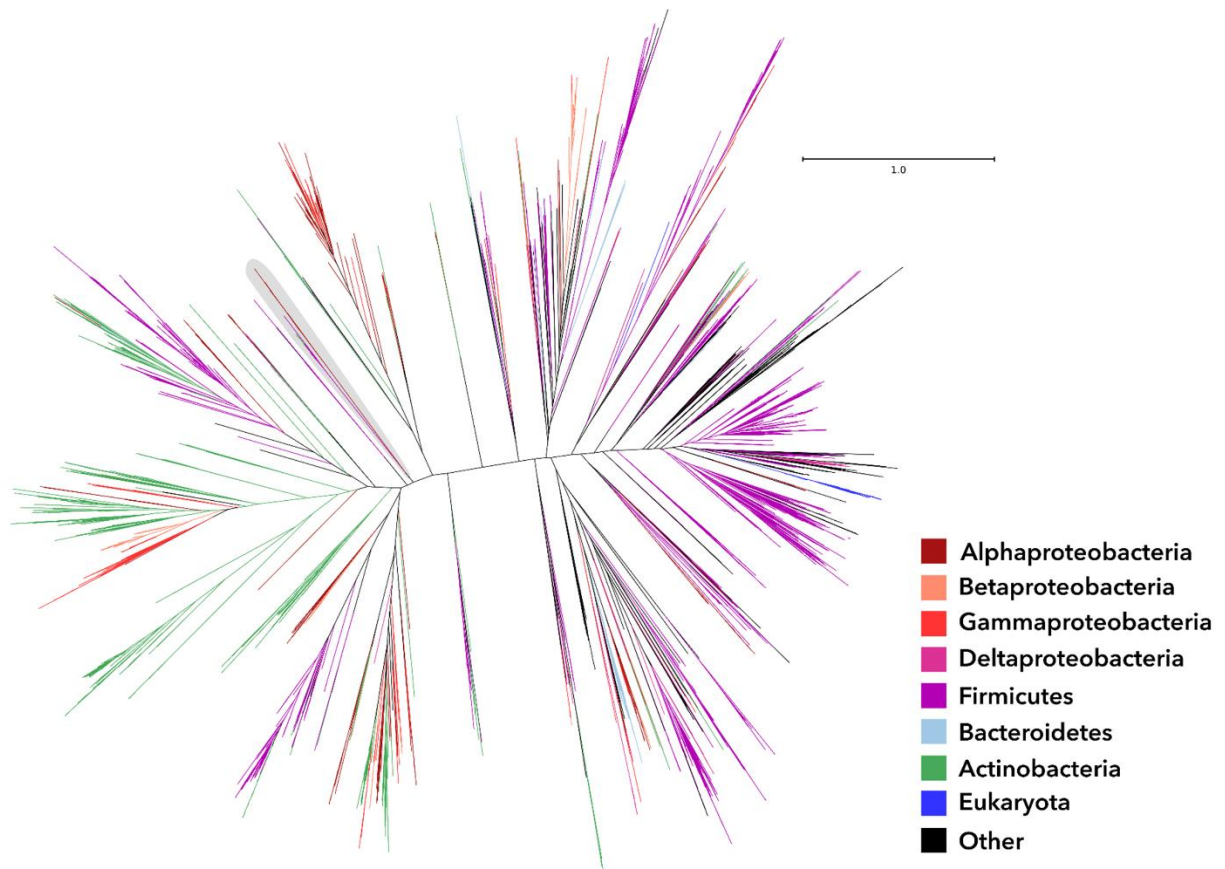
Supplementary Figure 3.2. 12β-HSDH 5,000-member tree. Maximum likelihood phylogeny

for the five thousand sequences from the NR database that are most similar to *C. paraputrificum*

12β-HSDH (WP_027099077.1). Taxonomic affiliations are indicated by branch colors as

specified in the legend. The region highlighted in gray shows the sequences that were reanalyzed

to generate the phylogeny shown in Figure 3.5.
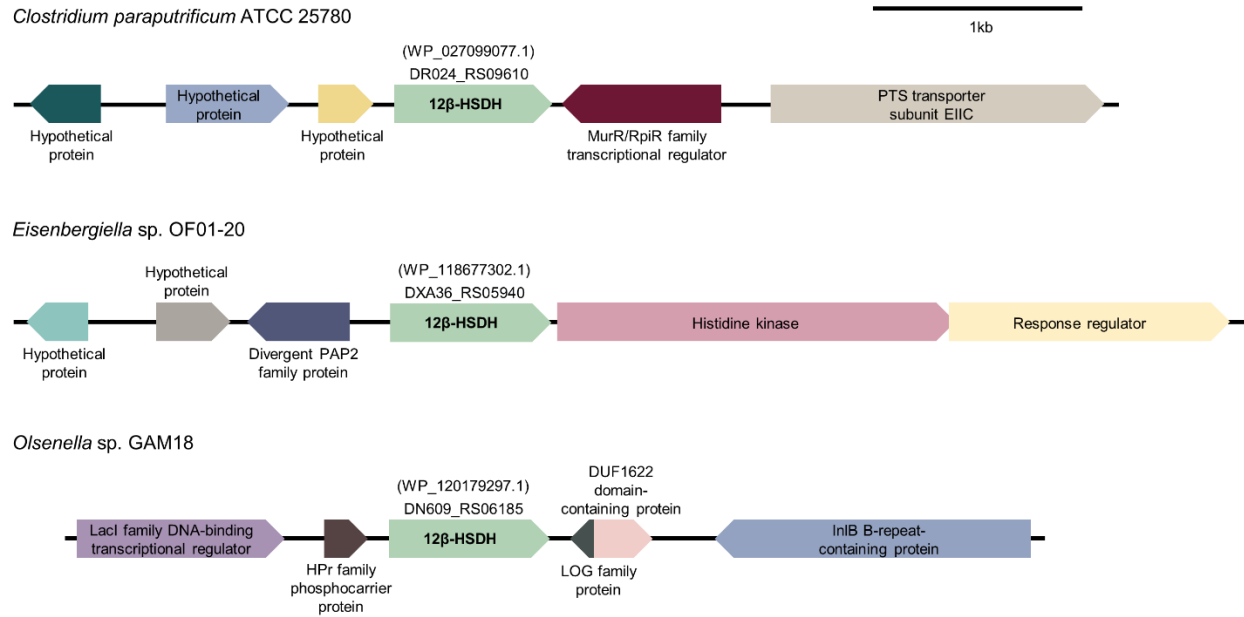
Supplementary Figure 3.3. Genomic context of 12β-HSDH genes from *Clostridium paraputrificum* ATCC 25780, *Eisenbergiella* sp. OF01-20, and *Olsenella* sp. GAM18. Genomic context was identified using the NCBI nucleotide graphics function.

SUPPLEMENTARY TABLES

Supplementary Table 3.1. Cloning primers for putative 12β-HSDHs and synthesized genes.

Gene ID Protein Accession	Primer pair	Vector	MW (kDa)[a]	Extinction coefficient (M^{-1}•cm^{-1})
DR024_RS09610 WP_027099077.1	ATATAT<u>CATATG</u>ATGAAAGAGTTAAATGAGAAAGTAGCTATTATTACAGG (5'-Forward Primer-3') ATATAT<u>CTCGAG</u>TTAAGGAGCTATTGAGTATCCACCATCC (5'-Reverse Primer-3')	pET-28a(+)	27.35	31,775
DR024_RS11955 WP_027098355.1	<u>GACGACGACAAG</u>ATGATTGAAAAAACGTTTCCAAAAACTTTTCCTACT <u>GAGGAGAAGCCCGGT</u>TTAACTATTAACTATTTCCCCTCCATTAACATGAAT	pET-46 Ek/LIC	31.39	19,160
DR024_RS09610 WP_027097937.1	ATATAT<u>GGATCC</u>ATGTCAAAGTTATTAGGTAAGGTAGCCATTATTAC ATATAT<u>CTCGAG</u>TTAAATATACCCTCCATCCACTCTAAGTATCTG	pET-28a(+)	26.26	14,565
DR024_RS15375 WP_027098604.1	<u>GACGACGACAAG</u>ATGAAAAATAAATTCACACTAATAACTGGTGGAAGTGA <u>GAGGAGAAGCCCGGT</u>CTAATACTTATTGCTTTTTAAGACCACTTTTCTACTTAATGCT	pET-46 Ek/LIC	28.75	18,005
DR024_RS02235 WP_027096909.1	ATATAT<u>CATATG</u>ATGAAGCAAATAAAAATAGCTAATACAGATATGGAAGT ATATAT<u>CTCGAG</u>TTAAGGAAGTGTATTTCCTGCTGCCTTA	pET-28a(+)	34.68	31,650
DR024_RS08875 WP_027099631.1	ATATAT<u>GCTAGC</u>ATGTCTGTATTGGAGAATAATTATAGGGATTTTAAATGT ATATAT<u>CTCGAG</u>TTAGCTATCTACCATAGTTCCACCATTAACA	pET-28a(+)	32.86	34,060
DXA36_RS05940 WP_118677302.1	ATATAT<u>CATATG</u>ATGAAACAGTTGAATGAGAAAGTAGCCAT ATATAT<u>CTCGAG</u>TTATGGCATAATCGAATAGGCTCCATCAAC ATGAAACAGTTGAATGAGAAAGTAGCCATCGTAACTGGTGCTGGTCAAGGCATTGGTCAAGGAATTGCATTATGTTTGGGGAAGCGCGGTGTTAAGGTAGTGTGCGTAGGTCGCCGTCCCGAACCTATTGAAGCTACTGCTAAAGAAATTCGTGACTTAGGTGGTGAGTCTTTCGCTATGACCTGTGACACAGCGGATCGTGATCGTGTTAAAGAGGTAGTCGCAAAAACCGTGGAAACGTACAAAACTGTAGATGTAATGATTAATAACGCGCAAAGTTTACCCGGATCCGCCCCTGTTGAGGAGGTCACTTATGAAATGATGTACACGGCATGGAGTACTGGAACACTTGGATCATTAAACTTTATGCAAGAATGTTTCCCTTACATGAAAGAGCAAGGGGAAGGGCGTGTAATCAACTTTGCGTCAGCCACTGGTATGTTCGGGTATGCTGGTAATCTTGCGTATGGTTGTAACAAAGAAGCCATTCGTGGCTTGACAAAGATTGCAGCGAAGGAGTGGGGCAAATACGGGATCTGTGTAAACTGTGTCCTGCCAGGCGCTGAGTCCCCAGCTGCTAAGATCTGGGCTGAAAAATTTCCAGAAAAGTATGCGGAGATCTTGGAGCAACAACCTATGAAACGCTTAGGAGACGCTGAAAAGGACATCGCGCCGGTGATTGCTTTCTTATCGGGTCCAGACAGCTGTTATTACAGTGGTCAATGCCTGTTGGTTGATGGAGCCTATTCGATTATGCCATAA[b]	pET-28a(+)	27.34	33,390
DN609_RS06185 WP_120179297.1	ATATAT<u>CATATG</u>ATGAAACAACTTAATGAAAAAGTTGCAATTGTTACT ATATAT<u>CTCGAG</u>CTAGGGCATGATACTATTAGCTCCGTC ATGAAACAACTTAATGAAAAAGTTGCAATTGTTACTGGTGCAGGGCAAGGGATCGGAAAAGGGATCGCCCTTTGTTTAGCTAAACGTGGTGTAAAGGTCGTGTGCACGGGACGCCGTGAAGCACCAATCCAACAGACTGTAGCAGAAATCGAAGAGTTAGGAGGTCAAGGACTTGCTATGACTTGTGACTCAGCAGACCGTGCGCGCGTCGAAGAGGTGGTCAAGGCCGCGGTCGATACGTTTGGCTCCATCGATGTGATTGTAAATAATGGACAAGCCATTGTCCCCTCTGCGCCTGTCGAGGACACTACATACGAGAACATGCTGGCGGCATGGCAAAGTGGCACTATCGGCTCGTTAAATTACATGCAGGCTGCATTTCCCCACATGAAGGAGCAGCACGAGGGGCGTATTATCAATTTCGCCTCCGCTACCGGAATGTTCGGCATTGCGGGACAACTGGCTTACGGCAGCAATAAAGAGGCACTTCGCGGTCTTACAAAGATCGCCGCTAAGGAATGGGGACAGTATGGAATTTGCGTTAACGTAGTATTGCCGGGGGGCGGAATCCCCAGCCGCCAAGGCGTGGGCCGAGAAATTTCCAGAAGAATACCAGAAACAGGTTATGTTAAATCCAATGCATCGTTTTGGGGACCCAGAAGATGACATCGCGCCAGTTGTGGCATTCCTGGCGGGTCCTGATAGTTGCTATTACTCCGGCCAATCAGTCATCGTGGACGGAGCTAATAGTATCATGCCCTAG[b]	pET-28a(+)	26.87	27,180

Supplementary Table 3.1 (cont.)

PMI02_RS03035 WP_007678535.1	ATATAT<u>GGATCC</u>ATGTCCAACGAGAGTCGTCGTTTAGA ATATAT<u>CTCGAG</u>TCAAAGCATTGTGGTGCCACCC	pET-28a(+)	27.05	22,590

ATGTCCAACGAGAGTCGTCGTTTAGAAGGTAAAGTTGCAATTGTCACGGGGGGCAGGTCAAGGGATCGGAGA
GGCTATCGCAGTTGGATACGCAGCTGCTGGAGCAAAAGTCTTGATCACGGGACGCACTCAGAGTAAGTTAG
ATGACGTTGTAAGTAAGATTGAAGCTGCCTCAGGGACAGCGATTGGGATGGAGGCATTGGCTGGTAACCGT
GAACATGCTGCGCTTACAGTTGATCGCGCCATTTCAGAATGGGGGCGCGTCGACGTACTGGTTAATAATGCG
CATACGTTTACGGATTTTTTGCCCATCGAGGACCCAAAGATGGAAGAGAATTGCAATACGGACATCCAATCA
GCTTTTTTTGGCAGTCTGCAGTTGATGCAATTGTGCCACCCCCATATGGCAGCCCAAGGTGGCGGCTCAATT
ATTAACATGGGTACTGGTTTTTCTATTCGTTGCGAGCCGGGTTTTTTGGCCTACGCCGCTACTAAGGAAGCGA
TCCGCGTATTGACCAAGACCGCAGCGAAGGAGTGGGGGAAACATAAGATCCGTGTGAATACGATCCTTCCT
TCTGCTCTGACGCCAAAAAGTATCTGGTATTTAGAGGACTCAGGAACATACGACCAGGAATTGGCTAAGGT
GGCTATGGGTCGTTTCGGCGAGCCGAAAGATATCGCGCCAACAGCGGTCTTTCTGGCCTCTGACGATTCAGA
TTTTGTGACGGGGCAGACTATCGGGGTGGAGGGTGGCACCACAATGCTTTGA[b]

[a]Deduced recombinant protein molecular weight.
[b]Gene synthesized by Integrated DNA Technologies with *E.coli* K12 codon optimization.

Supplementary Table 3.2. Sequences used in HSDH phylogenetic analysis.

Accession No.	Function	Organism	Reference
WP_027099077.1	12β-HSDH	*Clostridium paraputrificum* ATCC 25780	Current study
WP_007678535.1	putative 12β-HSDH	*Novoshingobium* sp. AP12	Current study
WP_120179297.1	12β-HSDH	*Olsenella* sp. GAM18	Current study
WP_009140706.1	putative 12β-HSDH	*Collinsella tanakaei*	Current study
WP_117888595.1	putative 12β-HSDH	*Ruminococcus* sp. AF14-10	Current study
WP_005612898.1	putative 12β-HSDH	*Ruminococcus lactaris*	Current study
WP_118677302.1	12β-HSDH	*Eisenbergiella* sp. OF01-20	Current study
EDS06338.1	12α-HSDH	*Clostridium scindens* ATCC 35704	Doden Appl. Environ. Microbiol. 2018
EEG75500.1	12α-HSDH	*Clostridium hylemonae* DSM 15053	Doden Appl. Environ. Microbiol. 2018
EEA85268.1	12α-HSDH	*Peptacetobacter hiranonis/Clostridium hiranonis* DSM 13275	Doden Appl. Environ. Microbiol. 2018
CDD59475.1	12α-HSDH	*Eggerthella* sp. CAG:298	Mythen Appl. Environ. Microbiol. 2018
ERJ00208.1	12α-HSDH	*Clostridium* sp. strain ATCC 29733	Macdonald J. Lipid Res. 1979; Aigner US patent 2011
WP_096515955.1	putative 12α-HSDH	*Clostridium perfringens*	Doden Appl. Environ. Microbiol. 2018
WP_011406312.1	putative 12α-HSDH	*Methanosphaera stadtmanae* DSM 3091	Kisiela J. Steroid Biochem. Mol. Biol. 2012
ABQ87936.1	putative 12α-HSDH	*Methanobrevibacter smithii* ATCC 35061	Kisiela J. Steroid Biochem. Mol. Biol. 2012
EBA39192.1	putative 12α-HSDH	*Collinsella aerofaciens* ATCC 25986	Kisiela J. Steroid Biochem. Mol. Biol. 2012
WP_006235414.1	putative 12α-HSDH	*Collinsella aerofaciens* ATCC 25986	Kisiela J. Steroid Biochem. Mol. Biol. 2012
EEG90138.1	putative 12α-HSDH	*Coprococcus comes* ATCC 27758	Kisiela J. Ster. Biochem. Mol. Biol. 2012
CDD59474.1	3α-HSDH	*Eggerthella* sp. CAG:298	Mythen Appl. Environ. Microbiol. 2018
ACV54671.1/ WP_009306474.1	3α-HSDH	*Eggerthella lenta* DSM 2243	Devlin Nat. Chem. Biol. 2015
EDN77529.1	3α-HSDH	*Ruminococcus gnavus* ATCC 29149	Devlin Nat. Chem. Biol.2015
P19337.1	3α-HSDH	*Clostridium scindens* ATCC 35704	Mallonee Curr. Microbiol. 1995; Bhowmik Proteins 2014
ACF20977.1	3α-HSDH	*Clostridium hylemonae* DSM 15053	Ridlon Anaerobe 2010
EEA86309.1	3α-HSDH	*Clostridium hiranonis* DSM 13275	Wells Appl. Environ. Microbiol. 2000
EEG48917.1	putative 3α-HSDH	*Blautia hydrogenotrophica* DSM 10507	Kisiela J. Steroid Biochem. Mol. Biol. 2012
EDR48431.1	putative 3α-HSDH	*Dorea formicigenerans* ATCC 27755	Kisiela J. Steroid Biochem. Mol. Biol. 2012
EDK24466.1	putative 3α-HSDH	*Ruminococcus torques* ATCC 27756	Kisiela J. Steroid Biochem. Mol. Biol. 2012
EDM85978.1	putative 3α-HSDH	*Ruminococcus obeum* ATCC 29174	Kisiela J. Steroid Biochem. Mol. Biol. 2012
EDT24590.1	putative 3α-HSDH	*Clostridium perfringens* B str. ATCC 3626	Macdonald Biochim. Biophys. Acta 1976; Hirano Appl. Environ. Microbiol. 1981
CDD59473.1	3β-HSDH	*Eggerthella* sp. CAG:298	Mythen Appl. Environ. Microbiol. 2018
ACV55294.1	3β-HSDH	*Eggerthella lenta* DSM 2243	Devlin Nat. Chem. Biol. 2015
ACV54192.1	3β-HSDH	*Eggerthella lenta* DSM 2243	Devlin Nat. Chem. Biol. 2015
EDN78833.1	3β-HSDH	*Ruminococcus gnavus* ATCC 29149	Devlin Nat. Chem. Biol. 2015
BAA01384.1	7α-HSDH	*Escherichia coli*	Yoshimoto J. Bacteriol. 1991
AAA53556.1	7α-HSDH	*Clostridium sordellii* ATCC 9714	Coleman J. Bacteriol. 1994
AAB61151.1	7α-HSDH	*Clostridium scindens* VPI 12708	Baron J. Bacteriol. 1991
WP_005792012.1	3-oxoacyl-[acyl-carrier-protein] reductase	*Bacteroides fragilis* ATCC 25285	Doden Appl. Environ. Microbiol. 2018
WP_006440226.1	putative 7α-HSDH	*Clostridium hiranonis* DSM 13275/*Peptacetobacter hiranonis*	Kisiela J. Steroid Biochem. Mol. Biol. 2012
WP_007751324.1	putative 7α-HSDH	*Bacteroides finegoldii* DSM 17565	Kisiela J. Steroid Biochem. Mol. Biol. 2012
WP_011860631.1	putative 7α-HSDH	*Clostridioides difficile* 630	Kisiela J. Steroid Biochem. Mol. Biol. 2012
NP_810824.1	putative 7α-HSDH	*Bacteroides thetaiotaomicron* VPI-5482	Ferrandi Appl. Microbiol. Biotechnol. 2012; Kisiela J. Steroid Biochem. Mol. Biol. 2012; Sherrod Biochim. Biophys. Acta 1977
WP_006236005.1	7β-HSDH	*Collinsella aerofaciens* ATCC 25986	Liu Appl. Microbiol. Biotechnol. 2011
WP_004843516.1	7β-HSDH	*Ruminococcus gnavus* ATCC 29149	Lee J. Lipid Res. 2013
AET80684.1	7β-HSDH	*Clostridium absonum/Clostridium sardiniense*	Ferrandi Appl. Microbiol. Biotechnol. 2011

Supplementary Table 3.2 (cont.)

EDS07887.1	20α-HSDH	*Clostridium scindens* ATCC35704	Ridlon J. Lipid Res. 2013
WP_107631222.1	putative 20α-HSDH	*Intestinibacillus* sp. Marseille-P4005	Current study
WP_145772380.1	putative 20α-HSDH	*Denitratisoma oestradiolicum*	Current study
WP_003810233.1	20β-HSDH	*Bifidobacterium adolescentis* L2-32	Doden J. Biol. Chem. 2019
WP_051643274.1	20β-HSDH	*Butyricicoccus desmolans* ATCC 43058	Devendran J. Lipid Res. 2017
WP_051196374.1	putative 20β-HSDH	*Clostridium cadaveris* AGR2141	Devendran J. Lipid Res. 2017
WP_027640050.1	putative 20β-HSDH	*Clostridium cadaveris* AGR2141	Devendran J. Lipid Res. 2017
BAQ31198.1	putative 20β-HSDH	*Bifidobacterium scardovii* DSM 13734	Devendran J. Lipid Res. 2017
WP_024111275.1	putative 20β-HSDH	*Propionimicrobium* sp. BV2F7	Devendran J. Lipid Res. 2017
AMB99905.1	putative 20β-HSDH	*Aerococcus urinaehominis*	Devendran J. Lipid Res. 2017
WP_040693668.1	putative 20β-HSDH	*Propionimicrobium lymphophilum* ACS-093-V-SCH5	Ly J. Steroid Biochem. Mol. Biol. 2020
WP_073996553.1	putative 20β-HSDH	*Arcanobacterium urinimassiliense*	Ly J. Steroid Biochem. Mol. Biol. 2020
WP_025022266.1	aldo/keto reductase	*Lactobacillus hayakitensis*	Mythen Appl. Environ. Microbiol. 2018
KRM19149.1	oxidoreductase	*Lactobacillus hayakitensis* DSM 18933	Mythen Appl. Environ. Microbiol. 2018
KRM65245.1	oxidoreductase	*Lactobacillus agilis* DSM 20509	Mythen Appl. Environ. Microbiol. 2018
WP_050611824.1	aldo/keto reductase	*Lactobacillus agilis*	Mythen Appl. Environ. Microbiol. 2018
WP_019206370.1	aldo/keto reductase	*Lactobacillus ingluviei*	Mythen Appl. Environ. Microbiol. 2018
KRL88436.1	organophosphate reductase	*Lactobacillus ingluviei* DSM 15946	Mythen Appl. Environ. Microbiol. 2018
ASN60792.1	aldo/keto reductase	*Lactobacillus curvatus*	Mythen Appl. Environ. Microbiol. 2018
WP_085844664.1	aldo/keto reductase	*Lactobacillus curvatus*	Mythen Appl. Environ. Microbiol. 2018
WP_054778020.1	aldo/keto reductase	*Lactobacillus saniviri*	Mythen Appl. Environ. Microbiol. 2018
BAN77682.1	conserved hypothetical protein	*Adlercreutzia equolifaciens* DSM 19450	Mythen Appl. Environ. Microbiol. 2018
WP_041715350.1	SDR family oxidoreductase	*Adlercreutzia equolifaciens*	Mythen Appl. Environ. Microbiol. 2018
CDD77593.1	putative uncharacterized protein	*Cryptobacterium* sp. CAG:338	Mythen Appl. Environ. Microbiol. 2018
CDD68244.1	putative uncharacterized protein	*Eggerthella* sp. CAG:368	Mythen Appl. Environ. Microbiol. 2018
CDB33831.1	putative uncharacterized protein	*Eggerthella* sp. CAG:209	Mythen Appl. Environ. Microbiol. 2018
CCY06514.1	putative uncharacterized protein	*Eggerthella* sp. CAG:1427	Mythen Appl. Environ. Microbiol. 2018
CDD56334.1	putative uncharacterized protein	*Bacteroides pectinophilus* CAG:437	Mythen Appl. Environ. Microbiol. 2018
WP_044925038.1	short-chain dehydrogenase	*Anaerostipes hadrus*	Mythen Appl. Environ. Microbiol. 2018
WP_008116638.1	NAD(P)-dependent oxidoreductase	*Bacteroides pectinophilus*	Mythen Appl. Environ. Microbiol. 2018
WP_077326212.1	short-chain dehydrogenase	*Anaerostipes hadrus*	Mythen Appl. Environ. Microbiol. 2018
	short-chain dehydrogenase	*Clostridiales bacterium* Nov_37_41	Mythen Appl. Environ. Microbiol. 2018
WP_055161728.1	short-chain dehydrogenase	*Anaerostipes hadrus*	Mythen Appl. Environ. Microbiol. 2018
WP_055258939.1	short-chain dehydrogenase	*Anaerostipes hadrus*	Mythen Appl. Environ. Microbiol. 2018
CCX88977.1	dehydrogenases with different specificities	*Clostridium* sp. CAG:590	Mythen Appl. Environ. Microbiol. 2018
WP_009265642.1	short-chain dehydrogenase	Lachnospiraceae bacterium 5_1_63FAA	Mythen Appl. Environ. Microbiol. 2018
AAB41916.1	3α-HSD	*Homo sapiens*	Khanna J. Biol. Chem. 1995
NP_612556.1	3α-HSD	*Rattus norvegicus*	Cheng Mol. Endocrinol. 1991
NP_001315544.1	3β-HSD/$\Delta^5 \to \Delta^4$/isomerase type 1	*Homo sapiens*	Thomas J. Steroid Biochem. 1989
NP_001007720.3	3β-HSD/$\Delta^5 \to \Delta^4$/isomerase type 1	*Rattus norvegicus*	de Launoit J. Biol. Chem. 1992
NP_001159592.1	3β-HSD/$\Delta^5 \to \Delta^4$/isomerase type 2	*Homo sapiens*	Rheaume Mol. Endocrinol. 1991
Q04828.1	AKR1C1 (20α-HSDH)	*Homo sapiens*	Zhang J. Mol. Endocrinol. 2000
NP_001748.1	Carbonyl reductase 1 (20β-HSDH)	*Homo sapiens*	Morgan Sci. Rep. 2017
AAA30980.1	20β-HSDH	*Sus scrofa*	Tanaka J. Biol. Chem. 1992
NP_861420.1	11β-HSD type 1	*Homo sapiens*	Tannin J. Biol. Chem. 1991
NP_058776.2	11β-HSD type 1	*Rattus norvegicus*	Agarwal J. Biol. Chem. 1989
NP_000187.3	11β-HSD type 2	*Homo sapiens*	Wilson J. Clin. Endocrinol. Metab. 1995
NP_000404.2	17β-HSD type 1	*Homo sapiens*	Winqvist Hum. Genet. 1990
NP_002144.1	17β-HSD type 2	*Homo sapiens*	Labrie DNA Cell Biol. 1995
AAC50066.1	17β-HSD type 3	*Homo sapiens*	Geissler Nat. Genet. 1994
AAC15902.1	17β-HSD type 10	*Homo sapiens*	He J. Biol. Chem. 1999

REFERENCES

1.   Ridlon JM, Kang D-J, Hylemon PB. Bile salt biotransformations by human intestinal bacteria. J Lipid Res 2006; 47:241–59.
2.   Vlahcevic ZR, Heuman DM, Hylemon PB. Physiology and pathophysiology of enterohepatic circulation of bile acids. In: Zakim D, Boyer T, editors. Hepatology: A Textbook of Liver Disease. Philadelphia: Saunders; 1996. page 376–417.
3.   Dawson PA, Karpen SJ. Intestinal transport and metabolism of bile acids. J Lipid Res 2015; 56:1085–99.
4.   Jones B V., Begley M, Hill C, Gahan CGM, Marchesi JR. Functional and comparative metagenomic analysis of bile salt hydrolase activity in the human gut microbiome. Proc Natl Acad Sci U S A 2008; 105:13580–5.
5.   Ridlon JM, Harris SC, Bhowmik S, Kang DJ, Hylemon PB. Consequences of bile salt biotransformations by intestinal bacteria. Gut Microbes 2016; 7:22–39.
6.   Watanabe M, Fukiya S, Yokota A. Comprehensive evaluation of the bactericidal activities of free bile acids in the large intestine of humans and rodents. J Lipid Res 2017; 58:1143–52.
7.   Bernstein C, Holubec H, Bhattacharyya AK, Nguyen H, Payne CM, Zaitlin B, Bernstein H. Carcinogenicity of deoxycholate, a secondary bile acid. Arch Toxicol 2011; 85:863–71.
8.   Yoshimoto S, Loo TM, Atarashi K, Kanda H, Sato S, Oyadomari S, Iwakura Y, Oshima K, Morita H, Hattori M, et al. Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome. Nature 2013; 499:97–101.
9.   Wu JT, Gong J, Geng J, Song YX. Deoxycholic acid induces the overexpression of intestinal mucin, MUC2, via NF-kB signaling pathway in human esophageal adenocarcinoma cells. BMC Cancer 2008; 8:1–10.
10.   Doden H, Sallam LA, Devendran S, Ly L, Doden G, Daniel SL, Ridlon JM. Metabolism of oxo-bile acids and characterization of recombinant 12α-hydroxysteroid dehydrogenases from bile acid 7α-dehydroxylating human gut bacteria. Appl Environ Microbiol 2018; 84:e00235-18.
11.   Macdonald IA, Jellett JF, Mahony DE. 12alpha-Hydroxysteroid dehydrogenase from *Clostridium* group P strain C48-50 ATCC #29733: partial purification and characterization. J Lipid Res 1979; 20:234–9.
12.   Harris JN, Hylemon PB. Partial purification and characterization of NADP-dependent 12α-hydroxysteroid dehydrogenase from *Clostridium leptum*. Biochim Biophys Acta 1978; 528:148–57.
13.   Macdonald IA, Jellett JF, Mahony DE, Holdeman L V. Bile salt 3α- and 12α-hydroxysteroid dehydrogenases from *Eubacterium lentum* and related organisms. Appl Environ Microbiol 1979; 37:992–1000.
14.   Macdonald IA, Meier EC, Mahony DE, Costain GA. 3α-, 7α- And 12α-hydroxysteroid dehydrogenase activities from *Clostridium perfringens*. Biochim Biophys Acta 1976; 450:142–53.
15.   Mythen SM, Devendran S, Méndez-García C, Cann I, Ridlon JM. Targeted synthesis and characterization of a gene cluster encoding NAD(P)H-dependent 3α-, 3β-, and 12α-hydroxysteroid dehydrogenases from *Eggerthella* CAG:298, a gut metagenomic sequence. Appl Environ Microbiol 2018; 84:e02475-17.

16. Ali SS, Kuksis A, Beveridge JM. Excretion of bile acids by three men on corn oil and butterfat diets. Can J Biochem 1966; 44:1377–88.
17. Ali SS, Kuksis A, Beveridge JM. Excretion of bile acids by three men on a fat-free diet. Can J Biochem 1966; 44:957–69.
18. Eneroth P, Gordon B, Ryhage R, Sjövall J. Identification of mono- and dihydroxy bile acids in human feces by gas-liquid chromatography and mass spectrometry. J Lipid Res 1966; 7:511–23.
19. Edenharder R, Schneider J. 12β-Dehydrogenation of bile acids by *Clostridium paraputrificum*, *C. tertium*, and *C. difficile* and epimerization at carbon-12 of deoxycholic acid by cocultivation with 12α-dehydrogenating *Eubacterium lentum*. Appl Environ Microbiol 1985; 49:964–8.
20. Edenharder R, Pfützner A. Characterization of NADP-dependent 12β-hydroxysteroid dehydrogenase from *Clostridium paraputrificum*. Biochim Biophys Acta 1988; 962:362–70.
21. Penning TM. Human hydroxysteroid dehydrogenases and pre-receptor regulation: Insights into inhibitor design and evaluation. J Steroid Biochem Mol Biol 2011; 125:46–56.
22. Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J Mol Biol 2001; 305:567–80.
23. Kiu R, Caim S, Alcon-Giner C, Belteki G, Clarke P, Pickard D, Dougan G, Hall LJ. Preterm infant-associated *Clostridium tertium*, *Clostridium cadaveris*, and *Clostridium paraputrificum* strains: Genomic and evolutionary insights. Genome Biol Evol 2017; 9:2707–14.
24. Muñoz M, Restrepo-Montoya D, Kumar N, Iraola G, Herrera G, Ríos-Chaparro DI, Díaz-Arévalo D, Patarroyo MA, Lawley TD, Ramírez JD. Comparative genomics identifies potential virulence factors in *Clostridium tertium* and *C. paraputrificum*. Virulence 2019; 10:657–76.
25. Liu L, Aigner A, Schmid RD. Identification, cloning, heterologous expression, and characterization of a NADPH-dependent 7β-hydroxysteroid dehydrogenase from *Collinsella aerofaciens*. Appl Microbiol Biotechnol 2011; 90:127–35.
26. Wegner K, Just S, Gau L, Mueller H, Gérard P, Lepage P, Clavel T, Rohn S. Rapid analysis of bile acids in different biological matrices using LC-ESI-MS/MS for the investigation of bile acid transformation by mammalian gut bacteria. Anal Bioanal Chem 2017; 409:1231–45.
27. Sohn JH, Kwon KK, Kang JH, Jung HB, Kim SJ. *Novosphingobium pentaromativorans* sp. nov., a high-molecular-mass polycyclic aromatic hydrocarbon-degrading bacterium isolated from estuarine sediment. Int J Syst Evol Microbiol 2004; 54:1483–7.
28. Hashimoto T, Onda K, Morita T, Luxmy BS, Tada K, Miya A, Murakami T. Contribution of the estrogen-degrading bacterium *Novosphingobium* sp. strain JEM-1 to estrogen removal in wastewater treatment. J Environ Eng 2010; 136:890–6.
29. Gan HM, Hudson AO, Rahman AYA, Chan KG, Savka MA. Comparative genomic analysis of six bacteria belonging to the genus *Novosphingobium*: Insights into marine adaptation, cell-cell signaling and bioremediation. BMC Genomics 2013; 14:431.
30. Yu J, Feng Q, Wong SH, Zhang D, Yi Liang Q, Qin Y, Tang L, Zhao H, Stenvang J, Li Y, et al. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. Gut 2017; 66:70–8.

31. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, Amiot A, Böhm J, Brunetti F, Habermann N, et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. Mol Syst Biol 2014; 10:1–18.

32. Vogtmann E, Hua X, Zeller G, Sunagawa S, Voigt AY, Hercog R, Goedert JJ, Shi J, Bork P, Sinha R. Colorectal cancer and the human gut microbiome: Reproducibility with whole-genome shotgun sequencing. PLoS One 2016; 11:e0155362.

33. Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, Lan Z, Zhang D, Xia H, Xu X, Jie Z, et al. Gut microbiome development along the colorectal adenoma-carcinoma sequence. Nat Commun 2015; 6:1–13.

34. Chung WSF, Meijerink M, Zeuner B, Holck J, Louis P, Meyer AS, Wells JM, Flint HJ, Duncan SH. Prebiotic potential of pectin and pectic oligosaccharides to promote anti-inflammatory commensal bacteria in the human colon. FEMS Microbiol Ecol 2017; 93:1–9.

35. Liu S, Zhao W, Liu X, Cheng L. Metagenomic analysis of the gut microbiome in atherosclerosis patients identify cross-cohort microbial signatures and potential therapeutic target. FASEB J 2020; 34:14166–81.

36. Aigner A, Gross R, Schmid R, Braun M, Mauer S. Novel 12α-hydroxysteroid dehydrogenases, production and use therof. 2011; :US patent 20110091921A1.

37. Baron SF, Franklund C V., Hylemon PB. Cloning, sequencing, and expression of the gene coding for bile acid 7α-hydroxysteroid dehydrogenase from *Eubacterium* sp. strain VPI 12708. J Bacteriol 1991; 173:4558–69.

38. Yoshimoto T, Higashi H, Kanatani A, Lin XS, Nagai H, Oyama H, Kurazono K, Tsuru D. Cloning and sequencing of the 7α-hydroxysteroid dehydrogenase gene from *Escherichia coli* HB101 and characterization of the expressed enzyme. J Bacteriol 1991; 173:2173–9.

39. Lee JY, Arai H, Nakamura Y, Fukiya S, Wada M, Yokota A. Contribution of the 7β-hydroxysteroid dehydrogenase from *Ruminococcus gnavus* N53 to ursodeoxycholic acid formation in the human colon. J Lipid Res 2013; 54:3062–9.

40. Ferrandi EE, Bertolesi GM, Polentini F, Negri A, Riva S, Monti D. In search of sustainable chemical processes: Cloning, recombinant expression, and functional characterization of the 7α- and 7β-hydroxysteroid dehydrogenases from *Clostridium absonum*. Appl Microbiol Biotechnol 2012; 95:1221–33.

41. Devlin AS, Fischbach MA. A biosynthetic pathway for a prominent class of microbiota-derived bile acids. Nat Chem Biol 2015; 11:685–90.

42. Doden HL, Pollet RM, Mythen SM, Wawrzak Z, Devendran S, Cann I, Koropatkin NM, Ridlon JM. Structural and biochemical characterization of 20β-hydroxysteroid dehydrogenase from *Bifidobacterium adolescentis* strain L2-32. J Biol Chem 2019; 294:12040–53.

43. Devendran S, Méndez-García C, Ridlon JM. Identification and characterization of a 20β-HSDH from the anaerobic gut bacterium *Butyricicoccus desmolans* ATCC 43058. J Lipid Res 2017; 58:916–25.

44. Ridlon JM, Ikegawa S, Alves JMP, Zhou B, Kobayashi A, Iida T, Mitamura K, Tanabe G, Serrano M, De Guzman A, et al. *Clostridium scindens*: a human gut microbe with a high potential to convert glucocorticoids into androgens. J Lipid Res 2013; 54:2437–49.

45. Bernardi R, Doden H, Melo M, Devendran S, Pollet R, Mythen S, Bhowmik S, Lesley S, Cann I, Luthey-Schulten Z, et al. Bacteria on steroids: the enzymatic mechanism of an NADH-dependent dehydrogenase that regulates the conversion of cortisol to androgen in the gut microbiome. 2020; :bioRxiv 2020.06.12.149468. Available from: https://doi.org/10.1101/2020.06.12.149468

46. Fahrbach M, Kuever J, Meinke R, Kämpfer P, Hollender J. *Denitratisoma oestradiolicum* gen. nov., sp. nov., a 17β-oestradiol-degrading, denitrifying betaproteobacterium. Int J Syst Evol Microbiol 2006; 56:1547–52.

47. Ricaboni D, Mailhe M, Vitton V, Cadoret F, Fournier PE, Raoult D. *'Intestinibacillus massiliensis'* gen. nov., sp. nov., isolated from human left colon. New Microbes New Infect 2017; 17:18–20.

48. Morris DJ, Ridlon JM. Glucocorticoids and gut bacteria: "The GALF Hypothesis" in the metagenomic era. Steroids 2017; 125:1–13.

49. Sherrod JA, Hylemon PB. Partial purification and characterization of NAD-dependent 7α-hydroxysteroid dehydrogenase from *Bacteroides thetaiotaomicron*. 1977; 486:351–8.

50. Edenharder R, Mielek K. Epimerization, oxidation and reduction of bile acids by *Eubacterium lentum*. Syst Appl Microbiol 1984; 5:287–98.

51. Snyder ML. The serologic agglutinatin of the obligate anaerobes *Clostridium paraputrificum* (Beinstock) and *Clostridium capitovalis* (Snyder and Hall). J Bacteriol 1936; 32:401–10.

52. Harris SC, Devendran S, Méndez- García C, Mythen SM, Wright CL, Fields CJ, Hernandez AG, Cann I, Hylemon PB, Ridlon JM. Bile acid oxidation by *Eggerthella lenta* strains C592 and DSM 2243. Gut Microbes 2018; 9:523–39.

53. Filling C, Berndt KD, Benach J, Knapp S, Prozorovski T, Nordling E, Ladenstein R, Jörnvall H, Oppermann U. Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. J Biol Chem 2002; 277:25677–84.

54. Grimm C, Maseri E, Möbusi E, Klebe G, Reuter K, Ficner R. The crystal structure of 3α-hydroxysteroid dehydrogenase/carbonyl reductase from *Comamonas testosteroni* shows a novel oligomerization pattern within the short chain dehydrogenase/reductase family. J Biol Chem 2000; 275:41333–9.

55. Savino S, Ferrandi EE, Forneris F, Rovida S, Riva S, Monti D, Mattevi A. Structural and biochemical insights into 7β-hydroxysteroid dehydrogenase stereoselectivity. Proteins 2016; 84:859–65.

56. Jang LG, Choi G, Kim SW, Kim BY, Lee S, Park H. The combination of sport and sport-specific diet is associated with characteristics of gut microbiota: An observational study. J Int Soc Sports Nutr 2019; 16:1–10.

57. Ma S, You Y, Huang L, Long S, Zhang J, Guo C, Zhang N, Wu X, Xiao Y, Tan H. Alterations in gut microbiota of gestational diabetes patients during the first trimester of pregnancy. Front Cell Infect Microbiol 2020; 10:1–14.

58. Amaruddin AI, Hamid F, Koopman JPR, Muhammad M, Brienen EAT, van Lieshout L, Geelen AR, Wahyuni S, Kuijper EJ, Sartono E, et al. The bacterial gut microbiota of schoolchildren from high and low socioeconomic status: A study in an urban area of Makassar, Indonesia. Microorganisms 2020; 8:1–12.

59. Doumatey AP, Adeyemo A, Zhou J, Lei L, Adebamowo SN, Adebamowo C, Rotimi CN. Gut microbiome profiles are associated with type 2 diabetes in urban africans. Front Cell Infect Microbiol 2020; 10:1–13.

60. Turpin W, Espin-Garcia O, Xu W, Silverberg MS, Kevans D, Smith MI, Guttman DS, Griffiths A, Panaccione R, Otley A, et al. Association of host genome with intestinal microbial composition in a large healthy cohort. Nat Genet 2016; 48:1413–7.

61. Labbé A, Ganopolsky JG, Martoni CJ, Prakash S, Jones ML. Bacterial bile metabolising gene abundance in Crohn's, ulcerative colitis and type 2 diabetes metagenomes. PLoS One 2014; 9:e115175.

62. Lepercq P, Gérard P, Béguet F, Raibaud P, Grill JP, Relano P, Cayuela C, Juste C. Epimerization of chenodeoxycholic acid to ursodeoxycholic acid by *Clostridium baratii* isolated from human feces. FEMS Microbiol Lett 2004; 235:65–72.

63. Horinouchi M, Hayashi T, Koshino H, Malon M, Yamamoto T, Kudo T. Identification of genes involved in inversion of stereochemistry of a C-12 hydroxyl group in the catabolism of cholic acid by *Comamonas testosteroni* TA441. J Bacteriol 2008; 190:5545–54.

64. Holert J, Kulić Ž, Yücel O, Suvekbala V, Suter MJF, Möller HM, Philipp B. Degradation of the acyl side chain of the steroid compound cholate in *Pseudomonas* sp. strain Chol1 proceeds via an aldehyde intermediate. J Bacteriol 2013; 195:585–95.

65. Edenharder R, Pfützner M, Hammann R. NADP-dependent 3β-, 7α- and 7β-hydroxysteroid dehydrogenase activities from a lecithinase-lipase-negative *Clostridium* species 25.11.c. Biochim Biophys Acta 1989; 1002:37–44.

66. Ridlon JM, Devendran S, Alves JM, Doden H, Wolf PG, Pereira G V., Ly L, Volland A, Takei H, Nittono H, et al. The 'in vivo lifestyle' of bile acid 7α-dehydroxylating bacteria: comparative genomics, metatranscriptomic, and bile acid metabolomics analysis of a defined microbial community in gnotobiotic mice. Gut Microbes 2020; 11:381–404.

67. Wahlström A, Kovatcheva-Datchary P, Ståhlman M, Bäckhed F, Marschall HU. Crosstalk between bile acids and gut microbiota and its impact on farnesoid X receptor signalling. Dig Dis 2017; 35:246–50.

68. Song X, Sun X, Oh SF, Wu M, Zhang Y, Zheng W, Geva-Zatorsky N, Jupp R, Mathis D, Benoist C, et al. Microbial bile acid metabolites modulate gut RORγ+ regulatory T cell homeostasis. Nature 2020; 577:410–5.

69. Campbell C, McKenney PT, Konstantinovsky D, Isaeva OI, Schizas M, Verter J, Mai C, Jin WB, Guo CJ, Violante S, et al. Bacterial metabolism of bile acids promotes generation of peripheral regulatory T cells. Nature 2020; 581:475–9.

70. Sutherland JD, Williams CN. Bile acid induction of 7α- and 7β-hydroxysteroid dehydrogenases in *Clostridium limosum*. J Lipid Res 1985; 26:344–50.

71. Hofmann AF, Roda A. Physicochemical properties of bile acids and their relationship to biological properties: An overview of the problem. J Lipid Res 1984; 25:1477–89.

72. Goossens J, Bailly C. Ursodeoxycholic acid and cancer: From chemoprevention to chemotherapy. Pharmacol Ther 2019; 203:107396.

73. Liu Y, Rong Z, Xiang D, Zhang C, Liu D. Detection technologies and metabolic profiling of bile acids: A comprehensive review. Lipids Health Dis 2018; 17:121.

74. Franco P, Porru E, Fiori J, Gioiello A, Cerra B, Roda G, Caliceti C, Simoni P, Roda A. Identification and quantification of oxo-bile acids in human faeces with liquid chromatography–mass spectrometry: A potent tool for human gut acidic sterolbiome studies. J Chromatogr A 2019; 1585:70–81.

75. Hofmann AF, Hagey LR, Krasowski MD. Bile salts of vertebrates: Structural variation and possible evolutionary significance. J Lipid Res 2010; 51:226–46.

76. Chang FC. Potential bile acid metabolites. 5.1 12β-hydroxy acids by stereoselective reduction. Synth Commun 1981; 11:875–9.
77. Iida T, Momose T, Chang FC, Nambara T. Potential bile acid metabolites. XI. Syntheses of stereoisomeric 7,12-dihydroxy-5α-cholanic acids. Chem Pharm Bull 1986; 34:1934–8.
78. Borgström B, Barrowman J, Krabisch L, Lindström M, Lillienau J. Effects of cholic acid, 7β-hydroxy- and 12β-hydroxy-isocholic acid on bile flow, lipid secretion and bile acid synthesis in the rat. Scand J Clin Lab Invest 1986; 46:167–75.
79. Eneroth P. Thin-layer chromatography of bile acids. J Lipid Res 1963; 4:11–6.
80. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. BLAST+: Architecture and applications. BMC Bioinformatics 2009; 10:1–9.
81. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 2004; 32:1792–7.
82. Stamatakis A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014; 30:1312–3.
83. Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R. Dendroscope: An interactive viewer for large phylogenetic trees. BMC Bioinformatics 2007; 8:460.
84. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, Beghini F, Manghi P, Tett A, Ghensi P, et al. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. Cell 2019; 176:649–62.
85. Eddy SR. Accelerated profile HMM searches. PLoS Comput Biol 2011; 7:e1002195.

CHAPTER 4

STRUCTURAL AND BIOCHEMICAL CHARACTERIZATION OF 20β-HYDROXYSTEROID DEHYDROGENASE FROM *BIFIDOBACTERIUM ADOLESCENTIS* STRAIN L2-32[1]

ABSTRACT

Anaerobic bacteria inhabiting the human gastrointestinal tract have evolved various enzymes that modify host-derived steroids. The bacterial steroid-17,20-desmolase pathway cleaves the cortisol side chain, forming pro-androgens predicted to impact host physiology. Bacterial 20β-hydroxysteroid dehydrogenase (20β-HSDH) regulates cortisol side-chain cleavage by reducing the C-20 carboxyl group on cortisol, yielding 20β-dihydrocortisol. Recently, the gene encoding 20β-HSDH in *Butyricicoccus desmolans* ATCC 43058 was reported, and a nonredundant protein search yielded a candidate *20β-HSDH* gene in *Bifidobacterium adolescentis* strain L2-32. *B. adolescentis* 20β-HSDH could regulate cortisol side-chain cleavage by limiting pro-androgen formation in bacteria such as *Clostridium scindens* and 21-dehydroxylation by *Eggerthella lenta*. Here, the putative *B. adolescentis* 20β-HSDH was cloned, overexpressed, and purified. 20β-HSDH activity was confirmed through whole-cell and pure enzymatic assays, and it is specific for cortisol. Next, we solved the structures of recombinant 20β-HSDH in both the apo- and holo-forms at 2.0–2.2Å resolutions, revealing close overlap

except for rearrangements near the active site. Interestingly, the structures contain a large, flexible N-terminal region that was investigated by gel-filtration chromatography and CD spectroscopy. This extended N terminus is important for protein stability because deletions of varying lengths caused structural changes and reduced enzymatic activity. A nonconserved extended N terminus was also observed in several short-chain dehydrogenase/reductase family members. *B. adolescentis* strains capable of 20β-HSDH activity could alter glucocorticoid metabolism in the gut and thereby serve as potential probiotics for the management of androgen-dependent diseases.

INTRODUCTION

Hydroxysteroid dehydrogenases (HSDH) are pyridine nucleotide-dependent enzymes classified into three superfamilies, including aldo-keto reductase (AKR), long/medium-chain dehydrogenase/reductase (MDR), and short-chain dehydrogenase/ reductase (SDR), which interconvert between enol and keto forms of hydroxyl groups on diverse small molecules.[1,2] The gut microbiome is exposed to a mixture of glucocorticoids and bile acids. Recent investigations identified bacterial strains of diverse taxa expressing stereo- and regio-specific HSDHs in the SDR and AKR families capable of oxidation and epimerization of bile acid hydroxyl groups at C-3, C-7, and C-12.[3–5] Cortisol and corticosterone are human steroid hormone glucocorticoids that can be metabolized by both host and bacterial enzymes.[6] Cortisol is the major glucocorticoid in humans and is present in concentrations ranging from 5 to 20 µg/dl throughout the day.[7] The various bacterial reactions altering steroid structure impact the host by generating potent signaling molecules capable of modulating host physiology.

Recent attention has been focused on the bacterial steroid- 17,20-desmolase pathway (*desABCD*) converting cortisol to 11β-hydroxyandrostenedione (11β-OHAD) in *Clostridium scindens* and the zinc-dependent MDR family 20α-HSDH (*desC*), a predicted enzymatic regulator of the pathway (**Figure 4.1**).[2,8,9] 11β- OHAD is a pro-androgen whose formation by gut bacteria[10] as well as urinary tract bacteria[9] may contribute to castration-resistant prostate cancer[11] and polycystic ovary syndrome[12]. The bacterial steroid-17,20-desmolase pathway has also been reported in *Clostridium cadavaris* and *Butyricicoccus desmolans* (formerly *Eubacterium desmolans*) and is associated with NADH-dependent cortisol 20β-HSDH (*desE*) activity.[9,13] Previous studies also identified strains of *Bifidobacterium adolescentis* isolated from human and rat fecal dilutions expressing cortisol 20β- HSDH activity alone.[14] Recently, the gene encoding gut bacterial NADH-dependent 20β-HSDH activity was reported in *B. desmolans.*[9] Phylogenetic analysis identified numerous *Bifidobacterium* spp. strains, including strains of *B. adolescentis*, encoding putative cortisol 20β-HSDH.[9]

Gut bacterial 20β-HSDH is a member of the SDR family of proteins.[9] The SDR family constitutes one of the largest protein superfamilies distributed across all three domains of life[15], comprising functionally diverse members.[16] SDR family enzymes include several Enzyme Commission (EC) classifications, including lyases and isomerases, along with pyridine nucleotide-dependent oxidoreductases, which constitute the majority of enzymes. Currently, in Uniprot there are 266,639 SDR entries, and members typically share only ~20–30% amino acid residue identity in pairwise comparisons. Most SDR family members are 250–350 amino acids in length, with secondary structures consisting of a Rossmann fold domain containing 6–7 β-strands flanked by three α-helices.[17] Members usually contain a conserved Gly-rich sequence critical in NAD(P)(H) binding, as well as a catalytic tetrad composed of Ser, Tyr, Lys, and Asn residues.[15]

The SDR family enzymes vary in substrate recognition, including sugars, dyes, prostaglandins, porphyrins, alcohols, and steroids.[16] Consequently, substrate-binding residues are not highly conserved. Differentiating HSDHs from other functional classes of SDR family enzymes is important for identifying steroid biochemical pathways in gut metagenomes. At present, this task is impeded by a lack of primary amino acid sequences for several HSDHs whose activities have been reported in bacterial isolates[18], but also the need to determine residues important in differentiating HSDHs into functional categories (*i.e.* 7α-HSDH from 7β-HSDH and both from 20β-HSDH).

The first SDR family enzyme with HSDH activity crystallized was 3α,20β-HSDH from the soil bacterium *Streptomyces hydrogenans.*[19] The 3α,20β-HSDH reversibly oxidizes the 3β-hydroxyl and 20β-hydroxyl groups of androgens and progestins. The *S. hydrogenans* 3α,20β-HSDH shares 31% amino acid sequence identity with the 20β-HSDH from *B. adolescentis*, which does not appear to function as a dual 3β-HSDH.[9] Humans express glucocorticoid 20β-HSDH activity as well via carbonyl reductase 1 (CBR1).[20] 20β-Dihydrocortisol is detected in the urine of patients with Cushing's disease[21] and the plasma of hypertensive individuals[22]. It is thus important to obtain a mechanistic understanding of gut bacterial 20β- HSDH because this enzyme is capable of generating 20β-dihydrocortisol and regulating both major microbial side-chain cleavages of cortisol: steroid-17,20-desmolase activity resulting in pro-androgens[6], and cortisol 21-dehydroxylation producing 21-deoxycortisol, a hypertensive compound (**Figure 4.1**).[6,23] By understanding the structure and function of 20β-HSDH, the ability of *B. adolescentis* strain L2-32 to shift glucocorticoid metabolism in the gut may be improved. 20β-HSDH activity could reduce formation and absorption of 11β-hydroxyandrogens and 21-deoxycortisol, which are implicated in chronic human disease.

Herein, we report the structures of NADH-dependent 20β- HSDH from *B. adolescentis* strain L2-32 at 2.2 Å for the apo-form and 2.0 Å for the binary complex with NADH. Through site-directed mutagenesis and binding-affinity studies, we have identified residues comprising the substrate-binding pocket and propose a catalytic mechanism for NADH-dependent C-20 reduction. Along, flexible N-terminal domain involved in overall 20β-HSDH stability was discovered, having widespread implications for other members of the diverse SDR family.

RESULTS

B. adolescentis expresses 20β-HSDH activity

The 20β-HSDH sequence from *B. desmolans* ATCC 43058 was used to query the nonredundant protein sequences database (BLASTP).[9] An annotated SDR family oxidoreductase (WP_003810233.1) from *B. adolescentis* was found with a sequence identity of 59%. This protein was selected for further study because it has the potential to be of great pharmaceutical interest, as strains of *B. adolescentis* have potential probiotic function, *i.e.* capable of conferring beneficial effects on the host. 20β-HSDH catalyzes the conversion of cortisol to 20β-dihydrocortisol in the *boxed reaction* depicted in **Figure 4.1**, and to determine whether *B. adolescentis* strain L2-32 possesses 20β- HSDH activity, whole-cell extracts were screened for the activity at 48 h. The result, as shown in **Figure 4.2**, was net conversion of cortisol to 20β-dihydrocortisol. As controls, *B. desmolans* ATCC 43058 and *C. scindens* ATCC 35704 were both shown to metabolize cortisol, as well as 20β-dihydrocortisol and 20α-dihydrocortisol, respectively. 20α-Dihydrocortisol is not metabolized by *B. adolescentis* strain L2-32 or *B. desmolans* ATCC 43058, both of which possess *desE*, a gene previously shown to encode NAD(H)-dependent 20β-HSDH in *B. desmolans* and *Bifidobacterium scardovii*.[9] 20β-

Dihydrocortisol is not a metabolite for *C. scindens* ATCC 35704, which was previously shown to harbor the *desC* gene, expressing NAD(H)-dependent 20α-HSDH.[8] The *desAB* genes, encoding steroid- 17,20-desmolase, are present in both *C. scindens* ATCC 35704 and *B. desmolans*ATCC43058. As expected, net conversions of 20α-dihydrocortisol to 11β-OHAD and 20β-dihydrocortisol to 11β-OHAD were observed in cultures of *C. scindens* ATCC 35704 and *B. desmolans* ATCC 43058, respectively. In contrast, *B. adolescentis* strain L2-32 lacks the *desAB* genes, and incubation with cortisol resulted in net conversion to 20β-dihydrocortisol under anaerobic conditions.

Biochemical characterization of recombinant 20β-HSDH

To determine whether the SDR family oxidoreductase identified in *B. adolescentis* (WP_003810233.1) encodes a 20β- HSDH (DesE), the gene was cloned into the pET-28a(+) expression vector. The N-terminal His6-tagged recombinant protein was overexpressed in *Escherichia coli* BL21(DE3)-RIPL and purified via metal-affinity chromatography as a single band on SDS-PAGE, and the molecular mass was similar to the theoretical calculated subunit molecular mass of 31.7 kDa (**Figure 4.3***A*). The native molecular mass of the recombinant 20β-HSDH (r20β-HSDH) protein was determined by gel-filtration chromatography with reference to protein standards. r20β-HSDH displayed an elution volume of 13.9 ml, corresponding to a molecular mass of 130.5 β 2.8 kDa (**Figure 4.3***B*). Thus, the gel filtration data, together with the subunit molecular mass determined by SDS-PAGE, suggest that r20β-HSDH forms a homotetrameric quaternary structure in solution. A tetrameric structure was also seen in the structural data, and the proposed tetramer is shown in **Figure 4.5***B*. The purified r20β-HSDH

exhibited a pH optimum of 5.0 in the reductive direction, and in the oxidative direction the optimum was pH 5.5 (**Figure 4.4***A*).

The Michaelis-Menten saturation curve was determined from the primary reaction of cortisol reduction performed by r20β-HSDH (**Figure 4.4***B*). The saturation curve was used to estimate the kinetic constants that are reported in **Table 4.1**. The purified r20β-HSDH has a K_m of 24.07 μM when cortisol is the substrate. The V_{max}, k_{cat}, and k_{cat}/K_m values for cortisol were determined to be 21.08 μmol•min^{-1}•μmg^{-1}, 668.3 min^{-1}, and 27.77 μM^{-1}•min^{-1}, respectively. Substrate-specificity studies revealed that cortisol was the primary substrate utilized by 20β-HSDH (**Table 4.2**). Hydroxyl groups on rings C and D are important for substrate specificity by r20β-HSDH. We observed an ~34% reduction in relative activity with 11-deoxycortisol and an ~61% reduction in activity with corticosterone, which lacks a 17-hydroxyl group. Reduction of the 3-oxo-Δ^4 to 3α-, 5β-H (tetrahydrocortisol) results in an ~56% reduction in activity relative to cortisol. It was previously determined that recombinant steroid-17,20-desmolase from *C. scindens* ATCC 35704 also showed significant reduction in side-chain cleavage activity with tetrahydrocortisol[8] but not allotetrahydrocortisol (3α-,5α-H). Consistent with the whole-cell conversion data from *B. adolescentis* strain L2-32 (**Figure 4.2***A*), the oxidative direction with 20β-dihydrocortisol and NAD$^+$ has only 0.23% of the relative activity compared with cortisol and NADH. As expected, 20α- dihydrocortisol did not serve as a substrate, nor did we detect formation of 20α-derivatives with purified recombinant r20β-HSDH in the presence of NADH. NADP(H) was tested, but we did not observe activity with this cofactor.

Overall crystallographic structure of 20β-HSDH

The apo crystal structure of r20β-HSDH was solved using molecular replacement with the structure of an SDR from *Mycobacterium marinum* (PDB code 3R1I) and refined at a resolution of 2.2 Å ($R_w = 0.1960$ and $R_f = 0.2310$) (**Table 4.3**). The tertiary structure contains motifs characteristic of the SDR superfamily, including a Rossmann fold consisting of seven central β-strands (β3–β2–β1–β4–β5–β6–β7) flanked by six α-helices (**Figure 4.5A**, *left*). The structure shows two monomers in the asymmetric unit that align with an RMSD of 0.339 Å. The largest differences between the monomers are the 2 Å shift in the α4 helix near the β4–α4 loop, the length of the β4, β5, and β6 strands, and the α6 helix. These differences are likely due to the flexibility of the adjacent loops that could not be modeled in either monomer and may play a role in substrate binding. The monomer with the most unambiguous density in this region is shown in **Figure 4.5A**. In addition to the Rossman fold, r20β-HSDH contains a partial α-helical N terminus of ~40 residues that extends outward from the core protein structure and interacts with other monomers when the protein is in its tetramer formation (**Figure 4.5**, *A*, *right*, and *B*).

We attempted to crystallize the ternary complex of NADH and cortisol bound to the inactive S181A mutant of r20β-HSDH (**Table 4.1** and **Supplementary Figure 4.1**). Crystals diffracted to 2.0 Å ($R_w = 0.187$ and $R_f = 0.238$) and show NADH clearly bound in the active site, but no density for cortisol is seen (**Supplementary Figure 4.2**, *A* and *B*). These crystals were of the C121 space group, with eight monomers in the asymmetric unit (**Table 4.3**). The monomers showed little variation and aligned with RMSD values ranging from 0.14 to 0.24 Å. The largest difference between monomers is the helical character of the N-terminal extension and a 1 Å shift in a flexible loop (residues 235–245) close to where cortisol is predicted to bind (**Supplementary Figure 4.2C**). This loop is near the α6 helix and was not resolved in the apo structure. The

inability to fully resolve this loop in all monomers suggests flexibility when cortisol is not bound.

The NADH was bound within a deep pocket and makes five side-chain and three peptidyl backbone hydrogen-bonding or electrostatic interactions (**Figure 4.5, _B–D_**). The carboxylic acid of Asp-101 is positioned 2.9 Å away from the adenine nucleotide, whereas the carboxylic acid of Asp-72 is positioned 2.5 and 2.8 Å from the O2 and O3 hydroxyls of the adenine ribose, respectively. The side-chain hydroxyl of Ser-235 is 2.5 Å from the α-phosphate group as well as within 3.3 Å of the carboxamide N of the nicotinamide. The N^ϵ of Lys-204 coordinates the O2 and O3 (3.3 and 3.1 Å) of the nicotinamide ribose, and Tyr-200, part of the catalytic tetrad, is directed toward the N of the pyridine ring and the adjacent ribose (3.4 and 2.6 Å). Additionally, the O3 of the nicotinamide ribose interacts with the peptidic N of Gly-130 (3.5 Å) and peptidic O of Asn-128 (2.7 Å), whereas the nicotinamide carboxamide is positioned 2.9 and 3.0 Å from Val-233 (**Figure 4.5_C_**). van der Waals interactions are seen between NADH and Cys-100, Val-102, the Asn-128 side chain, Val-131, Thr-179, Gly-231, Ile-237, and Phe-238 (_yellow_ in **Figure 4.5_D_**) and, most notably, Gly-48, -51, and -52, which fall within the Gly-rich region characteristic of SDR family enzymes (_pink_ in **Figure 4.5_D_**).

The overall structure and NADH-binding pocket of r20β-HSDH are similar to the NADH-bound structure of 3α,20β-HSDH from _S. hydrogenans_, which is the only structure of an SDR enzyme with known 20β-HSDH activity (**Supplementary Figure 4.2_D_**). The two structures align with an RMSD ~1 Å for 975 atoms. Significant differences are seen in the nicotinamide position and the enzyme's predicted substrate-binding site, especially around the 20β-HSDH α6-helix, likely allowing for the 3α-HSDH activity seen in the _S. hydrogenans_ enzyme (**Supplementary Figure 4.2_D_**). Despite these changes, most of the protein–cofactor

159

interactions are conserved. Notable differences are Arg-16 that interacts with the bis-phosphate in the 3α,20β-HSDH is replaced by a glycine in our structure. In addition, Thr-187 in 3α,20β-HSDH is replaced by Ser-235 in our structure and is too far (>3.5 Å) from the cofactor to make a meaningful interaction. Except for Arg-16 in 3α,20β-HSDH, the Gly-rich region characteristic of SDR enzymes provides a similar docking site for the adenine side of NADH (GAAGGLG in 20β-HSDH and GGARGLG in 3α,20β-HSDH), and the catalytic tetrad is completely conserved (**Supplementary Figure 4.2*D***).

Although the apo structure aligns well with the NADH-bound form (RMSD 0.49–0.56 Å across the NADH-bound monomers), significant rearrangements occur near the active site to accommodate NADH binding (**Supplementary Figure 4.2*E***). The β3 to α3 loop shifts up to 1.4 Å, allowing Asp-101 to interact with the adenine ring. Most striking is the loss of the discrete secondary structure toward the end of β4 and β5 allowing the nicotinamide ribose to sit in the space occupied by the 5th β-sheet in the apo structure. Along with the shortening of β4 and β5, the α4 and α5 helices shift ~2 and 3 Å, respectively. Movement of α5 allows a 3.8 Å shift of Tyr-200, which then interacts with the nicotinamide and the adjacent ribose. Neither the loops between β4 and α4 nor between β5 and α5 were resolved in the apo structure; however, these loops were well-defined in the NADH-bound structure and may help form the binding site for cortisol. Most significantly, the shortening of β4 allows the loop connecting to α5 to swing toward the α6 helix and C terminus of the protein. This movement elongates the binding cleft and positions Ser-181 (Ala in our NADH-bound structure) for potential interaction with cortisol.

Catalytic residues and predicted binding of NADH and cortisol

The conserved catalytic tetrad (NSYK), characteristic of the SDR superfamily, is conserved in 20β-HSDH (*green* in **Supplementary Figure 4.3**). Mutation of the putative catalytic serine (Ser-181) and tyrosine (Tyr-200) to an alanine completely abolished activity based on both spectrophotometric assay and TLC separation of overnight reaction products (**Table 4.1** and **Supplementary Figure 4.1***A*). Wildtype (WT) r20β-HSDH was incubated with cortisol, forming 20β-dihydrocortisol. 11β-OHAD showed no observable conversion, confirming *B. adolescentis* 20β-HSDH does not have 3α-HSDH activity (**Supplementary Figure 4.1***A*). To determine whether cortisol binds to the S181A mutant, we performed isothermal titration calorimetry (ITC) in the presence of NADH (2 mM) or a combination of NADH + cortisol (1mM) (**Supplementary Figure 4.1***B*). NADH bound tightly, with a K_d of 28.97 μM. However, cortisol was shown to bind with much less affinity, suggesting that Ser-181 is important for cortisol binding and explaining the absence of cortisol binding in attempts to obtain the ternary complex. The S183A mutant retained 56% activity relative to the WT. The Y200A mutant showed no activity, and ITC experiments with 2 mM NADH suggest loss of binding, confirming the importance of Tyr-200 in hydrogen bonding to the nicotinamide nitrogen. Comparison of circular dichroism (CD) spectra between WT and 20β-HSDH mutants indicate these alanine substitutions did not significantly affect secondary structural elements (**Supplementary Figure 4.1***C*).

Extended N terminus is important for secondary and quaternary structures

In both the apo and holo structures of r20_-HSDH, we were not able to fully resolve the N terminus (**Figure 4.5***A*, *right*). In the most complete monomers found in the NADH-bound

161

structure, there was no electron density for the first 10 residues, suggesting a flexible N terminus. Other SDR family proteins whose structures have been solved, and which contain a large, extended N-terminal region, include a quinuclidinone reductase (RrQR) from *Rhodotorula rubra*[24] and an *S*-specific carbonyl reductase from *Candida parapsilosis*[25]. There was no electron density for the first six amino acid residues of RrQR, indicating a flexible region. An extended structure was observed from residues 7–19 and observed hydrophobic stacking between Pro-10 and Tyr-229 of the protomer indicating possible quaternary stability.[24] The *S*-specific carbonyl reductase from *C. parapsilosis* is another example of a pyridine nucleotide-dependent SDR member that contains an extended N-terminal peptide. Comparison between WT and the N-terminally truncated mutant form (Asp-31) showed no observable change in catalytic activity, and only a marginal decrease in tetrameric stability in the truncated form.[25] To assess the magnitude of the extended N-terminal abundance throughout the SDR family, a representative full-sequence alignment was produced (**Supplementary Figure 4.4**). The *B. adolescentis* 20β-HSDH sequence and 555 sequences annotated as SDR family with predicted HSDH function and bacterial in origin were obtained from the UniProt database. Within these predicted bacterial HSDHs, about 10% (39/556) have an extended N terminus of greater than or equal to 10 amino acid residues. The extended N-terminal regions were not conserved throughout the alignment. Based on this subset of SDR sequences along with the functionally different quinuclidinone reductase and carbonyl reductase, a significant portion of SDR family members likely have an extended N terminus.

The functional role, if any, of this N-terminal region of 20β- HSDH and the nearest homologs are understudied. We generated the N-terminal truncations of varying lengths (r20β-HSDH$_{\Delta 1-17}$, r20β-HSDH$_{\Delta 1-21}$, and r20β-HSDH$_{\Delta 1-38}$) to determine the effect of truncation on

oligomerization, monomeric secondary structure, and enzymatic activity. The locations of these truncations are indicated on the structure in **Figure 4.5A**, *right*, although there is no density for any residues before 17 in the structure. The purified r20β-HSDH and mutants were separated by gel-filtration chromatography, resulting in a calculated mass of 130.5 ± 2.8 kDa (subunit 31.69 kDa) for r20β-HSDH–WT, 123.2 ± 5.8 (subunit 29.9 kDa) for r20β-HSDH$_{\Delta 1-17}$, 95.1 ± 8.6 (subunit 29.4 kDa) for r20β-HSDH$_{\Delta 1-21}$, and 43.4 ± 5.4 (subunit 28.7 kDa) for r20β-HSDH$_{\Delta 1-38}$ (**Supplementary Figure 4.5**). r20β-HSDHβ1–17 is predicted to be tetrameric much like r20β-HSDH–WT, which was supported by the crystal structure (**Figure 4.5B**). However, r20β-HSDH$_{\Delta 1-21}$ and r20β- HSDH$_{\Delta 1-38}$ were estimated to be trimeric and a mix of dimeric or monomeric forms, respectively.

We next determined the effect of extended N-terminal truncation on enzyme activity and thermal stability. Enzymatic activities of truncated derivatives were compared relative to the WT (**Figure 4.6A**). r20β-HSDH$_{\Delta 1-17}$ retained 62% activity when cortisol was the substrate and NADH co-substrate relative to r20β-HSDH–WT. Temperature-dependent CD of r20β-HSDH$_{\Delta 1-17}$ resulted in a transition temperature (ΔT_m) of 55.70 ± 0.099 °C (**Figure 4.6B**). r20β-HSDH$_{\Delta 1-21}$ retained only 2.9% activity relative to r20β-HSDH–WT with a ΔT_m of 54.7 ± 0.052 °C. r20β-HSDH$_{\Delta 1-38}$ lost complete enzyme activity, and the CD of r20β-HSDH–WT *versus* r20β-HSDH$_{\Delta 1-38}$ resulted in a shift in transition temperature from 56.16 ± 0.10 to 43.3 ± 0.11 °C. These results indicate an important role of the 38-residue extended N terminus in protein stability as deletion of the first 21 residues effectively abolished enzyme activity with a 1.45 °C shift in T_m.

CD in the far-UV region was performed to assess the impact of N-terminal truncation on secondary protein structure (**Figure 4.6, C** and **D**). CD spectra were analyzed using DichroWeb,

producing relative composition of α-helices, β-sheets, β-turns, and unordered regions. The relative amount of α-helix was reduced in each of the truncated proteins compared with r20β-HSDH–WT. r20β-HSDH$_{\Delta1-21}$ composition was most different with ~10% increase in β-sheet and β-turn, and 30% decrease in α-helix compared with WT. The percent of unordered regions was largely stable across all truncated proteins (21.3–28.3%). These results indicate an important role of the extended N terminus in overall protein stability illustrated by changes in secondary structure, depletion of enzyme activity, and loss of oligomerization (**Supplementary Figure 4.6**).

DISCUSSION

Bacterial 20β-HSDH is an NADH-dependent oxidoreductase belonging to the SDR family and is expressed by some strains of *B. adolescentis*.[9] The gene encoding NAD(H)-dependent 20β-HSDH from *B. adolescentis* strain L2-32 was cloned, overexpressed, and structurally characterized in the apo and NADH-bound forms. The 20β-HSDH crystal structure had a tetrameric configuration with the catalytic tetrad and coenzyme-binding domains characteristic of the SDR family (**Figure 4.5**).[15] Ser-181 is predicted to stabilize cortisol, and mutation to an alanine results in loss of substrate binding. The enzymatic reaction is proposed to initiate as a proton transfer from Ser-181 to the C-20 carbonyl of cortisol. Next, a hydride is predicted to be transferred to C-20 from NADH. From there, a subsequent proton relay likely occurs that follows the order: NADH → Ser-181, Tyr-200 → 2'OH ribose, and bulk solvent → Tyr-200. Determination of the structure of the ternary complex will be critical to confirm the catalytic mechanism and cortisol binding. Multiple attempts to crystallize with both NADH and cortisol bound at varying concentrations were unsuccessful. A possible avenue for future

experiments is to use NAD$^+$ and 20β-dihydrocortisol rather than NADH and cortisol. Alternatively, a cortisol derivative such as carbenoxolone could be utilized. Carbenoxolone was bound in the *S. hydrogenans* 3α,20β-HSDH structure, although this substrate seemed to inhibit NADH binding as well.[26]

Short-chain dehydrogenase/reductase family enzymes are extremely diverse, currently spanning 266,639 entries in the Uniprot database. SDR family proteins recognize diverse substrates, from steroids to sugars, alcohols, and xenobiotics.[16] At present, there are 475 SDR family structure entries in PDB and about 216 of those are HSDHs. To determine how widespread this extended N terminus may be, a full sequence alignment was created composed of bacterial HSDHs in the SDR family from the UniProt database (**Supplementary Figure 4.4**). A total of 555 predicted HSDH sequences were compiled, along with the *B. adolescentis* 20β-HSDH sequence, and 39 of them exhibited an extended N terminus with an unconserved amino acid sequence. Many of these 39 sequences were only annotated based on inferred homology, except for 053547_MYCTU and 3BHDP_RUMGV. The probable SDR dehydrogenase/reductase and the possible 17β-hydroxysteroid dehydrogenase, 053547_MYCTU, were first identified from the complete genome sequence of *Mycobacterium tuberculosis* strain H37Rv, although the function has not been experimentally determined.[27] Notably, 3BHDP_RUMGV in *Ruminococcus gnavus* ATCC 29149 was functionally studied and shown to have bile acid 3β-HSDH activity by TLC and GC coupled to MS.[5] Unfortunately, neither of these proteins have been crystallized and can corroborate the extended N termini demonstrated by sequence alignment. Nonetheless, extended N-terminal regions seem to be frequent (~10%) across putative bacterial HSDHs and are likely prevalent throughout the entire SDR family, as well. Takeshita *et al.*[24] noted the flexible N terminus in a quinuclidinone reductase, but it was not

studied in depth (PDB code 4O0L, **Supplementary Figure 4.3**). Zhang *et al.*[35] investigated a

carbonyl reductase extended N terminus in more detail and concluded that it conferred some

stability to the oligomer; however it did not impact enzymatic activity (PDB code 3CTM,

Supplementary Figure 4.3). This contrasts the extensive analysis performed in this study on the

B. adolescentis 20β-HSDH N terminus, which displays complete abolition of activity when fully

truncated, as well as marked changes in oligomerization and secondary structure.

Supplementary Figure 4.6 represents a potential model for the effect of N-terminal truncation

on r20β-HSDH structure/function where truncation of the first 17 amino acid residues resulted in

40% loss of activity relative to WT while retaining a tetrameric structure. Truncation of the next

four residues (r20β- HSDH$_{\Delta1-21}$) resulted in a precipitous drop in enzymatic activity to 2% of

WT and exhibited a trimeric configuration. Finally, the full truncation of the 38-residue extended

N terminus resulted in complete loss of activity and a mixed dimeric and monomeric form.

Future studies could utilize targeted site-directed mutagenesis to determine which residues along

the flexible N terminus, specifically between residues 17 and 21, are essential for stability. The

role of the N-terminal region in 20β-HSDH stability and activity could have broad impacts

across proteins of diverse functional classes within the vast SDR family.

The corticosteroids β-cortol and β-cortolone have been identified in human urine

samples.[28,29] β-Cortol is formed from tetrahydrocortisol, and β-cortolone is produced from

tetrahydrocortisone, both by the proposed action of a human 20β-HSDH.[28] Recent work

identified at least one enzyme, CBR1, with NADPH-dependent 20β-HSDH activity.[20] It was

determined that 20β-dihydrocortisol displayed weak glucocorticoid receptor agonism, indicating

that 20β-dihydrocortisol formation by gut bacteria may have important physiological effects on

the host. In zebrafish, *Danio rerio*, the gene encoding 20β-HSDH is distinct from that in

mammals.[30] Zebrafish 20β-HSDH type 2 catalyzes the conversion of cortisone to 20β-hydroxycortisone, which is important in zebrafish stress response through inactivation and excretion of both endogenous and exogenous cortisol.[30] Importantly, bacterial 20β-HSDH enzymes are becoming more well-studied, as evidenced by previous work in *B. desmolans*[9] and the current study in *B. adolescentis*. Further work will be needed to determine the contribution of gut bacteria to formation of 20β-derivatives of human glucocorticoids.

B. adolescentis is among the first bacteria to colonize the gastrointestinal tract of breast-fed infants; the species was first isolated from the stool of a healthy 2-year-old infant in 1996.[31,32] The presence of *B. adolescentis* has been associated with numerous health benefits; however, the mechanisms by which this species and other bifidobacteria confer these benefits on the host have yet to be fully understood. There are currently 31 known species of bifidobacteria[33] with few among a limited selection of commonly used probiotic organisms.[34] *B. adolescentis* strain L2-32 encoding a 20β-HSDH has the potential to be a novel therapeutic probiotic based on the important regulatory role of 20β-HSDH in the steroid-17,20-desmolase pathway.[9]

Bacterial steroid-17,20-desmolase (DesAB) converts cortisol into 11β-OHAD (**Figure 4.1**).[8] This compound is pro-androgenic due to the ability of both host[35] and microbial[6] enzymes to biotransform it into 11-ketotestosterone (11-KT). The androgen 11-KT has been shown previously to activate androgen receptor (AR) with similar effectiveness to testosterone itself.[11] AR is a ligand-dependent transcription factor that regulates genes for cell proliferation and death.[36] This recently discovered pro-androgenic microbe-mediated cortisol metabolism could directly impact androgen-dependent diseases, such as castration-resistant prostate cancer and polycystic ovary syndrome.[12]

Bacterial 20β-HSDH can be thought of as acting as an enzymatic "switch" by converting cortisol to 20β-dihydrocortisol, either regulating bacterial cortisol 21-dehydroxylation or effectively blocking the steroid-17,20-desmolase pathway (**Figure 4.1**).[9] *B. desmolans* and a urinary isolate, *Propionimicrobium lymphophilum*, have both 20β-HSDH and steroid-17,20-desmolase activity[9], whereas *C. scindens* expresses steroid-17,20- desmolase and 20α-HSDH[8]. *B. adolescentis* encodes only 20β-HSDH; therefore, *B. adolescentis* could be a potential inter-species modulator in gut environments with steroid-17, 20-desmolase and 20α-HSDH activities. Another competing cortisol biotransformation is 21-dehydroxylase activity in *Eggerthella*.[37] *E. lenta* lacks both 20α-HSDH and 20β-HSDH, and modification of the cortisol/corticosterone side chain blocks 21-dehydroxylation.[38] 21-Dehydroxylation of cortisol produces 21-deoxycortisol, a potent inhibitor of the host kidney 11β-hydroxysteroid dehydrogenase-2 (11β-HSD2) enzyme.[23] Inhibition of 11β-HSD2 was shown to cause severe hypertension in humans with an inactivating mutation in this gene.[6] Thus, a *B. adolescentis* probiotic with the 20β-HSDH "switch" could alleviate the inhibitory effects of 21-deoxycortisol by shifting cortisol metabolism toward 20β-dihydrocortisol. The *B. adolescentis* 20β-HSDH is about 400 times more active in the reductive direction with cortisol as substrate, the preferred direction for a probiotic that seeks to reduce androgen excess or concentrations of 21-deoxycortisol. However, the solved crystal structure supports targeted mutagenesis of binding residues to optimize activity in either direction. Future *in vivo* investigations are needed to assess the efficacy and safety of 20β-HSDH–encoding *B. adolescentis* strain L2-32 as a probiotic organism.

MATERIALS AND METHODS

Bacterial strains, enzymes, reagents, and materials

$B.\ adolescentis$, strain L2-32, was purchased through BEI Resources (catalog no. HM-633). Turbo Competent *E. coli* DH5α (high efficiency) cells were purchased from New England Biolabs (Ipswich, MA). BL21-CodonPlus (DE3)-RIPL–competent *E. coli* cells were purchased from Agilent Technologies (catalog no. 230280). Phusion High-Fidelity DNA polymerases, restriction endonucleases, and T4 DNA ligase were purchased from New England Biolabs. Isopropyl β-$_D$-1-thiogalactopyranoside (IPTG) was purchased from Gold Biotechnology. TALON His-tag purification resin was purchased from Takara Bio USA, Inc. Corning Spin-X UF concentrators with a 10-kDaMWCOwere purchased from MilliporeSigma. Steroids were purchased from Steraloids. All other materials were purchased from Thermo Fisher Scientific and MilliporeSigma.

Whole-cell steroid conversion assay

B. adolescentis strain L2-32, *B. desmolans* ATCC 43058, and *C. scindens* ATCC 35704 were cultivated in anaerobic brain heart infusion (BHI) broth. Steroids were dissolved in methanol and added to sterile anaerobic BHI medium at a concentration of 50 µM. The samples were then seeded separately with 0.1 ml of each bacterial culture and incubated at 37 °C for 48 h. After incubation, the products were extracted by vortexing the samples with 2 volumes of ethyl acetate for 3 min followed by recovery of the organic phase. The organic phase was then evaporated under nitrogen gas. The remaining residues were dissolved in 100µl of methanol and analyzed using a high-performance LC (HPLC) system (Shimadzu, Japan) equipped with a C18 analytical column (Capcell Pak c18, Shiseido, Japan). The mobile phase contained acetonitrile/

water with 0.01% formic acid, and the flow rate was maintained at 0.2 ml.min^{-1}. A DAD detector was used at a wavelength of 254 nm. Peak retention times and peak areas of samples were compared with standard steroid molecules.

Cloning and formation of the 20β-HSDH expression vector

Genomic DNA was extracted and purified from *B. adolescentis*, strain L2-32, using the FastDNA spin kit fromMPBiomedicals. The gene encoding 20β-HSDH (WP_003810233.1) was amplified using Phusion High-Fidelity DNA polymerase and the following primers: 5'-ATATATcatatgATGGCAGTGAATCATCGCAGATTCCA-3' and 5'-ATATATctcgagTTAGAAGACGCTTAGCCGCCATCTAC-3'. The restriction endonuclease sites for NdeI and XhoI are shown in lowercase, respectively. The PCR product was purified and double-digested with NdeI and XhoI restriction endonucleases. The digested PCR product was purified and cloned into the pET-28a(+) (Novagen, Madison, WI) vector using T4 DNA ligase to create the expression construct. The pET-28a(+) vector used had previously been re-engineered by replacing the kanamycin resistance gene with the ampicillin resistance gene. The resulting construct was transformed into Turbo Competent *E. coli* DH5α (high efficiency) cells by heat-shock at 43 °C for 35 s and outgrown for 1 h in Super Optimal Broth with Catabolite Repression (SOC) (Ipswich, MA) before growing on LB agar plates supplemented with ampicillin (100 μg/ml) for 16 h at 37 °C. Single colonies from the LB agar plates were inoculated into 5 ml of LB medium containing ampicillin (100 μg/ml) and grown overnight at 37 °C at 220 rpm. The cells were then centrifuged at 4000 x *g* for 10 min at 4 °C, and plasmids were extracted and purified from the pellets using the QIAprep Spin miniprep kit. Both strands of the plasmid DNA were sequenced for confirmation of the insert and absence of mutations using T7 promoter and

T7 terminator primers (W. M. Keck Center for Comparative and Functional Genomics at the University of Illinois at Urbana-Champaign). Once overexpressed, the expected protein should include an N-terminus His$_6$-tag and a thrombin cleavage site.

Site-directed mutagenesis

Plasmids containing the 20β-HSDH gene construct were mutated using the QuikChange Lightning site-directed mutagenesis kit. Primers containing the desired mutations were designed as follows: 5'-CTGATCTTCACAGCCgcgCTGTCCGACACAA-3' and 5'-TTGTGTCCGGACAGcgcGGCTGTGAAGATCAG-3' for serine 181 to alanine (S181A); and 5'-TCTTCACAGCCTCCCTGgcgGGACACAACGCGAACTA-3' and 5'-TATTCGCGTGTGTCCcgcCAGGGAGGCTGTGAAGA-3' for serine 183 to alanine (S183A). Mutated bases are shown in lowercase. The resulting mutated constructs were transformed into Turbo Competent *E. coli* DH5α (high efficiency) cells by heat shock at 43 °C for 35 s and outgrown for 1 h in Super Optimal Broth with Catabolite Repression (SOC) (New England Biolabs, Ipswich, MA) before growing on LB agar plates supplemented with ampicillin (100 µg/ml) for 16 h at 37 °C. Single colonies from the LB agar plates were inoculated into 5 ml of LB medium containing ampicillin (100 µg/ml) and grown overnight at 37 °C at 220 rpm. The cells were then centrifuged at 4000 x *g* for 10 min at 4 °C, and plasmids were extracted and purified from the pellets using the QIAprep Spin miniprep kit. Mutations were confirmed by sequencing of the mutated plasmid using T7 promoter and T7 terminator primers (W. M. Keck Center for Comparative and Functional Genomics at the University of Illinois at Urbana-Champaign).

Overexpression and purification of WT and mutant 20β-HSDH proteins

For protein expression, recombinant plasmids extracted from the *E. coli* DH5α cells were transformed into BL21- CodonPlus (DE3)-RIPL Competent *E. coli* cells by heat shock at 43 °C for 35 s and outgrown for 1 h in Super Optimal Broth with Catabolite Repression (SOC) (New England Biolabs, Ipswich, MA) before growing on LB agar plates supplemented with ampicillin (100 µg/ml) and chloramphenicol (50 µg/ml). After 16 h of growth at 37 °C, five colonies were picked to inoculate 10 ml of LB medium containing ampicillin (100 µg/ml) and chloramphenicol (50 µg/ml) and were grown for 6 h at 37 °C with shaking at 220 rpm. The turbid pre-culture was then added to 1 liter of LB medium containing the same antibiotics and shaken at 37 °C at 220 rpm. At an A_{600} of 0.6, the culture was induced with 0.25 mM IPTG, after which the culture was incubated at 16 °C at 220 rpm for 16 h. Cells were pelleted by centrifugation at 4000 x *g* for 30 min at 4 °C. The pellets were resuspended in 25 ml of buffer containing 50mM Tris-Cl and 300mM NaCl at pH 7.9. The resuspended cells were homogenized using an EmulsiFlex C-3 homogenizer by passing the samples through the homogenizer four times, and the cell debris was removed by centrifugation at 20,000 x *g* for 30 min at 4 °C. The recombinant protein in the soluble lysate was then purified using metal chelate chromatography followed by removal of the His6 tag by use of thrombin. The recombinant protein was further purified using anion-exchange chromatography to remove the excess thrombin. The resulting purified protein was analyzed using SDS-PAGE.

TLC

Reaction mixtures were made using 50 µM substrate, 150 µM cofactor, and 5 nM enzyme in 150 mM NaCl, 50 mM Tris-Cl buffer at pH 7.8. They were incubated overnight at room temperature and extracted by vortexing 2 volumes of ethyl acetate twice. The organic layer

172

was recovered and evaporated under nitrogen gas. The products were dissolved in 30 µl of methanol and spotted on a TLC plate (silica gel IB2-F flexible TLC sheet, 20 x 20 cm, 250-µm analytical layer; Avantor Performance Materials, LLC, PA). The steroids were separated with an isooctane/ethyl acetate/acetic acid (5:25:0.2, v/v) mobile phase and visualized by spraying with 50% sulfuric acid in ethanol and heating for 10 min at 100 °C.

Gel-filtration chromatography

Gel-filtration chromatography was performed on a Superose 6 10/300 GL analytical column (GE Healthcare) attached to an ÄKTAxpress chromatography system (GE Healthcare) at 4 °C. The purified protein was concentrated to 10 mg/ml and loaded onto the analytical column equilibrated with 50mMTris-Cl and 150 mM NaCl at pH 7.5 and was eluted at a flow rate of 0.3 ml/min. The native molecular mass of the 20β-HSDH proteins were determined by comparing elution volume with that of Gel Filtration Standard proteins (Bio-Rad): thyroglobulin, γ-globulin, ovalbumin, myoglobin, vitamin B_{12}.

Optimal pH determination

The 20β-HSDH pH optimum was measured spectrophotometrically by monitoring the reductive biotransformation of cortisol to 20β-dihydrocortisol as well as the reverse oxidative biotransformation of 20β-dihydrocortisol to cortisol. For the reductive direction, measurements were taken by analyzing the decrease of NADH at 340 nm. For the oxidative direction, measurements were taken by analyzing the increase of NADH at 340 nm. The buffers used in the pH profiling contained 50 mM buffering agent and 150 mM NaCl. Buffering agents used were as follows: sodium acetate trihydrate (pH 3.5–5.5), sodium phosphate monobasic monohydrate (pH

173

6.5–7.5), Tris (pH 8.0–9.0), glycine (pH 10–11). For the reductive reaction, 50 nM enzyme was added to 50µM cortisol and 150µM NADH in 200µl of buffer. For the oxidative reaction, 50 nM enzyme was added to 50 µM 20β-dihydrocortisol and 150 µM NAD^+ in 200 µl of buffer. For each reaction, the decrease or increase in NADH in the reaction mixture was measured continuously for 5 min, and the data were used in calculating the initial velocities.

Enzyme activity assays and determination of kinetic parameters

r20β-HSDH WT and mutant activities were measured spectrophotometrically by monitoring the reductive biotransformation of cortisol to 20β-dihydrocortisol as well as the reverse oxidative biotransformation of 20β-dihydrocortisol to cortisol by measuring the levels of NADH at 340 nm. The standard reaction mixture contained 50 mM sodium acetate trihydrate buffer (pH 5.0), 150 µM cofactor (NADH), 10 nM enzyme and a varied substrate concentration between 5 and 90 µM. GraphPad Prism was used to plot data, create Michaelis-Menten saturation curves, and Lineweaver-Burk plots and to calculate kinetic parameters.

Substrate-specificity assay

A saturating concentration of 50µM substrate, as determined by the enzyme activity analyses, was used as the concentration for analyzing various other steroid substrates and their activities. Similar to the enzyme activity assay, 20β-HSDH activity was measured spectrophotometrically by monitoring the biotransformation of substrate by measuring the levels of NADH at 340 nm. The standard reaction mixtures contained sodium acetate trihydrate buffer (pH 5.0 or pH 5.5), 150 µM cofactor (NADH or NAD^+), 50 nM enzyme, and 50 µM substrate.

CD spectroscopy

Determination of CD spectra for 20β-HSDH WT, its site-directed mutants, and the truncated derivatives was carried out using a J-815 CD spectropolarimeter (Jasco, Tokyo, Japan). Protein samples were prepared at a concentration of 0.2 mg/ml in 10 mM KH2PO4 buffer (pH 7.9). For the measurements, a quartz cell with a path length of 0.1 cm was used. CD scans were carried out at 25 °C from 190 to 260 nm at a speed of 50 nm/min with a 0.1-nm wavelength pitch, with five accumulations. Data files were analyzed on the DICHROWEB on-line server (http://www.cryst.ac.uk/cdweb/html/home.html)[39] using the CDSSTR algorithm with reference set 4, which is optimized for analysis of data recorded in the range of 190 to 240 nm. Mean residue ellipticity was calculated using millidegrees recorded, molecular weight, number of amino acids, and concentration of protein.

Thermal denaturation studies were performed by monitoring CD (millidegrees) at 220.5 nm from 25 to 95 °C at a rate of 1 °C/min. Each protein was run in triplicate with 0.2 mg/ml protein and 10 mM KH2PO4 buffer at pH 7.9.

Isothermal titration calorimetry

ITC analysis was performed with a VP-ITC microcalorimeter and 1.94-ml cell volume from MicroCal, Inc. The enzymes were dialyzed with 150 mM NaCl, 50 mM Tris buffer at pH 7.8. The pyridine nucleotide (NADH) was dissolved in this buffer, and cortisol was dissolved in methanol. 50 µM protein was injected with 28 successive 10-µl aliquots of 2 mM NADH at 300-s intervals. The protein and 2 mM NADH were then injected similarly with 1 mM cortisol. The data were fitted to a nonlinear regression model using a single binding site with the MicroCal

Origin software. The thermodynamic parameters were calculated using the Gibbs free energy equation ($\Delta G = \Delta H - T\Delta S$) and $\Delta G = -RT\ln(K_a)$.

Crystallization

WT 20β-HSDH crystals were obtained using the PEGRx HT kit from Hampton-Research. Purified 20β-HSDH protein was concentrated between a range of 15 and 20 mg/ml. Sitting drops were prepared consisting of 30 µl of well-condition and 300 nl of protein mixed with 300 nl of well-condition. Crystallization screening was performed in 96-well trays using Art Robbins PHOENIX robot. Plates were placed in 4 °C and single crystals formed after 2 days using condition 69 of PEGRx HT containing 20% v/v 2-propanol, 0.1 M MES monohydrate (pH 6.0), and 20% w/v PEG monomethyl ether 2000. The crystals continued to grow up to day 7. No further optimization was necessary. The crystals were overlaid with paraffin oil to prevent evaporation and 20% ethylene glycol as a cryoprotectant. Crystals were picked and immediately flash-cooled in liquid nitrogen.

For the NADH-bound structure, mutant S181A 20β-HSDH– purified recombinant protein was incubated with 2.5 mM NADH and 0.25 mM or 0.5 mM cortisol for 2 h at 4 °C. Crystals were grown in condition 86 of the Hampton PEG/Ion screen containing 0.05 M citric acid, 0.05 M Bis-tris propane (pH 5.0), and 16% w/v PEG 3350. The condition was then optimized in hanging-drop format using 18–20% w/v PEG3350. S181A crystals were frozen using the same process outlined for the WT 20β-HSDH.

176

X-ray data collection, processing, structure determination, and refinement

X-ray data were collected on the ID-D beamline at the Life Sciences Collaborative

Access Team of the Advanced Photon Source at Argonne National Laboratory. Data for the apo

structure were processed and scaled in HKL2000.[40] The crystal structure of the short-chain type

dehydrogenase/reductase from *M. marinum* (PDB code 3R1I) was selected using the MORDA

pipeline; Molrep and Refmac from CCP4 are used in this automation to test and refine initial

models.[41] The final molecular replacement model was rebuilt in Autobuild of the Phenix

package.[42] For the NADH-bound structure, data were processed in Autopro applying an

anisotropy correction in Staraniso.[43,44] Because of the highly anisotropic data, the completeness

for this structure appears low when calculated conventionally (isotropic) but the Staraniso

program utilizes ellipsoidal completeness to determine the resolution cutoff. Ellipsoidal

completeness for these data are high, 92.1% overall and 63.3% for the outer shell. Additionally,

the electron density maps are well-defined as shown in **Supplementary Figure 4.2,** *A* and *B*.

The apo structure was used as the molecular replacement model in Phaser-MR of the Phenix

package.[45] Each structure was refined via multiple rounds of manual model building in Coot and

refinement with Phenix.refine.[46,47] The X-ray processing and refinement statistics are presented

in **Table 4.3**.

Protein Data Bank accession code

The apo 20β-HSDH structure can be found in the Research Collaboratory for Structural

Bioinformatics (RCSB) Protein Data Bank (PDB) under the accession number 6M9U. The

NADH-bound structure can be found under the accession number 6OW4.

Multiple sequence alignment

The SDR family sequences were aligned using Clustal Omega (1.2.4) (https://www.ebi.ac.uk/Tools/msa/clustalo/).[48] The secondary structural elements were rendered using the ESPript 3.0 web server.[49]

FIGURES

Figure 4.1. Cortisol-metabolizing reactions by *B. adolescentis*, *C. scindens*, *B. desmolans*, *P. lymphophilum*, and *E. lenta*. *Boxed reaction* shows 20β-HSDH oxidizes the coenzyme NADH to transform cortisol into 20β-dihydrocortisol. Cortisol + NADH ↔ 20β-dihydrocortisol + NAD$^+$. This reaction reduces the double-bonded oxygen at C-20 (*highlighted in red*) on cortisol and is fully reversible.
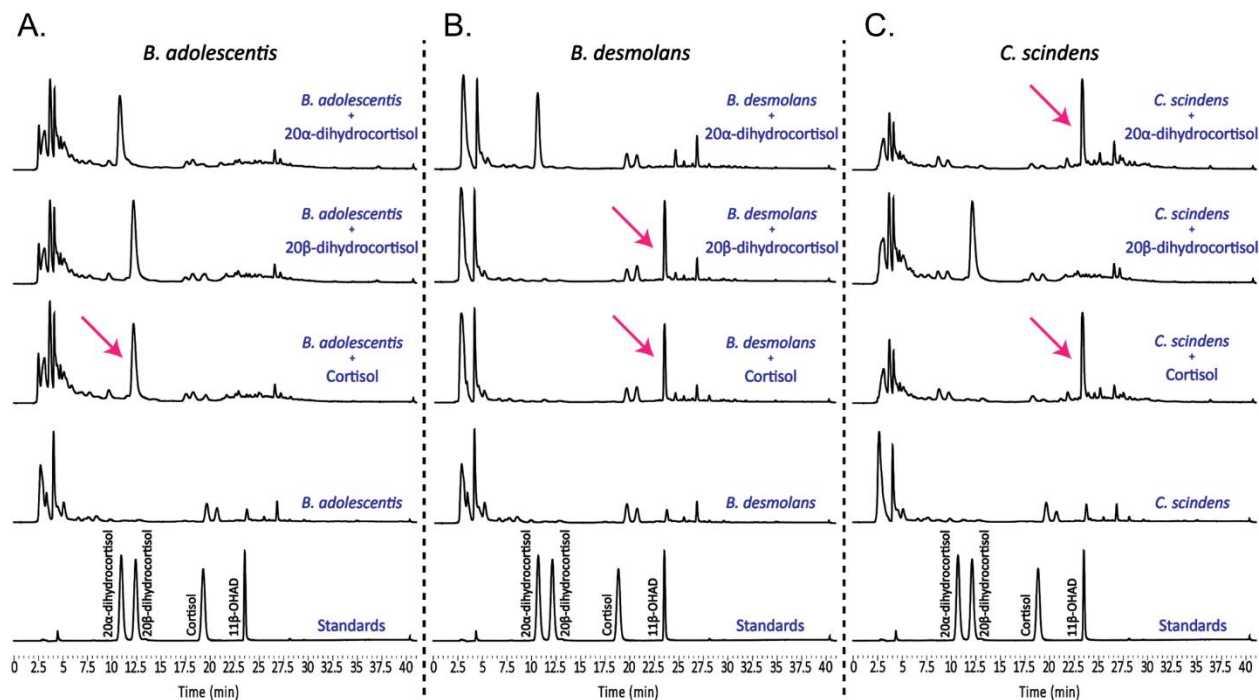
Figure 4.2. Whole-cell assays of *B. adolescentis* strain L2-32, *B. desmolans* ATCC 43058, and *C. scindens* ATCC 35704. *A, B. adolescentis* completely converts cortisol into 20β-dihydrocortisol. 20α-Dihydrocortisol is not metabolized. In a highly reductive environment and without pressure from a steroid-17,20-desmolase, *B. adolescentis* does not convert 20β-dihydrocortisol into cortisol. *B, B. desmolans* completely converts cortisol into 11β-OHAD and 20β-dihydrocortisol into 11β-OHAD. 20α-Dihydrocortisol is not metabolized. *C, C. scindens* completely converts cortisol into 11β-OHAD and 20α-dihydrocortisol into 11β-OHAD. 20β-Dihydrocortisol is not metabolized. Standards are reused in each panel for ease of comparison to reaction products above.
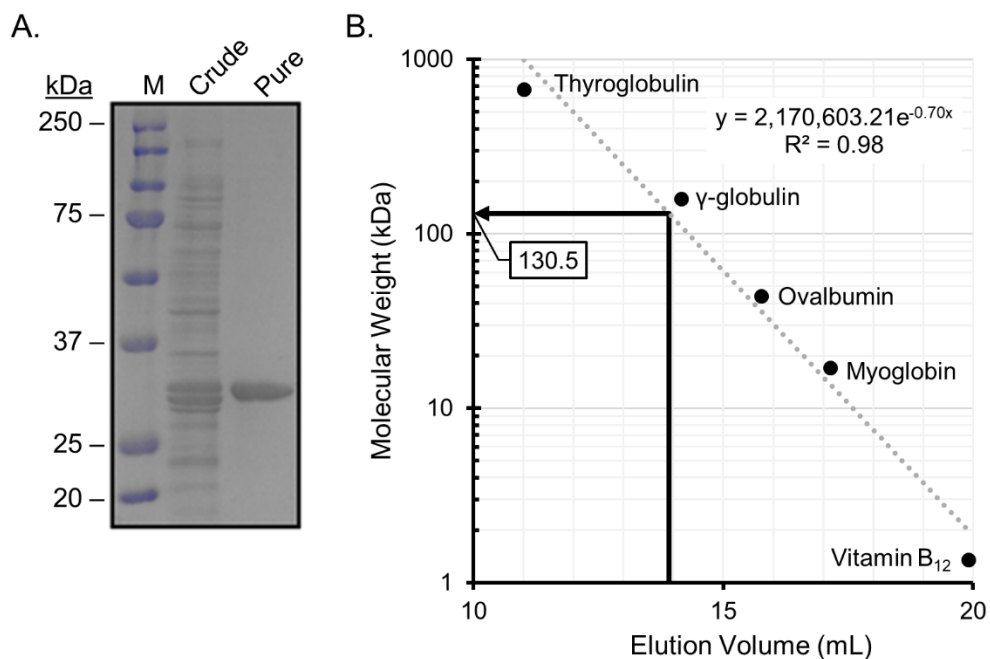
Figure 4.3. Purified 20β-HSDH from *B. adolescentis*, strain L2-32. *A,* SDS-polyacrylamide gel of crude and purified 20β-HSDH showing a subunit size of 32 ± 0.12 kDa. *Lane M,* molecular mass markers. *B,* native molecular size analysis of purified 20β-HSDH via gel-filtration chromatography.
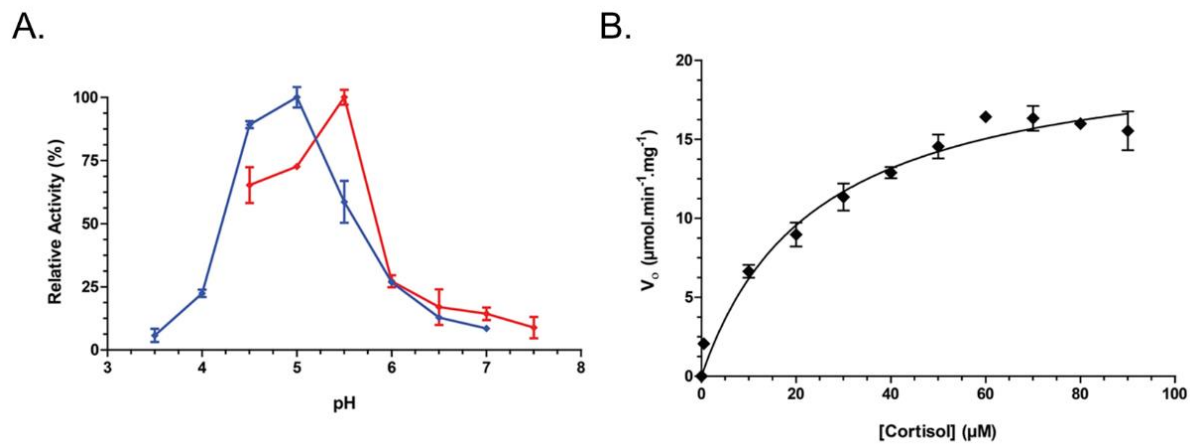
Figure 4.4. pH optimum and Michaelis-Menten saturation curve of WT 20β-HSDH. *A,* pH optimization of 20β-HSDH in the reductive (*blue*) and oxidative (*red*) direction. *B,* saturation curve showing the reduction of cortisol by WT 20β-HSDH.
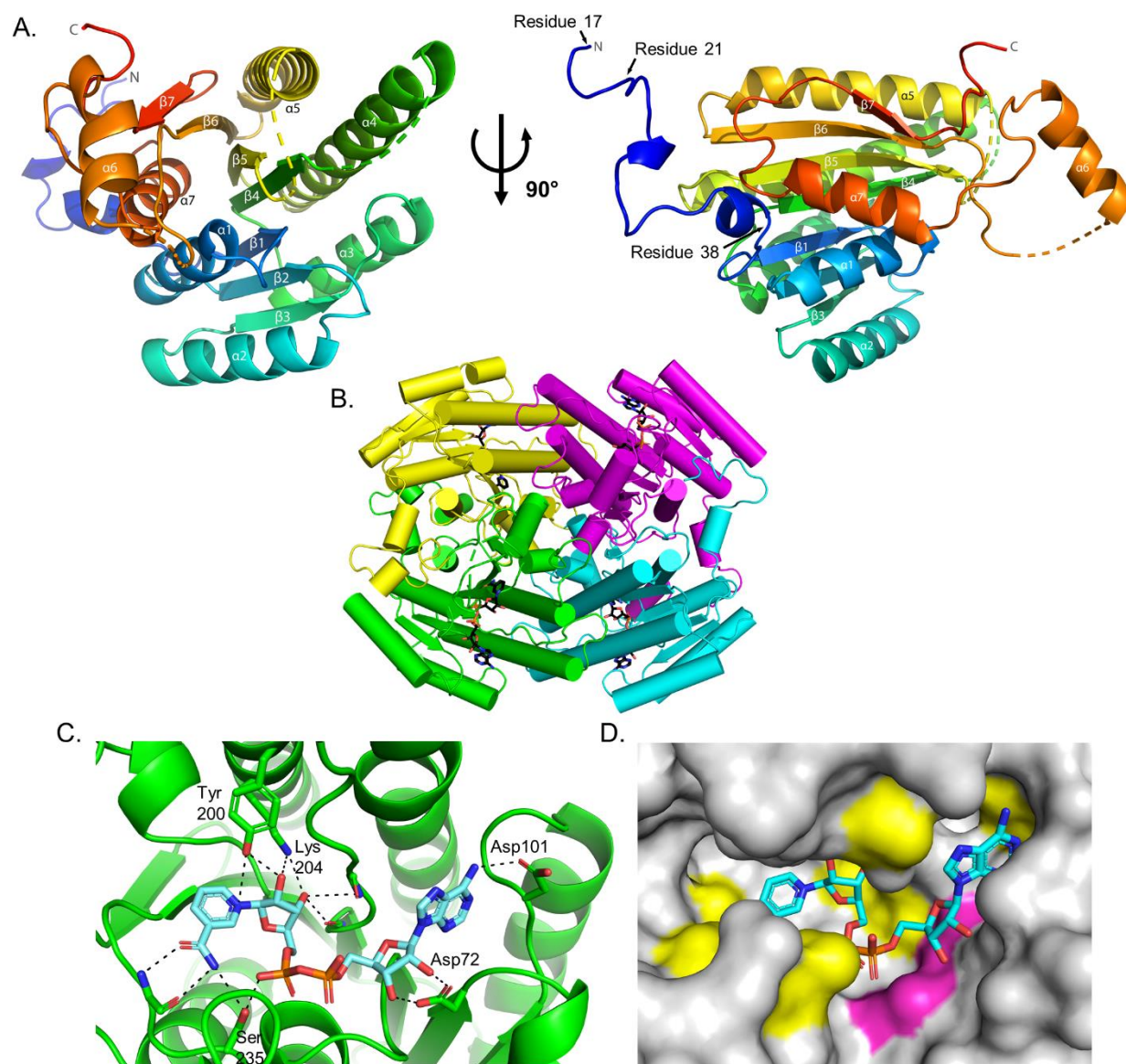
Figure 4.5. Structural characterization of 20β-HSDH. *A,* apo structure of 20β-HSDH showing the SDR-characteristic Rossmann fold consisting of seven central β-strands (β3–β2–β1–β4–β5–β6–β7) flanked by six α-helices. *Right panel* depicts the extended N terminus with no electron density before residue 17 and truncation sites. *B,* proposed tetramer of 20β-HSDH based on crystallographic arrangement. *C,* NADH-binding pocket showing five side chains and three peptidyl backbone interactions. *D,* surface representation of the NADH-binding pocket

Figure 4.5 (cont.)

highlighting residues that make van der Waals contacts to NADH (*yellow*). The Gly-rich region

is highlighted in *pink* and also makes van der Waals contacts to NADH.
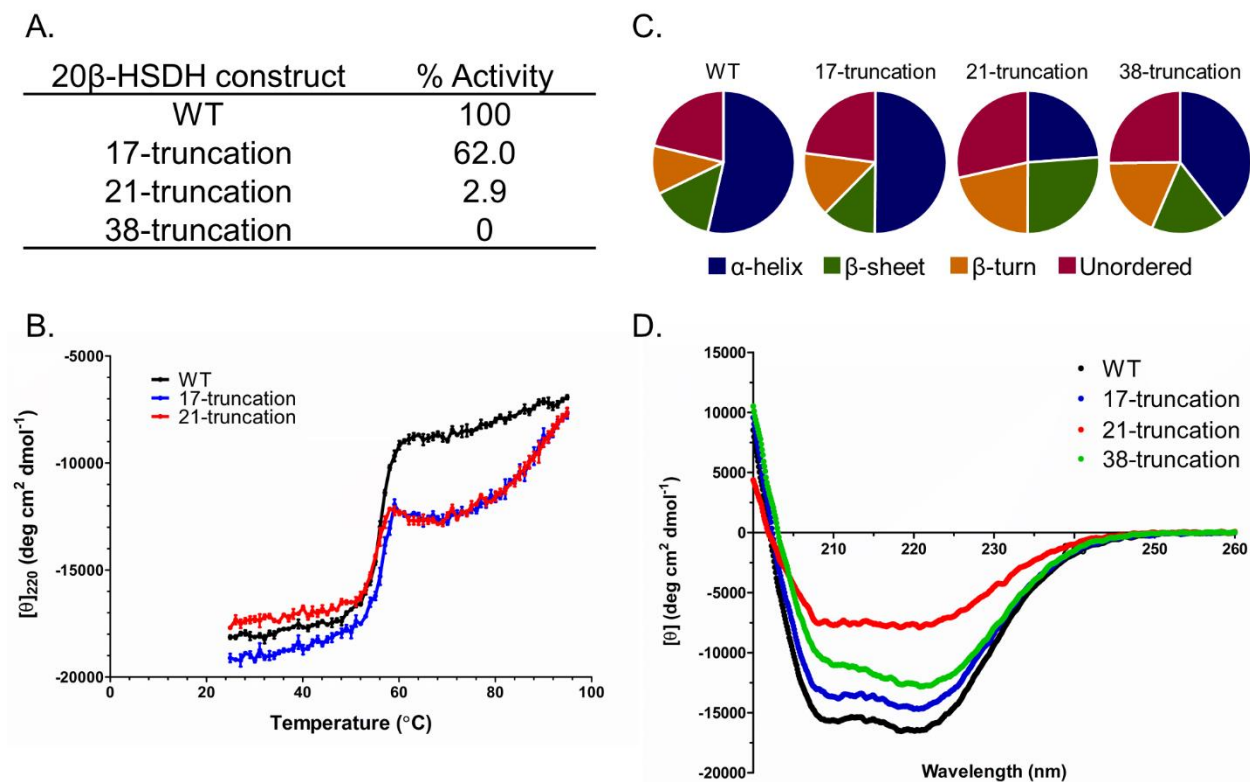
Figure 4.6. CD analysis and relative activity of WT and truncated 20β-HSDH. *A,* relative activity based on spectrophotometric assay. *B,* thermal stability of WT, 17-, and 21-truncated 20β-HSDH by temperature-dependent CD. *C,* percentage secondary structure calculated with DichroWeb according to *D* and CD spectra of WT and truncated proteins.

TABLES

Table 4.1. Steady-state kinetic parameters of 20β-HSDH and active site mutants

Protein	V_{max} (µmole•min^{-1}•mg^{-1})	K_m (µM)	k_{cat} (min^{-1})	Relative k_{cat} (%)	k_{cat}/K_m (µM^{-1}•min^{-1})	Relative k_{cat}/K_m (%)
Wild-type	21.08 ± 1.13^a	24.07 ± 3.80	668.30 ± 50.54	100	27.77 ± 3.65	100
S181A	-^{b}	-	-	-	-	-
S183A	4.70 ± 0.10	9.49 ± 2.60	148.70 ± 11.24	22	15.68 ± 2.06	56

[a] Values represent the means $\pm$ SD based on three or more replications.

[b] -, no activity detected.

Table 4.2. Substrate specificity of 20β-HSDH

Steroid	Trivial Name	Coenzyme	Relative Activity (%)
4-pregnen-11β,17,21-triol-3,20-dione	Cortisol	NADH	100.00 ± 0.77^a
4-pregnen-17,21-diol-3,20-dione	11-Desoxycortisol	NADH	66.03 ± 0.75
4-pregnen-11β,21-diol-3,20-dione	Corticosterone	NADH	39.32 ± 0.25
5β-pregnan-3α,11β,17,21-tetrol-20-one	Tetrahydrocortisol	NADH	44.29 ± 0.29
4-pregnen-11β,17,20β,21-tetrol-3-one	20β-Dihydrocortisol	NAD^+	0.23 ± 0.04
4-pregnen-17,20β-diol-3-one	17α,20β-Dihydroxyprogesterone	NAD^+	0.09 ± 0.05
4-pregnen-11β,20β,21-triol-3-one	20β-Dihydrocorticosterone	NAD^+	$-^b$
4-pregnen-17,20α-diol-3-one	17α,20α-Dihydroxyprogesterone	NAD^+	-
4-pregnen-11β,17,20α,21-tetrol-3-dione	20α-Dihydrocortisol	NAD^+	-

[a] Values represent the means ± SD from three or more replications.
[b] -, no activity detected.

Table 4.3. Crystallographic data collection and refinement statistics

Structure	Apo	NADH-bound
PDB ID	6M9U	6OW4
Resolution range (Å)	50.69 - 2.2 (2.279 - 2.2)	103.71 - 2 (2.045-2)
Space group	I 4 2 2	C 1 2 1
Unit Cell	168.371 168.371 126.961 90 90 90	165.078 134.71 96.396 90 100.15 90
Total no. of reflections	338597 (33468)	353360 (16341)
No. of unique reflections	46287 (4574)	73226 (3662)
Multiplicity	7.3 (7.3)	4.8 (4.5)
Completeness (%)	99.57 (99.72)	54.3 (8.7)
Mean I/sigma (I)	9.52 (1.66)	6.3 (1.7)
Wilson B-factor	41.11	16.22
R-merge	0.140 (1.715)	0.147 (0.806)
R-meas	0.151 (1.845)	0.178 (0.976)
R-pim	0.055 (0.673)	0.072 (0.412)
CC1/2	0.998 (0.492)	0.995 (0.670)
Reflections used in refinement	46144 (4561)	69583 (174)
Reflections used for R-free	1997 (198)	3673 (9)
R-work	0.196 (0.381)	0.187 (0.279)
R-free	0.231 (0.410)	0.238 (0.146)
Number of non-hydrogen atoms	4192	17642
macromolecules	3853	16839
ligands	179	352
solvent	160	451
Protein residues	501	2227
RMS(bonds)	0.011	0.013
RMS(angles)	1.46	1.64
Ramachandran favored (%)	96.70	94.43
Ramachandran allowed (%)	3.09	5.44
Ramachandran outliers (%)	0.21	0.14
Rotamer outliers (%)	3.24	0.80
Clashscore	6.46	2.37
Average B-factor	59.99	26.73
macromolecules	58.89	26.85
ligands	84.00	25.91
solvent	59.50	23.01

SUPPLEMENTARY FIGURES

A.

B.

C.
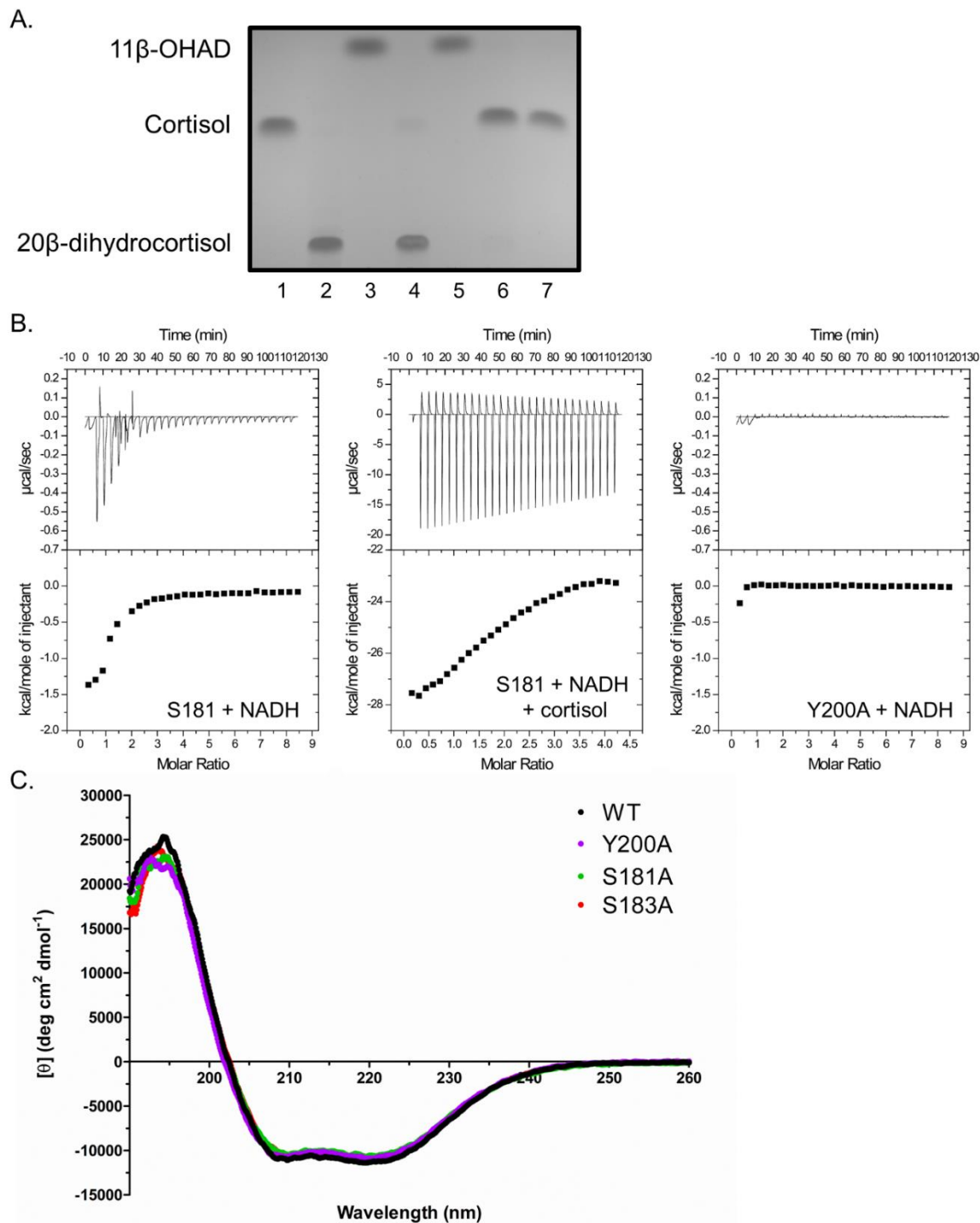
Supplementary Figure 4.1. Active site mutant thin layer chromatography (TLC) isothermal titration calorimetry (ITC) and circular dichroism (CD). (A) Wild-type and active site

Supplementary Figure 4.1 (cont.)

mutant 20β-HSDH overnight reaction products separated by TLC. (1) Cortisol standard, (2) 20β-dihydrocortisol standard, (3) 11β-hydroxyandrostenedione standard, (4) WT + NADH + cortisol, (5) WT + NADH + 11β-OHAD, (6) S181A + NADH + cortisol, (7) Y200A + NADH + cortisol. (B) Ligand binding order of S181A and Y200A studied by ITC. Left panel is 2 mM NADH binding to S181A, middle panel is 1 mM cortisol binding to S181A with 2 mM NADH, right panel is 2 mM NADH binding to Y200A. (C) CD spectra of purified recombinant 20β-HSDH and its active site mutants.
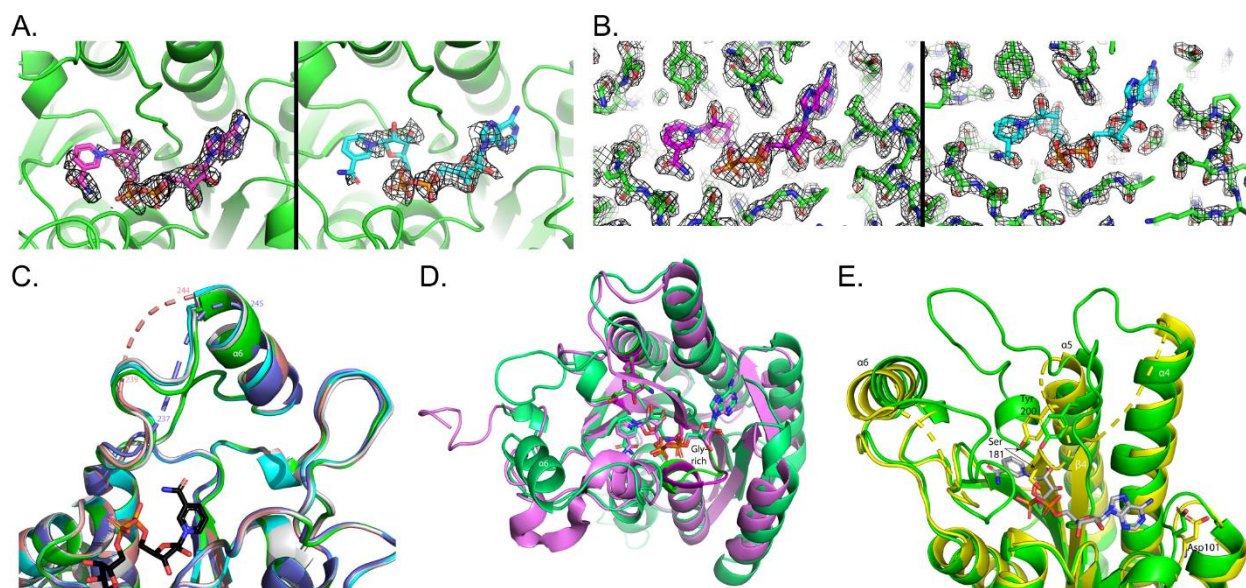
Supplementary Figure 4.2. Structural analysis of 20β-HSDH. (A) Fo-Fc omit map of NADH from Chain A and E shown at 2 sigma. (B) 2Fo-Fc map of NADH and surrounding backbone chains from Chain A and E shown at 2 sigma. (C) Alignment of 5 of the monomers from the NADH-bound structure showing the flexibility of loop 235-245 near where cortisol is predicted to bind. The 3 monomers not shown look similar to those shown. (D) Comparison of 20β-HSDH (green) to 3α,20β-HSDH from *Streptomyces hydrogenans (pink)*. The Gly-rich region is highlighted in brighter colors near the adenine of NADH and the side chains of the residues making up the catalytic tetrad are shown. Ser181 of the catalytic tetrad has been mutated to an alanine in 20β-HSDH. (E) Superposition of apo (yellow) and S181A mutant holo structure with NADH bound (green).

191
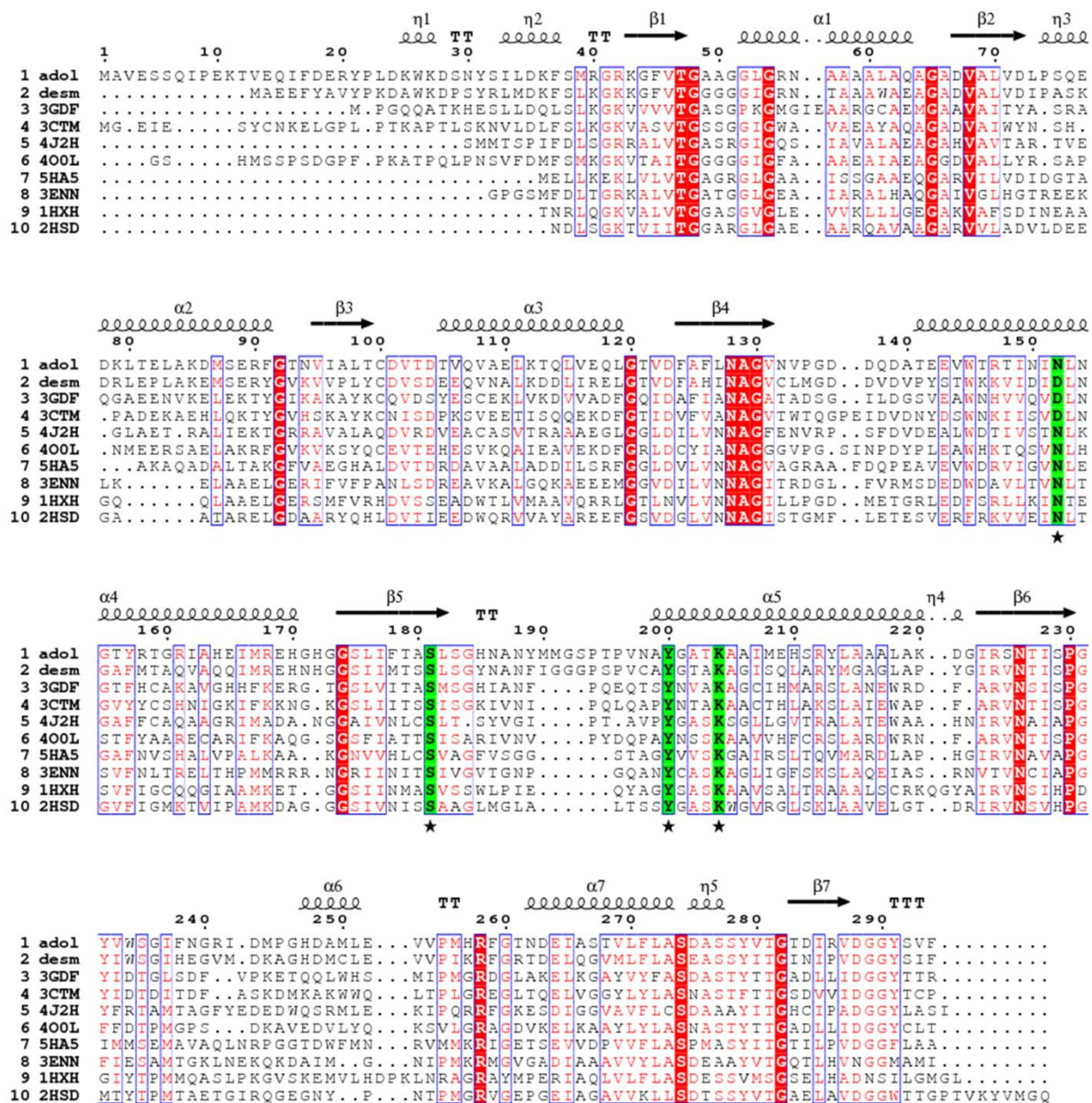
Supplementary Figure 4.3. Sequence alignment of 20β-HSDH found in *B. adolescentis* and other structurally similar SDR members. Red highlights indicate identical residues, green highlights and stars indicate active site residues, blue boxes indicate conserved residues. Secondary structural elements of apo-20β-HSDH are displayed on the top of the alignment. The abbreviations of protein names are as follows: adol, 20β-HSDH from *Bifidobacterium adolescentis* strain L2-32; desm, 20β-HSDH from *Butyricicoccus desmolans* ATCC 43058;

Supplementary Figure 4.3 (cont.)

remaining labels correspond to PDB IDs. The sequences were aligned with Clustal Omega

(1.2.4) and secondary structure rendered by ESPript 3.0 web server.

```
UniProt ID            Extended N-terminus

Ado1 WP_003810233.1   AVE---SSQI--PEKTVEQI-FDERYPLDKWKDSNYSILDKFSMRGR-KGFVTGAAGGLG 54
A0A1A9H573_HELPX      N--------------IRGIIKIIRGLAL----DNGR---WREKESQK-VAVITGASSGIG 39
U4WVT8_HELPX          ---------------------------------MGVGEKEEKKESQK-VAVITGASSGIG 26
A0A083YF40_HELPX      ---------------------MG-------VGEK---EEKKESQK-VAIITGASSGIG 26
T2SNM1_HELPX          ---------------------MG-------VGEK---EEKKESQK-VAVITGASSGIG 26
A0A0B2EPQ3_HELPX      ---------------------MG-------VGEK---EEKKESQK-VAIITGASSGIG 26
A0A0U1A5J2_9MYCO      GHWLPFSHPQVLAAATTELIDAVSGNQPGRGLRRAEMGKSRRPFEDQ-LVVITGGGSGIG 235
A0A0N8HB56_9ACTN      -----------------------MAEHAEHAEHAEQAEYPTGRAR-SVVITGASRGLG 34
A0A101BC04_9MYCO      --------------------------------MAEPRSGDERSCGKR-TVVITGASRGLG 27
A0A2D6MTV7_9DELT      SGM---------TRRSRSP----------ARISEEDKMKGALGYEGK-TVVITGAASGMG 51
A0A2E5YH49_9DELT      PGM---------KRKRRSS----------TRISEGYKMKNALGYEGK-TVVLTGAASGMG 51
A0A1X1T1G8_9MYCO      --------------------------------MTGIDGLWRHLGYHCR-RVVVTGCASGIG 28
A0A379BZG6_9NOCA      -----------------------------MRI-LGHGYPGIDLKGA-RVLITGAGRGIG 28
A0A3P8L1B1_TSUPA      -----------------------------MNL-FSSRDHLARLDGA-LVVVTGGARGIG 28
A0A2Z5YFI0_MYCMR      ------------------------------M-ADSTTIGVRVRDK-VIVITGGARGIG 26
A4ETZ5_9RHOB          A-R---GWTV--P-----PK-------QSTTFNSGDLEIKMMDIKGK-TVVITGASRGIG 55
A3JLM7_9RHOB          ------------------------------MNFEGMKMSNMQGK-VVFITGASRGIG 26
A3STW2_SULSN          ------------P--------------RAGLIDDEGNYTMDMTGK-TVMITGASRGIG 40
A0A0A3XQW2_BRAJP      ------------------------------MTRNAHHLRQAMTR-TILITGTSSGFG 26
A0A103KGN7_9BURK      -----------------------------MRLRLSNLACWKTIDMSK-TILITGASSGFG 30
A0A174GG33_9BACT      GHRER-GDKA--VTAPCEPAEVQGGTGE---RHGRRLERGAVAPGSA-WALVTGAGSGIG 113
U4E9H0_9VIBR          -------------------------------MKTSTDKTEVNIMK-TAFITGATSGIG 26
U4KB55_9VIBR          -------------------------------MKTSTDKTEVNIMK-TAFITGATSGIG 26
A0A1E3LA45_9BACL      ------------------------------MQYVYCDNMVKLANK-VVLITGASSGLG 27
A0A3S4VN67_MYCAU      -----------------------------MTKWTAADVPDQSGR-VAIVTGANTGIG 27
A0A2X1S640_MYCXE      ------------------------------MRWTAADLPSFAGR-TVVVTGANSGLG 26
A0A378YLY9_9NOCA      -----------------------------MAWKPSEIPDQSGR-TVVITGANGGIG 26
A0A174QRR8_BACVU      GIISKIKSKL--SYKEVTPL-YMDDLL---RDAYQTTSVVGGSLKGR-IALVTGATSGIG 57
D6D469_9BACE          SIKKYLKRAF--V------F-LLHGIPERHVIANITKLAPNEMLKGR-TALITGATSGIG 51
O53547_MYCTU          --------------------MK---LTESNRSPR-TTNTTDLSGK-VAVVTGAAAGLG 33
A0A1R3Y4F5_MYCBO      --------------------MK---LTESNRSPR-TTNTTDLSGK-VAVVTGAAAGLG 33
Q0S7K5_RHOJR          ------------------------------MNAV-ADRDVNVDGK-VAVVTGAGSGLG 26
J1RDG0_9NOCA          ------------------------------MFMNAV-ADRDVNVDGK-VAVVTGAGSGLG 28
A0A376F856_9MICO      ------------------------------MKLS---RRTAPSHAGR-CVLVTGDASGLG 26
A0A0M3C339_9SPHI      -------------------MFAF---IIDYNRTGMDISKLFDLSGK-TAIITGGAAGIG 36
3BHDP_RUMGV           ------------------------------MNFGGFIMGRFDEK-IMLVTGATSGIG 26
A2WJD3_9BURK          ------------------------------MHVNGTHDPAQLPLAGR-TALVTGGGRGLG 29
A0A378YJ86_9NOCA      ------------------------------MRLNP--FGGSRRTRYA-DAVVTGAGSGIG 27
A0A3B8M323_9ACTN      --------------------MV---LVAFWINT--GDMSNRLNGL-TAIISGGARGQG 32
```

Supplementary Figure 4.4. SDR family subset with extended N-terminus multiple sequence alignment. A Clustal Omega (1.2.4) alignment was performed with 555 sequences labeled as SDR family, hydroxysteroid dehydrogenase, and of bacterial origin from the UniProt database. This excerpt from the total sequence alignment includes 39 sequences with a ≥10 amino acid residue extended N-terminus. Conserved residues are highlighted, bolded residues indicate the *B. adolescentis* 20β-HSDH extended N-terminus, and the line indicates the beginning of the *B. adolescentis* 38-residue extended N-terminus.
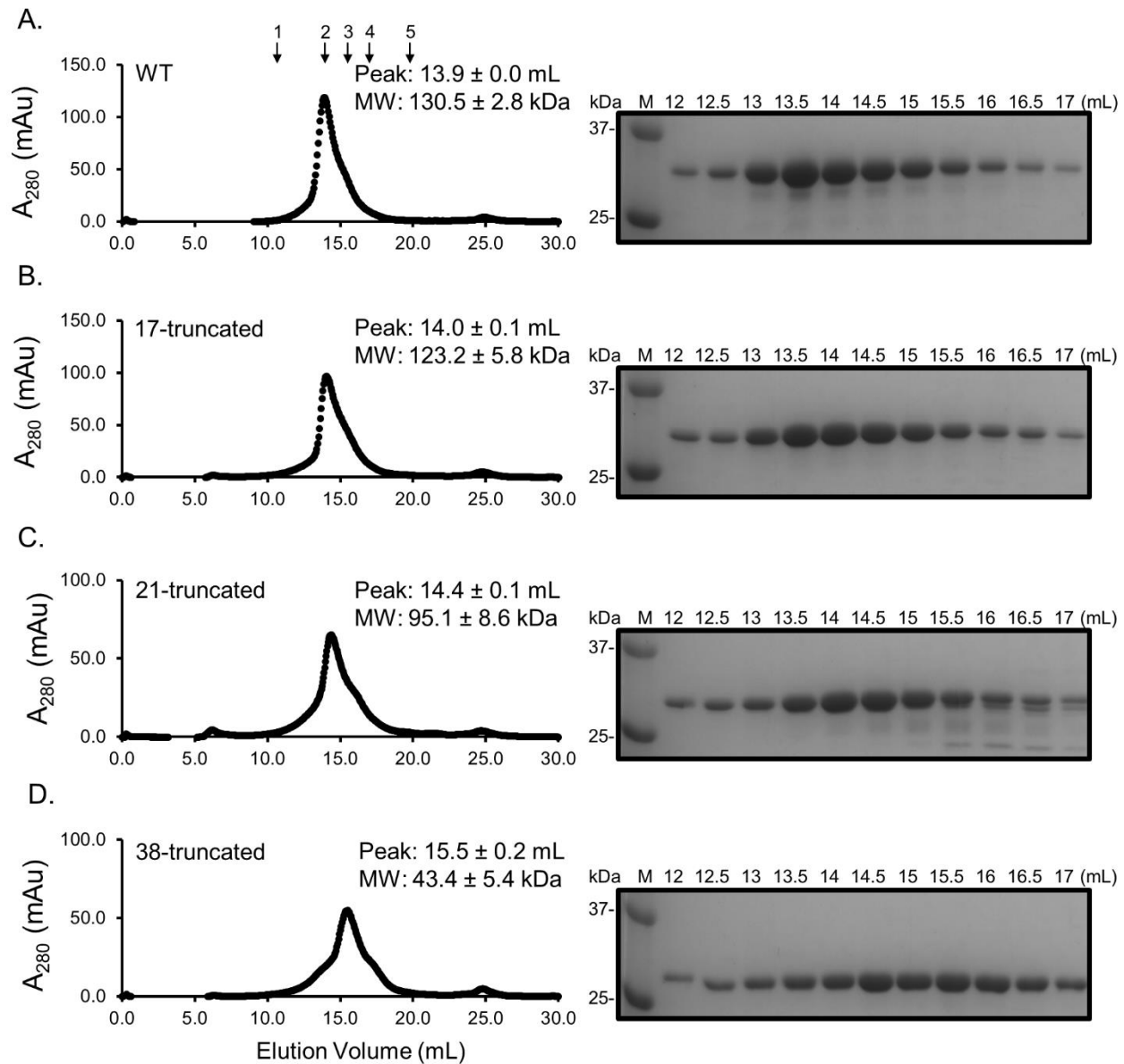
Supplementary Figure 4.5. Gel filtration chromatography of 20β-HSDH truncation mutants. Native molecular weight estimates were based on elution time of 10 mg/ml WT and truncated 20β-HSDH from a Superose 6 10/300 GL analytical column. Fractions were collected and analyzed on SDS-PAGE for visualization. (A) WT, (B) 17-truncated, (C) 21-truncated, (D) 38-truncated. Numbers indicate elution time of standard proteins (1) thyroglobulin, (2) γ-globulin, (3) ovalbumin, (4) myoglobin, (5) vitamin B_{12}.
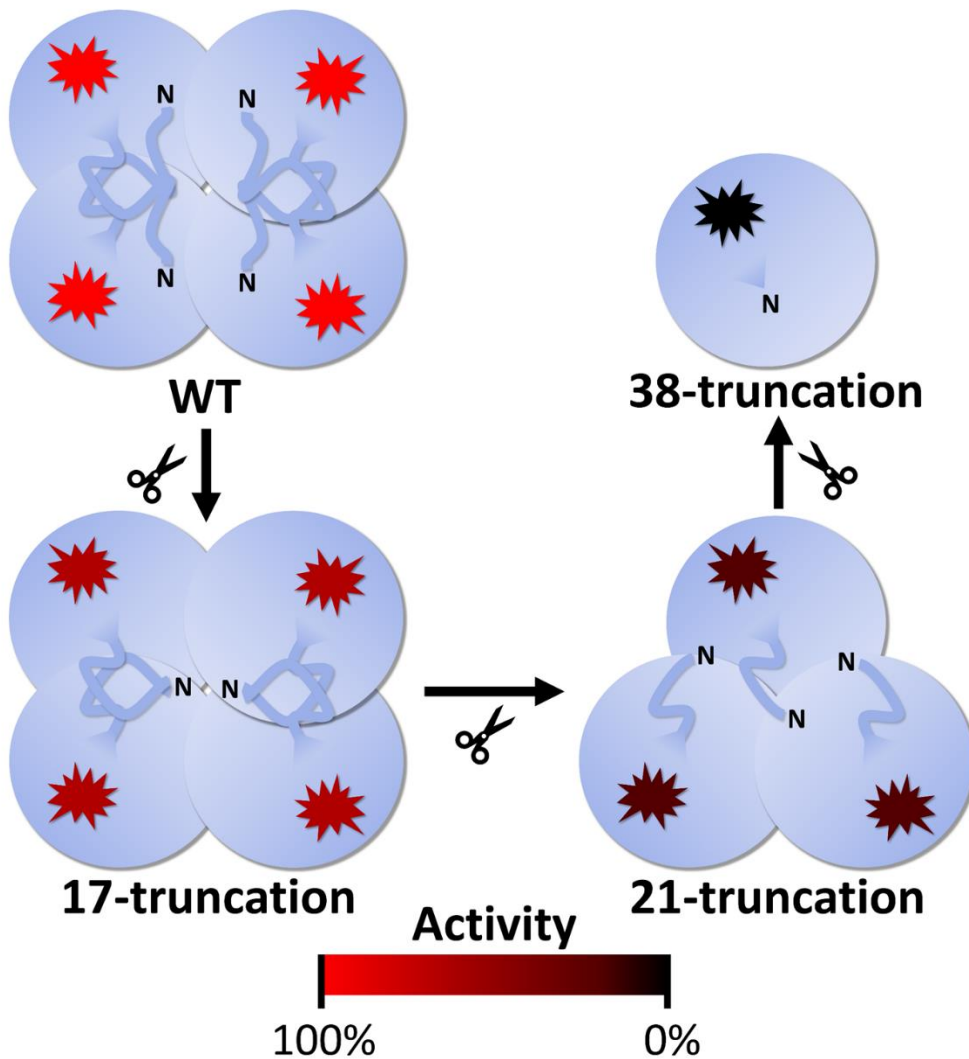
Supplementary Figure 4.6. Model of 20β-HSDH N-terminal truncation effects.

Oligomerization is shown by the number of blue circles, or subunits, and enzymatic activity is depicted by the color of the explosion shape. WT 20β-HSDH has 100% relative activity and is likely tetrameric. When the first 17 residues are truncated, the protein exhibits 62% activity, but remains a tetramer. When 21 residues are truncated, the protein loses quaternary structure and has only 2% activity. The full 38-residue truncation results in no activity and a mixed dimeric and monomeric form.

REFERENCES

1. Penning TM. Molecular endocrinology of hydroxysteroid dehydrogenases. Endocr Rev 1997; 18:281–305.
2. Ridlon JM, Ikegawa S, Alves JM, Zhou B, Kobayashi A, Iida T, Mitamura K, Tanabe G, Serrano M, De Guzman A, Cooper P, Buck G A, and Hylemon PB. *Clostridium scindens*: a human gut microbe with a high potential to convert glucocorticoids into androgens. J Lipid Res 2013*;* 54:2437–2449.
3. Ridlon JM, Harris SC, Bhowmik S, Kang DJ, and Hylemon PB. Consequences of bile salt biotransformations by intestinal bacteria. Gut Microbes 2016; 7:22–39.
4. Mythen SM, Devendran S, Méndez-García C, Cann I, and Ridlon JM. Targeted synthesis and characterization of a gene cluster encoding NAD(P)H-dependent 3α-, 3β-, and 12α-hydroxysteroid dehydrogenases from *Eggerthella* CAG:298, a gut metagenomic sequence. Appl Environ Microbiol 2018; 84:e02475–17.
5. Devlin AS, and Fischbach MA. A biosynthetic pathway for a prominent class of microbiota-derived bile acids. Nat Chem Biol 2015; 11:685–690.
6. Morris DJ, and Ridlon JM. Glucocorticoids and gut bacteria: "The GALF Hypothesis" in the metagenomic era. Steroids 2017; 125:1–13.
7. Krieger DT, Allen W, Rizzo F, and Krieger HP. Characterization of the normal temporal pattern of plasma corticosteroid levels. J Clin Endocrinol Metab 1971; 32:266–284.
8. Devendran S, Mythen SM, and Ridlon JM. The desA and desB genes from *Clostridium scindens* ATCC 35704 encode steroid-17,20-desmolase. J Lipid Res 2018; 59:1005–1014.
9. Devendran S, Méndez-García C, and Ridlon JM. Identification and characterization of a 20β-HSDH from the anaerobic gut bacterium *Butyricicoccus desmolans* ATCC 43058. J Lipid Res 2017; 58:916–925.
10. Bokkenheuser VD, Morris GN, Ritchie AE, Holdeman LV, and Winter J. Biosynthesis of androgen from cortisol by a species of *Clostridium* recovered from human fecal flora. J Infect Dis 1984; 149:489–494.
11. Pretorius E, Africander DJ, Vlok M, Perkins MS, Quanson J, and Storbeck K-H. 11-Ketodihydrotestosterone in castration-resistant prostate cancer: potent androgens which can no longer be ignored. PLoS ONE 2016; 11:e0159867.
12. Pretorius E, Arlt W, and Storbeck K-H. A new dawn for androgens: novel lessons from 11-oxygenated C19 steroids. Mol Cell Endocrinol 2017; 441:76–85.
13. Bokkenheuser VD, Winter J, Morris GN, and Locascio S. Steroid desmolase synthesis by *Eubacterium desmolans* and *Clostridium cadavaris*. Appl Environ Microbiol 1986; 52:1153–1156.
14. Winter J, Cerone-McLernon A, O'Rourke S, Ponticorvo L, and Bokkenheuser VD. Formation of 20β-dihydrosteroids by anaerobic bacteria. J Steroid Biochem 1982; 17:661–667.
15. Filling C, Berndt KD, Benach J, Knapp S, Prozorovski T, Nordling E, Ladenstein R, Jörnvall H, and Oppermann U. Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. J Biol Chem 2002; 277:25677–25684.
16. Kallberg Y, Oppermann U, Jörnvall H, and Persson B. Short-chain dehydrogenases/reductases (SDRs). Coenzyme-based functional assignments in completed genomes. Eur J Biochem 2002; 269:4409–4417.

17. Rossmann MG, Ford GC, Watson HC, and Banaszak LJ. Molecular symmetry of glyceraldehyde-3-phosphate dehydrogenase. J Mol Biol 1972; 64:237–245.
18. Edenharder R, and Schneider J. 12β-Dehydrogenation of bile acids by *Clostridium paraputrificum*, *C. tertium,* and *C. difficile* and epimerization at carbon-12 of deoxycholic acid by cocultivation with 12α-dehydrogenating *Eubacterium lentum*. Appl Environ Microbiol 1985; 49:964–968.
19. Ghosh D, Weeks CM, Grochulski P, Duax WL, Erman M, Rimsay RL, and Orr JC. Three-dimensional structure of holo 3α,20β-hydroxysteroid dehydrogenase: a member of a short-chain dehydrogenase family. Proc Natl Acad Sci U.S.A. 1991; 88:10064–10068.
20. Morgan RA, Beck KR, Nixon M, Homer NZM, Crawford AA, Melchers D, Houtman R, Meijer OC, Stomby A, Anderson AJ, Upreti R, Stimson RH, Olsson T, Michoel T, Cohain A, et al. Carbonyl reductase 1 catalyzes 20β-reduction of glucocorticoids, modulating receptor activation and metabolic complications of obesity. Sci Rep 2017; 7:1–11.
21. Schöneshofer M, Weber B, and Nigam S. Increased urinary excretion of free 20α- and 20β-dihydrocortisol in a hypercortisolemic but hypocortisoluric patient with Cushing's disease. Clin Chem 1983; 29:385–389.
22. Kornel L, Miyabo S, Saito Z, Cha R-W, and Wu F-T. Corticosteroids in human blood. VIII. Cortisol metabolites in plasma of normotensive subjects and patients with essential hypertension. J Clin Endocrinol Metab 1975; 40:949–958.
23. Latif SA, Pardo HA, Hardy MP, and Morris DJ. Endogenous selective inhibitors of 11β-hydroxysteroid dehydrogenase isoforms 1 and 2 of adrenal origin. Mol Cell Endocrinol 2005; 243:43–50.
24. Takeshita D, Kataoka M, Miyakawa T, Miyazono K, Kumashiro S, Nagai T, Urano N, Uzura A, Nagata K, Shimizu S, and Tanokura M. Structural basis of stereospecific reduction by quinuclidinone reductase. AMB Express 2014; 4:6.
25. Zhang R, Zhu G, Zhang W, Cao S, Ou X, Li X, Bartlam M, Xu Y, Zhang XC, and Rao Z. Crystal structure of a carbonyl reductase from *Candida parapsilosis* with anti-Prelog stereospecificity. Protein Sci 2008; 17:1412–1423.
26. Ghosh D, Erman M, Wawrzak Z, Duax WL, and Pangborn W. Mechanism of inhibition of 3α,20β-hydroxysteroid dehydrogenase by a licorice-derived steroidal inhibitor. Structure 1994; 2:973–980.
27. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE 3rd., Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 1998; 393:537–544.
28. Cho HJ, Kim JD, Lee WY, Chung BC, and Choi MH. Quantitative metabolic profiling of 21 endogenous corticosteroids in urine by liquid chromatography-triple quadrupole-mass spectrometry. Anal Chim Acta 2009; 632:101–108.
29. Romanoff LP, Parent C, Rodriguez RM, and Pincus G. Urinary excretion of β-cortolone (3α,17α,20β,21-tetrahydroxypregnane-11-one) in young and elderly men and women. J Clin Endocrinol Metab 1959; 19:819–826.
30. Tokarz J, Norton W, Möller G, Hrabé de Angelis M, and Adamski J. Zebrafish 20β-hydroxysteroid dehydrogenase type 2 is important for glucocorticoid catabolism in stress response. PLoS ONE 2013; 8:e54851.

31. Yoshioka H, Fujita K, Sakata H, Murono K, and Iseki K. Development of the normal intestinal flora and clinical significance in infants and children. Bifidobacteria Microflora 1991; 10:11–17.
32. Tissier H. Recherche Sur La Flore Intestinale Des Nourissons (Etat normal et pathologique). 1990. Ph.D. thesis, Univ of Paris, Paris, France.
33. Lee J-H, and O'Sullivan DJ. Genomic insights into *Bifidobacteria*. Microbiol Mol Biol Rev 2010; 74:378–416.
34. O'Toole PW, Marchesi JR, and Hill C. Next-generation probiotics: the spectrum from probiotics to live biotherapeutics. Nat Microbiol 2017; 2:17057.
35. Storbeck KH, Bloem LM, Africander D, Schloms L, Swart P, and Swart AC. 11β-Hydroxydihydrotestosterone and 11-ketodihydrotestosterone, novel C19 steroids with androgenic activity: a putative role in castration resistant prostate cancer? Mol Cell Endocrinol 2013; 377:135–146.
36. D'Errico I, and Moschetta A. Nuclear receptors, intestinal architecture and colon cancer: an intriguing link. Cell Mol Life Sci 2008; 65:1523–1543.
37. Feighner SD, Bokkenheuser VD, Winter J, and Hylemon PB. Characterization of a C21 neutral steroid hormone transforming enzyme, 21-dehydroxylase, in crude cell extracts of *Eubacterium lentum*. Biochim Biophys Acta 1979; 574:154–163.
38. Feighner SD, and Hylemon PB. Characterization of a corticosteroid 21-dehydroxylase from the intestinal anaerobic bacterium, *Eubacterium lentum*. J Lipid Res 1980; 21:585–593.
39. Whitmore L, and Wallace BA. Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. Biopolymers 2008; 89:392–400.
40. Otwinowski Z, and Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 1997; 276:307–326.
41. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr 2011; 67:235–242.
42. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, Mc-Coy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, et al. PHENIX:Acomprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 2010; 66:213–221.
43. Vonrhein C, Flensburg C, Keller P, Sharff A, Smart O, Paciorek W, Womack T, and Bricogne G. Data processing and analysis with the autoPROC toolbox. Acta Crystallogr D Biol Crystallogr 2011; 67:293–302.
44. Tickle IJ, Flensburg C, Keller P, Paciorek W, Sharff A, Vonrhein C, Bricogne G. STARANISO (http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi). Global Phasing Ltd., Cambridge, UK. 2018.
45. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, and Read RJ. Phaser crystallographic software. J Appl Crystallogr 2007; 40:658–674.
46. Emsley P, Lohkamp B, Scott WG, and Cowtan K. Features and development of Coot. Acta Crystallogr D Biol Crystallogr 2010*;* 66:486–501.
47. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, and Adams PD. Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr D Biol Crystallogr 2012; 68:352–367.

48. Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, Park YM, Buso N, and Lopez R. The EMBL-EBI bioinformatics web and programmatic tools framework. Nucleic Acids Res 2015; 43:W580–W584.
49. Robert X, and Gouet P. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 2014; 42:W320–W324.

CHAPTER 5

BACTERIA ON STEROIDS: THE ENZYMATIC MECHANISM OF AN NADH-DEPENDENT DEHYDROGENASE THAT REGULATES THE CONVERSION OF CORTISOL TO ANDROGEN IN THE GUT MICROBIOME

ABSTRACT

The ability to metabolize both endogenous and exogenous compounds to a variety of metabolic products is not exclusive to our human cells. In fact, the bacterial communities that inhabit our digestive system are responsible for a network of steroid transformations that can produce hormones in the gut, which are then absorbed by the host, potentially affecting host physiology. These communities have been shown to impact our health in numerous ways, affecting disease predisposition, pathogenesis, physical fitness, and dietary responsiveness. Steroid biotransformations by gut bacteria are predicted to impact the host endocrine system. A particular set of transformations facilitated by microbial enzymes has been shown to result in the formation of 11-oxy-androgens from host-derived cortisol. Since androgens have been implicated in disease and immune modulations, understanding the structure and catalytic mechanism of enzymes involved in cortisol metabolism is a key step towards developing strategies to reduce disease-promoting bioactive steroids in certain individuals. Here, we combine experimental and computational techniques to describe DesC, a 20α-hydroxysteroid dehydrogenase (HSDH) capable of creating 20α-dihydrocortisol and siphoning cortisol away from pathways that produce androgens. DesC diverges significantly from previously described bacterial and eukaryotic counterparts, catalyzing an NADH-dependent 20α-HSDH reaction but

presenting little sequence and structure similarity to them. The structural information obtained by X-ray crystallography and hybrid quantum mechanics/molecular mechanics (QM/MM) simulations, validated through mutagenesis studies, suggest the reaction occurs through a multi-step proton relay mechanism. Free energy calculations were then used to describe the kinetics of the reaction mechanism. The mechanistic information presented here can be employed in the development of therapeutics to divert microbial pathways away from disease-promoting steroids.

INTRODUCTION

Bacterial communities have a wide influence on human health and play important roles in disease predisposition, pathogenesis, immune regulation, and dietary responsiveness.[1–4] Gut microbes are estimated to be numerically equal to mammalian host cells, although their gene content is estimated to dwarf our own with 99% of functional genes in the human body being microbial. It is no surprise then that the gut microbial consortium has evolved biochemical pathways to biotransform, synthesize, or regulate the production of host endogenous signaling molecules. The contribution of host-associated bacterial metabolism of steroid hormones is coming into focus. Steroid hormones are essential for regulation of various physiological functions, including cellular communication, metabolism, inflammatory response, and stress response.[5] One class in particular, the endogenous glucocorticoids, is synthesized from cholesterol in the adrenal cortex[6] and released into the blood to influence cells throughout the body. Upon binding to host nuclear receptors they affect cellular processes through genomic and non-genomic mechanisms.[5] Thus, investigating the chemistry of steroid biotransformation in the gut microbiome is crucial for a better understanding of the symbiotic relationship between us and our microbiota.[7]

Cortisol, the major glucocorticoid in humans, is converted to various derivatives by enzymes throughout our body. The local concentrations of cortisol are regulated by enzymes that reversibly modify functional groups on the steroid nucleus and side chain, saturate the steroid rings, and conjugate the steroid with sulfate or glucuronide.[8–10] In this way, ratios of active vs. inactive glucocorticoids can be locally and rapidly altered in peripheral tissues. Intriguingly, gut microbes possess multiple cortisol-altering enzymes that have co-evolved separately from eukaryotic counterparts.[10–15] Microbes are therefore expected to play an important role in the complex steroid metabolome of the host. Indeed, accumulating evidence demonstrates human gut microbes are an important component of the host endocrine system.[9] Of particular relevance is the observation that peripheral side-chain cleavage of cortisol to 11β-hydroxyandrostenedione (11β-OHAD) occurs by some combination of host and microbial input.[16–18] The androgens produced from cortisol side-chain cleavage function similarly to testosterone by binding and activating androgen receptor.[19] Recently, these novel androgens have been shown to be relevant to human physiology and pathology, including castration-resistant prostate cancer[18,20], polycystic ovary syndrome[19,21], and potentially hypertension[11,22–25].

The human gut bacterium *Clostridium scindens* ATCC 35704 was previously shown to side-chain cleave cortisol[26]. However, the lack of knowledge about the metabolic pathway responsible for androgen production severely limits our ability to mechanistically study androgen formation by the human microbiome *in vivo*. We previously discovered a cortisol-inducible gene cluster (*desABCD*) in *C. scindens* ATCC 35704, encoding NADH-dependent 20α-hydroxysteroid dehydrogenase (20α-HSDH; DesC)[27] and heterotetrameric steroid-17,20-desmolase (DesAB)[28] which produce 20α-dihydrocortisol and 11β-OHAD from cortisol, respectively. Furthermore, the DesC product, 20α-dihydrocortisol, is not a substrate for DesAB[28], suggesting DesC acts as a

metabolic "switch" regulating side-chain cleavage of cortisol. DesC activity is relatively rare and the only gene reported thus far is from *C. scindens*.[27,29] Importantly, pharmaceutical glucocorticoids such as prednisone are substrates for side-chain cleavage, generating metabolites that drive prostate cancer cell proliferation.[30] Therefore, understanding the structure and catalytic mechanism of enzymes involved in cortisol metabolism is key to hasten the development of strategies to reduce drug metabolism and alter androgen formation by the human microbiome.

Here, we combine multiple experimental and computational approaches to investigate the enzymatic mechanism of *C. scindens* ATCC 35704 DesC, a potential regulator of the pro-androgenic DesAB pathway (**Figure 5.1**). Our findings reveal the structure of this enzyme with atomistic detail as well as its interaction with both NADH and cortisol substrates. Classical and hybrid quantum mechanics/molecular mechanics (QM/MM) molecular dynamics simulations indicate an intricate reaction mechanism including a hydride transfer from NADH to cortisol and a proton relay that finally leads to a 20α-dihydrocortisol enzymatic product.

RESULTS AND DISCUSSION

X-ray crystallography

To understand the structure and mechanism of DesC, the apo-form of the recombinant enzyme (rDesC) was crystallized via vapor diffusion sitting drop at pH 6.75 in a buffer containing 0.2M $MgCl_2$, 24% PEG 400, 0.1M HEPES, resulting in a resolution of 2.0Å. X-ray diffraction data can be found in **Supplementary Table 5.1**. Our attempts to obtain crystals of the binary and ternary DesC complex were unsuccessful. DesC is a member of the zinc-dependent medium-chain dehydrogenase/reductase (MDR) family that catalyzes zinc-dependent oxidation of primary and secondary alcohols using $NAD(P)^+$ as the proton and hydride acceptor.[31] DesC

displays the typical bilobal topology comprised of a nucleotide-binding Rossmann fold domain and a catalytic domain[32,33] (**Figure 5.2A**). Across the alcohol dehydrogenase superfamily, the size and architecture of the catalytic domain varies greatly, with the MDRs having a larger and more complex catalytic domain.[31] While this family of enzymes varies in sequence conservation, the Rossmann fold is the most conserved feature among short-, medium-, and long-chain dehydrogenases.[31] In the two-domain structure of DesC, the Rossmann fold is located proximal to the C-terminus, and comprised of β-strands 11-16 and flanked by α-helices 6-11 (**Figure 5.2B**). The predicted catalytic Zn^{2+} ion (**Figure 5.2C**) is octahedrally coordinated by two water oxygens and residues of the catalytic domain including the side chains of His73, Glu74, Glu156, and a cysteinesulfonic acid at Cys45.

Like many prokaryotic alcohol dehydrogenases, DesC is a tetramer (**Figure 5.2A**). The enzyme crystallized with one monomer in the asymmetric unit, and the tetrameric assembly is formed by crystallographic symmetry; therefore, no differences are observed among the monomers that comprise the biological assembly of DesC in this model.

The tetramer displays the catalytic domains at the periphery of the complex. Two Rossmann folds align along the first two-fold axis to create a continuous β-sheet comprised of 6 β-strands from each monomer across the dimeric interface. The second two-fold axis situates two adjacent catalytic domains such that a helical loop created by α4 and a 3_{10} helix and stabilized by a second Zn^{2+} (Zn finger) extends towards the adjacent monomer (**Figure 5.2D**). Referred to as the lobe loop, this structural feature is conserved among many alcohol dehydrogenases including those from mammals.[34–36] The Zn^{2+} at this position is located ~21Å from the catalytic cleft and is tetrahedrally coordinated by 4 Cys (105/108/111/119). Single alanine substitutions of Cys105, Cys108, Cys111, Cys115 residues composing the zinc-finger were sufficient to result in loss of

quaternary structure (**Supplementary Table 5.2**), confirming the structural role of the zinc-finger in dimer-dimer interaction.

The MDRs that are close structural and sequence homologs of DesC include the threonine dehydrogenases (TDH) from *Pyrococcus horikoshii* (PhTDH; Protein Data Bank (PDB) 2DFV, 2.0Å RMSD for 333 Ca, Z=40.5) and *Thermococcus kodakaraensis* (TkTDH; PDB 3GFB, 2.3Å RMSD for 334 Ca, Z=39.4), see **Supplementary Figure 5.1**. The mechanism for the TDHs and many MDRs is believed to proceed via an ordered bi-bi reaction whereby NAD^+ binds first and is the last to leave.[34,35,37] Binding of NAD^+ likely initiates the conformational change that in part closes the active site to engage the catalytic machinery. The crystal structures of PhTDH and TkTDH were determined with NAD^+.[34,35] An overlay of DesC and TDH structures (**Supplementary Figure 5.1**) demonstrates conservation of the structural and catalytic Zn^{2+}-binding residues (**Supplementary Figure 5.2**) as well as the NADH binding pocket. Not shared between these two is a hydrophobic patch of four tryptophan residues that is positioned near the lobe loop and conserved among the thermophilic TDHs.[34,35]

Conserved features among the NADH binding pocket include Asp204 in DesC which is Glu199 in PhTDH and hydrogen bonds with the O-2 and O-3 of the adenosine ribose. This residue is shared among bacterial alcohol dehydrogenases and imparts specificity for the cofactor.[34,38] Also of importance is Arg204 of PhTDH, which extends towards the pyrophosphate and is equivalent with K209 in DesC, though in our model this residue is positioned somewhat further from the predicted location of the pyrophosphates. Mutagenesis of Glu199 and Arg204 in PhTDH demonstrates these residues are significant determinants in NAD(H) binding. The nucleotide-binding GxGxxG motif is also conserved among these enzymes although Pro178 of PhTDH is replaced by Gly183 in DesC. Finally, in PhTDH, Thr44 and His50 are believed to be

part of the proton relay that connects the active site with bulk solvent, and these are conserved in DesC as Ser47 and His50. However, in the current crystal structure of DesC Ser47 and His50 are not aligned to interact. While the structures of DesC, PhTDH and TkTDH share a very similar fold and conservation of important catalytic residues, our overlay reveals a longer extension of α-helix (residues 58-66) and the loop connecting β1/2 (residues 9-19) in DesC. This region is near the entrance to the active site cleft, but ~20Å from the predicted site of hydride transfer to the nicotinamide ring. It is possible that this loop plays some role in selecting or restricting substrate. The structure of DesC was deposited to the protein data bank[39] (PDB) with accession code 4OH1.

Docking

Since structural data was obtained only for apo-DesC, we used homologous structures available within the PDB in order to fit an NADH molecule to the binding pocket of a monomer of DesC, employing a previously established protocol.[40] With BLAST[41], we obtained homologous structures (PDB IDs: 4ILK, 4EJ6, 4A2C, 3QE3, 3GFB, 2DQ4, 2DFV, 2D8A, 1PL7, 1E3J) from the PDB. The alignment (**Figure 5.3A,B**) and placing of the NADH molecule on the DesC binding site was performed using VMD.[42] Leveraging advanced run options of QwikMD[43], the structure of the ligand was minimized in the pocket together with atoms of nearby DesC residues, while maintaining a static DesC structure as much as possible.

The structure of cortisol was also fitted to the most probable binding site using VMD (**Figure 5.3C**). The docking was performed by a combined manual and computational approach[44], where VMD was used to position some of the reference atoms of cortisol, while NAMD[45], through its QwikMD interface[43], was used to minimize changes in structure of the

complex. As reference atoms we chose the cortisol atoms that were known to participate in the enzymatic reaction, positioning them near the Zn^{2+} and the NADH hydride donor (**Figure 5.3E**). Also, cortisol was aligned in a cleft (**Figure 5.3D**), which was shown to contain a more hydrophobic character, typical of ligand binding sites. The complex was then solvated, and the net charge of the system was neutralized in a 0.15 mol/L sodium chloride solution, using QwikMD to follow predetermined protocols employed in previous enzymatic studies.[40] The solvated system was subjected to 100 ns of equilibrium molecular dynamics (MD) simulation, using the CHARMM36 force field[46], along with the TIP3 water model[47]. The simulations were performed with periodic boundary conditions, in the NpT ensemble, with temperature maintained at 300 K and pressure at 1 bar. Long-range electrostatic interactions were treated using the particle-mesh Ewald (PME) method.[48] All MD simulations were executed with the GPU-accelerated NAMD molecular dynamics package.[45]

A MD simulation protocol can be used to investigate the stability of a molecular system, as well as predict the behavior and function of proteins or protein domains.[49] Here, after 100 ns of MD simulation, we observed that cortisol was stable in the predicted catalytic cleft. After analyzing the MD trajectory, we observed that the cortisol molecule was stabilized by interactions with two tyrosine residues, namely Tyr101, and Tyr124 (**Figure 5.3F**). Simulations revealed that the mostly hydrophobic cleft was fundamental in maintaining the complex structure stability, with cortisol fluctuating very little in the cleft. To validate the importance of Tyr101 and Tyr124 in substrate-binding, we performed site-directed mutagenesis, observing loss of enzyme activity at $96.01 \pm 0.67\%$ and $81.82 \pm 0.31\%$ activity, respectively, without altering secondary structural features as determined by circular dichroism (CD) spectroscopy (**Figure 5.3G**). Isothermal titration calorimetry (ITC) experiments showed that there was no effect on

NADH binding with rDesCY101R (K_d = 35.08 µM) and rDesCY124R (K_d = 39.06 µM).

However, binding affinity of cortisol to rDesCY101R and rDesCY124R was ablated, based on

ITC data (**Figure 5.3**). Additionally, mutation of tyrosine to arginine in the putative binding

pocket altered activity and kinetic parameters (**Supplementary Table 5.3**). The experimental

data support the binding of cortisol to the hydrophobic pocket predicted by the computational

study. Particularly, the experimental data shows that by mutating the aforementioned residues,

the structure did not change, but interactions with cortisol did as evidenced by altered ITC and

kinetic constants which were not observed with NADH.

Reaction mechanism

After examining the enzyme active site, we sought to determine the catalytic mechanism

involved in 20α-dehydrogenation. For that, we employed a combination of hybrid quantum

mechanics (QM) and classical molecular mechanics (MM) simulations. We also integrated a

state-of-the-art enhanced sampling method, which was recently implemented in NAMD.[50]

Hybrid QM/MM has been the main tool used to study processes that cannot be explained by

classical MM methods alone.[51–53] In biological contexts, QM/MM calculations are often used in

enzymology[54,55] as well as to investigate polarizable molecules in different environments.[56,57] In

the QM/MM molecular dynamics (MD) approach, the system of interest is typically investigated

in a water bath (**Figure 5.4A**), as in a regular classical MD simulation, with the exception that

part of the system is treated at the quantum-mechanical level.[50] This approach offers an

augmented MD simulation, where a selection of atoms is investigated at sub-atomic resolution in

order to account for electronic effects. Specifically, we employed an electrostatic embedded

QM/MM approach, where the classical part of the system was treated with the CHARMM36

force field[46], while the QM part of the system was treated at semiempirical PM7 level[58]. The QM region was chosen to be relatively large, with 573 atoms, because of the elongated shape of the catalytic pocket, and large presence of aromatic groups that could affect the electronic structure near the reaction site. To ensure the presence of a shell of QM atoms around the reaction site, a total of 36 QM-MM bonds were created using a link-atom approach.[59] Our computational studies were then supported by biochemical assays.

To investigate possible reaction mechanisms, we ran a 20 ps long QM/MM MD equilibration to prepare the system, and used a QM/MM steered MD (SMD) approach[50] to induce the reaction to occur according to the chosen mechanisms we describe below. These selected reaction mechanisms were designed to determine multiple aspects of this complex reaction, one being the participation of a water molecule to mediate the interaction between the Zn^{2+} ion and substrates, as previously suggested for similar enzymes.[60] Our QM/MM SMD only forced the initial steps of the reactions, allowing the remaining steps to occur freely in case they were energetically favorable. The simulations revealed that the most probable scenario would be the reaction mechanism without the participation of a water molecule, having a concerted proton transfer from Ser47 and a hydride from NADH to the C-20 ketone oxygen of cortisol. The QM/MM SMD simulations revealed a proton relay that was responsible for reestablishing the amino acid protonation states in the catalytic site (**Figure 5.4B**). The proton relay was composed of the replenishing of Ser47 by a proton transfer from the ribosyl of NAD^+, followed by the protonation of the ribosyl group by His50 (which was previously in a protonated state). His50 can once again become protonated when NAD^+ leaves the catalytic pocket for a new enzymatic cycle, a situation in which His50 becomes largely exposed to the bulk solvent.

To test this mechanism, we generated S47A and S47T mutations that resulted in loss of

catalytic function (100% vs. 98.52 ± 0.08 % loss of activity, respectively) (**Figure 5.5A,E**).

While S47T mutant bound NADH (K_d = 66.67 µM), cortisol lost binding affinity (**Figure 5.5D**).

By contrast, S47A showed high affinity for cortisol based on ITC measurements (K_d =11.56 µM)

in the presence of NADH (K_d = 97.0 nM) (**Figure 5.5C**). Hydrogen-bonding between the C-

21hydroxy group of cortisol and the NADH is predicted to be responsible for the orders of

magnitude difference in binding affinity between S47A, which binds cortisol, and S47T, which

does not. There are no statistically significant differences in CD spectra between DesCWT and

S47A and S47T (**Figure 5.5B**).

With the experimental validation of the possible mechanism, we moved to optimize the

reaction pathway and understand the atomistic details of the enzymatic mechanism. The string

method[61] was chosen for the task. This method is an optimization algorithm that relies on

iterations of biased QM/MM MD simulations to find the chemical reaction path with the smallest

energetic barrier. It does so by tracking "collective variables", a set of measurements made on

the system to describe the transformation being examined. In our case, the collective variables

were chosen to be distances between relevant atoms in the QM region of the simulation.

The string method was initialized with the QM/MM SMD simulation as a first hypothesis

for the reaction path. A set of representative snapshots of the chemical reaction was selected to

describe the entire process, from initial to final state, creating a set of "images" that represent a

copy of the entire simulated system at a different stage of the reaction. At each iteration, multiple

independent MD simulations are initiated from each image, allowing the atomic systems to

explore the energy surface and drift to local minima. Next, average values for the collective

variables are determined. Here, these values are cartesian coordinates, that are then used to

constrain the systems to keep consecutive images approximately equidistant in the collective

variables space, smoothing the reaction coordinate. Convergence is achieved when the iterations

do not produce significant changes in the mean values of the collective variables. Taking

advantage of NCSA's Blue Waters supercomputer, we ran 750 replicas for our string

optimization, revealing the pathway of the reaction (**Figure 5.4C**). The string optimization

shows that the initial reaction that triggers the enzymatic process is the hydride donation to the

cortisol side-chain C-20, followed by the Ser47 proton donation to cortisol. The formation of

20α-dihydrocortisol triggers the proton relay we described above.

Free energy calculation

To investigate the free energy barrier that limits the start of the reaction, namely the

activation free energy (ΔG_{act}), we perform a parallel extended-Adaptive Biasing Force (eABF)

calculation[62], following established protocols.[50,63] Using the final images of the optimized string

we can define a continuous path that represents the entire enzymatic reaction. The eABF method

can then be used to calculate the free energy changes along this path by constraining the system

to path collective variables S and Z, where S indicates progression along the reaction path, and Z

indicates a perpendicular distance to the path. Using NCSA's Blue Waters supercomputer, we

used a parallel strategy that initiated 600 walkers from the different images and were thus able to

conduct extensive sampling over the defined path. The multiple-walker method is so-called due

to the fact that each replica executes a random walk along the path.[64] Here, 10,000 MD steps

were performed by each walker, accumulating 6,000,000 QM/MM ABF steps, or 3 ns of

QM/MM MD simulation. The simulations revealed that only the initial step of the reaction,

namely the hydride transfer from NADH to cortisol C-20, imposes an energetic barrier in the

order of 6 kcal/mol (**Figure 5.4D**). The 3 proton transfers that follow the hydride transfer were found to be energetically favorable. The entire reaction process was found to be exergonic, with a free energy difference between reactants and products (ΔG) in the order of -25 kcal/mol. It is worth mentioning that this free energy difference only accounts for the reaction itself, and not for binding and rearrangements of the structure that occur prior and after the catalytic reaction. The ITC data however has also shown that the binding is exothermic, making the initial binding event, as expected, part of the overall favorable energy of the reaction.

The enhanced sampling protocol adopted here greatly accelerated the dynamics of the system, allowing us to balance the gain in precision and the increased cost of QM/MM simulations. The simulations revealed that all reaction steps can occur in quick succession due to a stabilization of the reaction site, which is partially imposed by the presence of a Zn^{2+} ion. After the initial reaction step, all reactions in the proton relay are highly favorable because all protons are already aligned in hydrogen bonds. This stability is noticeable in the snapshots shown in **Figure 5.4E**, which reveal very little difference in the catalytic region during the entire reaction process.

CONCLUSIONS

Here we report the crystal structure and propose a catalytic mechanism for DesC, a glucocorticoid NADH-dependent 20α-HSDH expressed by *Clostridium scindens* ATCC 35704. Bacterial metabolism of host and synthetic glucocorticoids may have wide ranging effects on the immune system[65–68], the cardiovascular system[9], the structure of the gut microbiome[69,70], and the urogenital tract.[30,71] Their importance for human health brings enzymes biotransforming 20-hydroxysteroids significant pharmaceutical value.

The expression of DesC by *C. scindens* ATCC 35704 highlights its importance as a member of the human 'sterolbiome'[28,72], the collective gene pathways responsible for steroid biotransformations by host-associated bacteria. In addition, DesC is part of a gene cluster encoding steroid-17,20-desmolase encoded by *desA* and *desB* genes.[27] Steroid-17,20-desmolase catalyzes side-chain 17,20-lyase activity, which is the first step to converting cortisone (inactive form of cortisol) to 11-keto-androstenedione, an androgen-receptor ligand[30]. Thus, *C. scindens* ATCC 35704 contributes to the sterolbiome through the formation of 11-oxy-androgens and 20-oxy-steroids. Importantly, *C. scindens* steroid-17,20-desmolase (DesAB)[28] and 20α-HSDH (DesC) emerge as potential therapeutic targets.

In nature, reversible conversion of the C-20 oxygen of C_{21} glucocorticoids is catalyzed by diverse pyridine nucleotide-dependent 20α-hydroxysteroid dehydrogenases. DesC, the NAD(H)-dependent 20α-HSDH described in the present study, diverges significantly from previously described bacterial counterparts. Strains of *Escherichia coli* (DH5α, K12, E132) encode a homotetrameric (28 kDa monomer) complex with NADH-dependent 20α-HSDH activity belonging to the classical short-chain dehydrogenase/reductase (SDR) family, which does not require metal ions[73]. The enzyme was previously reported as a 2-dehydro-3-deoxy-D-gluconate 5-dehydrogenase (KduD) and has broad substrate specificity for both sugars and C_{21}-glucocorticoids. Interestingly, KduD and DesC share only 16% amino acid identity. A second example is the dimeric (34 kDa monomer) NADPH-dependent 20α-HSDH purified from the protozoan *Tetrahymena pyriformis,* which was structurally and functionally distinct from bacterial as well as mammalian 20α-HSDHs.[74] The *T. pyriformis* 20α-HSDH was highly specific for 17α-progesterone, which inhibits the growth of this protozoan. Mammalian enzymes reported to have 20α-HSDH activity are monomeric with M_r values between 33 kDa to 55 kDa, except for

estradiol 17α-dehydrogenase (dual 20α-HSDH activity), which is a dimer. Many mammalian enzymes are polyfunctional, catalyzing several regio- and stereospecific reactions at more than one steroid keto group (e.g. 3α,20α-HSDH from human liver dihydrodiol dehydrogenase). Thus, the DesC represents a unique and now well-characterized bacterial 20α-HSDH.

In summary, we have employed a combination of state-of-the-art computational tools and experiments to reveal the structure and proposed catalytic mechanism of DesC, a key steroid-metabolizing enzyme found in the gut microbiome (**Figure 5.1**). X-ray crystallography was utilized to obtain the apo-DesC structure (**Figure 5.2**), which was then used as the starting point of a computational study that involved the identification of the DesC catalytic pocket, as well as the docking of both NADH and cortisol to this pocket (**Figure 5.3**). To confirm the proposed reaction mechanism, additional structural information with cofactor and substrate bound is needed. However, attempts to crystallize the binary and ternary complex were unsuccessful. Therefore, site-directed mutagenesis and biochemical assays were then employed to validate the key amino acid residues in the catalytic pocket. As it was hypothesized for the TDHs and many MDRs, our ITC data shows that NADH is the first substrate to bind the enzyme, given that cortisol does not bind to DesC unless NADH is present at the catalytic pocket. With the structure of DesC and its ligands now available, a hybrid QM/MM MD approach[50] was applied to investigate the reaction mechanism of DesC. The approach, which combined steered molecular dynamics, string method optimization, and extended adaptive biasing force, with hybrid QM/MM calculations, then revealed that NADH first donates a hydride to cortisol, which triggers a proton relay that provides a proton to cortisol, forming the 20α-dihydrocortisol enzymatic product. The proton relay then stabilizes the reaction pocket by a series of proton transfers that starts in a protonated histidine (**Figure 5.4**). The mechanism was validated by

biochemical assays, where amino acids that were revealed critical in the simulations were mutated (**Figure 5.5**). The enzymatic process was found to be exergonic, although the free energy difference calculated between reactants and products only accounts for the reaction itself and not for other steps of the DesC mechanism. Our findings can be applied in the development of approaches to modulate DesC activity, while our protocols can be employed to investigate other enzymes that are biomedically and biotechnologically relevant. The successful combination of experimental and computational approaches presented here reinforces the power and the need of combining these techniques.

MATERIALS AND METHODS

Bacterial Strains and Materials

Clostridium scindens ATCC 35704 has been maintained as -80°C glycerol stocks in our laboratory. *Escherichia coli* DH5α (turbo) competent cells were from New England Biolab (Ipswich, MA), *E. coli* BL21 CodonPlus (DE3)-RIPL was purchased from Stratagene (La Jolla, CA). The pET-51b vector was obtained from Novagen (San Diego, CA). Restriction enzymes were purchased from New England Biolab (Ipswich, MA), QIAprep Spin Miniprep kit was obtained from Qiagen (Valencia, CA). Isopropyl β-D-1-thiogalactopyranoside (IPTG) was purchased from Gold Biotechnology (St. Louis, MO). Strep-Tactin® resin was purchased from IBA GmbH, (Goettingen, Germany). Cortisol and 20a-dihydrocortisol were purchased from Steraloids (Newport, RI). Amicon Ultra centrifugal filter units with 10-kDa MWCO were obtained from Millipore (Billerica, MA). All other reagents were of the highest possible purity and were purchased from Fisher Scientific (Pittsburgh, PA).

Gene cloning, expression, and protein purification

The open reading frame (EDS07887) encoding 20α-hydroxysteroid dehydrogenase (20α-HSDH) amplified from *C. scindens* genomic DNA with primers reported in **Supplementary Table 5.4**, using high fidelity Phusion polymerase and cloned into pET51b as previously described. The sequence correctness of DNA insert in the recombinant plasmid was confirmed by DNA sequencing at W. M. Keck Center for Comparative and Functional Genomics at the University of Illinois at Urbana-Champaign.

For protein expression, the correct recombinant plasmid extracted from the *E. coli* DH5α cells was transformed into *E. coli* BL21 CodonPlus (DE3)-RIPL chemically competent cells by the heat shock method and grown overnight at 37 °C on LB agar plates supplemented with ampicillin (100 µg/mL) and chloramphenicol (50 µg/mL). After 16 hours, five isolated colonies were used to inoculate 10 mL of fresh LB medium supplemented with antibiotics and grown at 37 °C for 6 hours with vigorous aeration. The pre-cultures were then added to fresh LB medium (1 L), supplemented with the same antibiotics at the same concentrations, and grown with vigorous aeration at 37 °C. At OD600 of 0.3, isopropyl β-D-thio-galactopyranoside (IPTG) was added to each culture at a final concentration of 0.1 mM and the temperature was decreased to 16 °C. Following 16 hours of culturing, cells were pelleted by centrifugation (4,000 x g, 30 min, 4 °C) and re-suspended in 30 mL of binding buffer (20 mM Tris-HCl, 150 mM NaCl, 20% glycerol, 10mM 2-Mercaptoethanol pH 7.9). The cell suspension was subjected to four passages through an EmulsiFlex C-3 cell homogenizer (Avestin, Ottawa, Canada), and the cell lysate was clarified by centrifugation at 20,000 x g for 30 min at 4°C.

The recombinant 20α–HSDH was then purified using Strep-Tactin® resin (IBA GmbH, Goettingen, Germany) as per manufacturer's protocol. The recombinant protein was eluted using

an elution buffer composed of 20 mM Tris-HCl, 150 mM NaCl, 20% glycerol, 10mM 2-mercaptoethanol pH 7.9 and 2.5 mM d-desthiobiotin. The protein purity was assessed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and protein bands were visualized by staining with Coomassie brilliant blue G-250. The protein concentrations were calculated based on the molecular mass and computed extinction coefficient.

Protein Crystallization

Selenomethionine derivatized DesC protein from *C. scindens* ATCC 35704 yielded crystals from JCSG Core Suite (Qiagen Sciences, MD) using a drop volume of 0.2 mL and a 1:1 (vol/vol) ratio of reservoir to protein in 20 mM HEPES pH 8.0 solution equilibrated against 100 mL of reservoir solution. The apo-form of the recombinant DesC crystallized via sitting drop vapor diffusion at 20ºC against a crystallization solution containing 0.2M MgCl2, 24% PEG 400, 0.1M HEPES at pH 6.75. The crystals were cryoprotected with addition of 10% 1,2 ethanediol before harvesting and flash- freezing with liquid nitrogen. X-ray data were collected at BL14-1 at the Stanford Synchrotron Radiation Lightsource. The x-ray data were indexed and integrated with XDS[75] and scaled with XSCALE[76]. Multiwavelength anomalous dispersion (MAD) from the selenium signal used to determine the structure using the programs with SHELX[77,78]. Model refinement was performed with Refmac[79,80].

Mutational analysis of DesC to identify key residues for its activity and binding to substrate

In order to identify the key amino acid residues that may be involved in both activity and binding to cortisol, an amino acid sequence alignment was carried out with known zinc dependent medium chain dehydrogenase sequences obtained from NCBI. Mutations were made

in the pET-51b containing the wild type DesC using QuikChange Lightning Site-Directed Mutagenesis Kit (Agilent, Santa Clara) according to the manufacture's protocol. The primers used in the site-directed mutagenesis study are summarized in **Supplementary Table 5.4**. Mutated plasmids were transformed into *E. coli* XL10 competent cells by heat-shock and plated onto lysogeny broth (LB) solidified with Bacto-agar (Difco) containing 100 µg/mL ampicillin sodium salt, and the plates were incubated at 37 °C overnight. Individual colonies were cultivated in 5ml of LB medium supplemented with ampicillin (100 µg/mL) overnight at 37 °C, and plasmids were extracted using a QIAprep® Spin Miniprep Kit (Qiagen, Valencia, CA). Plasmids containing the expected mutations were confirmed by DNA sequencing (W.M. Keck Center for Comparative and Functional Genomics, University of Illinois at Urbana-Champaign). The mutated proteins were expressed and purified using the same procedure described above for wild type. The proteins were quantified based on their molecular mass and theoretical extinction coefficient.

Determination of structural integrity of the DesC wild type and mutants by circular dichroism (CD) spectra

Determination of circular dichroism (CD) spectra for both DesC wild type and its site-directed mutant proteins was carried out using a J-815 circular dichroism spectropolarimeter (Jasco, Tokyo, Japan). Protein samples were prepared at a concentration of 0.2 mg/ml in 10 mM KH_2PO_4 buffer (pH 7.5). For the measurements, a quartz cell with a path length of 0.1 cm was used. CD scans were carried out at 25°C from 190 nm to 260 nm at a speed of 50 nm/min with a 0.1-nm wavelength pitch, with five accumulations. Data files were analyzed on the DICHROWEB online server (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) using the

CDSSTR algorithm with reference set 4, which is optimized for analysis of data recorded in the range of 190 nm to 240 nm.

Isothermal titration calorimetric analysis of the DesC

Isothermal titration calorimetric (ITC) analysis was performed using a VP-ITC microcalorimeter with a 1.4 mL cell volume from MicroCal, Inc. The proteins were dialyzed with phosphate buffer (50 mM sodium phosphate, 150 mM NaCl, 10% glycerol, pH 7.5) and the pyridine nucleotide (NADH) and cortisol were dissolved in the same buffer. The proteins (50 μM) with and without pyridine nucleotide were then injected with 28 successive 10 μL aliquots of cortisol (0.5 mM) at 300-s intervals. The data were fitted to a nonlinear regression model using a single binding site (MicroCal Origin software). The thermodynamic parameters were calculated using the Gibbs free energy equation ($\Delta G = \Delta H - T\Delta S$), and the relationship $\Delta G = -RT\ln(Ka)$.

DesC enzyme assay

DesC enzyme activity for both wild type and mutants was determined as described previously. Linearity of enzyme activity with respect to time and enzyme concentration was determined aerobically by monitoring the oxidation of NADH at 340 nM ($\varepsilon = 6{,}220$ M-1•cm-1) in the presence of cortisol. The standard reaction mixture contained 50 mM phosphate buffer, pH 7.5, 50 μM of cortisol, 150 μM NADH, 0.05 μM enzyme, and buffer to a final volume of 0.5 ml. The reaction was started by the addition of the enzyme. The initial velocities of enzyme were plotted against the substrate concentrations, and the kinetic parameters were estimated by fitting

the data to the Michaelis-Menten equation by non-linear regression method using the enzyme kinetics module in GraphPad Prism (GraphPad Software, La Jolla, CA).

NADH and cortisol docking to DesC

Using BLAST[41], we obtained homologous structures (PDB IDs 4ILK, 4EJ6, 4A2C, 3QE3, 3GFB, 2DQ4, 2DFV, 2D8A, 1PL7, 1E3J) within the PDB. The alignment and placing of both NADH and cortisol molecules on their binding sites was performed using VMD[42]. Employing advanced run options of QwikMD[43], the structure of the ligands was minimized in the pocket together with the nearby DesC residues while maintaining the structure of most of DesC static. PyContact[81] was then used to analyze the contact interface.

Classical molecular dynamics simulations

MD simulations were performed employing the GPU-accelerated NAMD molecular dynamics package.[45] The simulations were performed assuming periodic boundary conditions in the NpT ensemble with temperature maintained at 300 K using Langevin dynamics for temperature and pressure coupling, the latter kept at 1 bar. A distance cut-off of 12.0 Å was applied to short-range non-bonded interactions, whereas long-range electrostatic interactions were treated using the particle-mesh Ewald (PME) method.[48] The equations of motion were integrated using the r-RESPA multiple time step scheme[82] to update the van der Waals interactions every step and electrostatic interactions every two steps. The time step of integration was chosen to be 2 fs, and before MD simulations, the system was submitted to an energy minimization protocol for 5,000 steps. An MD simulation with position restraints in the protein backbone atoms and ligands non-hydrogen atoms was performed for 10 ns. To allow for a total

relaxation of the system and to make sure ligands were stable in the DesC pocket, a 100 ns simulation in equilibrium, where no external forces were applied, was performed. The MD protocol served to pre-equilibrate the system before the hybrid QM/MM simulations.

Hybrid QM/MM calculations

Quantum mechanics (QM) is crucial to investigate subatomic mechanisms that occur in biology. However, studying entire biomolecular systems quantum mechanically is computationally prohibitive. Traditionally, NAMD[45] employs classical molecular mechanics (MM) to determine the movement of a molecule, solving Newton's equations of motion and treating atoms as spheres and bonds as springs. Combining both approaches, hybrid QM/MM methods employ the quantum mechanics formalism to key regions of the biological system, while using molecular mechanics approach to include the effects of the surrounding area. NAMD's QM/MM interface[50] can be combined with many molecular dynamics protocols, such as enhanced sampling and free energy calculations.[63] This combination was crucial in solving the mechanism of DesC. Employing advanced run options of VMD's[42] QwikMD[43] plugin, our in silico approach followed established protocols that were previously employed to investigate the mechanism of setting of the genetic code.[50]

In summary, we first preform a short 20 ps long hybrid QM/MM MD simulation using NAMD[50] and MOPAC[83,84], with a 2.0 fs integration time step. This equilibration simulation was followed by a hybrid QM/MM SMD simulation, with 0.5 fs integration time step, where possible reaction mechanisms were probed. The most favorable reaction mechanism was them tested experimentally and also using more advanced enhanced sampling techniques, namely string simulations. All simulations were carried out using the charge-shift method to treat link atoms,

222

and a "shift" function was applied to surrounding classical partial charges in the electrostatic embedding scheme.[50]

Hybrid QM/MM string simulations

To study a transformation that occurs in a biomolecular system, such as a chemical reaction[50] or a conformational change[85,86], one defines collective variables such as distances between atoms or between centers of mass of groups of atoms, or even angles between subdomains of a molecular structure[87,88]. The collective variables are used to track the changes in the system as it undergoes the transformation being studied and can be used to define a reaction coordinate.

The reaction coordinate defined in this study describes the chemical reaction carried out by DesC, and it was optimized by the string method, an iterative method that adapts the reaction coordinate as to fit it to the path of least resistance. Collective variables were defined as distances between atoms relevant for the reaction. From the initial state, before the chemical reaction has occurred, to the final state, after all transformations have taken place, multiple "images" are selected as representative stages of the overall transformation. Over multiple iterations, the string method initiates sets of short, unbiased, and independent QM/MM MD simulations from each image, creating full copies of the system that explore conformations the system can explore along the reaction coordinate. Then, each set of QM/MM MD simulations is used to re-define the image, leading to a smooth path from initial to final state. Iterations are performed until average consecutive changes converge to a stable value.

Here we have launched 750 multiple independent MD simulations per image, for a total of 25 images. The total simulation time was 4.5 ns, which was enough to converge the

simulations in the collective variables space. Collective variables were analyzed using in-house python scripts implemented in Jupyter notebooks.

Hybrid QM/MM free energy calculations

After optimizing the reaction coordinate, we applied the extended adaptive biasing force (eABF) method to estimate the free energy change along the reaction. In eABF, the QM/MM MD simulation is constrained to occur along the reaction coordinate using two new collective variables called "path collective variables" S and Z. The former describes the transformation of the system from start to end, and latter describes transformations "perpendicular" to the reaction coordinate, allowing minor adjustments and optimizations not obtained during the string optimization. We leveraged NAMD's scalability to initiate multiple QM/MM MD simulations along the reaction coordinate, leading to a multiple-walker eABF calculation that produced extensive sampling over the chemical reaction path.
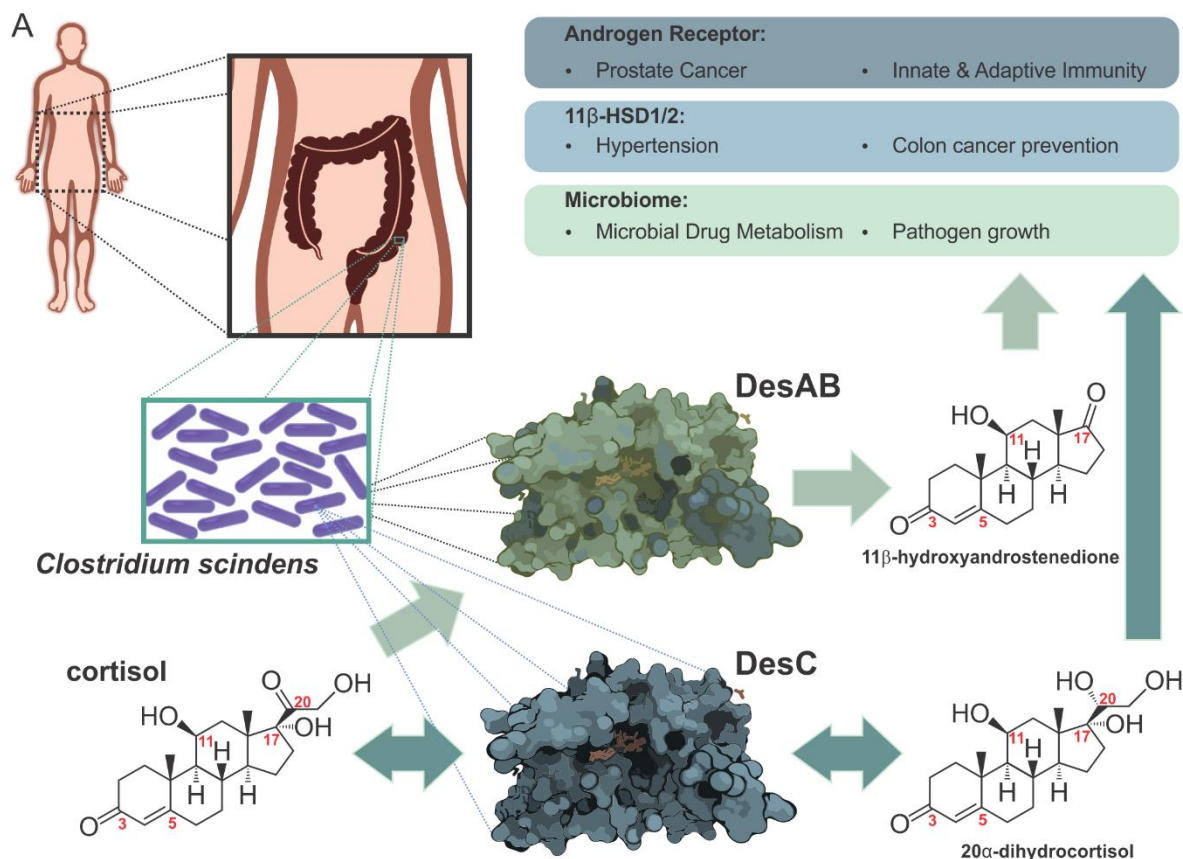
This approach allowed us to perform 3 ns of hybrid QM/MM simulation, which would not be feasible without the parallel distribution of eABF sampling. Specifically, 375 NCSA Blue Waters XE nodes were used to simulate 750 replicas (walkers) of the system in order to streamline the free energy calculation process.

ACKNOWLEDGMENTS

FIGURES

Figure 5.1. Overview of the metabolism of cortisol by the gut bacterium *Clostridium scindens*. The obligate anaerobic bacterium, *C. scindens* ATCC 35704, resides in the large intestine. This bacterium expresses a gene cluster (*desABCD*) encoding DesAB (no structure available) which removes the side-chain of cortisol, forming 11β-hydroxyandrostenedione. DesC (new structure presented in this manuscript) is an NAD(H)-dependent 20α−hydroxysteroid dehydrogenase that catalyzes the reversible oxidoreduction of the C20 ketone. The androgenic end product of DesAB is implicated in prostate cancer and may affect innate and adaptive immune function because many immune cells express androgen receptor. By inhibiting the host enzyme 11β-hydroxysteroid dehydrogenase 1 & 2 isoforms, these metabolites may also affect

Figure 5.1 (cont.)

blood pressure and inflammation in the GI tract. The formation of 20α-dihydrocortisol in the host

is associated with disorders such as Cushing's syndrome; however, little is known about the

physiological or pathophysiological role of 20-reduced forms of cortisol. Moreover, DesC

appears to regulate substrate availability to DesAB.
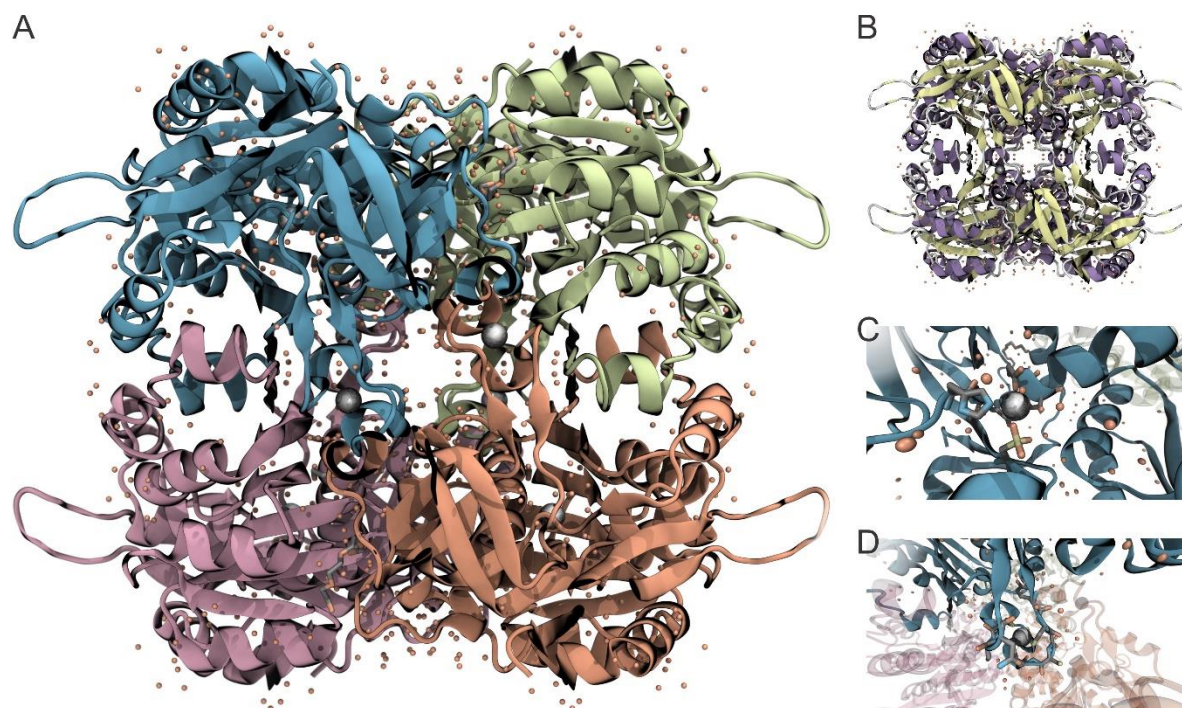
Figure 5.2. Depiction of DesC highlighting functional features. A. DesC structure with amino acid chains shown in cartoon representation and colored differently to show the bilobal topology. Structural waters are shown in red spheres, and two catalytic Zn^{2+} ions are shown as gray spheres. **B.** Same view of the complex colored by secondary structures: a-helices in purple and b-strands in green. **C.** Detailed view of the catalytic Zn^{2+} ion in octahedral coordination by water molecules and catalytic site residues. Amino acid residues are shown in licorice representation where carbon atoms are colored in gray, oxygen in red, nitrogen in blue and phosphorous in yellow. **D.** Detailed view of the helical loop that extends towards the adjacent monomer.
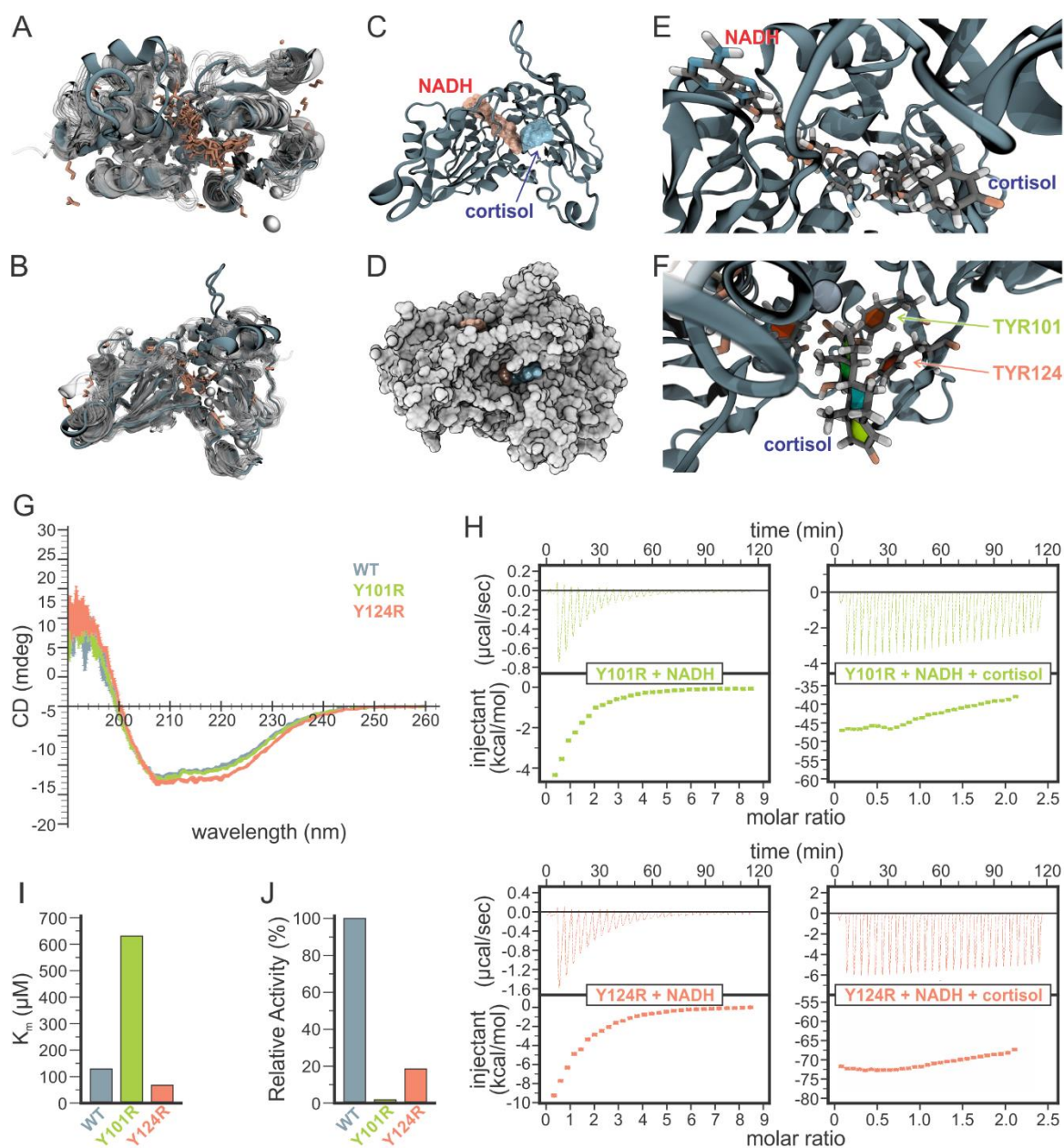
Figure 5.3. Characterization of the cortisol binding pocket of DesC. A. Homologous proteins

from the PDB were superimposed to highlight structure similarity and binding pose of multiple

ligands. All proteins are shown in cartoon representation while ligands are shown in orange

licorice representation. **B.** Same as previous, with alternate view of the binding site. **C.** Final

binding poses of NADH and cortisol to DesC. The protein is shown in cartoon representation. **D.**

Space-filling representation of DesC shows the alignment of cortisol deep in a cleft. **E.** Detailed

Figure 5.3 (cont.)

view of cortisol and NADH binding near the Zn^{2+} in the active site. Substrates are shown in ball-and-stick representations. Carbon atoms in gray, oxygen in red, nitrogen in blue and hydrogen in white. **F.** Detailed view of tyrosines 101 and 124 stabilizing cortisol's binding pose. **G.** Comparison of circular dichroism spectra (190-260 nm) between wild type (WT) recombinant DesC, Y101R mutant, and Y124R mutant. **H.** Isothermal titration calorimetry of Y101R mutant in the presence of 1.5 mM NADH (upper left panel), and 1.5 mM NADH + 0.5 mM cortisol (upper right panel) and Y124R mutant in the presence of 1.5 mM NADH (lower left panel), and 1.5 mM NADH + 0.5 mM cortisol (lower right panel). **I.** Effect of binding pocket mutations on K_m for cortisol in the presence of saturating NADH. **J.** Relative activity (%) of WT vs. binding pocket mutants when cortisol and NADH are saturating.
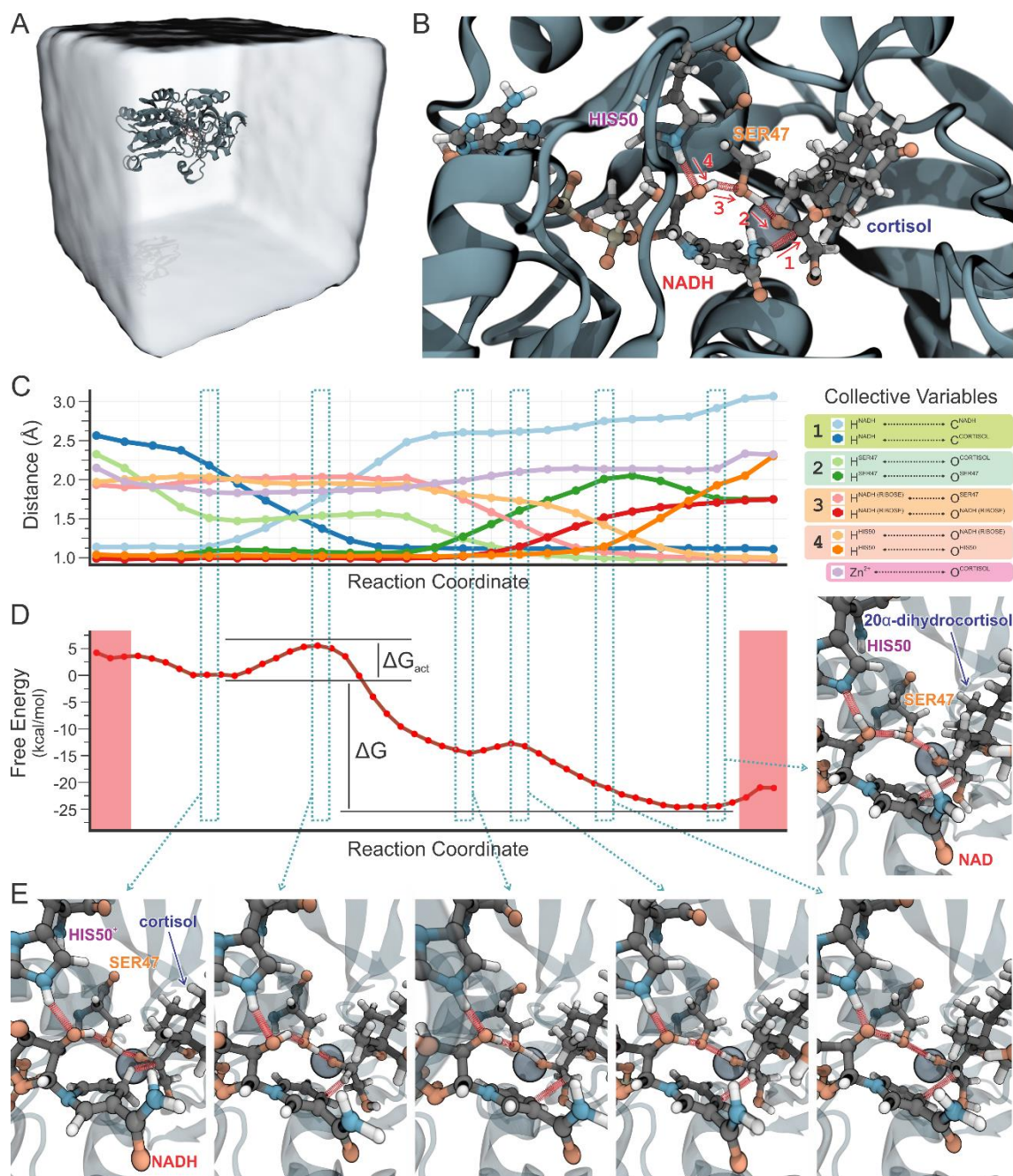
Figure 5.4. Description of QM/MM calculation of the reaction catalyzed by DesC. A.
Depiction of the entire simulated system, showing DesC in a simplified water box. **B.** Detailed
view of the active site showing all proton transfers that take place during the reaction. DesC is
shown in cartoon representation, while atoms in the QM region are shown in ball-and-stick
representations. Carbon atoms in gray, oxygen in red, nitrogen in blue and hydrogen in white. **C.**
Plot showing the progression of all collective variables tracked during the string method

Figure 5.4 (cont.)

optimization of the reaction coordinate. Collective variable pairs labeled with shades of the same color describe the distance between a hydrogen atom and its donor and acceptor carbon or oxygen atoms. The numbers correspond to the labelled reactions in section B of the figure. The interatomic distances show the proton relay progression during the catalytic process. **D.** Matching plot showing the same reaction coordinate and the results of the eABF free energy calculation. The first energy barrier corresponds to the proton exchange from NADH to cortisol. **E.** Multiple detailed views of the active site show the progression of the overall reaction. Blue dotted rectangles and arrows indicate which stage of the reaction is depicted.
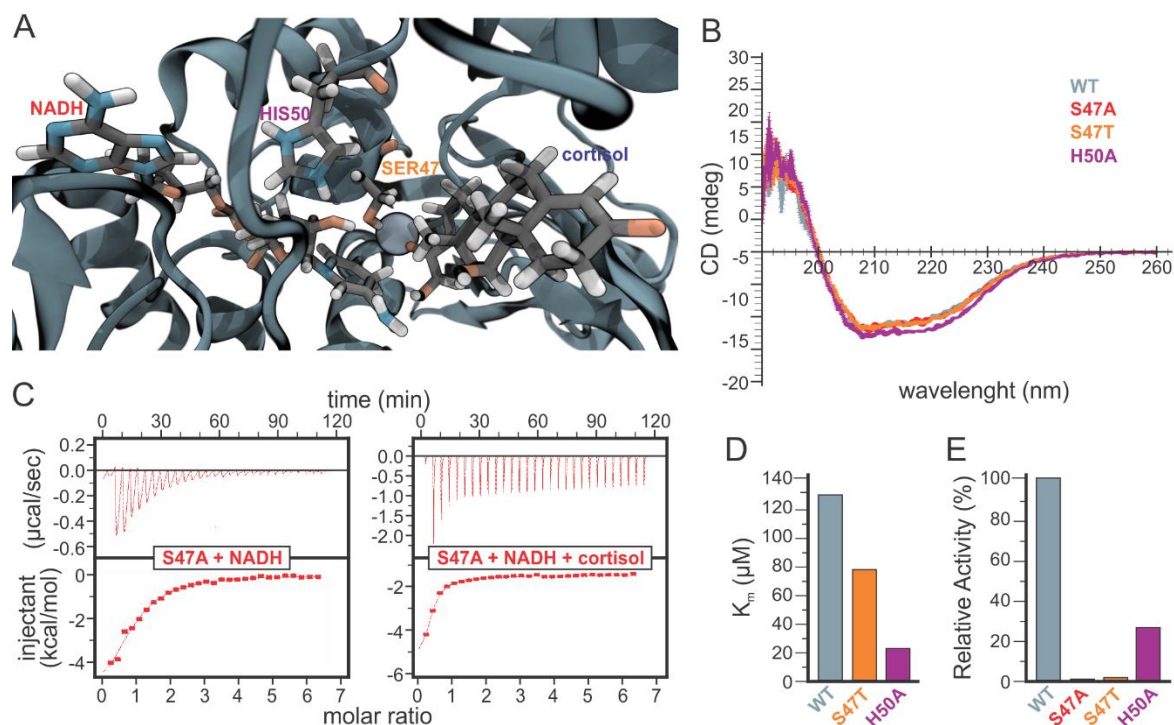
Figure 5.5. Role of Serine 47 in catalysis. A. Depiction of active site with bound reactants. The two residues highlighted were mutated to confirm their importance to binding and enzymatic activity. **B.** Comparison of circular dichroism spectra (190-260 nm) between wild type (WT) recombinant DesC, S47A mutant, S47T mutant, and H50A mutant. **C.** Isothermal titration calorimetry of S47A mutant in the presence of 1.5 mM NADH (left panel), and 1.5 mM NADH + 0.5 mM cortisol (right panel). **D.** Effect of active-site mutations on K_m for cortisol in the presence of saturating NADH. E. Relative activity (%) of WT vs. active-site mutants when cortisol and NADH are saturating.
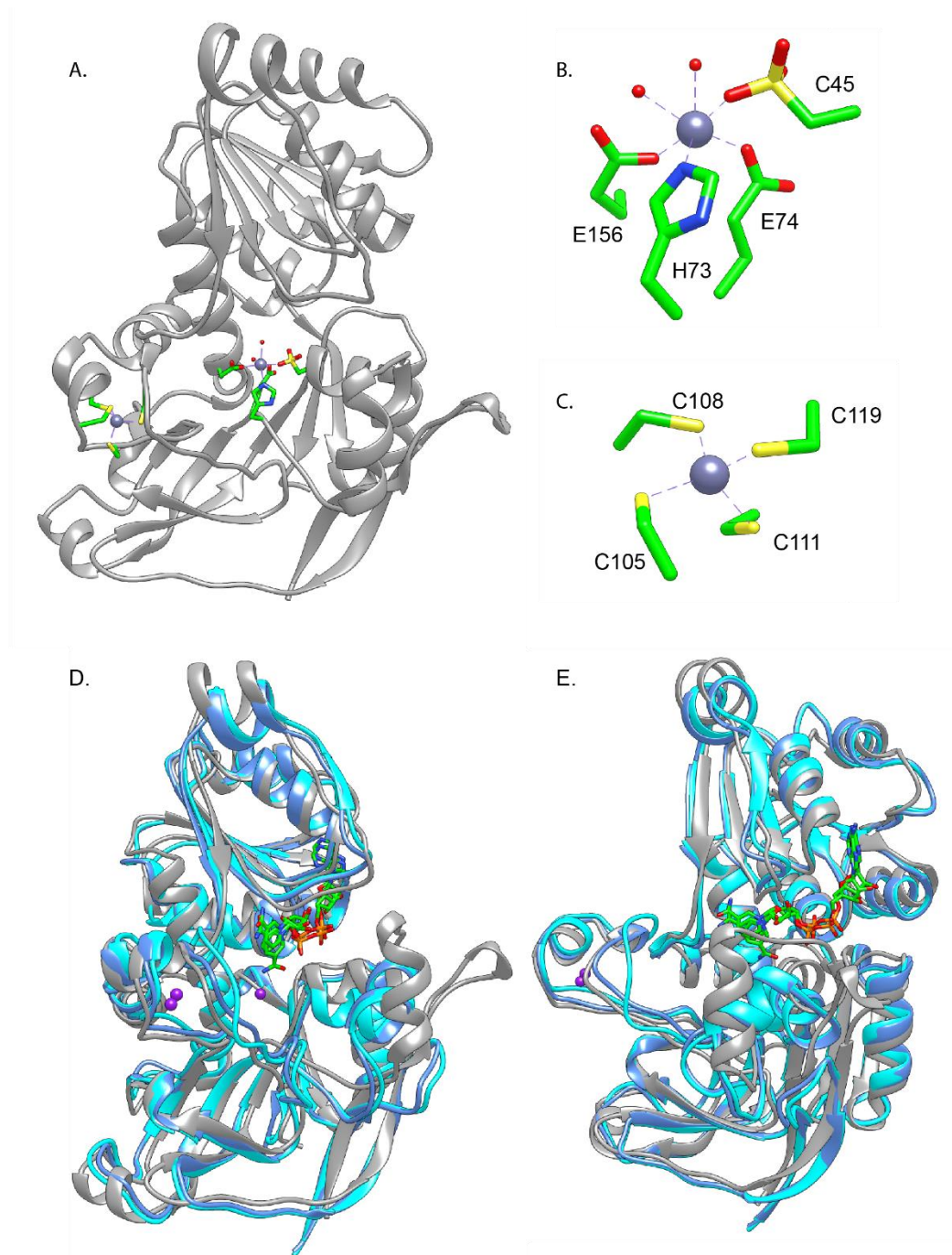
Supplementary Figure 5.1. Structure refinement. A. Ribbon structure of DesC. Zinc ions are displayed as purple spheres with coordinating residues in green sticks. B. Close-up of Zn^{2+} within the catalytic site. C. Close-up of Zn^{2+} at the lobe loop extension. D. Overlay of DesC

Supplementary Figure 5.1 (cont.)

(gray ribbon), with *Pyrococcus horikoshii* threonine dehydrogenase (2DFV, blue ribbon) and *Thermococcus kodakaraensis* threonine dehydrogenase (3GFB, cyan ribbon). Zinc is displayed as purple spheres and NAD+ bound to the threonine dehydrogenases is shown in green sticks. E. Overlay of DesC as in panel D, rotated 90° to show the extended loops of DesC surrounding the active site.
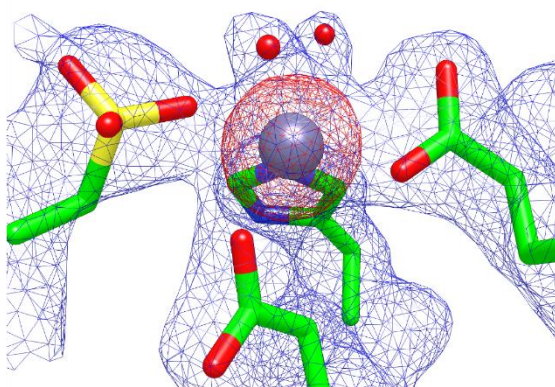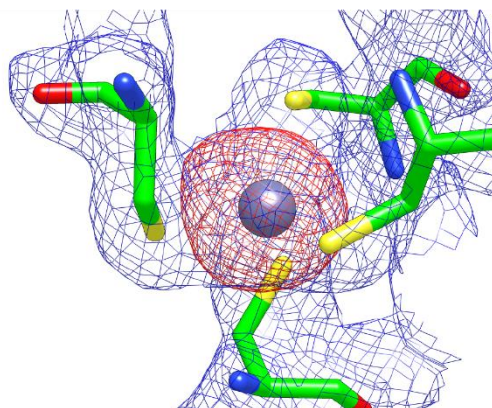
Supplementary Figure 5.2. Electron density. A. Electron density for Zn^{2+} bound at the active site. The $|Fo\text{-}Fc|$ density for Zn^{2+} is colored red and contoured at 7 σ. The $|2Fo\text{-}Fc|$ density for Zn and the coordinating residues is colored blue and contoured at 1.3 σ. B. Electron density for Zn^{2+} bound at the lobe loop. The $|Fo\text{-}Fc|$ density for Zn^{2+} is colored red and contoured at 5 σ. The $|2Fo\text{-}Fc|$ density for Zn^{2+} and the coordinating residues is colored blue and contoured at 1.0 σ.

SUPPLEMENTARY TABLES

Supplementary Table 5.1. Data collection and refinement statistics.

PDB code	4OH1
Space group	I222
Unit cell (Å)	a= 63.87, b=85.60, c=149.09
Resolution (Å)	$29.4 - 2.0 (2.05 - 2.00)^a$
Total Reflections	27918 (1995)
Completeness (%)	99.5 (98.1)
Mean I/sigma (I)	10.15 (1.4)
R_{merge}	0.05 (0.61)
$R_{crystal}/R_{free}$	0.168/0.216 (0.311, 0.297)
R_{free} test set	1403 reflections (5.00%)
Wilson B-factor ($Å^2$)	35.5
Fo-Fc correlation(free)	0.97 (0.95)
Total number of atoms	2918
Average B, all atoms ($Å^2$)	49
Ramachandran outliers	0
RMS bonds (Å)	0.02
RMS angles (º)	1.8
Sidechain outliers	1.40%
RSRZ outliers	7.90%
Numbers in final Model	
Residues	351
Protein Atoms	2704
Catalytic Zn^{2+}	2
Water Molecules	193
Chloride	3
1,2-ethanediol	1
tetraethylene glycol	1
Mean B-values ($Å^2$)	
Protein atoms	36.2
tetraethylene glycol	67.75
1,2-ethanediol	72.75
Zinc 1	47
Zinc 2	49
Chloride 1	45
Chloride 2	57

^{a} Values in parenthesis are for the highest resolution shell unless otherwise indicated.

Supplementary Table 5.2. Elution profile of zinc-finger Cys mutants by gel filtration chromatography.

Standards	MW (Kda)	Ve/Vo		DesC mutants	Ve/Vo	Estimated MW (Kda)
Blue dextran	2000	1		C105A	0.997871	>2000
Thyroglobulin	670	1.00795		C108A	0.997445	>2000
Gamma-globulin	158	1.028038		C111A	0.997658	>2000
Ovalbumin	44	1.045358		C119A	0.998935	>2000
Myoglobin	17	1.057496		C105A-C119A	0.999148	>2000
Vitamin B12	1.35	1.079074		C108A-C119A	0.999219	>2000
				WT	1.0264	164

Supplementary Table 5.3. Kinetic parameters for DesC activity.

Enzyme	K_m (mM)	V_{max} (nmol.min^{-1}.mg^{-1})	K_{cat} (min^{-1})	K_{cat}/K_m (min^{-1}. mM^{-1})	Activity loss (%)
WT	128.34 ± 5.26	1122.3 ± 64.19	44.33 ± 2.54	0.35 ± 0.03	0
S47A	NA	NA	NA	NA	100
H50A	23.99 ± 0.26	316.36 ± 67.55	12.5 ± 0.89	0.52 ± 0.04	71.81 ± 6.02
S47T	77.78 ± 1.78	16.66 ± 0.90	0.66 ± 0.04	0.0085 ± 0.0006	98.52 ± 0.08
Y101R	628.65 ± 29.96	44.80 ± 7.54	1.77 ± 0.30	0.0028 ± 0.0006	96.01 ± 0.67
Y124R	61.01 ± 2.71	204 ± 3.45	8.06 ± 0.14	0.13 ± 0.01	81.82 ± 0.31

Supplementary Table 5.4. Primers employed in the study.

Gene	5'-forward primer-3' 5'-reverse primer-3'
DesC WT	ATATAT*CCATGG*CAATGAGACAATTATTTGTTACTTCCA ATATAT*AAGCTT*CTACTTTTCGAACTGCGGGTGGCTCCATTCGTCCATCTTAATTACG ATC
DesC S47A	GTTTATGCGTCCATTTGTGGTGCGGATACACATATTCTTACCGG CCGGTAAGAATATGTGTATCCGCACCACAAATGGACGCATAAAC
DesC S47T	GCGTCCATTTGTGGTACCGATACACATATTC GAATATGTGTATCGGTACCACAAATGGACGC
DesC H50A	CATTTGTGGTTCCGATACAGCGATTCTTACCGGAAATCTG CAGATTTCCGGTAAGAATCGCTGTATCGGAACCACAAATG
DesC Y101R	CAGAAGGTTGTTGCAAACCGTGCAAAATACTGTGGATG CATCCACAGTATTTTGCACGGTTTGCAACAACCTTCTG
DesC Y124R	CTCTGCTCCAATATGGGACGTCGGATGAATGGATTTTC GAAAATCCATTCATCCGACGTCCCATATTGGAGCAGAG

REFERENCES

1.   Cann I, Bernardi RC, Mackie RI. Cellulose degradation in the human gut: Ruminococcus champanellensis expands the cellulosome paradigm. Environ Microbiol 2016; 18:307–10.
2.   David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling A V, Devlin AS, Varma Y, Fischbach MA, et al. Diet rapidly and reproducibly alters the human gut microbiome. Nature [Internet] 2014; 505:559–63. Available from: http://dx.doi.org/10.1038/nature12820
3.   Turnbaugh PJ, Gordon JI. The core gut microbiome, energy balance and obesity. J Physiol 2009; 587:4153–8.
4.   Milles LF, Schulten K, Gaub HE, Bernardi RC. Molecular mechanism of extreme mechanostability in a pathogen adhesin. Science 2018; 359:1527–33.
5.   Bereshchenko O, Bruscoli S, Riccardi C. Glucocorticoids, Sex Hormones, and Immunity. Front Immunol 2018; 9:1332.
6.   Cain DW, Cidlowski JA. Immune regulation by glucocorticoids. Nat Rev Immunol 2017; 17:233–47.
7.   Ridlon JM. Conceptualizing the vertebrate sterolbiome. Appl Environ Microbiol 2020; 86:e00641-20.
8.   Woods C, Tomlinson JW. The Dehydrogenase Hypothesis. In: Advances in Experimental Medicine and Biology. Springer New York LLC; 2015. page 353–80.
9.   Morris DJ, Ridlon JM. Glucocorticoids and gut bacteria: "The GALF Hypothesis" in the metagenomic era. Steroids 2017; 125:1–13.
10.   Macdonald IA, Bokkenheuser VD, Winter J, McLernon AM, Mosbach EH. Degradation of steroids in the human gut. J Lipid Res 1983; 24:675–700.
11.   Benach J, Filling C, Oppermann UCT, Roversi P, Bricogne G, Berndt KD, Jörnvall H, Ladenstein R. Structure of bacterial 3beta/17beta-hydroxysteroid dehydrogenase at 1.2 A resolution: a model for multiple steroid recognition. Biochemistry 2002; 41:14659–68.
12.   Ghosh D, Wawrzak Z, Weeks CM, Duax WL, Erman M. The refined three-dimensional structure of 3α,20β-hydroxysteroid dehydrogenase and possible roles of the residues conserved in short-chain dehydrogenases. Structure 1994; 2:629–40.
13.   Devendran S, Méndez-García C, Ridlon JM. Identification and characterization of a 20β-HSDH from the anaerobic gut bacterium *Butyricicoccus desmolans* ATCC 43058. J Lipid Res 2017; 58:916–25.
14.   Doden HL, Pollet RM, Mythen SM, Wawrzak Z, Devendran S, Cann I, Koropatkin NM, Ridlon JM. Structural and biochemical characterization of 20β-hydroxysteroid dehydrogenase from Bifidobacterium adolescentis strain L2-32. J Biol Chem 2019; 294:12040–53.
15.   Stokes NA, Hylemon PB. Characterization of delta 4-3-ketosteroid-5β-reductase and 3β-hydroxysteroid dehydrogenase in cell extracts of *Clostridium innocuum*. Biochim Biophys Acta 1985; 836:255–61.
16.   Swart AC, Storbeck KH. 11β-hydroxyandrostenedione: Downstream metabolism by 11βHSD, 17βHSD and SRD5A produces novel substrates in familiar pathways. Mol Cell Endocrinol 2015; 408:114–23.
17.   Shackleton CHL, Neres MS, Hughes BA, Stewart PM, Kater CE. 17-Hydroxylase/C17,20-lyase (CYP17) is not the enzyme responsible for side-chain cleavage of cortisol and its metabolites. Steroids 2008; 73:652–6.

18.   Storbeck K, Bloem LM, Africander D, Schloms L, Swart P, Swart AC. 11β-Hydroxydihydrotestosterone and 11-ketodihydrotestosterone, novel C19 steroids with androgenic activity: A putative role in castration resistant prostate cancer? Mol Cell Endocrinol 2013; 377:135–46.

19.   Pretorius E, Arlt W, Storbeck K-H. A new dawn for androgens: Novel lessons from 11-oxygenated C19 steroids. Mol Cell Endocrinol 2017; 441:76–85.

20.   Pretorius E, Africander DJ, Vlok M, Perkins MS, Quanson J, Storbeck K-H. 11-Ketodihydrotestosterone in castration resistant prostate cancer: Potent androgens which can no longer be ignored. PLoS One 2016; 11:e0159867.

21.   Turcu AF, Auchus RJ. Clinical significance of 11-oxygenated androgens. Curr Opin Endocrinol Diabetes Obes 2017; 24:252–9.

22.   Morris DJ, Latif SA, Hardy MP, Brem AS. Endogenous inhibitors (GALFs) of 11β-hydroxysteroid dehydrogenase isoforms 1 and 2: derivatives of adrenally produced corticosterone and cortisol. J Steroid Biochem Mol Biol 2007; 104:161–8.

23.   Wade AP, Slater JD, Kellie AE, Holliday ME. Urinary excretion of 17-ketosteroids following rectal infusion of cortisol. J Clin Endocrinol Metab 1959; 19:444–53.

24.   Harris SC, Devendran S, Méndez-García C, Mythen SM, Wright CL, Fields CJ, Hernandez AG, Cann I, Hylemon PB, Ridlon JM. Bile acid oxidation by *Eggerthella lenta* strains C592 and DSM 2243T. Gut Microbes 2018; 9:523–39.

25.   de Prada P, Setchell KD, Hylemon PB. Purification and characterization of a novel 17 alpha-hydroxysteroid dehydrogenase from an intestinal *Eubacterium* sp. VPI 12708. J Lipid Res 1994; 35:922–9.

26.   Bokkenheuser VD, Morris GN, Ritchie AE, Holdeman L V., Winter J. Biosynthesis of androgen from cortisol by a species of *Clostridium* recovered from human fecal flora. J Infect Dis 1984; 149:489–94.

27.   Ridlon JM, Ikegawa S, Alves JMP, Zhou B, Kobayashi A, Iida T, Mitamura K, Tanabe G, Serrano M, De Guzman A, et al. *Clostridium scindens*: a human gut microbe with a high potential to convert glucocorticoids into androgens. J Lipid Res 2013; 54:2437–49.

28.   Devendran S, Mythen SM, Ridlon JM. The desA and desB genes from *Clostridium scindens* ATCC 35704 encode steroid-17,20-desmolase. J Lipid Res 2018; 59:1005–14.

29.   Javdan B, Lopez JG, Chankhamjon P, Lee YCJ, Hull R, Wu Q, Wang X, Chatterjee S, Donia MS. Personalized mapping of drug metabolism by the human gut microbiome. Cell 2020; 181:1661–79.

30.   Ly LK, Rowles JL, Paul HM, Alves JMP, Yemm C, Wolf PM, Devendran S, Hudson ME, Morris DJ, Erdman JW, et al. Bacterial steroid-17,20-desmolase is a taxonomically rare enzymatic pathway that converts prednisone to 1,4-androstanediene-3,11,17-trione, a metabolite that causes proliferation of prostate cancer cells. J Steroid Biochem Mol Biol 2020; 199:105567.

31.   Kavanagh KL, Jörnvall H, Persson B, Oppermann U. Medium- and short-chain dehydrogenase/reductase gene and protein families. Cell Mol Life Sci 2008; 65:3895–906.

32.   Baker PJ, Britton KL, Fisher M, Esclapez J, Pire C, Bonete MJ, Ferrer J, Rice DW. Active site dynamics in the zinc-dependent medium chain alcohol dehydrogenase superfamily. Proc Natl Acad Sci 2009; 106:779–84.

33.   Esposito L, Bruno I, Sica F, Raia CA, Giordano A, Rossi M, Mazzarella L, Zagari A. Crystal Structure of a ternary complex of the alcohol dehydrogenase from *Sulfolobus solfataricus*. Biochemistry 2003; 42:14397–407.

34. Ishikawa K, Higashi N, Nakamura T, Matsuura T, Nakagawa A. The first crystal structure of L-threonine dehydrogenase. J Mol Biol 2007; 366:857–67.

35. Bowyer A, Mikolajek H, Stuart JW, Wood SP, Jamil F, Rashid N, Akhtar M, Cooper JB. Structure and function of the L-threonine dehydrogenase (TkTDH) from the hyperthermophilic archaeon *Thermococcus kodakaraensis*. J Struct Biol 2009; 168:294–304.

36. Eklund H, Nordström B, Zeppezauer E, Söderlund G, Ohlsson I, Boiwe T, Söderberg B-O, Tapia O, Brändén C-I, Åkeson Å. Three-dimensional structure of horse liver alcohol dehydrogenase at 2.4 Å resolution. J Mol Biol 1976; 102:27–59.

37. Tressel T, Thompson R, Zieske LR, Menendez MI, Davis L. Interaction between L-threonine dehydrogenase and aminoacetone synthetase and mechanism of aminoacetone production. J Biol Chem 1986; 261:16428–37.

38. Rossmann MG, Moras D, Olsen KW. Chemical and biological evolution of a nucleotide-binding protein. Nature 1974; 250:194–9.

39. Berman HM. The Protein Data Bank. Nucleic Acids Res 2000; 28:235–42.

40. Bernardi RC, Cann I, Schulten K. Molecular dynamics study of enhanced Man5B enzymatic activity. Biotechnol Biofuels 2014; 7:1–8.

41. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990; 215:403–10.

42. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. J Mol Graph 1996; 14:33–8.

43. Ribeiro J V, Bernardi RC, Rudack T, Stone JE, Phillips JC, Freddolino PL, Schulten K. QwikMD — Integrative molecular dynamics toolkit for novices and experts. Sci Rep 2016; 6:26536.

44. Goh BC, Hadden JA, Bernardi RC, Singharoy A, McGreevy R, Rudack T, Cassidy CK, Schulten K. Computational methodologies for real-space structural refinement of large macromolecular complexes. Annu Rev Biophys 2016; 45.

45. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K, et al. Scalable molecular dynamics with NAMD. J Comput Chem 2005; 26:1781–802.

46. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, Mackerell AD, MacKerell Jr. AD. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ1 and χ2 dihedral angles. J Chem Theory Comput 2012; 8:3257–73.

47. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983; 79:926.

48. Darden T, York D, Pedersen L. Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. J Chem Phys 1993; 98:10089–92.

49. Bernardi RC, Durner E, Schoeler C, Malinowska KH, Carvalho BG, Bayer EA, Luthey-Schulten Z, Gaub HE, Nash MA. Mechanisms of nanonewton mechanostability in a protein complex revealed by molecular dynamics simulations and single-molecule force spectroscopy. J Am Chem Soc 2019; 141:14752–63.

50. Melo MCR, Bernardi RC, Rudack T, Scheurer M, Riplinger C, Phillips JC, Maia JDC, Rocha GB, Ribeiro J V., Stone JE, et al. NAMD goes quantum: An integrative suite for hybrid simulations. Nat Methods 2018; 15:351–4.

51. Senn HM, Thiel W. QM/MM methods for biological systems. At Approaches Mod Biol from Quantum Chem to Mol Simulations 2007; 268:173–290.
52. Van Der Kamp MW, Mulholland AJ. Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. Biochemistry 2013; 52:2708–28.
53. Rosta E, Klähn M, Warshel A. Towards accurate ab initio QM/MM calculations of free-energy profiles of enzymatic reactions. J Phys Chem B 2006; 110:2934–41.
54. Kaila VRI, Wikström M, Hummer G. Electrostatics, hydration, and proton transfer dynamics in the membrane domain of respiratory complex I. Proc Natl Acad Sci U S A 2014; 111:6988–93.
55. Senn HM, Thiel W. QM/MM studies of enzymes. Curr Opin Chem Biol 2007; 11:182–7.
56. Bernardi RC, Pascutti PG. Hybrid QM/MM molecular dynamics study of benzocaine in a membrane environment: How does a quantum mechanical treatment of both anesthetic and lipids affect their interaction. J Chem Theory Comput 2012; 8:2197–203.
57. Mendes YS, Alves NS, Souza TLF, Sousa IP, Bianconi ML, Bernardi RC, Pascutti PG, Silva JL, Gomes AMO, Oliveira AC. The structural dynamics of the flavivirus fusion peptide-membrane interaction. PLoS One 2012; 7:e47596.
58. Stewart JJP. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. J Mol Model 2013; 19:1–32.
59. Singh UC, Kollman PA. A combined ab initio quantum mechanical and molecular mechanical method for carrying out simulations on complex molecular systems: Applications to the CH3Cl + Cl? exchange reaction and gas phase protonation of polyethers. J Comput Chem 1986; 7:718–30.
60. Dhoke G V., Davari MD, Schwaneberg U, Bocola M. QM/MM calculations revealing the resting and catalytic states in zinc-dependent medium-chain dehydrogenases/reductases. ACS Catal 2015; 5:3207–15.
61. Pan AC, Sezer D, Roux BB. Finding transition pathways using the string method with swarms of trajectories. J Phys Chem B 2008; 112:3432–40.
62. Fu H, Shao X, Chipot C, Cai W. Extended Adaptive Biasing Force algorithm. An on-the-fly implementation for accurate free-energy calculations. J Chem Theory Comput 2016; 12:3506–13.
63. Bernardi RC, Melo MCR, Schulten K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. Biochim Biophys Acta 2015; 1850:872--877.
64. Comer J, Phillips JC, Schulten K, Chipot C. Multiple-replica strategies for free-energy calculations in NAMD: Multiple-walker adaptive biasing force and walker selection rules. J Chem Theory Comput 2014; 10:5276–85.
65. Thackray VG. Sex, microbes, and polycystic ovary syndrome. Trends Endocrinol Metab 2019; 30:54–65.
66. Yurkovetskiy L, Burrows M, Khan AA, Graham L, Volchkov P, Becker L, Antonopoulos D, Umesaki Y, Chervonsky A V. Gender bias in autoimmunity is influenced by microbiota. Immunity 2013; 39:400–12.
67. Markle JGM, Frank DN, Mortin-Toth S, Robertson CE, Feazel LM, Rolle-Kampczyk U, von Bergen M, McCoy KD, Macpherson AJ, Danska JS. Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity. Science 2013; 339:1084–8.

68. García-Gómez E, González-Pedrajo B, Camacho-Arroyo I. Role of sex steroid hormones in bacterial-host interactions. Biomed Res Int 2013; 2013.

69. Menon R, Watson SE, Thomas LN, Allred CD, Dabney A, Azcarate-Peril MA, Sturino JM. Diet complexity and estrogen receptor β status affect the composition of the murine intestinal microbiota. Appl Environ Microbiol 2013; 79:5763–73.

70. Xiao L, Estellé J, Kiilerich P, Ramayo-Caldas Y, Xia Z, Feng Q, Liang S, Pedersen A, Kjeldsen NJ, Liu C, et al. A reference gene catalogue of the pig gut microbiome. Nat Microbiol 2016; 1:1–6.

71. Turcu AF, Rege J, Auchus RJ, Rainey WE. 11-Oxygenated androgens in health and disease. Nat. Rev. Endocrinol.2020; 16:284–96.

72. Ridlon JM, Bajaj JS. The human gut sterolbiome: Bile acid-microbiome endocrine aspects and therapeutics. Acta Pharm Sin B 2015; 5:99–105.

73. Tubeleviciute A, Teese MG, Jose J. *Escherichia coli* kduD encodes an oxidoreductase that converts both sugar and steroid substrates. Appl Microbiol Biotechnol 2014; 98:5471–85.

74. Inazu A, Sato K, Nakayama T, Deyashiki Y, Hara A, Nozawa Y. Purification and characterization of a novel dimeric 20α-hydroxysteroid dehydrogenase from Tetrahymena pyriformis. Biochem J 1994; 297:195–200.

75. Kabsch W. Integration, scaling, space-group assignment and post-refinement. Acta Crystallogr Sect D Biol Crystallogr 2010; 66:133–44.

76. Diederichs K. Some aspects of quantitative analysis and correction of radiation damage. Acta Crystallogr Sect D Biol Crystallogr 2006; 62:96–101.

77. Sheldrick GM. A short history of SHELX. Acta Crystallogr. Sect. A Found. Crystallogr.2008; 64:112–22.

78. Sheldrick GM. Experimental phasing with SHELXC/D/E: Combining chain tracing with density modification. Acta Crystallogr Sect D Biol Crystallogr 2010; 66:479–85.

79. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr Sect D Biol Crystallogr 1997; 53:240–55.

80. Winn MD, Murshudov GN, Papiz MZ. Macromolecular TLS refinement in REFMAC at moderate resolutions. Methods Enzymol 2003; 374:300–21.

81. Scheurer M, Rodenkirch P, Siggel M, Bernardi RC, Schulten K, Tajkhorshid E, Rudack T. PyContact: Rapid, customizable, and visual analysis of noncovalent interactions in MD simulations. Biophys J 2018; 114:577–83.

82. Bernardi R, Bhandarkar M, Bhatele A, Bohm E, Brunner R, Buelens F, Chipot C, Dalke A, Dixit S, Fiorin G, et al. NAMD User's Guide. Urbana, Illinois, USA Theor Comput Biophys Group, Beckman Institute, Univ Illinois 2018;

83. Stewart JJP. MOPAC: A semiempirical molecular orbital program. J Comput Aided Mol Des 1990; 4:1–103.

84. Maia JDC, Urquiza Carvalho GA, Mangueira CP, Santana SR, Cabral LAF, Rocha GB. GPU linear algebra libraries and GPGPU programming for accelerating MOPAC semiempirical quantum chemistry calculations. J Chem Theory Comput 2012; 8:3072–81.

85. Hoelz LVB, Bernardi RC, Horta BAC, Araújo JQ, Albuquerque MG, da Silva JFM, Pascutti PG, de Alencastro RB. Dynamical behaviour of the human β1-adrenoceptor under agonist binding. Mol Simul 2011; 37:907–13.

86. Hoelz LVB, Ribeiro AAST, Bernardi RC, Horta BAC, Albuquerque MG, da Silva JFM, Pascutti PG, de Alencastro RB. The role of helices 5 and 6 on the human β1-adrenoceptor activation mechanism. Mol Simul 2012; 38:236–40.

87. Fiorin G, Klein ML, Hénin J. Using collective variables to drive molecular dynamics simulations. Mol Phys 2013; 111:3345–62.

88. Hénin J, Fiorin G, Chipot C, Klein ML. Exploring multidimensional free energy landscapes using time-dependent biases on collective variables. J Chem Theory Comput 2010; 6:35–47.

CHAPTER 6

CONCLUSIONS AND FUTURE DIRECTIONS

BILE ACID 12α- AND 12β-HYDROXYSTEROID DEHYDROGENASE

Bile acid 12α-hydroxysteroid dehydrogenase (HSDH) activity was characterized in the high-activity 7α-dehydroxylating organisms *Clostridium scindens*, *C. hylemonae*, and *C. hiranonis* (reclassified as *Peptacetobacter hiranonis*).[1] The first gene encoding bile acid 12β-HSDH, which together with 12α-HSDH completes the epi-bile acid pathway, was also discovered from *Clostridium paraputrificum*.[2] Phylogenetic analysis of 12α-HSDH suggests that this activity is widespread across human gut microorganisms while 12β-HSDH is less ubiquitous. Through phylogenetic analysis of 12β-HSDH, two additional genes encoding this activity were validated. Importantly, by comparing sequences within the 12α- and 12β-HSDH trees, two *Collinsella* strains were identified that share both activities. These organisms represent the first identification of potential C-12 epimerizing strains.

12α- and 12β-HSDH may be of therapeutic importance in the future because the epi-pathway substrate, deoxycholic acid (DCA), has been mechanistically associated with cancers of the liver and colon.[3,4] The paired 12α- and 12β-HSDH activities in *Collinsella tanakaei* and *C. stercoris* strains evolved together and would likely be the best targets for favorable conversion of DCA to epiDCA. Thus, these enzymes may be used for future mechanistic studies on the effects of the epi-bile acid pathway and its products on the host. Currently, there are no crystal structures of 12α- or 12β-HSDH. Their structural biology will be important in the rational engineering of

these enzymes for optimal conversion of DCA to epiDCA, or for alteration of substrate specificity.

CORTISOL 20α- AND 20β-HYDROXYSTEROID DEHYDROGENASE

20β-HSDH from *Bifidobacterium adolescentis* strain L2-32 was biochemically and structurally studied.[5] The cortisol 20β-HSDH was crystallized in the apo-form and binary complex. A long, flexible N-terminal region was identified within the structure and investigated further, revealing its role in protein stability. The partner of 20β-HSDH with opposite stereospecificity, cortisol 20α-HSDH, from *Clostridium scindens* ATCC 35704 was also characterized and crystallized.[6] This HSDH structure represents the first of its specificity. A reaction mechanism was proposed using quantum mechanical simulations with validation through amino acid mutagenesis studies. Two additional putative 20α-HSDHs from *Denitratisoma oestradiolicum* DSM 16959 and *Intestinibacillus* sp. Marseille-P4005 were identified through BLASTP search based on high sequence similarity to *C. scindens* 20α-HSDH.[2]

Both 20α- and 20β-HSDH are putative regulators of the pro-androgenic cortisol steroid-17,20-desmolase (DesAB) pathway.[7,8] The product of the DesAB pathway can be further converted to potent androgens[9], which may be a previously unknown contributor to androgen-related diseases, such as castration-resistant prostate cancer and polycystic ovary syndrome[10]. While research must be done to first confirm the pro-androgenic role of microbial DesAB activity *in vivo*, 20α- and 20β-HSDH pose important therapeutic targets for androgen-related disease. Future research may entail utilizing their crystal structures and site-directed mutagenesis to optimize competitive conversion of cortisol away from the DesAB pathway. Subsequently,

gnotobiotic animal studies may be performed to determine the role of 20α- and 20β-HSDH products within the host and test competition with DesAB *in vivo*.

THE FUTURE OF HYDROXYSTEROID DEHYDROGENASES

The collection of studies herein identifying and characterizing new HSDH enzymes strengthens the foundation for future mechanistic work on the impacts of HSDHs on host physiology. Indeed, linking HSDH activity to host phenotypes is an important next step to the study of these diverse enzymes.[11,12] The development of HSDH gene knockouts, or alternatively integrating HSDH genes into genetically tractable microorganisms, along with gnotobiotic animal experiments will allow exploration of HSDH mechanisms.

The therapeutic HSDHs, confirmed by mechanistic studies, may then be leveraged against steroid-driven disease states. For example, recent advances in synthetic biology are allowing for the construction of genetically engineered bacteria for therapeutic and diagnostic purposes. Current applications of preclinically tested engineered bacteria include expression of drugs, drug-activating enzymes, and antimicrobial peptides.[13] These approaches generally focus on alleviating disease symptoms, not potential sources of disease progression within the body. In contrast, the therapeutic promise of HSDHs may come from their modulation of endogenous disease-promoting steroids. Thus, rationally designed HSDH-encoding probiotic microorganisms that exploit microbial metabolism of endogenous steroids may be developed for the treatment of steroid-driven disease.

REFERENCES

1.   Doden H, Sallam LA, Devendran S, Ly L, Doden G, Daniel SL, Alves JMP, Ridlon JM. Metabolism of oxo-bile acids and characterization of recombinant 12α-hydroxysteroid dehydrogenases from bile acid 7α-dehydroxylating human gut bacteria. Appl Environ Microbiol 2018; 84:e00235-18.

2.   Doden HL, Wolf PG, Gaskins HR, Anantharaman K, Alves JMP, Ridlon JM. Completion of the gut microbial epi-bile acid pathway. bioRxiv 2021.

3.   Bernstein C, Holubec H, Bhattacharyya AK, Nguyen H, Payne CM, Zaitlin B, Bernstein H. Carcinogenicity of deoxycholate, a secondary bile acid. Arch Toxicol 2011; 85:863–71.

4.   Yoshimoto S, Loo TM, Atarashi K, Kanda H, Sato S, Oyadomari S, Iwakura Y, Oshima K, Morita H, Hattori M, et al. Obesity-induced gut microbial metabolite promotes liver cancer through senescence secretome. Nature 2013; 499:97–101.

5.   Doden HL, Pollet RM, Mythen SM, Wawrzak Z, Devendran S, Cann I, Koropatkin NM, Ridlon JM. Structural and biochemical characterization of 20β-hydroxysteroid dehydrogenase from Bifidobacterium adolescentis strain L2-32. J Biol Chem 2019; 294:12040–53.

6.   Bernardi R, Doden H, Melo M, Devendran S, Pollet R, Mythen S, Bhowmik S, Lesley S, Cann I, Luthey-Schulten Z, et al. Bacteria on steroids: the enzymatic mechanism of an NADH-dependent dehydrogenase that regulates the conversion of cortisol to androgen in the gut microbiome. bioRxiv 2020;

7.   Ly LK, Doden HL, Ridlon JM. Gut feelings about bacterial steroid-17,20-desmolase. Mol Cell Endocrinol 2021; 525:111174.

8.   Doden HL, Ridlon JM. Microbial hydroxysteroid dehydrogenases: From alpha to omega. Microorganisms 2021; 9.

9.   Storbeck K, Bloem LM, Africander D, Schloms L, Swart P, Swart AC. 11β-Hydroxydihydrotestosterone and 11-ketodihydrotestosterone, novel C19 steroids with androgenic activity: A putative role in castration resistant prostate cancer? Mol Cell Endocrinol 2013; 377:135–46.

10. Turcu AF, Auchus RJ. Clinical significance of 11-oxygenated androgens. Curr Opin Endocrinol Diabetes Obes 2017; 24:252–9.

11. Fischbach MA. Microbiome: Focus on causation and mechanism. Cell 2018; 174:785–90.

12. Koppel N, Balskus EP. Exploring and understanding the biochemical diversity of the human microbiota. Cell Chem Biol 2016; 23:18–30.

13. Riglar DT, Silver PA. Engineering bacteria for diagnostic and therapeutic applications. Nat Rev Microbiol 2018; 16:214–25.