COMPUTATIONAL INVESTIGATION OF EARLY EVENTS IN PLANT GROWTH SIGNALING

BY

ALEXANDER SCOTT MOFFETT

DISSE­RTATION

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biophysics and Computational Biology in the Graduate College of the University of Illinois at Urbana-Champaign, 2019

Urbana, Illinois

Doctoral Committee:

Assistant Professor Diwakar Shukla, Chair
Professor Aleksei Aksimentiev
Assistant Professor Sarah Hind
Professor Huimin Zhao
Abstract

Brassinosteroids are a class of plant hormones critical for control of growth and development. While the signaling pathway through which brassinosteroid signals are processed has been the subject of intense study for the past 50 years, relatively little is known about the precise nanoscopic events underlying brassinosteroid signal transduction in terms of protein conformational dynamics. Focusing on the two primary brassinosteroid coreceptors, BRI1 and BAK1, we use a number of computational methods grounded in molecular dynamics simulations with the goal of understanding the physical mechanisms of brassinosteroid-induced BRI1-BAK1 association and activation.

First, we examine the brassinosteroid-induced association of the BRI1 and BAK1 extracellular domains. Using a replica exchange umbrella sampling scheme together with alchemical free energy calculations and conventional molecular dynamics simulations, we examine the mechanisms through which brassinosteroid binding to the BRI1 extracellular domain induces association with the BAK1 extracellular domain. We find that in addition to the stability provided by specific interactions between a brassinosteroid and both BRI1 and BAK1 in a manner depending on protonation of BAK1 residue H61, brassinosteroid binding stabilizes the BRI1 island domain while BRI1 can undergo a large conformational change which may further stabilize the BRI1-BAK1 complex.

The remaining work is focused on the BRI1 and BAK1 kinase domains. We performed adaptive sampling simulations on the isolated BRI1 and BAK1 kinase domains in their fully phosphorylated forms and built Markov state models from those simulations in order to examine the behavior of features characteristic of active protein kinases. We find that the BAK1 kinase domain displays local unfolding of the αC helix, while the BRI1 αC helix also unfolds, though to a lesser extent, while also swinging out away from the rest of the protein. Both behaviors are indicative of deactivation. Circular dichroism experiments and bioinformatic analysis indicate that the BRI1 and BAK1 αC helices are to some extent disordered, and that this may be a common feature in Arabidopsis thaliana protein kinases.

Next, we examine the mechanisms of phosphorylation-induced BAK1 activation, performing Gaussian accelerated molecular dynamics simulations on seven phosphorylation states likely to be important for BAK1
activation as well as one ATP-bound system. Again monitoring features known to be important for protein kinase activation, we find that phosphorylation of BAK1 T450 is critical for stabilization of the activation loop while T450 and T455 phosphorylation stabilize a more active αC helix conformation. Phosphorylation of T446 and T449 have much less of an effect, though both may have a role in interaction with BAK1 substrates.

Finally, we use adaptive sampling simulations and Markov state models to gain insight into the experimental finding that BAK1 is deactivated in vitro by S-glutathionylation. Simulating BAK1 with each of its three solvent-accessible cysteine residues individually S-glutathionylated, we find that modification of C408 induces further αC helix unfolding due to direct interactions, while causing global shifts in the BAK1 kinase domain dihedral angles. S-glutathionylation of C353 and C374 appears to have much less of an effect. This suggests that S-glutationylation of C408 may be responsible for the observed deactivation of BAK1.
To my parents.
Acknowledgments

First, I thank my advisor, Diwakar Shukla, and my lab mates, both past and present. I thank my dissertation committee members, Aleksei Aksimentiev, Sarah Hind, and Huimin Zhao. I also thank my experimental collaborators Kyle Bender and Steven Huber, the latter of whom also served on my preliminary exam committee. Both Kyle and Steve are authors on two of the published articles reproduced in this dissertation, and deserve all of the credit for the experimental work included. I thank the past and present Biophysics program coordinators, Marita Romine, Waad Ayoub, and Cindy Dodds. I thank my classmates in the Biophysics program. Finally but foremost, I thank my parents and my brother, as well as the rest of my family, both strictly and loosely defined.
# Table of Contents

List of Abbreviations ............................................................... vii

Chapter 1  Introduction .......................................................... 1
  1.1 Overview ........................................................................ 1
  1.2 A brief primer on brassinosteroid signaling ....................... 1
  1.3 A brief primer on molecular dynamics simulations .............. 5

Chapter 2  How do brassinosteroids activate their receptors? .... 11
  2.1 Introduction .................................................................... 11
  2.2 Results .......................................................................... 13
  2.3 Discussion ...................................................................... 19
  2.4 Methods ........................................................................ 22
  2.5 Supplementary Information ............................................ 27

Chapter 3  Exploring the dynamics of BRI1 and BAK1 kinase domains .................................................. 39
  3.1 Introduction .................................................................... 39
  3.2 Results .......................................................................... 41
  3.3 Discussion ...................................................................... 46
  3.4 Methods ........................................................................ 50
  3.5 Supplementary Information ............................................ 56

Chapter 4  How does multisite phosphorylation activate the BAK1 kinase domain? .................. 69
  4.1 Introduction .................................................................... 69
  4.2 Results .......................................................................... 70
  4.3 Discussion ...................................................................... 76
  4.4 Methods ........................................................................ 78

Chapter 5  The effects of S-glutationylation on the BAK1 kinase domain .............................................. 82
  5.1 Introduction .................................................................... 82
  5.2 Results .......................................................................... 85
  5.3 Discussion ...................................................................... 91
  5.4 Methods ........................................................................ 93
  5.5 Supplementary Information ............................................ 98

Chapter 6  A short outlook on the biophysics of brassinosteroid signaling ........................................... 102

References ................................................................................. 105
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAK1</td>
<td>BRI1-ASSOCIATED KINASE 1.</td>
</tr>
<tr>
<td>BL</td>
<td>Brassolinide.</td>
</tr>
<tr>
<td>BR</td>
<td>Brassinosteroid.</td>
</tr>
<tr>
<td>BRI1</td>
<td>BRASSINOSTEROID INSENSITIVE 1.</td>
</tr>
<tr>
<td>BRL</td>
<td>BRI1-like.</td>
</tr>
<tr>
<td>CV</td>
<td>Collective Variable.</td>
</tr>
<tr>
<td>ECD</td>
<td>Extracellular Domain.</td>
</tr>
<tr>
<td>EPK</td>
<td>Eukaryotic Protein Kinase.</td>
</tr>
<tr>
<td>FEL</td>
<td>Free Energy Landscape.</td>
</tr>
<tr>
<td>FLS2</td>
<td>FLAGELLIN-SENSITIVE 2.</td>
</tr>
<tr>
<td>KD</td>
<td>Kinase Domain.</td>
</tr>
<tr>
<td>LRR</td>
<td>Leucine-Rich Repeat.</td>
</tr>
<tr>
<td>LRR-RLK</td>
<td>Leucine-Rich Repeat Receptor-Like Kinase.</td>
</tr>
<tr>
<td>MAMP</td>
<td>Microbe-Associated Molecular Pattern.</td>
</tr>
<tr>
<td>MBAR</td>
<td>Multistate Bennett Acceptance Ratio.</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular Dynamics.</td>
</tr>
<tr>
<td>MSM</td>
<td>Markov State Model.</td>
</tr>
<tr>
<td>REUS</td>
<td>Replica-Exchange Umbrella Sampling.</td>
</tr>
<tr>
<td>SERK</td>
<td>Somatic Embryogenesis Receptor Kinase.</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Overview

Brassinosteroids are a class of plant growth hormone affecting a number of growth processes as well as immune responses. Brassinosteroids are perceived by BRASSINOSTEROID INSENSITIVE 1 (BRI1), an intrinsic membrane receptor, after which BRI1-ASSOCIATED KINASE 1 (BAK1), another intrinsic membrane receptor, associates with BRI1, activating the cytoplasmic kinase domains of both proteins. My work here focuses on understanding the events starting after a brassinosteroid has bound to BRI1 to phosphorylation of the BRI1 and BAK1 kinase domains. I have studied several steps of interest within this sequence of events, though even in this small “event frame” much more remains to be explored. In this chapter, I provide a brief introduction to brassinosteroid signaling, followed by a short primer on molecular dynamics simulations, partially adapted from a review paper I wrote [1].

1.2 A brief primer on brassinosteroid signaling

Brassinolide (BL) was first described by Grove et al. [2], who purified and crystalized it from pollen, yielding a structure from X-ray scattering. They found that application of purified BL promoted pinto bean internode elongation and cell division. A large number of brassinosteroids endogenous to a variety of plant species have since been discovered [3], and a great deal of research effort has gone into understanding how brassinosteroids are produced and how they induce growth.

Brassinosteroid signal transduction occurs primarily through a pair of leucine-rich repeat receptor-like kinases (LRR-RLKs) [4]. The LRR-RLKs are a large family of intrinsic membrane proteins characterized by their extracellular domains constructed from LRR domains [5, 6]. The extracellular LRR domains are structurally diverse across the LRR-RLKs, varying in length, curvature, and the presence of unique additional structural domains [7, 8]. LRR-RLKs are involved in a large number of signaling processes, including immune response [9] and response to abiotic stress [10] in addition to growth and development. From the N-
terminus to the C-terminus, the extracellular domain is followed by a transmembrane helix, an intracellular juxtamembrane region, and finally an intracellular kinase domain.

In the first step of brassinosteroid signal transduction, a brassinosteroid molecule binds to its cognate receptor BRASSINOSTEROID INSENSITIVE 1 (BRI1) [11, 12], named as such because the gene encoding BRI1 was originally identified in a growth deficient mutant which does not respond to brassinosteroid treatment. The extracellular region of BRI1 consists of 25 LRR domains in addition to a region called the island domain, which forms a pocket for brassinosteroid binding [13, 14]. A brassinosteroid binds with its side chain and C and D rings deeper into the pocket, with the A and B rings closer to the exterior of the pocket and the C-2 and C-3 hydroxyl groups extending out of the pocket [13, 14]. Brassinosteroid binding reduces conformational fluctuations in the BRI1 island domain [13, 14, 8] and promotes association of BRI1 with a second LRR-RLK, called BRI1-ASSOCIATED KINASE 1 [15, 16]. BAK1 plays a number of roles through its association with a number of larger LRR-RLKs in a variety of signaling pathways [17, 18, 19]. The BAK1 extracellular domain is far smaller than that of BRI1, consisting of 5 LRR domains [16]. The BRI1 LRRs on the C-terminal side of the island domain form a platform for BAK1 association [20, 16], with the C-2 and C-3 hydroxyl groups of a brassinosteroid bound to the BRI1 island domain completing the BRI1-BAK1 interface through interactions with BAK1 H61. The interactions between BRI1-bound brassinosteroids and BAK1 are the basis of the so-called molecular glue hypothesis, which posits that the interactions between BRI1-BAK1 mediated by BRI1-bound brassinosteroids are responsible for stabilizing the BRI1-BAK1 complex enough to cause activation.

The exact events that unfold in even these first steps of brassinosteroid signaling are unclear. Some evidence has suggested that BRI1 forms homodimeric or higher order complexes [21, 22, 23]. To complicate matters, some evidence suggests that BRI1 and BAK1 are pre-associated even in the absence of brassinosteroids [24] in seeming contention with the molecular glue hypothesis. What is clear is that brassinosteroid binding to the BRI1 island domain stabilizes the BRI1-BAK1 extracellular domain complex [25], increases the number of associated BRI1-BAK1 pairs in vivo [24], and induces phosphorylation of the BRI1 and BAK1 kinase domains [26, 27]. In its inactive form, the BRI1 kinase domain is inhibited by BRI1 Kinase Inhibitor 1 (BKI1) [21, 28], which is phosphorylated upon brassinosteroid-induced BRI1-BAK1 activation, subsequently unbinding from BRI1 [Figure 1.1] [29]. A number of phosphorylation sites on the BRI1 and BAK1 kinase domains have been identified [27, 30], with varying apparent degrees of functional importance. BAK1 has four phosphorylation sites on its activation loop, namely T446, T449, T450, and T455, as well as phosphorylation sites in the C-terminal tail and a number of other locations in the N-lobe [27, 31]. BRI1 is also phosphorylated on its activation loop, juxtamembrane region, and C-terminal tail [26]. While crystal
structures of the individual BRI1 and BAK1 kinase domains provide some insight into how phosphorylation of particular residues may promote phosphotransferase activity \cite{32, 33}, it is currently unknown exactly how phosphorylation activates BRI1 and BAK1.

![Figure 1.1: A simplified schematic of the brassinosteroid signaling pathway, with low (left) and high (right) brassinosteroid concentrations. This figure is inspired by work from Belkhadir and Jaillais \cite{34}.](image)

To further complicate matters, BRI1 and BAK1 are not the only receptors able to transduce brassinosteroid signals. The somatic embryogenesis receptor kinases (SERKs) are a group of LRR-RLKs of which BAK1 (also called SERK3) is a member \cite{35}. In addition to BAK1, SERK1 can associate with BRI1 \cite{36, 37}, although it is unclear if SERK1 plays any role in brassinosteroid signaling. BRI1 has several homologs, namely BRL1, BRL2, and BRL3, which are able to bind brassinosteroids to varying degrees \cite{38, 14}. More specifically, BRL1 and BRL3 are able to bind brassinolide with high affinity and are expressed in vascular...
cells, playing a role in vascular differentiation [35]. It appears that different brassinosteroid receptors play distinct roles, meaning that in actuality there is not one, but several brassinosteroid signaling pathways in a given plant, each with a different biological function [38].

The remaining steps in the brassinosteroid signaling pathway are also unclear in spatiotemporal terms, despite significant progress towards a more complete understanding of brassinosteroid signaling, as summarized in an excellent review by Belkhadir and Jaillais [34]. Activated BRI1 phosphorylates members of the BRI1 substrate kinases (BSKs) and CONSTITUTIVE DIFFERENTIAL GROWTH1 (CDG1) and CDG-like (CDL) families [Figure 1.1], which are receptor-like cytoplasmic kinases (RLCKs) [39, 40, 41, 42]. These proteins lack extracellular domains, though many are anchored to the intracellular side of the cell membrane. Phosphorylation of the BSK and CDG1/CDL targets activates their kinase domains, allowing them to phosphorylate BRI1 SUPPRESSOR 1 (BSU1) or BSU1-like (BSL) phosphatases [43, 40]. BSU1 and BSL phosphatases are activated by their own phosphorylation, allowing them to dephosphorylate the autophosphorylated Y200 on the cytoplasmic kinase BRASSINOSTEROID INSENSITIVE 2 (BIN2) [43]. BIN2 is active when Y200 is phosphorylated, allowing for phosphorylation of the transcription factors BRASSINAZOLE-RESISTANT 1 (BRZ1) and BRASSINAZOLE-RESISTANT 2 (BZR2), which subsequently reduces their ability to bind DNA and targets them for degradation via the 26S proteasome [44, 45, 46, 47, 48]. With BIN2 dephosphorylated, and therefore deactivated, BRZ1 and BZR2 are able to bind to DNA and affect gene expression. brassinosteroids can also promote expression of several protein phosphatase 2 (PP2A) variants, including PP2A B′η, which can dephosphorylate the BRI1 kinase domain [48], constituting a negative feedback loop, and PP2A B′α and PP2A B′β, which dephosphorylate and therefore activate BRZ1 and BZR2 [47, 48].

The brassinosteroid signaling pathway is highly integrated with other signaling pathways [49]. A number of studies have demonstrated crosstalk between brassinosteroid signaling and a number of other signaling pathways involving auxin [50, 51], gibberellin [52], abscisic acid [53], and immune response [54]. For example, although BAK1 is directly involved in initiating both brassinosteroid and several microbe-associated molecular pattern (MAMP) immune pathways [55, 56, 57], the pool of BAK1 that associates with BRI1 is spatially distinct from the pool interacting with the MAMP receptor FLAGELLIN-SENSITIVE 2 (FLS2) [57]. This indicates that interactions between MAMP and brassinosteroid signaling are unlikely to be due to competition between FLS2 and BRI1 for BAK1 as a binding partner, but rather are due to downstream crosstalk [57].
1.3 A brief primer on molecular dynamics simulations

The foundations of molecular dynamics simulation were initially developed by Enrico Fermi, John Pasta, Stanislaw Ulam, and Mary Tsingou in order to study the dynamics of relatively simple multiparticle systems with no closed-form mathematical solutions [58]. From then on, molecular dynamics simulations of gaseous argon [59], liquid water [60], and eventually proteins [61] were performed as the required software and hardware improved, and as protein crystal structures began to be solved [62]. Molecular dynamics simulation is often likened to a “computational microscope” [63, 64] in that it can give a detailed, atom-by-atom view of how proteins and nucleic acids move with time [65].

Modern molecular dynamics simulation of proteins requires several sources of input. First, an initial set of protein coordinates is needed from which to start simulations. These initial coordinates are typically taken from x-ray crystallography experiments or are inferred through homology modeling, which uses proteins with similar sequences and known structures to predict a possible structure. In principle, one could construct a computational model of an unfolded protein from sequence information alone and use molecular dynamics simulation to observe the folding process [66, 67], but this approach is usually impractical due to the enormous amount of computing time required to fold most proteins of biological interest.

The second input needed for molecular dynamics simulation is the set of mathematical functions defining forces between atoms in the protein. One must decide the forms of the functions used to calculate forces, thus approximating more accurate physical theory, as well as the parameters tuning those functions to match with experimental quantities and quantum calculations (Figure 1.2). In practice, for simulation of a protein consisting of the 20 standard amino acids, one must only choose from one of the many available biomolecular force fields [68, 69, 70, 71], which provide a functional form for force calculations and a set of parameters carefully tuned to reproduce certain experimental results.

---

1This section is adapted from: Moffett, AS and Shukla, D. Biochem. J. 475, 905-921 (2018).
Figure 1.2: Interatomic interactions in classical molecular dynamics (MD) force fields. a) A typical eukaryotic protein kinase (PK) structure (PDBID: 3TL8, chain A) prepared using VMD 1.9.2. The inset shows an amino acid with labeled atoms used to demonstrate force field terms. b)-f) Potential energy terms in MD force fields, with diagrammatic representations of the interactions and plots showing the relationship between the variables in the interaction diagrams and potential energy ($V$), shown on arbitrary, unitless scales. b) Covalent bonding is approximated by adding a spring between bonded atoms following Hooke’s law. The potential energy behaves quadratically with interatomic distance ($d_{ij}$). The parameter $k$ determines how stiff the “spring” is, where a high $k$ value yields low-amplitude vibrations and a low $k$ value yields high-amplitude vibrations. c) Angles between bonds on the same atom are also approximated with springs, where the potential energy instead depends on the angle between the two bonds ($\theta$). Diagonal angles are the angles ($\phi_{ijkl}$) between the planes on which the first and last three atoms of four sequentially bonded atoms fall. The potential energy is periodic with the value of a dihedral angle, where the $v$ parameter determines how high the potential energy barriers are, the $n$ parameter determines how many stable angles there are, and the $\phi_0$ parameter determines which angles are stable. d) Electrostatic interactions are approximated using Coulomb’s law, depending on interatomic distance ($d_{ij}$), while $q_i$ and $q_j$ are the electric charges of the two atoms and $\epsilon_0$ is the vacuum permittivity, which is replaced by $\epsilon_{Env}$ depending on the environment in which the interaction takes place. The electrostatic potential energy between positively and negatively charged atoms decreases proportionally to $1/d_{ij}$ as $d_{ij}$ decreases, but other force field terms account for repulsive interactions preventing two atoms from acquiring arbitrarily small distances from one another. e) Other non-bonded interactions are represented by the Lennard-Jones potential, depending on interatomic distance ($d_{ij}$). The $\sigma$ parameter determines the relative stability of the lowest potential energy distance, while $A$ and $B$ determine the balance between attractive [$B/d_{ij}$] and repulsive [$A/d_{ij}$] terms in the potential. The repulsive term provides atoms in molecular dynamics simulations with their space-filling features, as the potential energy between two atoms increases so rapidly as they come closer together past a certain point that the atoms effectively act like billiards balls.

Once the initial atomic coordinates have been acquired and the force field has been chosen, molecular dynamics simulation can begin. At their core, MD programs operate using the same physical principles that Isaac Newton developed over 300 years ago to describe the motion of far larger objects, hence the word “classical” in the title of this section. At the same time, methodological advances continue to improve the accuracy, efficiency, and capability of molecular dynamics simulation.

---

2For the sake of brevity, we avoid discussion of several crucial steps in between securing input information and actual molecular dynamics simulation as they are not important for the overall description of molecular dynamics simulation. At the same time, the omitted steps are oftentimes among the most challenging in the entire process and cannot be ignored in a more thorough discussion.
Figure 1.3: Two examples of simple systems with dynamics calculated using Newtonian mechanics. a) A single particle with mass \( m \) in one dimension has an initial position, \( x_1(t_1) \), and velocity, \( v_1(t_1) \). b) With no forces acting upon the particle, its motion can be calculated exactly according to \( x_1(t_2) = x_1(t_1) + (t_2 - t_1) v_1(t_1) \) and \( v_1(t_2) = v_1(t_1) \). c) For a one-dimensional system with two particles interacting through a spring potential, we must also consider the force that the two particles impart on one another. In this case, the initial positions of the two atoms, \( x_1(t_1) \) and \( x_2(t_1) \), are known along with the initial velocities \( v_1(t_1) \) and \( v_2(t_1) \). The two atoms feel an equal and opposite force, according to \( f(x_1, x_2) = -k|\Delta x| \) (where \( \Delta x = x_1 - x_2 \)), depending only on the positions. d) The new positions of the particles can be calculated as \( x_1(t_2) \approx x_1(t_1) + (t_2 - t_1) v_1(t_1) + (1/2m)(t_2 - t_1)^2 f_1(t_1) \) (where \( x_2(t_2) \) is calculated similarly) and new velocities as \( v_1(t_2) \approx v_1(t_1) + (1/m)(t_2 - t_1) f_1(t_1) \).

The motion of each atom is calculated in time steps, producing a sequence of protein conformations describing its motion with time. To give a simple illustrative example of the process, we consider a single atom moving along a straight line (Figure 1.3a). The position of the atom on the line at a certain time is denoted as \( x(t) \) with a velocity \( v(t) \) and acceleration \( a(t) \). Just as one might calculate the motion of an object in a secondary school physics problem, the position of the atom after a certain amount of time, which we call \( \Delta t \), can be calculated according to \( x(t + \Delta t) \approx x(t) + v(t) \Delta t + (1/2m)(\Delta t)^2 f(t) \). The initial position and velocity of the particle are provided at the start of the simulation. From Newton’s second law of motion, \( f = ma \), we know that the acceleration of an object is proportional to the force that it feels, meaning that the force fields described above will provide the \( a(t) \) term (where \( a(t) = f(t)/m \)). However, our system only has one particle with no forces acting on it and will perpetually continue motion in one direction (Figure 1.3b), which is not particularly useful for our purposes. If we add another atom so that the two atoms are attached by a spring and the force between them is \( f(x_1, x_2) = -k|x_1 - x_2| \) (Figure 1.3c), where \(|\cdot|\) is the absolute value of “\( \cdot \)”, we will have to calculate the motion of both atoms, taking into account the force that they impart on one another (Figure 1.3d). If we instead consider a system with many interacting atoms each moving in three dimensions, the dynamics can become very complicated and the number of force calculations

\[ \text{In practice, this simple algorithm for calculating trajectories is not used due to the high degree of error it incurs. This error arises due to the necessarily non-vanishing values of } \Delta t \text{ used in computer simulations, where } v(t) \text{ and } a(t) \text{ are approximated as constants over the period of this time step.} \]
between each pair of atoms will increase rapidly with the total number of atoms in the system.

**Analysis and extensions of molecular dynamics**

Once trajectories are completed, useful information has to be extracted in order to provide biological insight. One common method of analyzing molecular dynamics trajectories is to calculate a free energy landscape (FEL) over protein structural features. A FEL provides information about the relative stability of different protein conformations as well as the free energy barriers between stable states, which control how fast a protein can switch between different conformations. For example, if two conformations of a protein are compared, one can use the free energy difference between the two to determine their relative stabilities which can be translated into the relative amount of time a given protein will spend in either conformation.

A glaring limitation of the FEL is that proteins require a large number of variables to fully describe the motion of all atoms therein; one must choose several “important” variables to calculate a FEL which can be visualized. Deciding what makes a variable important and how to identify such variables is a challenge, and remains an active area of research.

Beyond the basic molecular dynamics simulations we have described above, often referred to as unbiased molecular dynamics for reasons which will become clear shortly, many other computational methods for studying protein conformational dynamics exist. One class of these methods is biased molecular dynamics simulation, where the potential energy function determining the motion of a protein is altered in some way from the realistic approximate energy function of unbiased molecular dynamics simulation, promoting conformational changes that could otherwise take prohibitively long times to occur. Because bias added to the potential energy function is a known quantity, one can account for the bias in the analysis stage and produce the same FEL that might have come from unbiased simulation, but with less computational time required. Despite the potential for increased efficiency in biased simulations, it can be a challenge to determine how exactly to add the biasing energy to a simulation in order to actually improve efficiency. Another method that can be used involves creating coarse-grained models of a protein structure, where multiple atoms are combined into representative units and the motion of these units is calculated. This strategy provides several advantages, decreasing the number of calculations a computer must do and thus allowing simulation of larger systems for longer periods of time, while sacrificing the detailed description of dynamics provided by all-atom molecular dynamics simulation.
What biology can we learn from molecular dynamics simulations?

Molecular dynamics simulations provide a nanoscopic interpretation to a number of experimentally observable quantities. For example, molecular dynamics simulations have been used to uncover the effects of phosphorylation and ATP binding on protein kinases, including induction of conformational change along activation pathways [78, 79, 80, 81, 82, 83, 84, 85], functional mechanisms of activating mutations [86, 87, 88], binding properties of inhibitors [89, 90, 91, 92, 93], and allosteric interactions between different structural domains [94, 95]. All of these properties are difficult to access experimentally but are critical for fully understanding the structure, dynamics, and function of proteins. At the same time, the successes of molecular dynamics simulations have largely come through integration with experimental results. Information can flow either way; molecular dynamics simulations are capable of providing a wide range of experimentally testable predictions, while experimental results can be used to improve the accuracy and efficiency of simulations. For example, observables from NMR spectroscopy experiments can be compared with the same quantities calculated from simulations or used to restrain atoms in simulations in such a way as to ensure agreement with experimental values [96, 97]. The idea behind introducing NMR restraints is to account for inaccuracies in the empirical force fields, with the intent of improving the predictive power of simulations for quantities not included in the restraints. As opposed to the quantitative physical properties measured in NMR, molecular dynamics simulations can also be used to make qualitative predictions, such as which residues are important for protein function or stability [98, 99, 100]. This type of prediction can be tested through site-directed mutagenesis coupled with functional assays. Inversely, deep mutational scanning [101]
can be used to experimentally identify functionally important residues in a protein, and molecular dynamics simulations can be used to understand why a particular residue is important [102].
Chapter 2

How do brassinosteroids activate their receptors?

2.1 Introduction

An atomistic picture of brassinosteroid (BR) signaling initiation has begun to emerge from a number of structural and computational studies [14, 13, 16, 10, 82, 103, 104, 105, 106, 107, 1, 25]. The predominant hypothesis concerning the mechanism of BR-induced BR1-BAK1 association is that a BR acts as a “molecular glue”, mediating interactions between the BRI1 and BAK1 extracellular domains (ECDs), as observed in crystal structures [36, 16]. Several researchers have suggested that BR binding stabilizes the BRI1 island domain, which could reduce the entropic cost of BRI1-BAK1 association [14, 13, 8]. However, the exact molecular events that unfold upon binding of a BR to BRI1 remain unclear, especially in light of evidence that BRI1 and BAK1 preform dimers in vivo in the absence of BRs [24, 23]. While recent work has provided quantitative thermodynamic data concerning BR binding to BRI1 and BRI1-BAK1 association [25], the nanoscale details of how hormone binding causes the BRI1-BAK1 complex to assemble and activate remain largely unknown.

In addition to the molecular glue and BRI1 island domain stabilization hypotheses, the possibility of BR-induced shifts in the BRI1 conformational equilibrium, thus stabilizing the BRI1-BAK1 complex, has not been conclusively ruled out. In fact, it remains possible that all three proposed mechanisms play important roles in BRI1-BAK1 activation. The BRI1 island domain B-factors are lower in BR-bound (holo) BRI1 crystal structures than in BR-unbound (apo) structures [14, 13], which suggests that BR binding to BRI1 reduces fluctuations in the BRI1 island domain, providing a stable platform for BAK1 interaction. In a crystal structure of the BRI1-BAK1 complexed with the common brassinosteroid brassinolide (BL) [16], the two BL hydroxyl groups of carbons 2 and 3 on the A steroid ring interact with the BAK1 backbone and H61 side chain, while the phenyl group of BAK1 F60 lies in a plane parallel to the BL steroid rings. In vitro experiments on the BRI1 and BAK1 ECDs found that their BL-dependent association was also dependent on pH [16], suggesting that the protonation state of H61 may play an important role in controlling BRI1-BAK1 association. Although no previous evidence of large BRI1 conformational changes exists, the structure and
dynamics of the BRI1 ECD has only been investigated in the context of low temperature crystal packing, and it remains unclear whether global conformational changes in the BRI1 ECD occur or if BL binding influences BRI1 conformational dynamics.

**Figure 2.1:** Analysis of BRI1 and BAK1 ECD crystal structures. **A** Structural alignment of the BRI1 ECD from apo (PDBID: 3RGX [14], shown in purple), BL-bound (PDBID: 3RGZ [14], shown in tan), and BL-BAK1-bound (PDBID: 4M7E, chain A [7], shown in grey) crystal structures. **B** A closeup of the island domains of structures shown in A). Note that part of the island domain of the apo structure (PDBID: 3RGX) is not resolved in the crystal structure, indicating disorder. **C** Crystal structure of the BL-bound BRI1-BAK1 ECD complex (PDBID: 4M7E, chains A and C). BRI1 is shown in grey, while BAK1 is shown in orange. **D** A closeup of the BRI1-BAK1-BL interface. The critical BAK1 residues F60 and H61 are shown interacting with BL and BRI1 N705. The BAK1 R146-BRI1 E749 interactions is also shown. This figure was produced using VMD 1.9.2 [108] and Inkscape 0.91 [109].

In this study, we used molecular dynamics (MD) simulations to examine at the atomic scale several proposed mechanisms as to how BL binding causes the association of BRI1 and BAK1. Starting from a crystal structure of the BRI1-BAK1-BL complex[16], we calculated the absolute apo and holo BRI1-BAK1 ECD association free energies for a truncated BRI1 ECD (tBRI1) containing LRRs 13-25 and the island domain using the replica-exchange umbrella sampling (REUS)[110 111] approach developed by Woo and Roux[112] and Gumbart _et al._ [113 114]. We found that using protonation states consistent with a pH 7 solution, BL binding only moderately stabilizes the tBRI1-BAK1 complex. By removing the N-terminal region of the BRI1 ECD, we quantified the impact of BL-mediated interactions on BRI1-BAK1 complex stability under the assumption that BRI1 LRRs 1-12 are static with respect to the rest of the BRI1 ECD and do not play any role in BRI1-BAK1 association. We used alchemical free energy calculations to explore the impact of pH on BL-induced tBRI1-BAK1 association, finding that double protonation of BAK1 H61 plays a central role in stabilizing the holo complex over the apo complex. We performed Gaussian accelerated molecular dynamics (GAMD) simulations of apo and holo tBRI1 in order to investigate the effect of BL
binding on island domain dynamics. The tBRI1 island domain was stabilized by the presence of BL, while the apo island domain exhibited a greater degree of flexibility. Finally, we ran unbiased simulations of the complete apo and holo BRI1 and BRI1-BAK1 ECDs in order to examine the role of BRI1 LRRs 1-12 in the association process and in stabilizing the complex. In the holo complex, BRI1 underwent a large conformational change, forming a secondary interface with BAK1 through the BRI1 N-terminal LRRs.

### 2.2 Results

**BL binding stabilizes the tBRI1-BAK1 complex in a pH-dependent manner**

Using a REUS sampling scheme, we calculated the standard free energy of tBRI1-BAK1 ECD association in the presence and absence of BL. We estimated the overall $\Delta G^\circ$ of association values to be $-2.120 \pm 0.380$ kcal-mol$^{-1}$ for the apo complex and $-3.630 \pm 0.453$ kcal-mol$^{-1}$ for the holo complex (Table 2.1). The presence of BL in the island domain of tBRI1 stabilized the tBRI1-BAK1 complex by a moderate $-1.510 \pm 0.591$ kcal-mol$^{-1}$, and shifted the minimum in the separation potential of mean force (PMF) from $\sim$26 Å to $\sim$27 Å (Figure 2.2). For both the apo and holo complexes, the separation PMFs were surprisingly shallow (Figure 2.2). Individually, the free energies of adding and removing conformational restraints (Table 2.1) contributed far less than in previous studies [114], almost certainly due to the fact that we restrained RMSD values to non-zero values, giving tBRI1 and BAK1 some flexibility at the cost of longer convergence times.

**Table 2.1:** Calculated contributions to the free energy of BRI1-BAK1 association with standard deviations.

<table>
<thead>
<tr>
<th>Contribution</th>
<th>Apo (kcal-mol$^{-1}$)</th>
<th>Holo (kcal-mol$^{-1}$)</th>
<th>$\Delta$ (kcal-mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta G^S_{BR}$</td>
<td>0.105 $\pm$ 0.011</td>
<td>0.530 $\pm$ 0.030</td>
<td>0.426 $\pm$ 0.032</td>
</tr>
<tr>
<td>$\Delta G^S_{BA}$</td>
<td>0.369 $\pm$ 0.028</td>
<td>0.118 $\pm$ 0.008</td>
<td>$-0.250 \pm 0.029$</td>
</tr>
<tr>
<td>$\Delta G^S_{B}$</td>
<td>0.304 $\pm$ 0.016</td>
<td>0.139 $\pm$ 0.008</td>
<td>$-0.165 \pm 0.018$</td>
</tr>
<tr>
<td>$\Delta G^S_{O}$</td>
<td>0.176 $\pm$ 0.012</td>
<td>0.315 $\pm$ 0.028</td>
<td>0.140 $\pm 0.031$</td>
</tr>
<tr>
<td>$\Delta G^S_{O}$</td>
<td>0.489 $\pm$ 0.044</td>
<td>0.749 $\pm$ 0.066</td>
<td>0.260 $\pm 0.079$</td>
</tr>
<tr>
<td>$\Delta G^S_{O}$</td>
<td>0.120 $\pm$ 0.008</td>
<td>0.290 $\pm$ 0.021</td>
<td>0.170 $\pm 0.022$</td>
</tr>
<tr>
<td>$\Delta G^S_{O}$</td>
<td>0.270 $\pm$ 0.018</td>
<td>0.091 $\pm$ 0.005</td>
<td>$-0.179 \pm 0.018$</td>
</tr>
<tr>
<td>$\Delta G^S_{BL}$</td>
<td>0.00</td>
<td>0.139 $\pm$ 0.008</td>
<td>0.139 $\pm 0.008$</td>
</tr>
<tr>
<td>$\Delta G^P_{BL}$</td>
<td>0.00</td>
<td>0.223 $\pm$ 0.017</td>
<td>0.223 $\pm 0.017$</td>
</tr>
<tr>
<td>$\Delta G^P_{O}$</td>
<td>6.63</td>
<td>6.62</td>
<td>$-0.007$</td>
</tr>
<tr>
<td>$\Delta G^P_{O}$</td>
<td>0.053 $\pm$ 0.004</td>
<td>0.142 $\pm$ 0.011</td>
<td>0.088 $\pm 0.012$</td>
</tr>
<tr>
<td>$\Delta G^P_{O}$</td>
<td>0.043 $\pm$ 0.004</td>
<td>0.039 $\pm$ 0.003</td>
<td>$-0.004 \pm 0.005$</td>
</tr>
<tr>
<td>$-\beta^{-1} \ln(O^T C^O)$</td>
<td>$-7.015 \pm 0.013$</td>
<td>$-8.285 \pm 0.012$</td>
<td>$-1.270 \pm 0.019$</td>
</tr>
<tr>
<td>$\Delta G^*$</td>
<td>$-2.120 \pm 0.380$</td>
<td>$-3.630 \pm 0.453$</td>
<td>$-1.510 \pm 0.591$</td>
</tr>
</tbody>
</table>
Results from analytical size exclusion chromatography have provided evidence that the ECDs of BRI1 and SERK1, a close homolog of BAK1, do not interact in the absence of BL [30]. This suggests a positive $\Delta G$ of association at relevant concentrations, in contrast to our $\Delta G^\circ$ of $-2.120 \pm 0.380$ kcal-mol$^{-1}$, though we cannot say how relevant this result is for SERK3 with any confidence. Previously reported experimental values of $\Delta G^\circ$ for BAK1 association with BL-bound BRI1 are $-8.44$ kcal-mol$^{-1}$ from grating-coupled interferometry and $-9.20$ kcal-mol$^{-1}$ from isothermal titration calorimetry [25] (see Supplementary Information for details), both far larger in magnitude than our result of $-3.630 \pm 0.453$ kcal-mol$^{-1}$.

![Figure 2.2](image-url)

**Figure 2.2:** The potential of mean force as a function of BRI1-BAK1 separation distance ($r$) from REUS simulations. Roman numerals indicate associated (I), intermediate (II), and dissociated (III) states. We estimated the PMF and error bars, representing standard deviations, using MBAR. This figure was produced using VMD 1.9.2 [108], Matplotlib 2.2.2 [115], and Inkscape 0.91 [109].

Both experimental association free energies were determined in solutions with pH 5 [25], whereas our protein models were constructed in the most likely protonation states at pH 7. BRI1 and BAK1 are known to interact more strongly in acidic conditions [16], possibly due to double protonation of the BAK1 H61 side chain. In order to explore the origins of this pH dependence we performed alchemical free energy calculations in order to estimate $\Delta \Delta G$ of association for apo and holo BRI1 and BAK1 with protonation of BAK1 H61 and the interfacial BRI1 E749.

For protonation of BAK1 H61 (Supplementary Figure 2.1 A-B), we estimated $\Delta \Delta G_{Apo} = 2.729 \pm 0.097$ kcal-mol$^{-1}$ and $\Delta \Delta G_{Holo} = -1.958 \pm 0.096$ kcal-mol$^{-1}$, the changes in tBRI1-BAK1 association free energy. Under the assumption that the internal and relative conformational dynamics of tBRI1 and BAK1 will not be significantly affected by BAK1 H61 protonation, we can add these values to our REUS $\Delta G^\circ$ estimates, yielding $\Delta G_{Apo}^\circ = 0.609 \pm 0.393$ kcal-mol$^{-1}$ and $\Delta G_{Holo}^\circ = -5.588 \pm 0.463$ kcal-mol$^{-1}$ at pH 5. These estimates are far closer to the experimental values, reflecting the important role of BAK1 H61 protonation...
in distinguishing apo from holo BRI1.

**BL binding stabilizes the BRI1 island domain**

It has been previously proposed that BR binding to the BRI1 ECD could alter the dynamics of the island domain, based on the decreased B-factors in the island domain of holo BRI1 ECD crystal structures as compared to apo BRI1 [14, 13, 8]. We estimated B-factors from conventional MD simulations of the apo and holo BRI1 ECDs according to $B_i = \frac{8\pi^2}{3} R_i^2$ [110], where $R_i$ is the root mean-squared fluctuation of residue $i$, and compared the estimated B-factors to crystallographic B-factors [PDBIDs: 3RGX (apo) and 3RGZ (holo)] (Figure 2.3). In the island domain region, the crystallographic B-factors are uniformly slightly higher for apo BRI1 than holo BRI1 (Figure 2.3 A), as has been noted previously [14, 13, 8]. In comparison, B-factors calculated from our simulations of the BRI1 ECD did not display noticeable differences in the island domain between apo and holo systems (Figure 2.3 B). The B-factors for both apo and holo BRI1 are much higher from our simulations than for the experimental systems, likely due to the fact that our simulations were performed at 300 K and in solution, in contrast to the conditions necessary for x-ray crystallography. Additionally, the particularly large B-factors in the terminal regions of BRI1 ECD from simulations suggest that there is likely large-scale flexibility in BRI1, which we further explore later.

![Figure 2.3: A Crystallographic B-factors for apo (PDBID: 3RGX [14]) and holo (PDBID: 3RGZ [14]) BRI1. The island domain is shown within the vertical dashed lines. Note the higher B-factors for the apo structure within the island domain. B B-factors calculated from unbiased molecular dynamics simulations of the apo and holo BRI1 ECD. The N-terminal and C-terminal regions display high B-factors as compared with their crystallographic counterparts, while there appears to be little difference between the apo and holo B-factors. This figure was produced using Matplotlib 2.2.2 [115].](image)

As it is difficult to make conclusions from the B-factor calculations, we performed GAMD on apo and holo tBAK1 and calculated reweighted PMFs over the island domain RMSD with respect to a crystal structure.
Both the apo and holo tBRI1 island domain PMFs had a minimum near 1.2 Å, with a second local minimum at 1.7 Å (apo) and 1.9 Å (holo). Most notably, the apo PMF is far flatter than the holo PMF. At an RMSD of around 1.6 Å, the holo PMF is over 2 kcal·mol$^{-1}$ larger than the apo PMF, and the difference in the PMFs largely increases at even larger RMSDs. These results indicate that binding of BL to the BRI1 island domain decreases fluctuations, stabilizing the island domain.

Figure 2.4: Potential of mean force for the RMSD of the apo and holo tBRI1 island domains with respect to a BRI1 ECD crystal structure (PDBID: 4M7E chain A [7]) from GAMD simulations. The holo island domain is stabilized with a single dominantly stable state at an RMSD of around 1.2 Å, while the apo island domain can attain a wider range of RMSDs. This figure was produced using Matplotlib 2.2.2 [115].

BRI1 LRRs 1-12 are highly flexible and interact with BAK1

In order to examine the overall flexibility of the BRI1 ECD, we first performed anisotropic normal mode (ANM) analysis [117, 118] on the apo ECD of BRI1 using ProDy [119]. The structure of the apo BRI1 ECD was taken from the PDB 4M7E chain A, with BL removed. The first two modes, shown in Figure 2.5, provide initial clues as to what global BRI1 dynamics look like. Mode I (Figure 2.5 A-D) corresponds to relative motion between LRRs 1-12 and the remainder of the ECD which takes BRI1 towards and away from a planar closed ring. Mode II (Figure 2.5 E-H) corresponds to a curling motion, analogous to the coiling of a spring. The angle $\Xi$ defined in the Methods section roughly corresponds to normal mode I while $\Omega$ roughly corresponds to normal mode II.

Next, we performed unbiased MD simulations of the apo and holo BRI1 ECD and BRI1-BAK1 ECD complex, for a total of four systems. In addition to the angles $\Xi$ and $\Omega$ described in the Methods section, we examined the distances between BRI1 R131 and BRI1 D651 for all four systems as well as the distance between BRI1 N108 and BAK1 K44 for the complexes. For the BRI1 ECD alone, LRRs 1-12 were flexible both in the presence and absence of BL. The BRI1 R131-D651 distance and both $\Xi$ and $\Omega$ fluctuated around
Figure 2.5: Anisotropic normal mode analysis of the BRI1 ECD taken from a crystal structure of the BRI1-BAK1-BL complex (PDBID: 4M7E chain A). A-D depict the first normal mode while E-H depict the second normal mode. A and C show the displacement vectors of α-carbons along normal mode I, where C is rotated 90° with respect to A. B and D depict conformations along the first normal mode, where again D is rotated 90° with respect to B. E-H can be described analogously to A-D. This figure was produced using VMD 1.9.2 [108], the VMD plugin Normal Mode Wizard [119], and Inkscape 0.91 [109].

the same mean value in both apo and holo BRI1 in a manner qualitatively consistent with harmonic dynamics (Figure 2.6), although the BRI1 R131-D651 distance distribution is apparently tighter for holo BRI1. While these results provide evidence that BRI1 LRRs 1-12 are in fact flexible, there appears to be little difference in apo and holo BRI1 ECD dynamics in the absence of the BAK1 ECD.

In simulations of the BRI1-BAK1 ECD complex, we found that without BL bound, the dynamics of BRI1 differed very little from BRI1 dynamics without BAK1 present (Figure 2.7). However, in the presence of BL, BRI1 underwent a conformational change after around 26 ns of simulation, forming a closed BRI1 conformation featuring a second BRI1-BAK1 interface. The closed conformation is characterized by formation of a contact within BRI1 between R131 and D651, which remained stable for the remaining 34 ns of the trajectory after the conformational change occurred (Figure 2.7). A contact between BRI1 N108 to BAK1 K44 also formed, although this interaction was able to break and reform several times (Figure 2.7 C) even after the conformational change, leading to a multi-peaked distribution (Figure 2.7 D). This conformational change corresponds with a change in BRI1 Ξ (Figure 2.7 C-D), demonstrated by the stability of Ξ at around 40° starting around the 26 ns mark and by the doubly peaked angle distribution. This suggests that this conformational change occurs roughly along normal mode I, while motion along normal mode II does not contribute as much, although Ω appears to be stabilized in the holo complex especially after the conformational change (Figure 2.7 C-D).

As mentioned before, in the secondary interface BRI1 D651 interacts with BRI1 R131 and BAK1 K44 interacts with BRI1 N108 as well as several backbone oxygens (Figure 2.8). Additionally, BRI1 N37 interacts...
Figure 2.6: Conformational dynamics of the apo and holo BRI1 ECD. For all the following time series, two independent simulations are shown. A Time series of R131-D651 distance in apo BRI1. B Time series of R131-D651 distance in holo BRI1. C Probability densities of the R131-D651 distance in apo and holo BRI1 from the time series in A and B estimated with histograms. D Time series of \( \Xi \) and \( \Omega \) in apo BRI1. E Time series of \( \Xi \) and \( \Omega \) in holo BRI1. F Probability densities of \( \Xi \) and \( \Omega \) in apo and holo BRI1 from the time series in E and F estimated with histograms. This figure was produced using Matplotlib 2.2.2 [115].

Figure 2.7: Conformational dynamics of the apo and holo BRI1-BAK1 ECD complex. A Time series of the BRI1 R131-D651 and BRI1 N108-BAK1 K44 distances in apo and holo systems. Note the conformational change near 26 ns in the holo system, and the subsequent fluctuations between two apparent states in the BRI1 R131-D651 and BRI1 N108-BAK1 K44 distance. B Probability densities of the BRI1 R131-D651 and BRI1 N108-BAK1 K44 distances in apo and holo systems estimated using histograms. C Time series of \( \Xi \) and \( \Omega \) in the apo and holo BRI1-BAK1 complexes. D Probability densities of \( \Xi \) and \( \Omega \) in the apo and holo BRI1-BAK1 complexes estimated using histograms. This figure was produced using Matplotlib 2.2.2 [115].
with the backbone oxygen of BRI1 S184 (Figure 2.8). We predicted the mean and mode of the BRI1 D651 pKₐ distribution from BRI1 apo simulations to be 3.712 and 3.944 respectively, using PROPKA 3.1 [120, 121], meaning that this interaction is likely as strong at pH 5 as it is at pH 7.

![Figure 2.8](image-url) The secondary interface between the BRI1 and BAK1 ECDs. The crystal structure (PDBID: 4M7E chains A and C [7]) is shown in a transparent representation. The inset subfigures A and B represent two alternative interfaces seen in simulations, with A being the state in Figure 2.7 C with BRI1 N108 and BAK1 K44 at the smallest distance and B being the state with a larger BRI1 N108-BAK1 K44 distance, where BAK1 K44 interacts with BRI1 backbone oxygens.

It is important to note that we cannot make claims with any confidence regarding the thermodynamics of BRI1 conformational changes from such short simulations. We have found that the closed BRI1 conformation can occur in simulations of the holo BRI1-BAK1 complex within a timespan of 60 ns, but it is difficult to make any statement about the stability of this conformation from our data. Similarly, BRI1 LRRs 1-12 are flexible, but the presented histograms in Figure 2.6 likely only represent local BRI1 dynamics and may not represent the global dynamics of the BRI1 ECD. Further investigation of BRI1 ECD dynamics is required, but is beyond the scope of this study.

2.3 Discussion

Our examination of three mechanisms of BL-induced BRI1-BAK1 association indicates that each one likely plays an important role. Our REUS-based $\Delta G^\circ$ estimates indicate that at pH 7 the BRI1-BAK1 ECD complex is weakly favored both with and without BL bound. Alchemical free energy calculations suggest that protonation of BAK1 H61, consistent with a solution pH of 5, is instrumental in differentiating apo and holo BRI1, increasing the stability gap between the holo and apo BRI1-BAK1 complex from $-1.510 \pm 0.591$ kcal·mol⁻¹ to $-6.197 \pm 0.607$ kcal·mol⁻¹. With the contributions of BAK1 H61 protonation, we estimate
tBRI1-BAK1 ECD association free energies of $\Delta G_{\text{apo}} = 0.609 \pm 0.393 \text{ kcal-mol}^{-1}$ and $\Delta G_{\text{holo}} = -5.588 \pm 0.463 \text{ kcal-mol}^{-1}$, closer to the experimental values of -8.44 kcal-mol$^{-1}$ or -9.20 kcal-mol$^{-1}$ for the holo complex, but still less stable by several kcal-mol$^{-1}$. Our GAMD simulations provide direct evidence that BL binding to BRI1 stabilizes the island domain, likely reducing the entropic cost of BRI1-BAK1 association. Finally, unbiased MD simulations revealed a conformational change in the BRI1 ECD when complexed with BAK1 and BL, forming a secondary interface between BRI1 and BAK1, which may further stabilize the BRI1-BAK1 complex.

Are our results consistent with experimental findings? The combined contributions of direct BL-BAK1 interactions, BL-induced BRI1 island domain stabilization, and formation of the second BRI1-BAK1 interface complicate direct comparison of our REUS-based association free energies with experimental quantities. Beyond the inaccuracies of force fields in describing protein-protein interaction [122], and the possibility of insufficient sampling, which does not appear to be an issue [Supplementary Figure 2.2], we found that direct BL-BAK1 interactions contribute significantly to the overall BRI1-BAK1 BL-dependent association free energy, but cannot entirely account for BRI1-BAK1 association. From alchemical free energy calculations, protonation of BAK1 H61 appears to play a critical role in distinguishing apo and holo BRI1, destabilizing the apo BRI1-BAK1 complex by 2.729 $\pm$ 0.096 kcal-mol$^{-1}$ while stabilizing the holo complex by 1.958 $\pm$ 0.096 kcal-mol$^{-1}$. The important role of pH in the stability of the BRI1-BAK1 complex is consistent with experimental findings [16].

Although it appears direct interactions between BRI1 and BAK1 mediated by BL do not account for the entire association free energy, it is difficult to claim with certainty that the difference between our estimates and the experimental association free energies implies other sources of BRI1-BAK1-BL complex stability. The fact that we observe stabilization of the BRI1 island domain upon BL binding and formation of a second BRI1-BAK1 interface with BL bound adds to the weight of evidence suggesting that the molecular glue hypothesis alone cannot explain BL-induced BRI1-BAK1 association. Although we did not observe a conformational change in apo BRI1 complexed with BAK1, we cannot be sure that BRI1 is less likely to undergo conformational change without BL bound than it is with BL bound. However, as BAK1 is far less likely to associate with apo BRI1 than holo BRI1, even an equally probable BRI1 conformational change in the apo and holo complexes would preferentially stabilize the favorable holo complex [Figure 2.9].

Experimental evidence already exists for the molecular glue and island domain stabilization hypotheses. In this paper, we have made the additional claim that BRI1 LRRs 1-12 are highly dynamic and play an important role in stabilizing the BRI1-BAK1 ECD complex. Supposing that this claim is true, BRI1 mutations D651A and R131A should increase the effective experimental BRI1-BAK1 dissociation rate, $k_d$,
and consequently decrease $K_A$. This should occur because BRI1 R131 and D651 interact and stabilize the closed BRI1 conformation and mutating either will destabilize the closed BRI1 conformation. Furthermore, BAK1 mutations N37A and K44A or BRI1 mutations N108A or D161A will increase BRI1-BAK1 $k_d$, while the double mutations BAK1 N37A-K44A or BRI1 N108A-D161A will further increase $k_d$. This should occur because any of these mutations will disrupt interactions in the secondary BRI1-BAK1 interface, which is only formed when BRI1 is in the closed conformation.

LRR-RLKs are an important group of receptors in plants with a common evolutionary ancestry and the shared structural features, namely the predominance of LRR domains in the ECD. Our finding that BRI1 undergoes a conformational change in order to form a second interface with BAK1 raises several questions about the LRR-RLKs in general. How flexible are the ECD domains of large LRR-RLKs? To what extent does this flexibility play a functional role in association with a coreceptor and receptor activation? How common is the formation of secondary interfaces between larger BRI1-like and smaller BAK1-like LRR-RLK ECDs? We believe that a combination of computational modeling and biophysical experiments is uniquely suited to address these questions, which could have significant impact on our understanding of plant growth, developmental, and immune signaling.

In summary, we find that direct BL-BAK1 interactions (i.e. the molecular glue hypothesis), BRI1 island domain stabilization, and a secondary BRI1-BAK1 interface all likely impact BRI1-BAK1 association. BL-induced association of the BRI1 and BAK1 ECDs may occur through the following sequence of events (Figure 2.9):
BL binds to the BRI1 island domain, stabilizing it in a conformation amenable to BAK1 association. With the BRI1 LRRs 1-12 far from the rest of the ECD, as in the crystal structures, BAK1 associates with BRI1 and interacts with BL through F60, H61, V62, and D74. BRI1 LRRs 1-12 can then undergo a conformational change, forming the secondary BRI1-BAK1 interface and further stabilizing the BRI1-BAK1 complex.

2.4 Methods

System setup

All systems were set up using the tleap program within AmberTools 15 [124] using the Amber ff14SB force field [70]. Initially, the protein atoms of the BRI1-BAK1 extracellular domains in complex with BL were taken from a crystal structure (PDBID: 4M7E, chains A and C [7]). Histidine protonation states were determined in accordance with a solution at pH 7 using the H++ 3.2 web server [125]. Disulfide bonds were added according to the crystal structure. All systems were solvated in boxes of TIP3P [126] water molecules, so that the edges of the box were at least 10 Å from any protein atom. Parameters for BL were assigned using the Antechamber program within AmberTools15 and the GAFF force field [127]. As Antechamber assigns a small, non-integer overall negative charge to BL, the absolute value of the overall charge was distributed as evenly as was practical across all BL atoms so as to obtain an overall neutral charge. Sodium and chloride ions were added to each system so as to neutralize the charge and bring the salt concentration to ~150 millimolar NaCl. Simulations using the full BRI1 ECD consisted of residues 34-766, while the tBRI1 ECD consisted of residues 378-766, and the BAK1 ECD consisted of residues 26-200.

General simulation details

All simulations, except alchemical free energy calculations (see below), were run in NAMD 2.12-2.13 [128] with a time step of 2 fs and with hydrogen-containing bonds constrained using the SHAKE algorithm [129]. All production runs were maintained at constant temperature of 300 K using a Langevin thermostat with a coupling time constant of 2 ps and at a constant pressure of 1 bar using a Berendsen barostat. The particle mesh Ewald method [130] was used to treat electrostatics and a cutoff of 10 Å was used for non-bonded interactions. Trajectory analysis was performed using MDTraj [131] and NumPy [132] within Jupyter Notebooks [133].
Determining BRI1-BAK1 extracellular domain $K_A$s

We calculated the equilibrium association constant ($K_A$) of BAK1 to BRI1 with and without BL bound to BRI1 using the REUS\[110, 111\] approach developed by Woo and Roux\[112\] and Gumbart et al.\[113, 114\]. For the apo complex, five restraints on the relative orientation of the BRI1 and BAK1 ectodomains were introduced in addition to the restraint on the PMF CV (Figure 2.10). An additional restraint was added in the holo complex to ensure that BL remained bound to BRI1 (Figure 2.10).

Replica-exchange umbrella sampling

Both the apo and holo BRI1-BAK1 complexes were set up as described above, subjected to 20,000 steps of energy minimization and equilibrated for 2 ns. The final 1 ns of equilibration for both systems was used to compute average values of each restraint CV in the complex. These average values were used as the reference values of each CV in the corresponding harmonic restraints (Supplementary Table 2.1). Next, the center of the C1-C’1 distance restraint was increased from the equilibrium value of both associated complexes to 45 Å over 1 ns simulations in order to generate starting structures for each umbrella sampling window. REUS along the C1-C’1 distance was performed with a total of 61 windows evenly spaced between 25 Å and 45 Å (0.33 Å intervals between window centers), where the starting structure for each window was equilibrated for 1 ns and each replica was simulated for 7 ns. An additional 7 ns of REUS was performed for each window of both the apo and holo systems to ensure convergence of the separation PMF (Supplementary Figure 2.2). A total of 854 ns of REUS was performed in order to estimate the separation PMF of each system.

We used REUS to calculate the free energies of applying and removing restraints on the relative orientations of BRI1 and BAK1 ($\Theta$, $\Phi$, $\theta$, $\phi$, and $\Psi$) in addition to RMSD and BL position restraints in both the apo and holo states. In order to generate wide ranges of each restrained CV for initiating each umbrella sampling window, we performed temperature-accelerated molecular dynamics simulations\[134, 135\], coupling each respective CV to a dummy particle experiencing a temperature of 2500 K. The number of windows for the REUS simulations of each CV varied, while each window was equilibrated for 1 ns, followed by 7 ns of REUS.

Rather than focusing on the PMFs estimated from each set of REUS simulations, we examined the integrands of the target ensemble averages for convergence (Supplementary Figure 2.3, Supplementary Figure 2.22). We ensured that each integrand qualitatively converged with time and that the integrand fell to zero at both edges of the sampled range for each restraint CV, indicating that the ensemble averages should be estimated well from our simulations. For the RMSD restraint on apo BAK1 in the bound state with respect to BRI1, we noticed that the range of RMSD sampled was insufficient to estimate the ensemble average, and
subsequently performed steered MD followed by standard umbrella sampling for an additional 6 windows (Supplementary Figure 2.4).

**Figure 2.10:** Collective variables harmonically restrained during umbrella sampling. C1 to C3 and C’1 to C’3 are the centers of mass of the α-carbons of different sets of residues in BRII and BAK1, respectively (the residues used are listed in the lower lefthand corner). The distance r between C1 and C’1 was used as the CV for umbrella sampling. The angular restraints are defined as: Θ (C’1-C1-C2) and θ (C1-C’1-C’2). The torsion restraints are defined as: Φ (C’1-C1-C2-C3), φ (C1-C’1-C’2-C’3), and Ψ (C2-C1-C’1-C’2). Additionally, for the BRII-BL-BAK1 simulations, the distance between the center of mass of all heavy atoms in BL and the α-carbon of BRII F658 was restrained to prevent BL unbinding. This figure was produced using VMD 1.9.2 [108] and Inkscape 0.91 [109].

**PMF estimation**

All unbiased PMFs were calculated using the multistate Bennett acceptance ratio (MBAR) estimator [136] as implemented in pymbar 3.0.4 [137]. All trajectories were subsampled to a frequency of 50 ns−1 to reduce correlations between samples while maintaining a sufficient number of samples (Supplementary Figure 2.23-Supplementary Figure 2.24). While this choice should improve the PMF estimates over samples chosen according to the correlation time [138] due to retention of more samples, it likely also resulted in underestimation of the standard deviations for bin free energies due to sample correlation [136].

**Overall KA calculations**

As per [113 114], the overall standard dissociation constant can be calculated as a product of terms, each calculated from simulations or analytically in the case of bulk alignment contributions. For simulations in the bound state (referred to as “site” in [113 114]), the distance between C1 and C’1 (Figure 2.10), r, was subjected to a harmonic restraint with a force constant of 10 kcal·mol−1 Å−2 if r increased beyond 29 Å. For simulations in the bulk state, a distance of \( r^* = 44 \) Å was chosen according to a distance where both separation PMFs had plateaued, and r was harmonically restrained at \( r^* \) with a force constant of 10
kcal·mol$^{-1}$Å$^{-2}$.

The overall calculation of $\Delta G^\circ$ of association involves calculating the free energy of adding each restraint one after another in the bulk state, calculating the free energy of translating the restrained BRI1 and BAK1 to the bound state, and then calculating the free energy of releasing each restraint in the bound state. The free energies of adding and removing restraints can conveniently be calculated using averages of exponentials of restraint potential energy (see Supplementary Information for full details).

The separation PMF integrals were approximated as

$$I^* = \int_{\text{bound}} e^{-\beta\left[W(r) - W(r^*)\right]} dr \approx \sum_{i=1}^{N_b} e^{-\beta\left[W_i - W^*\right]} \Delta r_i. \quad (2.1)$$

where $W(r)$ is the PMF at separation distance $r$, $N_b$ is the number of bins used to approximate $W(r)$ in the bound state, and $W_i$ is the free energy of occupying bin $i$. The bound region was defined as the set of 13 bins centered on the bin with the lowest overall PMF (roughly $r_0 \pm 1.32$ Å for the apo system and $r_0 \pm 1.35$ Å for the holo system). Assuming mutual independence of each $W_i$, we approximated the variance of $I^*$ ($\sigma^2_{I^*}$) using a Taylor series expansion truncated at the first order term as

$$\sigma^2_{I^*} \approx \sum_{i=1}^{N_b} \left(\beta e^{-\beta\left[W_i - W^*\right]} \Delta r_i\right)^2 \sigma^2_{W_i}. \quad (2.2)$$

where $\sigma^2_{W_i}$ is the MBAR variance of $W_i$. The exponential averages needed to calculate restraint free energy contributions were estimated as MBAR expectation values.

**BAK1 H61 protonation**

We performed alchemical free energy perturbation simulations on snapshots from separation PMF REUS windows 8 and 57 (corresponding to separation distances of 27.67 Å and 44 Å, respectively) in order to assess the contributions of residue protonation to association free energy. We focused on the protonation of BAK1 H61, which interacts with BRI1-bound BL and has been proposed to be an important factor in pH-dependent BRI1-BAK1 interaction [16]. Because BRI1 E749 is in the BAK1-BRI1 ECD interface and initially appeared to have a pK$_a$ above 5 based on predictions from H++ [125] on a crystal structure (PDBID: 4M7E), we also estimated the free energy of BRI1 E749 protonation through alchemical free energy calculations. However, PROPKA 3.1 [120, 121] calculations based on the whole apo BRI1 ECD simulations suggest that BRI E749 would probably be unprotonated at pH 5 [Supplementary Figure 2.25], meaning that BRI1 E749 protonation analysis is not particularly pertinent.
All alchemical free energy simulations were performed using Amber 18 [139, 140], using a Monte Carlo barostat. Due to the change in charge that occurs when protonating a residue, we included a co-alchemical Na\(^+\) ion in each simulation to maintain a net zero system charge [141], where we linearly scaled the charge to zero concurrently with linearly scaling the amino acid proton charge to one. No restraints on the co-alchemical ions were used. We used 41 windows in each alchemical transformation, where each window was run for 11 ns with the first 1 ns discarded as equilibration. In order to prevent association of BRI1 and BAK1, we applied restraints on the \(\alpha\)-carbons of BRI1 residues 651-666 and of BAK1 residues 86-90 in all simulations started from snapshots with \(r \approx 44\) Å. The free energy of protonation for each case was estimated using MBAR, with frames sampled at a rate of 100 ns\(^{-1}\).

**Accelerated MD simulations of apo and holo tBRI1**

We performed dual boost Gaussian accelerated MD (GAMD) [142] simulations of the apo and holo BRI1 ECD using NAMD 2.13 [143]. Restraints were added to the terminal regions of tBRI1 to prevent unrealistic unfolding and on the overall orientation of tBRI1 to allow for the use of a smaller periodic box. Each system was subjected to energy minimization for 20,000 steps, followed by 6 ns of equilibration. In the initial phase of simulations, potential energy statistics were recorded during an additional 11 ns of simulation, after which GAMD biases were applied. Each system was equilibrated for 1 ns with GAMD biases on according to the gathered statistics, after which 11 ns of further equilibration was performed where the bias parameters \(E\) and \(k_0\) were updated. In the final phase, ten independent GAMD trajectories were run for both apo and holo BRI1 each for a total of \(~90\) ns, saving frames every 200 fs. Each trajectory was run in six 15 ns segments, where the energy statistics and bias parameters were updated in the first and second nanoseconds of simulation, respectively, each with 400 ps of preparation simulations. While possibly not ideal for cumulant expansion reweighting, the entire 90 ns of each trajectory was used for final analysis. Unbiased PMFs were estimated using second-order cumulant expansions [144]:

\[
F_j = -\beta^{-1}\log(p_j^*)
\]

\[
\approx -\beta^{-1}\left[\log(p_j) + \log\left(\frac{\sum_{j=1}^{N_b} e^{\beta C_{1,j}} + \frac{\beta^2}{2} C_{2,j}}{\sum_{j=1}^{N_b} e^{\beta C_{1,j}} + \frac{\beta^2}{2} C_{2,j}}\right)\right]
\]

where \(F_j\) is the reweighted free energy of occupying bin \(j\), \(p_j\) is the unreweighted histogram count of bin \(j\), and \(C_{i,j}\) is the \(i\)th cumulant of the GAMD boost energy for frames in bin \(j\). Bins with counts fewer than 100 were removed from the final PMFs.
BRI1 conformational analysis

Two angles, \( \Xi \) and \( \Omega \), were defined in order to explore the conformational dynamics of the entire BRI1 ECD. We defined four centers of mass, from the backbone atoms of residues 34-104 (\( c_1 \in \mathbb{R}^3 \)), 253-293 (\( c_2 \)), 383-423 (\( c_3 \)), and 667-767 (\( c_4 \)). We then defined \( c_{21} \) as \( c_2 - c_1 \), \( c_{32} \) as \( c_3 - c_2 \), and \( c_{34} \) as \( c_3 - c_4 \). The angle \( \Xi \) was defined as

\[
\Xi = \frac{180^\circ}{\pi} \arccos \left( \frac{(c_{21} \cdot c_{32})}{||c_{21}|| ||c_{32}||} \right) \cdot \left( \frac{(c_{34} \cdot c_{32})}{||c_{34}|| ||c_{32}||} \right)
\]  
(2.5)

and the angle \( \Omega \) as

\[
\Omega = \frac{180^\circ}{\pi} \arccos \left( \frac{(c_{21} \cdot c_{32})}{||c_{21}|| ||c_{32}||} \right) \cdot \left( \frac{(c_{34} \cdot c_{32})}{||c_{34}|| ||c_{32}||} \right)
\]  
(2.6)

where caution was taken to differentiate distinct angles yielding identical values of \( \Xi \) and \( \Omega \), allowing for negative values of each angle.

We ran unbiased simulations of the apo and holo complete BRI1 ECD, as well as the apo and holo BRI1-BAK1 ECD complex. Two 100 ns simulations were run for each of the apo and holo BRI1 systems, while a single 60 ns simulation was run of both the apo and holo BRI1-BAK1 complex.

2.5 Supplementary Information

BRI1 E749 protonation

For the sake of completeness, we report results for BRI1 E749, although BRI1 E749 is unlikely to be protonated at pH 5 (Supplementary Figure 2.25). Counterintuitively, protonation of tBRI1 E749 appeared to strongly stabilize both the apo and holo tBRI1-BAK1 complex (Supplementary Figure 2.1 C-D). As BRI1 E749 interacts with BAK1 R146 in crystal structures (Figure 2.1), we expected protonation of BRI1 E749 to break this interaction and weaken the stability of the rBRI1-BAK1 complex. However, for protonation of tBRI1 E749 we estimated \( \Delta \Delta G_{\text{Apo}} = -12.225 \pm 0.101 \text{ kcal-mol}^{-1} \) and \( \Delta \Delta G_{\text{Holo}} = -7.113 \pm 0.102 \text{ kcal-mol}^{-1} \). It appears that the interaction partners of BRI1 E749 in BAK1 are involved in unrealistic interactions with the truncated C-terminus of tBRI1.
Calculation of association free energies from previously reported dissociation constants

Here, we show our calculations for binding free energy given the dissociation constants obtained by Hohmann and coworkers [25]. We assume a temperature of 300 K and use a standard concentration of $C^\circ = 1/(1661 \, \text{Å}^3)$. The standard free energy of association between the BAK1 and BL-bound BRI1 extracellular domains was calculated from grating-coupled interferometry results as follows:

$$K_D = 0.71 \, \mu \text{M} = \frac{0.71 \text{μmol}}{L} \cdot \frac{1 \text{ mol}}{10^6 \text{ μmol}} \cdot \frac{1 \text{ L}}{10^{27} \text{ Å}^3} \cdot 6.022 \cdot 10^{23} \text{ mol}^{-1} \approx 4.28 \cdot 10^{-10} \text{ Å}^{-3}$$

$$K_A = \frac{1}{K_D} \approx 2.34 \cdot 10^9 \, \text{Å}^3$$

$$\Delta G^\circ = -RT \log(K_A/1661 \, \text{Å}^3) \approx -8.44 \text{ kcal} \cdot \text{mol}^{-1}$$

The free energy of association from isothermal titration calorimetry (-9.20 kcal·mol$^{-1}$) was calculated accordingly.

Calculation of association free energies from REUS

For the sake of brevity, we use the following shorthand notation for averages: $\langle A(x) \rangle_{E}^{\{\alpha, \beta, \gamma \ldots \}}$, where $A(x)$ is the quantity, dependent on protein coordinates, to be averaged, the superscript $E$ denotes the state of the association progress, either at the binding site (S) or in bulk solution (B), and $\{\alpha, \beta, \gamma \ldots \}$ represents the set of collective variables restrained when the average is taken. For example,

$$\langle e^{-\beta u(x)} \rangle_{BR, BA}^S = \frac{\int_S e^{-\beta [U(x) + pV(x) + u_{BR}(x) + u_{BA}(x) + u_\Theta(x)]} dx}{\int_S e^{-\beta [U(x) + pV(x) + u_{BR}(x) + u_{BA}(x)]} dx}$$

(2.7)

is the average value of $e^{-\beta u(x)}$ in the NPT ensemble, $u_\Theta(x)$ being the restraint potential on $\Theta$, with BRI1 and BAK1 bound and with the BRI1 and BAK1 backbone atoms restrained. The overall calculations of standard association free energies proceed as follows:
\[
\begin{align*}
\Delta G_{BR}^S &= \beta^{-1} \ln \left( e^{-\beta u_{BR}(x)} \right)^S_B \\
\Delta G_{BA}^S &= \beta^{-1} \ln \left( e^{-\beta u_{BA}(x)} \right)^S_B \\
\Delta G_{BR}^\phi &= \beta^{-1} \ln \left( e^{-\beta u_{\phi}(x)} \right)^S_B \\
\Delta G_{BA}^\phi &= \beta^{-1} \ln \left( e^{-\beta u_{\phi}(x)} \right)^S_B \\
\Delta G_{BL}^S &= \beta^{-1} \ln \left( e^{-\beta u_{BL}(x)} \right)^S_B \\
\Delta G_{BL}^B &= \beta^{-1} \ln \left( e^{-\beta u_{BL}(x)} \right)^B_B \\
\Delta G_{BA}^B &= \beta^{-1} \ln \left( e^{-\beta u_{BA}(x)} \right)^B_B \\
\Delta G_{BR}^B &= \beta^{-1} \ln \left( e^{-\beta u_{BR}(x)} \right)^B_B \\
\Delta G_{\Theta}^S &= \Delta G_{\Theta}^S + \Delta G_{\Phi}^S + \Delta G_{\Phi}^S + \Delta G_{\phi}^S + \Delta G_{\phi}^S \\
\Delta G_{\Theta}^B &= \Delta G_{\Theta}^B + \Delta G_{\Phi}^B + \Delta G_{\Phi}^B = -\beta^{-1} \ln \left[ \frac{1}{8\pi^2} \int_0^\pi \int_0^{2\pi} \sin(\Theta) e^{-\beta u_0(\Theta,\Phi)} d\Psi d\Phi d\Theta \right] \\
I^* &= \int_{\text{bound}} e^{-\beta [W(r) - W(r^*)]} dr \\
O^* &= (r^*)^2 \int_0^\pi \int_0^{2\pi} \sin(\Theta) e^{-\beta u_0(\Theta,\Phi)} d\Psi d\Phi \\
K_A^{Ap0} &= O^* I^* e^{-\beta \left[ (\Delta G_{BR}^B - \Delta G_{BR}^S) + (\Delta G_{BA}^B - \Delta G_{BA}^S) + (\Delta G_{BL}^B - \Delta G_{BL}^S) \right]} \\
K_A^{BL} &= O^* I^* e^{-\beta \left[ (\Delta G_{BR}^B - \Delta G_{BR}^S) + (\Delta G_{BA}^B - \Delta G_{BA}^S) + (\Delta G_{BL}^B - \Delta G_{BL}^S) \right]} \\
\Delta G^0 &= -\beta^{-1} \ln [K_A C^0], \quad C^0 = \frac{1}{1661 \text{Å}^3}
\end{align*}
\]

Note that the \( O^* \) and \( \Delta G_{0}^B \) terms can be calculated analytically, as shown below, while each other term requires MD simulation.
Apo orientational restraint contribution

\[ O_{Apo}^O = (r^*)^2 \int_0^\pi \int_0^{2\pi} \sin(\Theta)e^{-\beta u_a(\Theta, \Phi)} d\Phi d\Theta \]

\[ = (44 \text{ Å})^2 \int_0^{2\pi} e^{-\beta(\Phi)}(\pi/2)^2(\Phi-16.8693(\pi/180))^2 d\Phi \int_0^\pi \sin(\Theta)e^{-\beta(\Phi, 1)(\pi/2)}(\Theta-120.4496(\pi/180))^2 d\Theta \]

\[ = (1936 \text{ Å}^2)(0.106819)(0.0920026) = 19.0262 \text{ Å}^2 \]

Apo bulk angular restraint contributions

\[ \Delta G_{B, Apo} = -\frac{1}{\beta} \ln \left[ \frac{1}{8\pi^2} \int_0^{2\pi} \int_0^{2\pi} \sin(\Theta) e^{-\beta u_0(\Theta, \Phi, \Psi)} d\Psi d\Phi d\Theta \right] \]

\[ = -\frac{1}{\beta} \ln \left[ \frac{1}{8\pi^2} \int_0^{2\pi} e^{-\beta(\Phi)}(\pi/2)^2(\Phi-33.0273(\pi/180))^2 d\Phi \int_0^\pi \sin(\Theta)e^{-\beta(\Phi, 1)(\pi/2)}(\theta-73.4478(\pi/180))^2 d\theta \right] \]

\[ = 6.63 \text{ kcal} \cdot \text{mol}^{-1} \]

Holo orientational restraint contribution

\[ O_{Holo}^O = (r^*)^2 \int_0^\pi \int_0^{2\pi} \sin(\Theta)e^{-\beta u_a(\Theta, \Phi)} d\Phi d\Theta \]

\[ = (44 \text{ Å})^2 \int_0^{2\pi} e^{-\beta(\Phi, 1)(\pi/2)}(\Phi-15.9178(\pi/180))^2 d\Phi \int_0^\pi \sin(\Theta)e^{-\beta(\Phi, 1)(\pi/2)}(\Theta-118.5483(\pi/180))^2 d\Theta \]

\[ = (1936 \text{ Å}^2)(0.106819)(0.0937143) = 19.3802 \text{ Å}^2 \]

Holo bulk angular restraint contributions

\[ \Delta G_{B, Holo} = -\frac{1}{\beta} \ln \left[ \frac{1}{8\pi^2} \int_0^{2\pi} \int_0^{2\pi} \sin(\Theta) e^{-\beta u_0(\Theta, \Phi, \Psi)} d\Psi d\Phi d\Theta \right] \]

\[ = -\frac{1}{\beta} \ln \left[ \frac{1}{8\pi^2} \int_0^{2\pi} e^{-\beta(\Phi)}(\pi/2)^2(\Phi-33.7587(\pi/180))^2 d\Phi \int_0^\pi \sin(\Theta)e^{-\beta(\Phi, 1)(\pi/2)}(\phi-267.2762(\pi/180))^2 d\phi \right] \]

\[ = 6.62 \text{ kcal} \cdot \text{mol}^{-1} \]
Supplementary Table 2.1: Force constants and reference collective variable values used for restraints in REUS PMF calculations.

<table>
<thead>
<tr>
<th>Collective variable</th>
<th>( k_{\text{force}} )</th>
<th>Apo reference</th>
<th>Holo reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>( tBRI1 ) RMSD</td>
<td>10.0(^*)</td>
<td>1.1825 Å</td>
<td>1.1492 Å</td>
</tr>
<tr>
<td>( BAK1 ) RMSD</td>
<td>10.0(^*)</td>
<td>1.1973 Å</td>
<td>1.3730 Å</td>
</tr>
<tr>
<td>( \Theta )</td>
<td>0.10(^\dagger)</td>
<td>120.4496°</td>
<td>118.5843°</td>
</tr>
<tr>
<td>( \Phi )</td>
<td>0.10(^\dagger)</td>
<td>16.8963°</td>
<td>15.9178°</td>
</tr>
<tr>
<td>( \Psi )</td>
<td>0.10(^\dagger)</td>
<td>33.0273°</td>
<td>33.7587°</td>
</tr>
<tr>
<td>( \phi )</td>
<td>0.10(^\dagger)</td>
<td>261.461°</td>
<td>267.2762°</td>
</tr>
<tr>
<td>( \theta )</td>
<td>0.10(^\dagger)</td>
<td>73.4478°</td>
<td>75.9999°</td>
</tr>
<tr>
<td>BL-( tBRI1 ) distance</td>
<td>10.0(^*)</td>
<td>N/A</td>
<td>18.7735 Å</td>
</tr>
</tbody>
</table>

\(^*\) kcal·mol\(^{-1}\)·Å\(^{-2}\), \(^\dagger\) kcal·mol\(^{-1}\)·degree\(^{-2}\)

Supplementary Figure 2.1: Results from alchemical free energy calculations describing protonation of \( tBAK1 \) H61 and BRI1 E749, each performed with \( tBRI1 \) and BAK1 close (window 8, \( r = 27.67 \) Å) and distant (window 57, \( r = 44.00 \) Å). On the x-axis is the alchemical parameter \( \lambda \) while on the y-axis is the MBAR-derived relative free energy of each state. A Protonation of apo \( tBAK1 \) H61. B Protonation of holo \( BAK1 \) H61. C Protonation of apo \( tBRI1 \) E749. D Protonation of holo \( tBRI1 \) E749. This figure was produced using Matplotlib 2.2.2 \cite{matplotlib}.
Supplementary Figure 2.2: Convergence of the A apo and B holo tBRI1-BAK1 separation PMFs with REUS sampling time per window. This figure was produced using Matplotlib 2.2.2 [115].

Supplementary Figure 2.3: tBRI1 conformational restraint contributions for bound apo tBRI1-BAK1. A The restraint PMF. B Convergence of the restraint PMF with simulation time. C Convergence of the integrand within the associated ensemble average estimation with time. This figure was produced using Matplotlib 2.2.2 [115].

Supplementary Figure 2.4: BAK1 conformational restraint contributions for bound apo tBRI1-BAK1. A The restraint PMF. The PMF including additional umbrella sampling is shown in orange. B Convergence of the restraint PMF with simulation time. The PMF including additional umbrella sampling is shown in grey. C Convergence of the integrand within the associated ensemble average estimation with time. The integrand including additional umbrella sampling is shown in orange. This figure was produced using Matplotlib 2.2.2 [115].

Supplementary Figure 2.5: Θ restraint contributions for bound apo tBRI1-BAK1. A The restraint PMF. B Convergence of the restraint PMF with simulation time. C Convergence of the integrand within the associated ensemble average estimation with time. This figure was produced using Matplotlib 2.2.2 [115].
Supplementary Figure 2.6: Φ restraint contributions for bound apo tBRI1-BAK1. A The restraint PMF. B Convergence of the restraint PMF with simulation time. C Convergence of the integrand within the associated ensemble average estimation with time. This figure was produced using Matplotlib 2.2.2 [115].

Supplementary Figure 2.7: Ψ restraint contributions for bound apo tBRI1-BAK1. A The restraint PMF. B Convergence of the restraint PMF with simulation time. C Convergence of the integrand within the associated ensemble average estimation with time. This figure was produced using Matplotlib 2.2.2 [115].

Supplementary Figure 2.8: φ restraint contributions for bound apo tBRI1-BAK1. A The restraint PMF. B Convergence of the restraint PMF with simulation time. C Convergence of the integrand within the associated ensemble average estimation with time. This figure was produced using Matplotlib 2.2.2 [115].

Supplementary Figure 2.9: θ restraint contributions for bound apo tBRI1-BAK1. A The restraint PMF. B Convergence of the restraint PMF with simulation time. C Convergence of the integrand within the associated ensemble average estimation with time. This figure was produced using Matplotlib 2.2.2 [115].
Supplementary Figure 2.10: BAK1 conformational restraint contributions for unbound apo tBRI1-BAK1. A The restraint PMF. B Convergence of the restraint PMF with simulation time. C Convergence of the integrand within the associated ensemble average estimation with time. This figure was produced using Matplotlib 2.2.2 [115].

Supplementary Figure 2.11: tBRI1 conformational restraint contributions for unbound apo tBRI1-BAK1. A The restraint PMF. B Convergence of the restraint PMF with simulation time. C Convergence of the integrand within the associated ensemble average estimation with time. This figure was produced using Matplotlib 2.2.2 [115].

Supplementary Figure 2.12: tBRI1 conformational restraint contributions for bound holo tBRI1-BAK1. A The restraint PMF. B Convergence of the restraint PMF with simulation time. C Convergence of the integrand within the associated ensemble average estimation with time. This figure was produced using Matplotlib 2.2.2 [115].

Supplementary Figure 2.13: BAK1 conformational restraint contributions for bound holo tBRI1-BAK1. A The restraint PMF. B Convergence of the restraint PMF with simulation time. C Convergence of the integrand within the associated ensemble average estimation with time. This figure was produced using Matplotlib 2.2.2 [115].

34
**Supplementary Figure 2.14:** Θ restraint contributions for bound holo tBRI1-BAK1. A The restraint PMF. B Convergence of the restraint PMF with simulation time. C Convergence of the integrand within the associated ensemble average estimation with time. This figure was produced using Matplotlib 2.2.2 [115].

**Supplementary Figure 2.15:** Φ restraint contributions for bound holo tBRI1-BAK1. A The restraint PMF. B Convergence of the restraint PMF with simulation time. C Convergence of the integrand within the associated ensemble average estimation with time. This figure was produced using Matplotlib 2.2.2 [115].

**Supplementary Figure 2.16:** Ψ restraint contributions for bound holo tBRI1-BAK1. A The restraint PMF. B Convergence of the restraint PMF with simulation time. C Convergence of the integrand within the associated ensemble average estimation with time. This figure was produced using Matplotlib 2.2.2 [115].

**Supplementary Figure 2.17:** φ restraint contributions for bound holo tBRI1-BAK1. A The restraint PMF. B Convergence of the restraint PMF with simulation time. C Convergence of the integrand within the associated ensemble average estimation with time. This figure was produced using Matplotlib 2.2.2 [115].
Supplementary Figure 2.18: θ restraint contributions for bound holo tBRI1-BAK1. A The restraint PMF. B Convergence of the restraint PMF with simulation time. C Convergence of the integrand within the associated ensemble average estimation with time. This figure was produced using Matplotlib 2.2.2 [115].

Supplementary Figure 2.19: BL restraint contributions for bound holo tBRI1-BAK1. A The restraint PMF. B Convergence of the restraint PMF with simulation time. C Convergence of the integrand within the associated ensemble average estimation with time. This figure was produced using Matplotlib 2.2.2 [115].

Supplementary Figure 2.20: BL restraint contributions for unbound holo tBRI1-BAK1. A The restraint PMF. B Convergence of the restraint PMF with simulation time. C Convergence of the integrand within the associated ensemble average estimation with time. This figure was produced using Matplotlib 2.2.2 [115].

Supplementary Figure 2.21: BAK1 conformational restraint contributions for unbound holo tBRI1-BAK1. A The restraint PMF. B Convergence of the restraint PMF with simulation time. C Convergence of the integrand within the associated ensemble average estimation with time. This figure was produced using Matplotlib 2.2.2 [115].
**Supplementary Figure 2.22:** tBRI1 conformational restraint contributions for unbound holo tBRI1-BAK1. A The restraint PMF. B Convergence of the restraint PMF with simulation time. C Convergence of the integrand within the associated ensemble average estimation with time. This figure was produced using Matplotlib 2.2.2 [115].

**Supplementary Figure 2.23:** Rationale for our choice of a uniform 50 ns⁻¹ subsampling rate for REUS simulations. A The apo and B holo separation PMFs using a uniform 50 ns⁻¹ subsampling rate and a subsampling rate determined individually for each window using the correlation time method implemented in pymbar [138, 137]. The number of samples remaining for each window for the C apo and D holo systems. Using the correlation time to subsample yields uncorrelated samples but at the same time causes highly uneven sampling across \( r \). Using a uniform subsampling time of 50 ns⁻¹ by definition yields even sampling over \( r \), but likely results in the use of correlated samples in some windows, leading to underestimation of error [136]. This figure was produced using Matplotlib 2.2.2 [115].

**Supplementary Figure 2.24:** The effects of subsampling rate on the A apo and B holo separation PMFs. We chose three subsampling rates to include with error bars shown for the C apo and D holo separation PMFs. Note the general insensitivity to both separation PMFs to the choice of uniform subsampling rate over all windows. This figure was produced using Matplotlib 2.2.2 [115].
Supplementary Figure 2.25: Violin plot of BRI1 side chain pKₐs for residues with side chain pKₐs close to 5, calculated from the simulations of the full, apo BRI1 ECD using PROPKA 3.1 \cite{120, 121}. Frames were taken from apo BRI1 simulations at a rate of 10 ns⁻¹. Bars represent the interval from the lowest to the highest calculated pKₐ over all frames. This figure was produced using Matplotlib 2.2.2 \cite{115}.
Chapter 3

Exploring the dynamics of BRI1 and BAK1 kinase domains

3.1 Introduction

Insight into kinase activation downstream of BR perception can be gleaned from the expansive structural information available from animal systems, given the highly conserved nature of kinase structures across the eukaryotes[145]. Eukaryotic protein kinases have been thoroughly reviewed[145, 146] and several features have been suggested as necessary for kinase activity (Figure 3.1), including proximal positioning and proper folding of the $\alpha$C helix, formation of a salt bridge between the conserved glutamate (BRI1: E927, BAK1: E334) on the $\alpha$C helix and the conserved lysine (BRI1: K911, BAK1: K317) on the $\beta$4 sheet (K-E salt bridge), formation of the regulatory and catalytic hydrophobic spines, outward positioning of the conserved phenylalanine (BRI1: F1028, BAK1: F435) in the DFG motif away from the active site, and unfolding of the activation loop. Careful positioning of these regulatory components is necessary for a kinase to perform catalysis, meaning that any conformation with a regulatory structure out of place is inactive. Despite the resultant enormous multiplicity of possible inactive states, several distinct conformations recur throughout the kinases. For example, the human cytoplasmic tyrosine kinase Src takes on an inactive conformation where the unphosphorylated activation loop folds and obscures the active site and the $\alpha$C helix swings away from the active site in concert with breakage of the K-E salt bridge[147, 79] (references to a “Src-like” state throughout this paper will exclusively denote distal swinging of the $\alpha$C helix, as folding of a phosphorylated activation loop is not expected).

In order to understand how the BRI1 and BAK1 kinase domains switch between catalytically competent and incompetent conformations, we carried out molecular dynamics (MD) simulations of the fully phosphorylated, isolated BRI1 and BAK1 kinase domains. We find that both kinase domains have high degrees of conformational heterogeneity in the two-dimensional space of K-E distance and $\alpha$C helix helical content, meaning that even in the purportedly active phosphorylation state the isolated kinase domains spend a significant amount of time in catalytically incompetent configurations with broken K-E salt bridges and

---

1This chapter is adapted from: Moffett, AS, Bender KW, Huber, SC, and Shukla, D. J. Biol. Chem. 292, 12643-12652 (2017).
unfolded αC helices. Such notable sampling of likely inactive conformations could be indicative of a role for BRI1-BAK1 dimerization in promoting catalytically competent conformations in both receptors, or perhaps another factor not considered in our simulations. In order to validate our computational predictions, we synthesized the αC helix of BRI1 and performed solution circular dichroism (CD) spectroscopy to determine its helical propensity and obtained experimental results consistent with our model. To further computationally test whether the BAK1 αC helix is intrinsically disordered or whether the local protein environment is responsible for its instability, we created an in silico chimeric kinase domain, substituting the αC helix sequence of human Src kinase, which has been demonstrated to be highly stable in previous studies, for the native BAK1 αC helix. Analysis of the entire A. thaliana kinome using disorder prediction software suggests that αC helix disorder may be a widely utilized regulatory mechanism in plant kinases.
3.2 Results

Conformational heterogeneity in the BRI1 phosphorylated kinase domain.

Despite the fact that the simulated BRI1 kinase domain was fully phosphorylated, so that one might expect it to be locked into a catalytically competent conformation, we observed a high degree of conformational diversity in structural features known to be important for kinase function. The BRI1 kinase domain exhibits a high degree of αC helix disorder and is permissive towards breaking of the conserved K-E salt bridge required for catalytic activity (Figure 3.2a). Still, the active-like state is the most stable region (Figure 3.2a, region I) although there does not appear to be a free energy barrier to unfolding of the αC helix with the K-E salt bridge formed (Figure 3.2a, region II) until there is already a significant degree of disorder. It is unclear how this unfolding behavior concurrent with a formed K-E salt bridge affects catalytic activity. The K-E salt bridge can be broken while the αC helix remains ordered and in an active-like conformation (Figure 3.2a, unlabeled region above region I) likely due to interaction between K911 and ATP, while further distance between K911 and E927 coincides with greater fluctuation in the same distance. The region characterized by a large distance between K911 and E927 (Figure 3.2a, regions III and IV) covers a similar range of αC helix disorder to the region with a formed K-E salt bridge (Figure 3.2a, regions I and II).

![Figure 3.2: BRI1 conformational heterogeneity. a) Markov state model-weighted free energy surface of the BRI1 kinase domain as a function of αC helix helical content and the distance between the glutamate δ-carbon atom and the lysine side-chain nitrogen in the conserved K-E salt bridge. The free energy is calculated by adding histograms of each state weighed by state equilibrium probability, and taking the natural logarithm of this sum, as follows: $F(x, y) = -\beta^{-1} \ln \left[ \sum_{i=1}^{N} \pi_i h_i(x, y) \right]$ b) (I-IV) BRI1 conformations representing the corresponding regions on the free energy surface in part a). I) Active-like structure, II) αC helix disordered structure with an intact K-E salt bridge, III) Src-like inactive structure characterized by a folded αC helix swung distally and a broken K-E salt bridge, IV) structure with a disordered αC helix and a broken K-E salt bridge.](image-url)
Presence of a Src-like inactive state in BRI1.

The distance between K911 and E927 is measured between the terminal nitrogen atom in lysine and the δ-carbon in glutamate, meaning that changes in this distance while the αC helix remains ordered could be caused by conformational changes in the side chains, swinging of the αC helix outward, or some combination of the two. Region III in Figure 3.2a, represents breaking of the K-E salt bridge coupled with distal movement of the αC helix (Supplementary Figure 3.2). This suggests the presence of a Src-like inactive state in BRI1, where the αC helix remains folded but swings outward from its active-like position, breaking the K-E salt bridge while preventing rapid reformation by conformational changes in the side chains alone. The Src-like inactive state represents the second most stable region of the free energy landscape, possibly suggesting that transitions to this state are the main mechanism of BRI1 deactivation when only considering its internal conformational dynamics.

Conformational heterogeneity in the BAK1 phosphorylated kinase domain.

Similarly to BRI1, BAK1 has an active-like state (Figure 3.3a, region I), a region with αC helix disorder but retaining the K-E salt bridge (Figure 3.3a, region II), and a state with a folded and active-like positioned αC helix and a broken K-E salt bridge due to interaction between K317 and ATP (Figure 3.3a, region III). However, BAK1 does not take on a Src-like inactive conformation, as there is no distal swinging of the folded αC helix (Supplementary Figure 3.3). The most striking feature of the BAK1 free energy landscape is the large, well-populated area where the αC helix is highly disordered (Figure 3.3a, region IV), allowing for a large range of distances between K317 and E334. Breaking of the K-E salt bridge allows for a greater extent of αC helix disorder in BAK1 than in BRI1.

Circular dichroism spectroscopy confirms the presence of disorder in the BRI1 αC helix peptide.

In order to validate our findings on the αC helix disorder of the simulated kinases, we performed circular dichroism experiments on αC helix peptides (Supplementary Table 3.1). The well characterized kinase domains of the human epidermal growth factor receptor (hEGFR) and Src (hSrc) are believed to represent relative extremes on the disorder-order spectrum of αC helices, with hEGFR on the disordered side and hSrc the ordered[86, 79]. For this reason, we used hEGFR and hSrc αC helix peptides both as positive controls and to provide a scale for comparison. Quality spectra from 180 to 260 nm were obtained for the hSrc, hEGFR, and BRI1 peptides (Supplementary Figure 3.1). Due to lack of solubility, and incompatibility of
Figure 3.3: BAK1 conformational heterogeneity. a) Markov state model-weighted free energy surface of the BAK1 kinase domain as a function of αC helix helical content and the distance between the glutamate δ-carbon atom and the lysine side-chain nitrogen in the conserved K-E salt bridge. The free energy is calculated by adding histograms of each state weighed by state equilibrium probability, and taking the natural logarithm of this sum, as follows: \( F(x, y) = -\beta^{-1} \ln \left[ \sum_{i=1}^{N} \pi_i h_i(x, y) \right] \). b) (I-IV) BAK1 conformations representing the corresponding regions on the free energy surface in part a). I) Active-like structure, II) αC helix disordered structure with an intact K-E salt bridge, III) structure with a folded and properly positioned (as in the active-like state) αC helix but with a broken K-E salt bridge, IV) structure with a disordered αC helix and a broken K-E salt bridge.

solubility agents (i.e. DMSO) with circular dichroism spectroscopy, no CD spectra could be obtained for the BAK1 peptide. The three peptides tested showed different responses to 2,2,2-trifluoroethanol (TFE), suggesting distinct helical propensities between peptides. In the absence of TFE, all three peptides appear largely unstructured, although the positive band at 190 nm and negative bands at 208 and 220 nm are indicative of a small amount of α-helix formation in the hSrc αC peptide (Supplementary Table 3.2). The hSrc αC peptide exhibited an approximately 14-fold increase in \([\theta]\) at 190 nm, and 4-fold and 5-fold decreases in \([\theta]\) at 208 and 220 nm, respectively when measured in the presence of 20% TFE (Supplementary Table 3.2) indicating a shift to an α-helical state. Further increase in the TFE concentration had minimal effect on the hSrc αC peptide CD spectrum (Supplementary Figure 3.1). By comparison, transition to an α-helical state was more gradual for both the hEGFR and BRI1 αC peptides (Supplementary Figure 3.1 & Supplementary Table 3.2). Collectively, the results are indicative of different helix forming propensities among the three peptides where hSrc > BRI1 > hEGFR.
In silico substitution of the hSrc αC helix into BAK1 suppresses disorder.

Given the fact that we were unable to characterize the BAK1 αC helix peptide experimentally, it is unclear whether the disorder in the BAK1 αC helix is primarily due to interaction with neighboring residues or intrinsic disorder of the αC helix itself. If the BAK1 kinase domain as simulated is inhospitable to folding of any αC helix sequence, we would expect the otherwise stable hSrc αC helix to unfold in a similar manner to the native sequence. We find that the hSrc αC helix remains stable relative to the native BAK1 helix, eliminating a large portion of the metastable unfolded region (Figure 3.4a, region IV) though still allowing for notable deviation from a purely helical form (Figure 3.4a, regions II and IV). In addition to the stabilized αC helix, the conserved K-E salt bridge is also stabilized, where the αC helix remains folded in the main region of breakage (Figure 3.4a, region III). As with native BAK1 there does not appear to be a Src-like inactive state (Supplementary Figure 3.4).

Although the extent of unfolding of the Src αC helix is less than the native BAK1 αC helix, there is little barrier to a certain degree of unfolding when the K-E salt bridge is formed. Regions I and II of Figure 3.4a appear to constitute a kinetically indistinct state where the BAK1-Src αC helix can partially unfold and refold while constrained by the K-E salt bridge, while breaking of the K-E salt bridge occurs primarily by way of region I into region III over a free energy barrier of approximately 2-3 kcal-mol$^{-1}$. This transition
is likely due to a rotation of the lysine side chain towards the negatively charged ATP phosphate groups as is seen in native BRI1 and BAK1.

Interestingly, there is a small, low free energy region near 5 Å αC helix root mean-squared deviation (RMSD) and 20 Å K-E salt bridge distance (Figure 3.4a, region IV) which is not as stable in the unweighted free energy landscape (Supplementary Figure 3.5). The Markov state model (MSM) states contributing to this apparent highly populated inactive region have comparatively high error in their equilibrium populations (Supplementary Figure 3.6) as calculated from Markov chain Monte Carlo sampling of the prior distribution of transition matrices[148], meaning that this low free energy region could be present due to sampling error (see Methods for further discussion).

Finally, in order to predict hydrogen-deuterium (HD) exchange results for possible future experiments from our simulations, we calculated MSM-weighted solvent accessible surface area (SASA) for the αC helices of BRI1, BAK1, and BAK1-Src (Supplementary Figure 3.7). While SASA does not directly represent the HD exchange signal, accessibility of a residue to solvent is related to the signal and represents an appropriate surrogate.

αC helix disorder across the A. thaliana kinome.

In order to investigate the universality of αC helix disorder in the A. thaliana kinome, we obtained a set of 930 A. thaliana kinase sequences identified by Zulawski et al.[149] and used the PONDR-Fit web server[150] to predict disorder propensities of residues in putative αC helix regions. We then mapped the predicted disorder propensities onto the A. thaliana kinome phylogenetic tree (Figure 3.5a)[149]. It is evident from this tree that αC helix disorder is predicted to be widespread across the A. thaliana kinome, suggesting that transitions between ordered and disordered states in this region could be a common regulatory mechanism. As further validation for our computational and experimental results, we find that the hSrc sequence is predicted to have low disorder across the αC helix region, while hEGFR and BRI1 are predicted to display disorder qualitatively similar to the average A. thaliana kinase (Figure 3.5b). BAK1 on the other hand displays disorder near the upper boundary of the interquartile range for all examined residues (Figure 3.5b), and is clearly on the upper tail of the distribution of average αC helix residue disorder (Figure 3.5c). Several kinase families are predicted to have a majority of members with high αC helix disorder, including the CRR, NEK, and SnRK1 families (Supplementary Figure 3.8). Similar analyses were performed for the rice (Oryza sativa) and human kinomes, revealing a wide range of disorder in the putative αC helix regions across kinase families in both organisms (Supplementary Figure 3.9-Supplementary Figure 3.12). Rice was chosen due to the importance of rice in global agriculture and humans due to extensive previous structural studies on
multiple human kinases and the relevance of the human kinome to medicine.

Figure 3.5: Predicted αC helix disorder across the *A. thaliana* kinome. a) Disorder propensity of 41 residues centered around the conserved E334 (BAK1 numbering) from a Clustal Omega multiple sequence alignment (MSA) of the 930 kinases identified and arranged into a maximum likelihood phylogenetic tree[149]. Disorder propensity was predicted using the PONDR-Fit web server, and the average disorder propensity of the 15 MSA residues corresponding the BAK1 αC helix are mapped onto the phylogenetic tree[150]. b) Profile of the predicted disorder across the 15 MSA residues corresponding the BAK1 αC helix for hSrc, hEGFR, BRI1, and BAK1. Also included is a box plot showing the means and quartiles of predicted disorder propensity distributions for each residue across the 930 included sequences of the *A. thaliana* kinome. The conserved E334 (BAK1 numbering) is denoted by the vertical dashed line. c) Histogram of the average αC helix predicted disorder propensities calculated for part a).

3.3 Discussion

Eukaryotic protein kinases act as dynamic switches, the units of cellular computation, optimized not for turnover of phosphorylated product, but rather for precise tuning of activation through post-translational modification and protein-protein interactions[145, 151, 152, 153]. This ability of kinases to respond to multiple inputs allows for adaptable programmed responses integrating numerous signals[154, 155, 155, 156]. It is no surprise then that two protein kinases of such great import in control of myriad aspects of plant growth...
and development as BRI1 and BAK1 would be heavily regulated[21, 28, 33, 157, 158, 49]. Here we present evidence for yet another regulatory mechanism in the initiation of brassinosteroid signaling: intrinsic disorder in the αC helix in the BRI1 and BAK1 kinase domains. In MD simulations, fully phosphorylated BRI1 and BAK1 kinase domains in isolation were able to spontaneously transition from an active-like conformation into an inactive state characterized by breaking of the conserved K-E salt bridge and unfolding of the αC helix (Figure 3.6). This finding could provide some insight into the physical mechanism of how association of BRI1 and BAK1 leads to activation of their kinase domains, suggesting that αC helix stabilization due to physical interaction between BRI1 and BAK1 may be necessary for kinase activation in addition to phosphorylation, drawing parallels with well studied metazoan receptor kinase systems such as EGFR[34, 86, 147].

Distal swinging of the αC helix away from the active site is a well established mechanism of deactivation in numerous animal kinases, where interactions with regulatory domains or other kinases causes proximal relocation of the αC helix and subsequent activation[145, 147, 159]. BRI1 is able to transition to a similar Src-like inactive state in our simulations, suggesting a possibly similar mechanism of control. A related regulatory mechanism was discovered in hEGFR, a receptor tyrosine kinase (RTK) implicated in multiple cancers, where disorder in the αC helix of one kinase domain is believed to be suppressed by ligand-induced homodimerization and formation of an asymmetric dimer in the kinase domains, with one kinase acting as an “activator” for the other[86, 147, 159]. By comparison with hEGFR, our discovery of highly populated states in large timescale MD simulations of the BRI1 and BAK1 kinase domains characterized by disorder in the αC helices suggests that dimerization of BRI1 and BAK1 might promote catalytically competent conformations of both proteins via stabilization of their respective αC helices.

Given the conditions of our simulations, it is somewhat surprising that an inactive state persists in any notable capacity: ATP is already bound preventing a DFG flip and phosphorylation of key residues on the activation loop would be expected to keep it in the unfolded state needed for activation in other kinases[145]. Indeed, a DFG flip is impossible with ATP bound and we see no Src-like folding of the activation loop (Supplementary Figure 3.13), although there is currently no evidence for a well defined, metastable inactive state of BAK1 and scant evidence of such in BRI1, consisting only of our discovery of a Src-like inactive state in terms of αC helix placement. However, supported by experimental evidence from CD spectra along with bioinformatic analysis, our simulations show that both the BRI1 and BAK1 kinase domains have highly populated states characterized by αC helix disorder (Figures Figure 3.2 & Figure 3.3), suggesting that the BRI1 and BAK1 kinase domains are partially inactive when isolated from all potentially interacting proteins and relieved of their juxtamembrane and C-terminal tail domains, even when in an otherwise “active” state (that is, with all relevant residues phosphorylated and with ATP bound). The
The possibility of formation of an EGFR-like asymmetric dimer has been suggested\cite{34} based on the similarity of the binding interface between hEGFR and its inhibitor MIG6 with that of BRI1 and BKI1. Evidence of EGFR-like αC helix disorder in BRI1 and BAK1 supports the formation of asymmetric BRI1-BAK1 dimers as a possible activation mechanism, although direct experimental evidence for this hypothesis is lacking.

We demonstrated that the BAK1 αC helix disorder can be suppressed by substitution of the stable hSrc helix for the native helix, suggesting the presence of intrinsic disorder rather than disorder due to structural context. Although we were unable to experimentally characterize the BAK1 αC helix, our finding that the BRI1 αC helix peptide is in fact disordered lends further credence to this line of reasoning. While substitution of a stable αC helix into the BRI1 or BAK1 kinase domains could potentially be an avenue for genetically engineering plants with enhanced brassinosteroid signaling, and therefore more robust growth, we note that in attempts to experimentally create these chimeric kinases we found a BRI1-Src chimera to lack any activity and the BAK1-Src chimera described in our simulations to be toxic to its \textit{E. coli} expression system. The overall effects of such a drastic change to a region potentially important for regulation of a major growth and development signaling pathway on a plant are unknown and could easily be deleterious without a careful system-level consideration.

Intrinsically disordered regions (IDRs) are thought to confer versatility to proteins in terms of their interaction partners, and there is evidence that IDRs are enriched in protein-protein interaction network hub proteins\cite{160, 161}, including hubs in kinase interaction networks\cite{162}. BAK1, which we have shown

\textbf{Figure 3.6: Conformational network of BRI1 and BAK1 kinases.} The conformational states obtained from molecular simulations reveal different mechanisms of catalytic domain inhibition due to the conformational change in the αC helix.
in our models to have considerable disorder in its αC helix, is known to have an important role not just in brassinosteroid signaling, but also in innate immunity, cell death control, and light response, associating with several other LRR-RLKs\cite{163, 18}. Therefore, we propose that if BAK1 does in fact interact with BRI1 through its αC helix, it is possible that BAK1 is able to interact with other RLKs by a similar mechanism with the disorder to order transition of the BAK1 αC helix both serving to activate BAK1 and mediate binding to a relatively diverse set of kinase interfaces. Furthermore, given our predictions showing αC helix disorder throughout a large portion of the A. thaliana kinome (Figure 3.5), it seems possible that this regulatory mechanism could be used in more than one instance by plant cells, and that other kinases regulated by αC helix disorder may also have a high degree of binding versatility.

A question that naturally arises regarding the assumptions needed for our interpretation of the presented simulations is this: how stable is the catalytically competent state in an active kinase relative to incompetent states? Our simulation, experimental, and bioinformatic results suggest that the fully phosphorylated BRI1 and BAK1 kinase domains in isolation both have noticeable degrees of conformational heterogeneity in their αC helices, which can be taken to mean that these proteins in the given conditions should be less catalytically active than their fully active counterparts. However, the kinase activity that can be measured in a laboratory results from an averaging over the thermally accessible conformations of the protein, which does not necessarily rule out significant deviations from the region of catalytically competent conformations with non-negligible probability. Directly mapping the conformation of a kinase to its catalytic activity remains a significant challenge, leaving interpretation of kinase structures as active or inactive by analogy as the only clear option not requiring expensive simulation at higher level of physical theory.

Formation of an asymmetric dimer between the BRI1 and BAK1 kinase domains is, of course, not the only other possible mechanism by which the plant cell could stabilize the BRI1 and BAK1 αC helices in an active-like conformation. Our simulations neglected the juxtamembrane domains of both receptors, which have been shown to activate their respective kinase domains\cite{158}, for the sake of saving simulation time given their unstructured nature. For similar reasons, in order to allow us to sample the conformational dynamics of the kinase domains extensively, we did not include the plasma membrane, which could potentially interact with the kinase domains and have some effect on their conformational ensemble. Either case, or potentially another unknown case, would still be an interesting mode of regulation and further studies are needed in order to determine exactly how BRI1 and BAK1 are able to suppress αC helix disorder (and in the case of BRI1, promote proximal swinging of the αC helix) and activate.

The activation pathways of BRI1 and BAK1 are still unknown, as there are no inactive, unphosphorylated structures available for either receptor (Supplementary Figure 3.14), and studying how phosphorylation leads
to activation using simulation methods would be a logical next step. However, in this study we have identified inactive conformations of the phosphorylated kinase domains and thus demonstrated the insufficiency of phosphorylation alone to fully activate either of the BRI1 and BAK1 kinase domains, suggesting the necessity of other factors, potentially BRI1/BRI1, BAK1/BAK1, or BRI1/BAK1 interactions between kinase domains and juxtamembrane domains or other kinase domains. This finding reveals a potential additional regulatory mechanism in the initiation of *A. thaliana* brassinosteroid signaling, and possibly provides a structural basis for the versatility of BAK1 in its roles in multiple signaling pathways.

### 3.4 Methods

**Simulation Details.**

All simulations were performed using the AMBER 14 molecular dynamics package and the CHARMM 36 force field on the Blue Waters petascale computing facility. All simulation systems were set up using the VMD plugin Psfgen 1.6 and converted to Amber format using the CHAMBER software tool. Starting coordinates for BRI1 and BAK1 kinase domains were taken from available crystal structures (BRI1 PDB ID: 4OA2; BAK1 PDB IDs: 3TL8 (chains A, D, G, and H), 3UIM, and 3ULZ) (see Supplementary Figure 3.15 for exact sequences and Supplementary Table 3.3 for phosphorylation states). All segments missing from the BRI1 and BAK1 crystal structures were modeled using the SwissModel webserver. In order to create the chimeric BAK1 kinase the αC helix of hSrc kinase (PDB ID: 1Y57) was grafted into the 3ULZ and 3UIM structures in place of the native αC helix (see Supplementary Figure 3.15 for exact sequence).

Starting structures were solvated in water boxes with dimensions of approximately 90 Å X 70 Å X 63 Å with TIP3P model molecules. Sodium and chloride ions were added to neutralize the charge of all systems and bring salt concentration to approximately 150 mM. An ATP molecule with two magnesium ions bound, taken from previous simulations, was inserted into the binding pocket for all structures in place of the modified adenosine-phosphate molecules used in crystallization, aligned with the adenosine ring of the corresponding ATP analogue. All systems were subjected to 10,000 steps of energy minimization and were equilibrated for 8-10 ns in an NPT ensemble at 300K and 1 atm, maintained using Langevin dynamics and a Berendsen barostat. Simulations were performed using a 2 fs time step, periodic boundary conditions, particle mesh Ewald electrostatics, and constraints of hydrogen-containing bonds using the SHAKE algorithm. Equilibrated structures were then equilibrated for another 10 ns to obtain average dihedral angle potential energies for calculation of accelerated MD (AMD) parameters.
Table 3.4) according to reference[173]. For BAK1, ten replicates of each of the six structures were simulated using AMD for a total time given in Supplementary Table 3.5. BRI1 AMD simulations were initiated from a single crystal structure, and two rounds of adaptive AMD sampling were performed, first with a single trajectory and for the second and third rounds with 25 trajectories run in parallel for a total simulation time given in Supplementary Table 3.5. Finally, 25 AMD simulations were run in parallel serially for a total time indicated in Supplementary Table 3.5. Similarly, BAK1-Src AMD simulations were initiated from each of the two structures indicated in Supplementary Table 3.4 and run for one round of adaptive AMD with 25 parallel trajectories, and finally run with 25 trajectories in parallel serially for total times indicated in Supplementary Table 3.5. Ideally the AMD simulation strategy for all three systems would be identical, and it is possible that some of the differences observed between the free energy landscapes of each system could be due in part to differences in sampling schemes. The final round of AMD sampling for each system was clustered in the space of the distance between the most distal side chain nitrogen in the lysine and carbon in glutamate within the conserved K-E salt bridge and αC helix RMSD from the crystal structures (or initial structure, for the BAK1-Src chimera) into 100 states, and the nearest neighbors of the cluster centroids were chosen as starting structures for unbiased simulation. An aggregate of 30-40 μs of unbiased MD simulation time was performed for each system [Supplementary Table 3.5].

Trajectory analysis

All figures were produced using Matplotlib 1.5[176] and VMD 1.9[72]. α-helical content (α, α ∈ [0, 1]) measures the degree to which a given sequence of N + 1 residues forms an α-helix from MD trajectories based on angles between α-carbons and hydrogen bonds between backbone nitrogens and oxygens separated by four residues. We used the function for α as defined in the NAMD collective variables[177]. Values used for each parameter are shown in parentheses. The angle function component of α is defined as follows:

$$\text{angf}(C^{(n)}_\alpha, C^{(n+1)}_\alpha, C^{(n+2)}_\alpha) = \frac{1 - \left(\theta(C^{(n)}_\alpha, C^{(n+1)}_\alpha, C^{(n+2)}_\alpha) - \theta_0\right)^2}{1 - \left(\theta(C^{(n)}_\alpha, C^{(n+1)}_\alpha, C^{(n+2)}_\alpha) - \theta_0\right)^4}/(\Delta\theta_{\text{tol}})^4$$  

(3.1)

where $\theta(C^{(n)}_\alpha, C^{(n+1)}_\alpha, C^{(n+2)}_\alpha)$ is the angle between the three adjacent α-carbons of residues n, n + 1, and n + 2, $\theta_0$ is the reference angle (88°), and $\Delta\theta_{\text{tol}}$ is the angle tolerance (15°). We modified the form of this function in order to remove discontinuities, taking:

$$x = \left(\theta(C^{(n)}_\alpha, C^{(n+1)}_\alpha, C^{(n+2)}_\alpha) - \theta_0\right)/(\Delta\theta_{\text{tol}})$$  

(3.2)
\[
\text{angf}(C^{(n)}_\alpha, C^{(n+1)}_\alpha, C^{(n+2)}_\alpha) = \frac{1 - x^2}{1 - x^4} = \frac{(1 - x)(1 + x)}{(1 - x)(1 + x)(1 + x^2)} = \frac{1}{1 + x^2} \quad (3.3)
\]

The hydrogen bond function is defined as follows:

\[
hbf(O^{(n)}, N^{(n+4)}) = \frac{1 - |x_{O^{(n)}} - x_{N^{(n+4)}}|/d_0}{1 - (|x_{O^{(n)}} - x_{N^{(n+4)}}|/d_0)^k} \quad (3.4)
\]

where \(|x_{O^{(n)}} - x_{N^{(n+4)}}|\) is the distance between \(O^{(n)}\), the backbone oxygen of residue \(n\), and \(N^{(n+4)}\), the backbone nitrogen of residue \(n + 4\), \(d_0\) is the cutoff distance (3.3 Å), and \(k\) (6) and \(m\) (8) are parameters controlling long-range behavior of the function. We also modified the form of this equation to remove discontinuities, for \(k = 6\) and \(m = 8\):

\[
y = |x_{O^{(n)}} - x_{N^{(n+4)}}|/d_0 \quad (3.5)
\]

\[
hbf(O^{(n)}, N^{(n+4)}) = \frac{1 - \frac{y^6}{1 - y^8}}{1 - \frac{y^6}{1 - y^8}} = \frac{(1 - y)(1 + y)(1 + y^4) + (1 - y)(1 + y)y^2}{(1 - y)(1 + y)(1 + y^2)(1 + y^4)} = \frac{1 + y^2 + y^4}{1 + y^2 + y^4 + y^6} \quad (3.6)
\]

The alpha helical content is then defined as:

\[
\alpha(C^{(N_0)}, \ldots, C^{(N_0+N)}_\alpha, O^{N_0}, \ldots, O^{N_0+N-4}, N^{N_0+4}, \ldots, N^{N_0+N}) =
\frac{1}{2(N - 2)} \sum_{n=N_0}^{N_0+N-2} \text{angf}(C^{(n)}_\alpha, C^{(n+1)}_\alpha, C^{(n+2)}_\alpha) + \frac{1}{2(N - 4)} \sum_{n=N_0}^{N_0+N-4} hbf(O^{(n)}, N^{(n+4)}) \quad (3.7)
\]

Markov State Model Theory and Construction.

MSMs treat protein conformational dynamics as a Markov process on discretized configurational space in order to estimate thermodynamic and kinetic properties, regardless of whether the simulation dataset represents an equilibrium distribution[178]. This allows for freedom of choice in starting structures, which need not be chosen from a Boltzmann distribution in order to ultimately calculate accurate thermodynamic quantities from a finite simulation set. Discretization is achieved through clustering of trajectories featurized into some function of the protein atomic coordinates in order to form states. Transitions between states in a period of time, called the lag time, are then counted in order to estimate maximum likelihood transition probabilities with the constraint of detailed balance imposed[179]. Ensemble dynamics on the resulting network of states and transition probabilities can be modeled as a master equation, as follows for an \(N\) state
model:

$$p_i(t + \tau) = \sum_{j=1}^{N} p_j(t)p_{ji}(\tau)$$  \hspace{1cm} (3.8)

Here \(p_i(t)\) and \(p_i(t + \tau)\) are the probabilities of finding a protein in state \(i\) at times \(t\) and \(t + \tau\), \(\tau\) is the lag time of the model, and \(p_{ji}(\tau)\) is the probability of a transition from state \(j\) to state \(i\) in the time \(\tau\). In a more compact matrix form:

$$\mathbf{p}(t + \tau) = \mathbf{p}(t)\mathbf{T}(\tau)$$  \hspace{1cm} (3.9)

where \(\mathbf{T} = T_{ij}\). The eigenvectors of \(\mathbf{T}(\tau)\) in ascending order represent the slowest decorrelating vectors in state space, where the first eigenvector is the equilibrium state probability vector with an eigenvalue of one and an infinite implied relaxation timescale. One can obtain implied relaxation timescales of the \(i^{th}\) eigenvector in the following manner:

$$t_i = -\frac{\tau}{\ln \lambda_i}$$  \hspace{1cm} (3.10)

The entire MSM construction process was performed using the MSMBuilder 3 Python package for all systems\[180\]. All trajectories were subsampled so that the time difference between consecutive frames in a trajectory was 200 ps. We chose the RMSD of the atoms contained in the \(\alpha\)C helix and separately the N-terminal lobe excluding the \(\alpha\)C helix with respect to the corresponding reference structure (BRI1, PDB: 4OA2\[104\]; BAK1, PDB: 3TL8, chain A\[32\]; initial structure of BAK1-Src, PDB: 1Y57\[170\] for the \(\alpha\)C helix, PDB: 3UIM\[33\] and 3ULZ\[168\] for the remainder of the kinase) as features for construction of MSMs for our systems of interest. Each of the two metrics for BRI1 and BAK1 trajectories were normalized by subtracting the mean and dividing by the standard deviation, where both statistics were calculated over all trajectories from the appropriate protein. The ranges of valid lag times for the BRI1, BAK1, and BAK1-Src models were both chosen to be 50-150 ns based on convergence of implied timescales across models built with a range of cluster numbers (Supplementary Figure 3.16-Supplementary Figure 3.18). We then used the Osprey variational cross-validation package to select lag times of 50 ns for all three models and cluster numbers of 278, 319, and 386 for the BRI1, BAK1, and BAK1-Src models, respectively, which maximized the model generalized matrix Rayleigh quotient cross-validation scores using shuffle-split cross-validation\[181\]. Using these MSM parameters, we estimated reversible maximum likelihood transition probability matrices for all models.

Free energy plots were constructed by first creating normalized a two-dimension histogram for each of \(N\) states, using the structures in each state projected onto the distance between the glutamate \(\delta\)-carbon atom and the lysine side-chain nitrogen in the conserved K-E salt bridge and the helical content of the \(\alpha\)C helix, as defined in the NAMD 2.11 manual\[177\] as detailed above. Each state histogram \([h_i(x, y)\) for state \(i\)] for
the given model was then weighed by the equilibrium probability ($\pi_i$) of the corresponding state calculated from the transition probability matrix and summed together, according to the following equation:

$$F(x, y) = -\beta^{-1} \sum_{i=1}^{N} \pi_i h_i(x, y)$$  \hspace{1cm} (3.11)

Unweighted free energy plots are shown in Supplementary Figure 3.21, Supplementary Figure 3.22 & Supplementary Figure 3.5.

**MSM error analysis**

Error analysis was conducted by constructing Bayesian MSMs\[148\] with the same parameters as the corresponding maximum likelihood MSMs, utilizing Metropolis Markov chain Monte Carlo to sample the prior distribution of transition matrices.

Because our sampling method involved initiating 100 independent trajectories obtained from adaptive AMD sampling, a method which should in principle roughly return a Boltzmann distribution even with limited sampling if properly reweighted\[144\], we expect the state sample size and MSM equilibrium population to covary in a roughly linear fashion (as should the logarithms of both). While deviations from a linear relationship are not necessarily a result of sampling error, we took deviant states as potential contributors to the MSM error and looked for a combination of deviation from the linear relationship above coupled with relatively high standard deviation in their MSM equilibrium population. Standard deviation was calculated for each transition probability, and for properties of MSMs which are functions of transition probabilities (Supplementary Figure 3.19, Supplementary Figure 3.20 & Supplementary Figure 3.6), using 500 samples generated by the Bayesian MSM method implemented in MSMBuilder 3.4\[180\] with Markov chain Monte Carlo (MCMC) sampling of the prior distribution of transition probability matrices\[148\].

We performed several tests for each MSM to determine the level of error due to sampling in the state equilibrium populations, including MCMC sampling of the prior distribution of transition probability matrices. Since our method of sampling involved initiating one or two hundred trajectories from structures produced by adaptive AMD without including the remaining adaptive AMD trajectories in our final analysis, we expect the raw distribution of conformations to very roughly approximate the expected underlying Boltzmann distribution, where this similarity to the equilibrium distribution is expected to decrease with more frequent adaptive sampling. Because of this expected similarity to the equilibrium distribution, we also expect that the relationship between the number of times each state has been visited and the MSM equilibrium population of each corresponding state to be roughly linear, and any deviation from linearity could
be an indication of high sampling error of the transition probabilities, although this is not necessarily the case. By combining this mode of analysis with standard deviation in the equilibrium populations, where a state deviating from linearity and displaying a high standard deviation is likely strongly affected by sampling error, affecting our level of confidence in its equilibrium population. We found that in the BAK1-Src simulations, two states (Supplementary Figure 3.6) both deviated from linearity in a log-log plot of state sample size versus MSM equilibrium population and had high standard deviations in their equilibrium populations (1.08 · 10^{-3} and 3.08 · 10^{-3} as compared with the highest standard deviations for BRI1 and BAK1 states, 7.11 · 10^{-4} and 7.33 · 10^{-4} respectively). Both assertions are, of course, of a somewhat subjective nature and prevent any quantitative statement of confidence level in these states versus others. It is worth noting that further sampling of transitions involving these two states would likely significantly change their equilibrium populations. These two states are the main contributors to the low free energy region of the BAK1-Src free energy landscape (Figure 5) where the αC helix has an RMSD of around 5.5 Å from the initial structure and the distance between K317 and E334 is around 20 Å, bringing into question the existence of this metastable (in the K-E distance, αC helix projection) state.

**Peptide synthesis.**

21-mer peptides (Table 1) corresponding to the αC-helix of different protein kinases were custom synthesized by Genscript (hSrc and BRI1) or LifeTein (hEGFR and BAK1). Peptides contained acetylation and amidation as modifications at the N and C-terminus of each peptide respectively. Peptides were solubilized in sterile water to a concentration of 10 mM.

**Circular dichroism spectroscopy.**

Circular dichroism spectroscopy was performed using a JASCO J-715 spectropolarimeter. Spectra (180 to 260 nm) were collected at room temperature in a 0.1 mm quartz cuvette on samples containing 250 µM peptide, 5 mM HEPES-NaOH pH 7.5 and different amounts of TFE as indicated. Spectra were collected at 1 nm steps across the 180 to 260 nm range. Instrument response time was set to 1 sec and four spectra were collected and averaged for each buffer baseline and peptide sample. Data were collected in millidegrees and spectra were smoothed using a Savitsky-Golay method, followed by subtraction of the appropriate buffer baseline. Smoothed, subtracted data were converted to units of molar ellipticity ([θ]).
Bioinformatic analysis.

The list of the 940 *A. thaliana* kinases identified by Zulawski et al. [149] was used to create a multiple sequence alignment (MSA) using Clustal Omega [182] after the kinase sequences were retrieved from the Arabidopsis Information Resource [183]. Using the Biopython package, [184] the position of the conserved glutamate in the αC helix (BAK1: 334, BRI1: 927) in the MSA was identified, and every kinase without a gap at that position was selected for further analysis (a total of 930 out of the original 940 sequences). Sequences putatively containing the αC helix of each kinase were produced by taking a subsequence consisting of the conserved glutamate position and 20 residues in both directions for a total of 41 residues. All 930 subsequences were submitted to the PONDR-Fit consensus disorder prediction web server [150], and the average disorder propensity over the 15 residue segment corresponding to the BAK1 αC helix for each sequence was calculated and mapped onto the *A. thaliana* kinome phylogenetic tree from Zulawski et al. [149] using FigTree 1.4.2 [185].

The human and rice (*Oryza sativa*) kinomes [186, 187] were analyzed in the same manner as the *A. thaliana* kinome as described above, except that the human kinome MSA was obtained online [2] as Clustal Omega yielded an MSA which appeared to display a low level of αC helix E310 (Src kinase numbering) conservation. Out of the human kinome, 485 kinases were used for analysis while 1780 rice kinases were used, where omissions were due to gaps in the MSA at the conserved glutamate site or a lack of 20 flanking residues on the N-terminal end.

### 3.5 Supplementary Information

![Supplementary Figure 3.1](image)

**Supplementary Figure 3.1:** Experimental validation of BRI1 αC helix disorder with CD spectroscopy. Far-UV circular dichroism spectra of αC helix peptides at different concentrations of 2,2,2-trifluoroethanol. a) hSrc kinase. b) hEGFR. c) BRI1. Spectra were collected on samples containing 250 µM of each peptide and are presented in units of molar ellipticity. Each spectrum is an average of 4 scans.

---

Supplementary Table 3.1: Properties of synthetic peptides used in this study.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>Mol. weight (Da)</th>
<th>MRW†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Src</td>
<td>Ac-GTMSPEALQEQVMKKLRHE-NH₂</td>
<td>2471.87</td>
<td>117.7081</td>
</tr>
<tr>
<td>EGFR</td>
<td>Ac-SPKANKEILDEAYVMASVDNP-NH₂</td>
<td>2332.57</td>
<td>111.0748</td>
</tr>
<tr>
<td>BRI1</td>
<td>Ac-SQQGDREFMAEMETIHKIKHR-NH₂</td>
<td>2461.79</td>
<td>117.2281</td>
</tr>
<tr>
<td>BAK1</td>
<td>Ac-TQGGELQFQTEVEMISMAVHR-NH₂</td>
<td>2432.75</td>
<td>115.8452</td>
</tr>
</tbody>
</table>

† MRW, mean residue weight (MW/# of residues). MRW is used in conversion of CD units.

Supplementary Table 3.2: Properties of CD spectra for αC-helix peptides at different concentrations of 2,2,2-trifluoroethanol.

<table>
<thead>
<tr>
<th>Molar Ellipticity (deg cm⁻² dmol⁻¹ · 10⁻³)</th>
<th>Src</th>
<th>BRI1</th>
</tr>
</thead>
<tbody>
<tr>
<td>TFE %</td>
<td>190 nm</td>
<td>208 nm</td>
</tr>
<tr>
<td>0</td>
<td>2.98</td>
<td>-4.44</td>
</tr>
<tr>
<td>40</td>
<td>43.64</td>
<td>-18.62</td>
</tr>
<tr>
<td>60</td>
<td>45.63</td>
<td>-19.4</td>
</tr>
<tr>
<td>80</td>
<td>47.14</td>
<td>-20.48</td>
</tr>
</tbody>
</table>

Supplementary Table 3.3: BRI1, BAK1, and BAK1-Src in silico phosphorylation states.

<table>
<thead>
<tr>
<th>Name</th>
<th>Phosphorylated residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRI1</td>
<td>pT1039, pS1042, pS1044, pS1049</td>
</tr>
<tr>
<td>BAK1</td>
<td>pS290, pT312, pT324, pT446, pT449, pT450, pT455</td>
</tr>
<tr>
<td>BAK1-Src</td>
<td>pS290, pT312, pT446, pT449, pT450, pT455</td>
</tr>
</tbody>
</table>

Supplementary Table 3.4: AMD parameters for all systems.

<table>
<thead>
<tr>
<th>Name</th>
<th>α_D</th>
<th>E_{Threshold,D}</th>
</tr>
</thead>
<tbody>
<tr>
<td>4OA2 (BRI1)</td>
<td>235</td>
<td>4045</td>
</tr>
<tr>
<td>3TL8A (BAK1)</td>
<td>241</td>
<td>4123</td>
</tr>
<tr>
<td>3TL8D (BAK1)</td>
<td>241</td>
<td>4186</td>
</tr>
<tr>
<td>3TL8G (BAK1)</td>
<td>243</td>
<td>4154</td>
</tr>
<tr>
<td>3TL8H (BAK1)</td>
<td>248</td>
<td>4180</td>
</tr>
<tr>
<td>3UM (BAK1)</td>
<td>244</td>
<td>4114</td>
</tr>
<tr>
<td>3ULZ (BAK1)</td>
<td>244</td>
<td>4117</td>
</tr>
<tr>
<td>3UM (BAK1-Src)</td>
<td>240</td>
<td>4100</td>
</tr>
<tr>
<td>3ULZ (BAK1-Src)</td>
<td>240</td>
<td>4100</td>
</tr>
</tbody>
</table>

Supplementary Table 3.5: Total simulation time for all systems.

<table>
<thead>
<tr>
<th>Name</th>
<th>Adaptive AMD time (µs)</th>
<th>Serial AMD time (µs)</th>
<th>Unbiased MD time (µs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRI1</td>
<td>1.5</td>
<td>4.5</td>
<td>29.9</td>
</tr>
<tr>
<td>BAK1</td>
<td>0.0</td>
<td>8.6</td>
<td>34.3</td>
</tr>
<tr>
<td>BAK1-Src</td>
<td>0.5</td>
<td>4.5</td>
<td>32.4</td>
</tr>
</tbody>
</table>
Supplementary Figure 3.2: Unweighted histogram of frames from simulation of BRI1 with an $\alpha$C helix RMSD of less than 1.5 Å from the reference used in the main text. The features used are the distances between the terminal nitrogen atom and the $\delta$-carbon of K911 and E927, respectively, and the $\alpha$-carbons of K911 and E927. The color bar scale represents raw counts.

Supplementary Figure 3.3: Unweighted histogram of frames from simulation of BAK1 with an $\alpha$C helix RMSD of less than 1.5 Å from the reference used in the main text. The features used are the distances between the terminal nitrogen atom and the $\delta$-carbon of K317 and E334, respectively, and the $\alpha$-carbons of K317 and E334. The color bar scale represents raw counts.
Supplementary Figure 3.4: Unweighted histogram of frames from simulation of BAK1-Src chimera with an αC helix RMSD of less than 3 Å from the reference used in the main text. The features used are the distances between the terminal nitrogen atom and the δ-carbon of K317 and E334, respectively, and the α-carbons of K317 and E334. The color bar scale represents raw counts.

Supplementary Figure 3.5: Unweighted free energy landscape of the BAK1-Src chimera kinase domain. A 2-dimensional histogram was created in the distance between K317 and E334 in the conserved K-E salt bridge and the RMSD of the αC helix with respect to the crystal structure. The negative logarithm of the normalized histogram was multiplied by k_B T in order to calculate a free energy landscape.
Supplementary Figure 3.6: Error analysis of the BAK1-Src MSM. a) Log state sample size versus log MSM equilibrium population. b) Mean equilibrium populations of all states in the BAK1-Src MSM with error bars representing standard deviation. Means and standard deviations were calculated from 500 samples generated by constructing a Bayesian MSM (with the same parameters as the corresponding maximum likelihood MSM) using Metropolis Markov chain Monte Carlo sampling of the prior distribution of transition probability matrices in MSMBuilder 3.4. It is of note that the two most deviant states in a) correspond to the states in b) with by far the largest standard deviations of any state in any MSM constructed in this study, and that these two states have extraordinarily high equilibrium populations in the maximum likelihood MSM which contribute to the low free energy region.

Supplementary Figure 3.7: MSM weighted average per-residue solvent accessible surface areas (SASA) for the αC helices of all three simulated systems. While SASA does not directly correspond to the signal obtained from hydrogen-deuterium (HD) exchange experiments, the two are related and these results or similar ones could be used to compare with potential future HD exchange experiments on the full kinase domains.
Supplementary Figure 3.8: Binned distribution of predicted αC helix disorder by *A. thaliana* kinase family identified in a previous study. [149] For each family, all member kinases were assigned to one of three bins representing a range of disorder propensity depending on their predicted disorder propensity values and the fraction in each bin was calculated by normalizing with the total number of kinases in that family.

Supplementary Figure 3.9: a) Binned distribution of predicted αC helix disorder in the *O. sativa* kinome, identified in a previous study. [186] b) Box plot of the predicted rice kinome putative αC helix region disorder by residue.
Supplementary Figure 3.10: Binned distribution of predicted αC helix disorder by *O. sativa* kinase family identified in a previous study. For each family, all member kinases were assigned to one of three bins representing a range of disorder propensity depending on their predicted disorder propensity values and the fraction in each bin was calculated by normalizing with the total number of kinases in that family.
Supplementary Figure 3.11: a) Binned distribution of predicted αC helix disorder in the human kinome, identified in a previous study \[187\]. b) Box plot of the predicted human kinome putative αC helix region disorder by residue.

Supplementary Figure 3.12: Binned distribution of predicted αC helix disorder by human kinase family identified in a previous study \[187\]. For each family, all member kinases were assigned to one of three bins representing a range of disorder propensity depending on their predicted disorder propensity values and the fraction in each bin was calculated by normalizing with the total number of kinases in that family.
Supplementary Figure 3.13: MSM-weighted probability density functions (PDFs) for α-helical content of the fully phosphorylated activation loops of BRI1 and BAK1 from MD simulations. Shown for comparison is the α-helical content of the human Src kinase activation loop from an inactive crystal structure (PDB: 2SRC[188]). α-helical content was calculated according to the NAMD Collective Variables method. Residues selected for measurement were as follows, starting from the glycine in the conserved DFG motif: BRI1, 1029-1054; BAK1, 436-461; hSrc, 406-431. PDFs for each state were approximated using Gaussian kernel density estimation implemented in the Python package Scipy[189], and the total PDFs were estimating using a sum of PDFs of each state weighted by the state equilibrium population.

Supplementary Figure 3.14: Comparison of all available crystal structures of the BAK1 (PDB IDs: 3TL8[32], 3UIM[33], 3ULZ[35]) and BRI1 (PDB IDs: 4OA2[104], 4OA6[104], 4OAH[104], 4OAC[104], 4OH4[103], 4Q5J[190]) kinase domains on the RCSB Protein Data Bank. Separate monomers in the crystallographic unit of the same PDB entry are considered separately, with the chain ID specified. a) To the left, the six available crystal structures of the BAK1 kinase domain superimposed over one another, and to the right, pairwise RMSDs of backbone atoms commonly present in all crystal structures calculated in the RMSD Trajectory tool in VMD[72]. b) To the left, the nine available crystal structures of the BRI1 kinase domain superimposed over one another, and to the right, pairwise RMSDs of backbone atoms commonly present in all crystal structures. c) All available BAK1 and BRI1 kinase domain crystal structures mapped onto the distance between K317 (BRI1: K911) and E334 (BRI1: E327) and the RMSD of the common set of backbone atoms in the αC helix, as defined in the main text.
Supplementary Figure 3.15: Multiple sequence alignment (MSA) of the BRI1, BAK1, and BAK1-Src chimera amino acid sequences used in simulations, numbered by the residue in the sequence and not by the residue number of the entire protein. The lysine of the conserved K-E salt bridge is indicated by a lower case k while the glutamate is indicated by a lower case e (the location in the MSA of this glutamate is different for BAK1-Src from BAK1 and BRI1), the region spanning all of the regions considered the αC helix for all three proteins is indicated by a lower-case c, the DFG motif is indicated by ∗, and the region spanning all of the regions considered the activation loop for all three proteins is indicated by a lower-case a. This MSA was generated using the MultiSeq VMD 1.9.2 plugin and the visualization was created using the TCoffee and BoxShade 3.2 servers.

Supplementary Figure 3.16: Scaling of the top five implied timescales with the MSM lag time for the BRI1 MSM. A lag time range of 50-150 ns was selected based on convergence of the timescales and a final lag time of 50 ns was selected from this range based on the Osprey cross-validation score.
Supplementary Figure 3.17: Scaling of the top five implied timescales with the MSM lag time for the BAK1 MSM. A lag time range of 50-150 ns was selected based on convergence of the timescales and a final lag time of 50 ns was selected from this range based on the Osprey cross-validation score[181].

Supplementary Figure 3.18: Scaling of the top five implied timescales with the MSM lag time for the BAK1-Src chimera MSM. A lag time range of 50-150 ns was selected based on convergence of the timescales and a final lag time of 50 ns was selected from this range based on the Osprey cross-validation score[181].
Supplementary Figure 3.19: Error analysis of the BRI1 MSM. a) Log state sample size versus log MSM equilibrium population. b) Mean equilibrium populations of all states in the BRI1 MSM with error bars representing standard deviation. Means and standard deviations were calculated from 500 samples generated by constructing a Bayesian MSM (with the same parameters as the corresponding maximum likelihood MSM) using Metropolis Markov chain Monte Carlo sampling of the prior distribution of transition probability matrices in MSMBuilder 3.4.

Supplementary Figure 3.20: Error analysis of the BAK1 MSM. a) Log state sample size versus log MSM equilibrium population. b) Mean equilibrium populations of all states in the BAK1 MSM with error bars representing standard deviation. Means and standard deviations were calculated from 500 samples generated by constructing a Bayesian MSM (with the same parameters as the corresponding maximum likelihood MSM) using Metropolis Markov chain Monte Carlo sampling of the prior distribution of transition probability matrices in MSMBuilder 3.4.
**Supplementary Figure 3.21:** Unweighted free energy landscape of the BRI1 kinase domain. A 2-dimensional histogram was created in the distance between K911 and E927 in the conserved K-E salt bridge and the RMSD of the αC helix with respect to the crystal structure. The negative logarithm of the normalized histogram was multiplied by $k_B T$ in order to calculate a free energy landscape.

**Supplementary Figure 3.22:** Unweighted free energy landscape of the BAK1 kinase domain. A 2-dimensional histogram was created in the distance between K317 and E334 in the conserved K-E salt bridge and the RMSD of the αC helix with respect to the crystal structure. The negative logarithm of the normalized histogram was multiplied by $k_B T$ in order to calculate a free energy landscape.
Chapter 4

How does multisite phosphorylation activate the BAK1 kinase domain?

4.1 Introduction

While a great deal of effort has gone into understanding the signaling pathways in which BAK1 is involved, the physico-chemical basis of BAK1 activation through multisite phosphorylation remains partially unclear, though crystal structures of the BAK1 kinase domain have provided important clues \[32, 33\] and our computational work has explored BAK1 kinase domain conformational dynamics \[106, 107, 1\].

Previously, Wang et al. used a combination of in vitro kinase activity assays and mass spectrometry to explore the effects of mutating BAK1 phosphorylation sites on its activity towards a number of its LRR-RLK substrates \[193\]. Their study revealed differential phosphorylation of LRR-RLK targets by BAK1 in vitro with different BAK1 phosphorylation patterns produced through mutations. Their results suggest that, at least for phosphorylation of the other kinase domains tested, the BAK1 activation loop residues T450 and T455 are the most important phosphorylation sites, followed by T446 and T449 and subsequently S290 and T312. These results raise the possibility that the phosphorylation pattern of BAK1 controls its specificity towards its many substrates, further supported by evidence that distinct sets of phosphorylation sites control BAK1 activity within brassinosteroid and immune signaling pathways \[194\]. Other results from kinase activity assays using a peptide substrate found that a T455A mutation of BAK1, preventing T455 phosphorylation, had the largest effect on BAK1 activity, while mutation of the other three activation loop threonine residues (T446, T449, T450) had less of an effect \[27\].

In our view, a complete explanation of the mechanisms of BAK1 phosphorylation-controlled activity and specificity would require exploration of changes in the phosphotransfer reaction itself through quantum mechanics / molecular mechanics (QM/MM) simulations, the BAK1 conformational ensemble through classical MD simulations, and the thermodynamics and kinetics of BAK1-substrate association through biased and/or coarse-grained MD. In this study, we focus on the second of these levels of biomolecular dynamics, exploring the impact of seven phosphorylation states, constructed from the four threonine phosphorylation sites on the activation loop, as well as ATP binding on the conformational ensemble of BAK1 through Gaussian...
accelerated molecular dynamics (GAMD) simulations [144].

**Figure 4.1:** A crystal structure of the BAK1 kinase domain in an active conformation (PDBID: 3TL8 [32], chain A). The structure is coarsely partitioned into important regions of note for our analysis and color coded in a manner consistent throughout this article. The four threonine phosphorylation sites on the BAK1 activation loop are shown. B Closeup views of the four threonine phosphorylation sites on the BAK1 activation loop and nearby residues. The threonine residues are phosphorylated as in the crystal structure. This figure was produced using VMD 1.9.2 [108] and Inkscape 0.91 [109].

Given the conserved structural features of protein kinases necessary for catalysis of phosphotransfer reactions [195], we compare the free energies over these features between all eight mod-forms, a term coined to generally describe modification states of proteins [196], in order to gain insight into how phosphorylation of BAK1 A-loop residues individually and collectively affect the conformational aspect of BAK1 activation. We find that the activation loop of the unphosphorylated BAK1 kinase domain “cracks” and forms what is likely an inactive conformation. Phosphorylation of T450 alone, which is surrounded by positive residues in crystal structures [32, 33], can prevent activation loop cracking, and hold the BAK1 activation loop in a largely active conformation. The αC helix is able to swing outward in the unphosphorylated state, while it appears unable to do so in the presence of any phosphorylation state we have examined. Additionally, the αC helix appears to remain largely intact except when all four activation loop threonine residues are phosphorylated and ATP is not present. Our results suggest that phosphorylation of T450 is the most important factor for controlling the features of activation that we have examined, and provide an expanded structural basis for the primary importance of T450 and T455 phosphorylation in BAK1 activation.

### 4.2 Results

**Phosphorylation, especially of T450, stabilizes the activation loop**

We calculated the root mean-squared fluctuations (RMSF) of the BAK1 α-carbons in each mod-form, in order to evaluate the effects of phosphorylation and ATP binding on overall backbone fluctuation. The BAK1
activation loop had large RMSF values in the unphosphorylated state Figure 4.2, indicating a high degree of structural fluctuations. In comparison, the RMSF of the activation loop and the region from residues 475-525 is greatly reduced in the pT450 state, in addition to every other state in which T450 is phosphorylated. Although the RMSF of the activation loop is slightly reduced in the pT455 state as compared with the unphosphorylated state, the pT455 state has a higher RMSF in this regions in comparison with all other states. These results suggest that phosphorylation of T450 may be critical for stabilizing the activation loop. Very little difference in the RMSF of the αC helix is apparent across all mod-forms. We also note that the RMSF of the last ~25 residues in the pT450-pT455 state is far higher than in any other state.

Figure 4.2: α-carbon RMSF of the BAK1 kinase domain for each mod-form calculated from GAMD simulations reweighted using MBAR. Phosphorylation of T450 reduces the large activation loop fluctuations seen in the unphosphorylated state, while phosphorylation of T455 is insufficient for activation loop stabilization. This figure was produced using Matplotlib 1.5 [176].

To further examine the effects of mod-form on activation loop dynamics, we calculated the PMF over activation loop RMSD to a crystal structure (PDBID: 3TL8 [32], chain A), as described in the Methods section. In agreement with our RMSF calculations, the BAK1 activation loop is able to explore a greater range of conformations in the unphosphorylated state, as indicated by the relatively flat PMF exhibiting one minimum around 0.4 nm and another at 1.0 nm Figure 4.3. Phosphorylation of T450 eliminates the second minimum and stabilizes the activation loop at a single minimum RMSD around 0.25 nm. Phosphorylation of T455 alone appears to stabilize the activation loop to a small extent as compared with the unphosphorylated state, again yielding two minima in the RMSD PMF, though the second minimum is shifted down to around
0.6 nm. All other mod-form RMSD PMFs are similar to that of the pT450 state, although in the pT450-
pT455-pT446 state a second minimum in the PMF is evident, although around 2 kcal·mol$^{-1}$ less stable than
the minimum at 0.25 nm.

Figure 4.3: PMFs over the BAK1 activation loop RMSD with respect to a crystal structure (PDBID: 3TL8 [32], chain A). The
activation loop RMSD of unphosphorylated BAK1 has a minimum in its PMF around 0.4 nm and a second minimum around 1.0 nm.
Except for the pT455 and pT450-pT455-pT446 states, the PMF over the activation loop has a single minimum around 0.25 nm. The
pT455 state displays larger activation loop fluctuations than any other phosphorylated state, but only has a second minimum near 0.6
nm, indicating reduced fluctuations from the unphosphorylated state. The PMFs were calculated from GAMD simulations reweighted
using MBAR. Error bars indicate standard deviations calculated using MBAR. This figure was produced using Matplotlib 1.5 [176].

The BAK1 $\alpha$C helix is stabilized in an active conformation by each mod-form

Figure 4.4: PMFs over the BAK1 $\alpha$C helix to N-lobe $\beta$-sheet center of mass distance. The PMFs were calculated from GAMD
simulations reweighted using MBAR. This figure was produced using Matplotlib 1.5 [176].
As with all previously described features, the αC helix swing distance, defined in the Methods section, can vary far more in the unphosphorylated BAK1 kinase domain than in any of the other mod-forms (Figure 4.4). The PMF over αC helix swing distance for unphosphorylated BAK1 displays two clear minima, around 1.45 nm and 1.8 nm, with the second minimum only slightly less stable than the first (< 1 kcal-mol\(^{-1}\)). In the pT450 and pT455 states a single free energy minimum exists around 1.55 nm, while in the pT450-pT455 state two minima exist, though separated by a small free energy barrier (< 1 kcal-mol\(^{-1}\)) and by roughly 0.1 nm. In the pT450-pT455-pT446, pT450-pT455-pT446-pT449, and pT450-pT455-pT446-pT449-ATP states, there is an overall shift towards larger αC helix swing distances, where the minimum for the pT450-pT455-pT446-pT449-ATP state is around 1.6 nm.

**Figure 4.5:** PMFs over the BAK1 αC helix helicity. The PMFs were calculated from GAMD simulations reweighted using MBAR. This figure was produced using Matplotlib 1.5 [176].

Phosphorylation of T450, T455, and T449 appear to either stabilize or have no effect on the helicity of the αC helix, while phosphorylation of T446 on top of phosphorylation of the other residues appears to induce unfolding (Figure 4.5). In all mod-forms, the αC helix can display a range of helicities, with PMF values of around 2 kcal-mol\(^{-1}\) consistently at 0.6 helicity and a global PMF minimum between 0.7 and 0.85 helicity. The presence of local unfolding of the αC helix is consistent with previous results [106], even though in this case we do not consider phosphorylation of T324 adjacent to the αC helix. It is interesting that phosphorylation of T446 has such a large apparent effect on helicity of the αC helix, even though no direct contacts between pT446 and the αC helix are formed (Figure 4.6).
Phosphosite contact probabilities

In order to better understand the physical mechanism through which phosphorylation affects BAK1 dynamics, we calculated MBAR reweighted contact probabilities for each of the activation loop threonine residues with all other protein residues, using a cutoff distance of 0.5 nm.

The contact probabilities of T446 reveal that in all but the unphosphorylated state, T446 is in contact with S470 more often than not (Figure 4.6). As S470 lies at the beginning of the stable αF helix, a broken T446-S470 contact implies that the tip of the activation loop, which we define as residues 442-447, is mobile in the unphosphorylated state and can break its interaction with the αF helix. Phosphorylation of T450 or T455 appears to be sufficient to prevent this type of activation loop motion, as the T446-S470 contact (or T446 contact with adjacent residues) remains intact most of the time for each of the mod-forms with either residue phosphorylated.

T449 ubiquitously interacts with K468, although less often in the unphosphorylated and pT455 states.
Other residues interacting with T449 are R415 and K439, both of which primarily interact with T450, especially with T450 phosphorylation. T449 only interacts with R415 when T450 is unphosphorylated, suggesting that T450 phosphorylation either causes T450 to outcompete T449 for a contact with R415, or pT450 prevents deformation of the activation loop necessary for interaction between T449 and R415. T449 interacts with K439 most strongly in the pT450-pT455 and pT450-pT455-pT446 states, though it is difficult to tell why T449 interacts with K439 in the mod-forms that it does.

As in crystal structures, T450 contacts R415, K439, Y463, and K468 in the active activation loop conformation. It is precisely the interactions between pT450 and the positively charged R415, K439, and K468 that is likely to be responsible for the ability of T450 to stabilize the activation loop in an active conformation. It is clear from the T450 contact probabilities that phosphorylation of T450 itself is alone responsible for promoting these interactions, as in the pT455 state the T450 contacts are far less often formed while at the same time the T450 contact probabilities in the pT450 state are essentially identical to all other mod-forms containing a phosphorylated T450.
Figure 4.8: Contact probabilities for BAK1 T450 in each mod-form, reweighted using MBAR. The black dotted line represents T450 while white dotted lines denote contacting residues of interest. This figure was produced using VMD 1.9.2 [108], Inkscape 0.91 [109], and Matplotlib 1.5 [176].

Finally, the contacts of T455 indicate that T455 phosphorylation directly plays some role in influencing the inter-lobe dynamics, by interacting with R297 on the N-lobe (Figure 4.9). This contact is weak in the unphosphorylated state, but is only absent in the pT450 and pT450-pT455-pT446-pT449-ATP state. Phosphorylation of T455 appears to promote interaction with R418, potentially explaining how T455 phosphorylation plays a role in stabilizing the activation loop, although interaction with R490 may also contribute to activation loop stabilization.

4.3 Discussion

We have described the effects of seven phosphorylation patterns (including the null, unphosphorylated pattern) and ATP binding on the conformational dynamics of the BAK1 kinase domain. From our simulations, it appears that phosphorylation of T450 is most important for maintaining an active activation loop conformation, while T455 phosphorylation further stabilizes the activation loop. All examined phosphorylation states promote proximal positioning of the αC helix to the remainder of the N-lobe as compared with the
unphosphorylated state, where the αC helix can move further away from the N-lobe. Interestingly, phosphorylation of T446 coincides with a higher degree of local αC helix unfolding than is observed in mod-forms with unphosphorylated T446.

It is difficult to make comparisons with the available experimental data in a meaningful way, due to the combination of means through which phosphorylation of a residue could influence the activity of BAK1 towards a given target. Purely in terms of BAK1 conformational dynamics, the T450 phosphorylation site appears to be most important for stabilizing a crystal structure-like conformation. Phosphorylation of T455 appears to be of secondary importance, significantly stabilizing the activation loop only when T450 is also phosphorylated. From our results, T446 and T449 phosphorylation do not appear to significantly contribute towards stabilizing an active BAK1 conformation. By mutating BAK1 phosphorylation sites to alanine and examining the amount of phosphorylated substrate in vitro, several groups have found that T450 and T455 are likely the most functionally important sites, while T446 and T449 are less important due to the smaller
effect of their mutation to alanine on substrate phosphorylation levels [27, 193]. However, it is unclear whether mutation of T455 has a larger effect than mutation of T450 [27] or if mutation of either has a similar effect, although this likely depends on the substrate.

As mentioned in the Introduction, we specifically focused on the effects of phosphorylation and ATP binding on the BAK1 kinase domain conformational dynamics, neglecting the possible effects of these changes to BAK1 on the catalyzed phosphotransfer reaction itself and on association with substrates. Although phosphorylation of T446 and T449 does not appear to affect BAK1 conformational dynamics significantly, it is possible that pT446 and pT449 play an important role in recognition of specific BAK1 substrates. Similarly, pT450 and pT455 may play some role in association with substrates, or may alter the free energy landscape associated with phosphotransfer from ATP to the substrate. Both of these possibilities require future investigation.

For any BAK1 conformation to be active, ATP must be bound. Ideally, we would have simulated each BAK1 phosphorylation state with and without ATP bound. However, we chose to examine ATP-bound BAK1 when all four activation loop sites were phosphorylated, in order to reduce the number of simulations required. We based this choice on the reasoning that mutation of any one of the activation loop phosphorylation sites reduces BAK1 activity [27, 193], so that it is reasonable to assume that the pT450-pT455-pT446-pT449 is a minimally phosphorylated, fully active kinase.

It is important to note that, as with any molecular dynamics simulations of proteins, the total amount of sampling we have done is likely inadequate to completely describe the global conformational dynamics of the BAK1 kinase domain relevant to its biological function, even with the GAMD boost potentials applied. Given the hierarchical structure of protein free energy landscapes [197, 198], we could continue to simulate BAK1 on longer timescales while continuing to observe new behaviors with slower associated kinetics to no end. Still, it is likely that the behavior we have observed is relevant to dynamics in the local region of conformational space around the crystal structure. Of course, we must then assume that the crystal structure is biologically relevant and actually resembles a typical conformation of an active BAK1 kinase domain, but making this assumption appears to be the option given the experimental data that is available.

4.4 Methods

System setup

All systems were set up using the Psfgen plugin within VMD [72]. The CHARMM 36 force field [199] was used for protein atoms. The protein atoms of the core BAK1 kinase domain were taken from a crystal structure
(PDBID: 3TL8, chain A [32]), and all phosphoryl group atoms were removed. The desired phosphoryl groups for each system were added within Psfgen, along with hydrogen atoms. The BAK1 structure was then solvated in a box of TIP3P [126] water molecules, so that the edges of the box were 10 Å from any protein atom. Cl\(^-\) and Na\(^+\) ions were added to neutralize each system and bring the salt concentration to ~150 mM. In order to add ATP and Mg\(^{2+}\) to BAK1, a PKA crystal structure (PDBID: 4HPU [200]) containing AMP-PNP and two associated Mg\(^{2+}\) ions was aligned to BAK1 (3TL8, chain A) using the MultiSeq VMD plugin [201], and the resulting AMP-PNP and Mg\(^{2+}\) coordinates were used to place ATP and Mg\(^{2+}\) into the BAK1 ATP binding pocket. The Parmed tool within AmberTools 18 [139] was used to convert X-PLOR format structure and topology files into Amber format.

**Simulation details**

All simulations were run in Amber 18 [139] with a time step of 2 fs and with hydrogen-containing bonds constrained using the SHAKE algorithm [173]. All production runs were maintained at constant temperature of 300K using a Langevin thermostat with a coupling time constant of 2 ps and at a constant pressure of 1 bar using a Berendsen barostat with a \(\tau\) of 1 ps. The particle mesh Ewald method [130] was used to treat electrostatics and a cutoff of 10 Å was used for non-bonded interactions. Frames were saved to trajectory files every 10 ps.

All systems were subjected to energy minimization. For each of the eight systems, we initiated ten independent simulations starting from the equilibration step and following the same procedure from then on. Each replicate was equilibrated for 20 ns, and then run for an additional 26 ns in order to collect statistics for determination of Gaussian accelerated MD (GAMD) [142] parameters. Using these parameters, independently estimated for each replicate, we ran 500 ns of GAMD simulations for each replicate, amounting to 5 \(\mu\)s of aggregate simulation time for each system.

**Definitions of collective variables**

In order to measure its helical content, the \(\alpha\)C-helix was defined as residues 324-339. Helical content was measured according to the definition in the NAMD colvars manual [177]. The activation loop was defined as residues 437-459 and the C-lobe excluding the activation loop, defined as residues 368-436 and residues 460-576, was used to align BAK1 to the crystal structure. The RMSD of the activation loop in each frame compared with the crystal structure was then measured. Swinging of the \(\alpha\)C-helix was described by the distance between the centers of mass of the sets of backbone atoms defining the \(\alpha\)C-helix and the N-lobe \(\beta\)-sheet (residues 302-307, 313-319, and 360-364).
Gaussian accelerated molecular dynamics reweighting

Typically when using MBAR, PMFs are estimated through expectations

\[ p_i = \langle \chi_i(x) \rangle \]  \hspace{1cm} (4.1)

where \( \chi_i(x) \) is a function taking a value of 1 if \( x \) falls into bin \( i \), and 0 otherwise. In order to produce smooth PMFs, we instead use Gaussian basis functions

\[ g_k(\xi, \xi_0) = \frac{1}{\sqrt{(2\pi)^k\sigma^2}} e^{-\frac{1}{2\sigma^2}(\xi-\xi_0)^T(\xi-\xi_0)} \]  \hspace{1cm} (4.2)

where \( \xi \in \mathbb{R}^k \) is the variable vector of interest, and \( \sigma \) is the bandwidth of the basis functions. In our analysis, \( k \in \{1, 2\} \). We assume zero local covariance between variables. The unbiased probability density is then estimated as an MBAR expectation value as

\[ p(\xi) = \langle g_k(\xi, \xi(x)) \rangle \]  \hspace{1cm} (4.3)

where \( \xi(x) \) is the function mapping the full configuration \( x \in \mathbb{R}^{3N} \) of a frame to the collective variable vector \( \xi \). We estimate the PMF as

\[ F(\xi) = -\beta^{-1} \log p(\xi), \]  \hspace{1cm} (4.4)

subtracting an estimate of \( \min_\xi [F(\xi)] \) from each \( F(\xi) \).

Root mean-squared fluctuations

The root mean-squared fluctuations (RMSF) of BAK1 \( \alpha \)-carbons were estimated from MD simulations by first aligning all frames to the crystal structure \( \alpha \)-carbons and calculating \( \alpha \)-carbon displacements according to

\[ \Delta d_i(t) = ||x_{C\alpha,i}(t) - x_{C\alpha,i}^{\text{crystal}}|| \]  \hspace{1cm} (4.5)

where \( x_{C\alpha,i}(t) \) is the coordinate vector for the \( \alpha \)-carbon of residue \( i \) at time \( t \) and \( x_{C\alpha,i}^{\text{crystal}} \) is the coordinate vector of the same residue \( \alpha \)-carbon in the crystal structure. The root mean-squared fluctuations (RMSF) of \( \alpha \)-carbons were then obtained according to

\[ R_i = \sqrt{\langle (\Delta d_i - \langle \Delta d_i \rangle)^2 \rangle} \]  \hspace{1cm} (4.6)
where the ensemble averages were estimated in the unbiased ensemble using MBAR.
Chapter 5

The effects of S-glutathionylation on the BAK1 kinase domain

5.1 Introduction

Reversible post-translational modifications (PTMs) are an essential mechanism for control of protein function in biochemical networks, allowing a cell to modulate protein function rapidly in response to changes in its metabolic state or in the local environment\(^1\). The phosphorylation-dephosphorylation cycle of serine, threonine, and tyrosine residues catalyzed by protein kinases and phosphatases, respectively, is a well studied example of this mode of control. Protein phosphorylation allows for feedback, oscillation, and multistability at the network scale as well as integration of multiple signals in a single protein through the possibility of a multisite phosphorylation code\(^1\).

Growing evidence supports the role of S-glutathionylation in control of protein function in response to oxidative stress\(^2\). Glutathione (GSH) is a tripeptide consisting of a glutamate bound through a \(\gamma\)-peptide linkage to a conventional cysteine-glycine dipeptide (Figure 5.1b). GSH is present in nearly all known life forms and the enzymes responsible for its synthesis are known to be essential for normal function in several animals and plants\(^2\). GSH is able to modulate protein function by forming a disulfide bond with cysteine residues (S-glutathionylation) under oxidative conditions (Figure 5.1b), a process believed to be catalyzed in both directions by glutaredoxin enzymes\(^2\). S-glutathionylation plays a role in control of numerous enzymes\(^2\), including protein kinases\(^2\), such as human AMP-activated protein kinase (AMPK) on C299 and C304 within the catalytic \(\alpha\)-subunit C-terminal tail increases enzymatic activity\(^2\). On the other hand, glutathionylation reduces the enzymatic activity of rat protein kinase C-\(\alpha\) (PKC-\(\alpha\))\(^2\), mouse cyclic AMP-dependent protein kinase (cAPK) on C199\(^2\), rat MAPK (mitogen-activated protein kinase)/ERK (extracellular-signal-regulated kinase) kinase 1 (MEKK1) on C1238\(^2\), mouse I\(\kappa\)B kinase subunit \(\beta\) (IKK-\(\beta\)) on C179\(^2\), rat Ca\(^2+\)/calmodulin-dependent kinase I (CaMKI) on C179\(^2\), and mouse p90-ribosomal S6 kinase 1 (p90-RSK1) on C223\(^2\). Glutathionylation of these and other protein kinases may serve to control their activity in response to oxidative stress.

\(^1\)This chapter is adapted from: Moffett, AS, Bender KW, Huber, SC, and Shukla, D. *Biophys. J.* 113, 2354-2363 (2017).
Figure 5.1: BAK1 S-glutathionylation. a) Structural and functional features of a protein kinase. b) Structure of glutathione in a mixed disulfide with a cysteine residue. Glutathione consists of a glutamate connected through a γ-peptide bond to a cysteine-glycine dipeptide. BAK1 structures taken from simulation with c) C353, d) C374, and e) C408 glutathionylated. ATP is shown in the ATP binding pocket for all three systems.

Recently, *in vitro* experiments demonstrated that the *Arabidopsis thaliana* leucine-rich repeat receptor-like kinase (LRR-RLK) BRASSINOSTEROID INSENSITIVE 1 (BRI1)-ASSOCIATED RECEPTOR-LIKE KINASE 1 (BAK1) interacts with the glutaredoxin GRXC2, catalyzing glutathionylation of BAK1 cysteine residues and reducing BAK1 phosphotransferase activity[217]. Three potential glutathionylation sites were identified, namely C353, C374, and C408 (Figure 5.1c-e), while single and double mutations of these residues to serine suggested that C353 and C408 are the dominant sites. However, the mechanistic details of how glutathionylation far from the active site is able to reduce BAK1 activity remain unclear.

The structural characteristics of a catalytically competent kinase have been studied at length and include formation of a salt bridge between a conserved glutamate-lysine pair (K-E salt bridge; E334 and K317 in BAK1), a folded αC helix, an unfolded activation loop, an inward-flipped phenylalanine in the DFG motif, and an ATP molecule bound to the ATP binding site[145, 104, 8] (Figure 5.1a). By monitoring these structural features, we can roughly map the observed protein kinase conformations to qualitative estimates
of phosphotransferase activity. For example, if the αC helix is completely unfolded and the KE salt bridge is broken, the conformation is likely inactive. By comparison, if another conformation exhibits a folded αC helix and a formed KE salt bridge in conjunction with a properly positioned DFG motif and activation loop, it is likely catalytically competent. We will refer to these structural features of kinases as activation-relevant collective variables. In our previous work on the BAK1 core kinase domain, we found that the αC helix displays considerable disorder \[106\]. This disorder could represent a regulatory mechanism where association of BAK1 with another LRR-RLK promotes folding of the αC helix in a similar manner to the metazoan epidermal growth factor receptor \[86, 54, 106\].

To determine the molecular mechanisms of BAK1 inhibition by glutathionylation and to gain insight into the site-specific function of BAK1 glutathionylation, we performed extensive molecular dynamics (MD) simulations on the BAK1 core kinase domain, singly glutathionylated on C353, C374, and C408 (BAK1-C353SG, BAK1-C374SG, and BAK1-C408SG respectively). Additionally, we used simulations of the non-glutathionylated BAK1 core kinase domain (BAK1-SH) from our previous work for comparison \[106\]. Using
the Kullback-Leibler (KL) divergence, a measure of similarity between probability distributions borrowed from information theory, we quantified the global effects of glutathionylation on the BAK1 conformational ensemble and found little change upon C353 or C374 glutathionylation in contrast to the large shifts in BAK1-C408SG behavior from the non-glutathionylated kinase. We found little change in the free energy landscapes of BAK1-C353SG or BAK1-C374SG and observed that S-glutathionylation of C408 eliminates the active-like state of BAK1 and stabilizes an outward-swung αC helix. As two of the three discussed BAK1 cysteine residues appear to be conserved to a certain extent in the A. thaliana kinome, glutathionylation may function similarly in regulation of other protein kinases during shifts in cellular redox state.

5.2 Results

S-glutathionylation of C408 alters the BAK1 conformational ensemble.

In order to assess the global effects of glutathionylation on the conformational ensemble of BAK1, we estimated discretized probability distributions for the $\phi$, $\psi$, and $\chi_1$ dihedral angles of all residues for BAK1-SH and singly glutathionylated forms, and calculated the local KL divergence between each of the glutathionylated BAK1 ensembles and the non-glutathionylated ensemble. We chose the $\phi$, $\psi$, and $\chi_1$ dihedral angles in order to allow comparison between residues with different-sized side chains while still capturing some side chain motion, although we note that glycine and alanine lack a $\chi_1$ angle. A high KL divergence for a given residue is an indication that glutathionylation has significantly altered the effective residue potential energy landscape while a low or vanishing value indicated that glutathionylation has had little or no effect on the residue.
Considering relative differences in KL divergences between residues within the same ensemble, BAK1-C353SG and BAK1-C374SG are clearly perturbed most near the site of glutathionylation (Figure 5.3a-b). In this intra-ensemble view, distant allosteric effects are also visible for both glutathionylation sites, although the KL divergence values are generally an order of magnitude smaller than those for BAK1-C408SG (Figure 5.3d). Interestingly, the KL divergence at C408 for BAK1-C408SG is fairly small relative to most other residues in the ensemble, indicating the role of C408 as a glutathionylation site with primarily long-range effects. The residues most perturbed by S-glutathionylation of C408 are near the N-terminus of the kinase domain, at the back of the ATP-binding pocket (with the opening of the ATP-binding pocket facing the viewer, as in Figure 5.3a-c), and in the catalytic loop (Figure 5.3c). However, BAK1-C408SG KL divergences from the un-glutathionylated ensemble were statistically significant for each residue, indicating a global effect of C408 glutathionylation on the structure of BAK1. These results are supported by the large change in residue fluctuations from BAK1-SH to BAK1-C408SG as compared to the changes upon C353 and C374.
glutathionylation (Supplementary Figure 5.5).

**S-glutathionylation of C408 promotes an inactive conformational state.**

Our simulations were conducted with ATP and Mg\(^{2+}\) bound, precluding a DFG flip, and a fully phosphorylated activation loop; consequently, the activation-relevant collective variables most likely to shift to values corresponding to an inactive kinase are the proper positioning and folding of the αC helix and the presence of the K-E salt bridge. To look for changes in these structural features we estimated two-dimensional free energy landscapes over the space of the α-helical content of the αC helix region and the inverse of the distance between the δ-carbon of E334 and the side chain nitrogen of K317. Inverse distance was chosen to better resolve small inter-atomic distances. In order to reduce bias introduced by initiating simulations from a non-Boltzmann distribution, Markov state models (MSMs) were constructed for each system and the free energy landscapes were weighted by state equilibrium populations. As shown in previous work [106], BAK1-SH exhibits a large degree of structural heterogeneity in these coordinates, with a stable active-like state, (Figure 5.4a, region I), a state with a formed αC helix, but a broken K-E salt bridge (Figure 5.4a, region III), and a number of metastable regions displaying high levels of αC helix disorder (Figure 5.4a, regions II, IV, and V).

The BAK1-C353SG free energy landscape retains most features of the BAK1-SH landscape, with the notable difference that S-glutathionylation of C353 increases the free energy of the active-like state by \(\sim 1\) kcal-mol\(^{-1}\) (Figure 5.4b, region I), although this is unlikely to represent a significant change. The BAK1-C374SG free energy surface is largely unchanged from the BAK1-SH surface, though it displays a reduced range of distances between K317 and E334 when the αC helix is folded (Figure 5.4b\&c, region III).
Figure 5.4: Effects of S-glutathionylation on prevalence of the active-like state in BAK1. MSM-weighted free energy plots projected onto the αC helix helical content and the inverse K317-E334 distance from simulations of a) non-glutathionylated BAK1, and BAK1 S-glutathionylated on b) C353, c) C374, and d) C408. Points on the plane with representative structures shown are demarcated by triangles labeled in black and correspond to: I) BAK1-SH; active-like state, II) BAK1-SH; formed K-E salt bridge and unfolded αC helix, III) BAK1-C353SG; slightly broken K-E salt bridge and formed αC helix, IV) BAK1-C408SG; broken K-E salt bridge and partially unfolded αC helix, V) BAK1-C408SG: broken K-E salt bridge and unfolded αC helix.

In contrast, there are several clear changes between the BAK1-C408SG and BAK1-SH free energy landscapes, most notably the near absence of both the active-like state (Figure 5.4d, region I) and the state characterized by a slightly broken K-E salt bridge (Figure 5.4d, region III) found in the BAK1-SH free energy surface. Instead, an unfolded αC helix and a broken K-E salt bridge are favored (Figure 5.4d, region V). An additional metastable state is greatly stabilized in BAK1-C408SG (Figure 5.4d, region IV) in regions with moderate αC helix helical content (≈0.65) and moderate K317-E334 inverse distances (≈0.8 nm⁻¹). The only viable pathway for breaking of the K-E salt bridge is through a highly unfolded αC helix (to the left of region II in Figure 5.4d), suggesting that states with both a moderately folded αC helix and a broken K-E salt bridge occur through unfolding and refolding of the αC helix from the active-like state rather than a simple swinging motion.
Figure 5.5: Glutathione interaction partners. Normalized contact frequencies between GSH heavy atoms and heavy atoms of all BAK1 residues with a cutoff of distance of 4 Å. More explicitly: \( f = \frac{\min_{x \in G, y \in R} d(x, y)}{n} \) where \( f \) denotes the normalized frequency over all trajectory frames, \( d(x, y) \) is the Euclidean distance between the two atoms with positions \( x \) and \( y \), \( G \) is the set of all heavy glutathione atoms, and \( R \) is the set of all heavy atoms in a given BAK1 residue. In each simulation set, the cysteine forming a disulfide bond with GSH has a contact frequency of 1.0 and is marked with a dotted vertical line. Secondary structure from the crystal structure is shown just below the frequencies [32, 224].

In order to investigate the mechanism of the structural disruption caused by C408 glutathionylation, we calculated the frequencies for each simulation set at which any heavy atom in GSH came within 4 Å of a heavy atom of each BAK1 residue (Figure 5.5). While spatial proximity between two residues is not necessarily caused by strong, direct physical interaction between them, we use it here as a simple proxy for interaction. C353SG interacts exclusively with the \( \alpha_N \) helix and \( \beta \)-sheets of the N-lobe primarily at residues 273 to 290 and 350 to 365. C374SG interacts with residues around the ATP binding pocket at residues 366 to 386 and 423 to 429 as well as R297 in the glycine-rich loop to a lesser extent. C408SG interacts most strongly with its flanking residues near the end of the \( \alpha_E \) helix around residues 405 to 410, the \( \alpha_C \) helix at residues 328 to 347, and to a lesser extent with L440 in the activation loop. It is worth noting that GSH at all three sites largely does not directly interact with both the catalytic loop (residues 416 to 422), containing the residues directly involved in catalysis, and the activation loop (residues 433 to 459), the regulatory sequence in kinases usually responsible for activation by phosphorylation [145].

Cysteine thiol pK\(_a\) predictions suggest low reactivity in isolation.

Although our simulations suggest that glutathionylation of C408 has strong allosteric effects on BAK1, it remains unclear which cysteine residues are likely to be reactive. In order to address this question, we
estimated pK\textsubscript{a} values for randomly selected structures in each MSM state using PropKa 1.3\cite{219,220} and created a pK\textsubscript{a} distribution estimate by way of a MSM-population-weighted sum of state distributions (Figure 5.6). We took this approach due to the fact that distinct local environments surrounding ionizable residues are known to have a strong effect on side chain pK\textsubscript{a} values\cite{225,226,227}.

All estimated mean pK\textsubscript{a} values are well above values considered reactive at physiological pH, with C353 predicted to be the most reactive (pK\textsubscript{a} 9.41), followed by C408 (pK\textsubscript{a} 10.23) and C374 (pK\textsubscript{a} 10.47) (Figure 5.6). As expected due to the fact that it is buried in the hydrophobic core of the kinase, C545 is predicted to be almost exclusively protonated (pK\textsubscript{a} 12.25). While all other cysteine residues have sharply peaked pK\textsubscript{a} distributions, C353 has a broad, nearly bimodal distribution, suggesting a highly variable local environment (Figure 5.6).

**Figure 5.6: Predicted distributions of BAK1 cysteine thiol pK\textsubscript{a}s.** We calculated the pK\textsubscript{a} of each BAK1 cysteine thiol using PropKa 3.1\cite{219,220} for 10 randomly chosen structures for each MSM state from the non-glutathionylated simulations of BAK1. We created kernel density estimates (KDEs) using the Scipy package\cite{189} for each state with the pK\textsubscript{a}s calculated for the corresponding 10 structures. The final plots are sums of the KDEs of each state weighted by MSM equilibrium populations. Estimated means of the MSM weighted KDEs are as follows, C353: 9.41, C374: 10.47, C408: 10.23, C545: 12.25.

**BAK1 C353 and C545 are well conserved in the A. thaliana kinome.**

Using a multiple sequence alignment of all known A. thaliana kinase sequences, we calculated the frequency at which cysteine appears at each alignment position. Cysteine is well conserved at the alignment positions of BAK1 C353 and C545 (frequencies of 0.49 and 0.57 respectively), while C408 is less conserved (0.19), and C374 is poorly conserved (0.01) although a nearby alignment position which could be a functional equivalent in other kinases has a cysteine frequency of 0.10 (Figure 5.7). Two other cysteine frequency peaks (0.15 and
0.16) further to the C-terminal end of the alignment are beyond the C-terminus of BAK1 in the alignment.

5.3 Discussion

We have demonstrated the dramatic effects that S-glutathionylation can have on the structure of a kinase, providing a plausible explanation for the effects of glutathionylation on the activity of BAK1 observed in previous experiments\[217\]. Furthermore, although S-glutathionylation affects the activity of numerous other protein kinases\[209, 210, 211, 212, 213, 214, 215, 216\] (Figure 5.2), to our knowledge this is the first study addressing the molecular mechanisms of protein kinase activity modulation through glutathionylation. Our findings provide evidence that S-glutathionylation can allosterically alter the activity of protein kinases.

Our simulations suggest that glutathionylation of BAK1 C408 may be responsible for the concentration-dependent decrease in activity in the presence of GSSG and GRXC2 \[217\]. However, we only address the effects of glutathionylation on the dynamics of the isolated kinase domain and cannot exclude other consequences of glutathionylation relating to interactions with other proteins or other processes important for BAK1 activity not considered here. Our simulations are consistent with experiments supporting C408 and C353 as the most likely sites for GRXC2-catalyzed S-glutathionylation\[217\]. Global changes in residue conformations as measured by the local KL divergence between each glutathionylated BAK1 simulation set and the BAK1-SH simulations are generally an order of magnitude larger for BAK1-C408SG than BAK1-C353SG or BAK1-C374SG (Figure 5.3). Surprisingly, most residues in BAK1-C408SG have a high KL divergence with respect to BAK1-SH, including distant residues in the C-lobe even though GSH only directly interacts with its flanking residues and the \( \alpha \)C helix in the N-lobe.
Using the 940 *A. thaliana* kinases identified in [149], we created a multiple sequence alignment using Clustal Omega [182]. Using the region of the alignment corresponding to the BAK1 kinase domain, we calculated the frequencies of cysteine at each alignment position. All four BAK1 cysteine residues are in regions of high cysteine frequency, most notably C353 and C545. However, BAK1 C374 and C408 have nearby alignment positions with high frequency, likely corresponding to comparable positions in the contributing kinases.

Differences in free energy surfaces calculated over a measure of αC helix helicity and the inverse distance between K317 and E334 for the four BAK1 constructs were consistent with calculated KL divergence values. The BAK1-C353SG and BAK1-C374SG free energy surfaces are similar to that for BAK1-SH, although subtle differences, including an increase in the free energy of the active-like state in BAK1-C353SG, are noticeable. The level of disorder in the αC helices of BAK1-C353SG and BAK1-C374SG is largely unchanged from BAK1-SH. In contrast, the BAK1-C408SG free energy landscape differs from BAK1-SH in several key ways. C408SG interacts directly with the αC helix, a region known to be involved in regulation of protein kinase activity [145, 147, 228, 229, 159, 230], causing the αC helix to more strongly favor both unfolded and folded but distally-swung conformations. The distally-swung conformation is similar to a well-defined, metastable inactive state found in human Src kinase where the αC helix also swings away from the active site [147, 79]. In BAK1, this likely occurs due to direct interactions with GSH stabilizing the outward-swung conformation. Unlike the proposed mechanisms of formation of a Src-like inactive state in other kinases [229, 228, 159], the αC helix in BAK1-C408SG must first unfold, followed by breakage of the K-E salt bridge and refolding of the αC helix. Furthermore, the active-like state (Figure 5.4a-e, region 1) is nearly eliminated in BAK1-C408SG simulations.

It should be noted that sampling of the BAK1-C408SG system was 1.5 times longer than any of the other systems on top of more extensive accelerated MD (AMD) to generate initial structures for unbiased
Thus it is possible that the differences between BAK1-C408SG and the other systems are (unintentionally) by construction due to the differences in sampling. However, this does not appear to be the case for several reasons. First, even with the most extensive sampling, BAK1-C408SG exhibits no breaking of the K-E salt bridge coincident with a folded αC helix, a transition well sampled by all other systems (Figure 5.4). Second, BAK1-C408SG samples the active-like state far less than the other systems, which seems unlikely to be due to oversampling BAK1-C408SG with respect to the other systems. Finally, there are not high free energy barriers impeding access to new states appearing in BAK1-C408SG, and the fringes of the areas spanned by these states are visited in other systems. If the differences in the free energy landscapes of BAK1-C408SG and the other systems were largely due to differences in sampling, we would expect BAK1-C408SG to have explored new areas not accessible to more poorly sampled systems.

Collectively, these results suggest that modification of C408 by S-glutathionylation allosterically deactivates BAK1 through interaction with the αC helix. The αC helix is in generally believed to be a focal point for protein kinase regulation, physically linking numerous regions important for kinase function[230]. This central positioning makes the αC helix a prime target for allosteric modulators and possibly explains the global effects of C408 glutathionylation shown in KL divergence measurements. There is evidence for the specific importance of the αC helix in regulation of BAK1[106], which is further supported by the apparent role of the αC helix in response to S-glutathionylation. While the predicted C408 pKₐ (10.23) indicates that it is overwhelmingly protonated and therefore marginally reactive, C408 is glutathionylated in vitro[217]; this inconsistency is at least partially explained by the fact that PropKa 3.1 systematically overestimates cysteine pKₐ values[231]. Interestingly, C353 is moderately conserved in the A. thaliana kinome, present in ~50% of the 940 known sequences, while C408 is present in ~20% (Figure 5.7). It is possible that the conservation of C353 is due to its function as a glutathionylation site, working to regulate kinase activity in some way not captured by our simulations, although there is currently no evidence to support this claim. The comparatively low level of conservation in C408 indicates that it is not universally essential for function of A. thaliana protein kinases. However, the functional importance of C408 for BAK1 is supported by the decrease in catalytic activity caused by mutation of C408 to tyrosine[232].

5.4 Methods

MD simulations

All simulations were performed using the AMBER 14 molecular dynamics package[124] and the CHARMM 36 force field[164][165][166] on the Blue Waters petascale computing facility. All simulation systems were set
up using the VMD\cite{72} plugin Psfgen 1.6 and converted to Amber format using the CHAMBER tool\cite{167}.

All simulations of BAK1-SH were taken from our previous work \cite{106}, which we describe in the following two paragraphs. Starting coordinates for the BAK1-SH kinase domain were taken from available crystal structures (PDB IDs: 3TL8 (chains A, D, G, and H)\cite{32}, 3UIM\cite{33}, and 3ULZ\cite{168}). Starting structures were solvated in water boxes with dimensions of approximately 90 Å X 70 Å X 63 Å with TIP3P model molecules\cite{126}. Sodium and chloride ions were added to neutralize the charge of all systems and bring salt concentration to approximately 150 mM. An ATP molecule with two magnesium ions bound, taken from previous simulations, was inserted into the binding pocket for all structures in place of the modified adenosine-phosphate molecules used in crystallization, aligned with the adenosine ring of the corresponding ATP analogue. All systems were subjected to 10,000 steps of energy minimization and were equilibrated for 8-10 ns in an NPT ensemble at 300K and 1 atm, maintained using Langevin dynamics and a Berendsen barostat. Simulations were performed using a 2 fs time step, periodic boundary conditions, particle mesh Ewald electrostatics,\cite{171} and constraints of hydrogen-containing bonds using the SHAKE algorithm\cite{172, 173}. Equilibrated structures were then equilibrated for another 10 ns to obtain average dihedral angle potential energies for calculation of AMD\cite{174} parameters (Supplementary Table 5.1) according to\cite{175}.

In order to generate a diverse set of starting structures for unbiased MD simulation, we performed AMD sampling for BAK1-SH, initiating 25 independent simulations from the result of equilibrating each crystal structure for an aggregate $\sim$8.6 $\mu$s (Supplementary Table 5.2). The final round of AMD sampling was clustered using the k-means algorithm in the space of the distance between the most distal side chain nitrogen in the lysine and carbon in glutamate within the conserved K-E salt bridge and $\alpha$C helix RMSD from the crystal structures into 100 states, and the nearest neighbors of the cluster centroids were chosen as starting structures for unbiased simulation. These structures were used to initiate 100 independent unbiased simulations for an aggregate $\sim$30 $\mu$s (Supplementary Table 5.2).

The BAK1-SH structures produced by AMD were similarly used as starting structures for simulations of BAK1-C353SG and BAK1-C374SG. To prepare the initial BAK1-C353SG and BAK1-C374SG structures from BAK1-SH AMD conformations, we took the same 100 cluster centroid nearest neighbors and added a randomly rotated GSH to the appropriate cysteine, and then performed energy minimization and equilibration steps before beginning production runs in the same manner as for BAK1-SH. Total BAK1-353SG and BAK1-374SG unbiased simulation times were $\sim$30 $\mu$s and $\sim$26.8 $\mu$s, respectively (Supplementary Table 5.2).

For BAK1-C408SG, separate AMD sampling was done due to the possibility of high energy structures resulting from adding GSH to C408 in BAK1-SH AMD conformations. Eight rotational conformations of GSH attached to C408 of a single BAK1 crystal structure (PDB ID: 3TL8, chain A)\cite{32}, were initially set up,
minimized, and equilibrated in the same manner as described for BAK1-SH. The same procedure used for generating BAK1-SH structures to initiate unbiased sampling was used for three rounds of AMD sampling (1.9 \(\mu\)s starting from 8 structures, 5.9 \(\mu\)s starting from 100 structures, and finally 6.0 \(\mu\)s starting from 100 structures) (Supplementary Table 5.2). After each round, the trajectories were clustered using the same metrics used for BAK1-SH and starting structures for the subsequent sampling round were generated using cluster centroid nearest neighbors. Finally, 150 cluster centroid nearest neighbors were chosen from the final round of AMD sampling and using to run 150 independent unbiased MD simulations for an aggregate 45 \(\mu\)s (Supplementary Table 5.2).

**Markov state model construction**

All trajectory analysis was done using MDTraj 1.7[131] except where otherwise noted while clustering and MSM construction was done using MSMBuilder 3.6[180]. We aligned each frame of every trajectory to a BAK1 crystal structure (PDB ID: 3TL8, chain A) and calculated the root mean-squared deviation (RMSD) of the N-terminal lobe and C-terminal lobe separately with respect to the crystal structure. We chose a lag time for each model by plotting implied time scales of MSMs built with increasing lag times and choosing the lag time at which the time scales began to plateau in order to improve the validity of the Markov assumption for our models (Figures Supplementary Figure 5.1-Supplementary Figure 5.4). We then used the Osprey variational cross-validation package to choose the number of clusters, ranging from 100 to 500, that maximized the mean cross-validation score of the MSM generalized matrix Raleigh quotient calculated from 5 equal partitions of the data into training and test sets[181]. The chosen lag time and number of clusters for each MSM are shown in Supplementary Table 5.3.

**Free energy landscape construction**

The free energy landscapes for each of the four systems were constructed by building normalized two-dimensional histograms from the conformations within each state of the corresponding MSM in the space of the inverse distance between the glutamate \(\delta\)-carbon atom and the lysine side-chain nitrogen in the KE salt bridge and the \(\alpha\) helical content of the \(\alpha\)C helix. The \(\alpha\) helical content was calculated according to the NAMD 2.11 collective variable “alpha” [177], as detailed in [106]. State histograms \(h_i(x, y)\) for state \(i\) were weighed by the MSM equilibrium probability \((\pi_i)\) of the corresponding state to estimate calculate MSM-weighted probabilities, which were then used to estimate free energy as follows:

\[
F(x, y) = -RT \log \left[ \sum_{i=1}^{N} \pi_i h_i(x, y) \right]
\]  

(5.1)
where $R$ is the gas constant and $T$ is the temperature. For each system, the lowest free energy was used as a reference, and set to zero.

**Local Kullback-Leibler divergence**

We used the local KL divergence measure introduced in [233], defined as follows for residue $i$:

$$K L_i = \sum_{d \in D_i} \sum_{X=1}^{N} p_d(X) \ln \frac{p_d(X)}{p_d^*(X)}$$  \hspace{1cm} (5.2)

This is equivalent to the sum of relative entropies between the reference ($p_d^*(X)$) and test ($p_d(X)$) probability distributions discretized into $N$ bins (where $\sum_{X=1}^{N} p_d(X) = 1$) over the $\phi$, $\psi$, and $\chi_1$ dihedral angles of residue $i$ ($D_i = \{\phi_i, \psi_i\}$ for glycine and alanine residues and $D_i = \{\phi_i, \psi_i, \chi_1 i\}$ for all others). Following [233], we calculated a bootstrap distribution of KL divergences for each system in order to test the null hypothesis that the KL divergence between designated test and reference ensembles is no greater than expected from variability within the reference ensemble. This is accomplished using the bootstrap KL divergence:

$$K L_{i,b}^B = \sum_{d \in D_i} \sum_{X=1}^{N} p_d^B(X) \ln \frac{p_d^B(X)}{p_d^B(X)}$$  \hspace{1cm} (5.3)

and the mean bootstrap KL divergence

$$\overline{K L}_i^B = \left(\frac{2k}{k}\right)^{-1} \sum_{b=1}^{2k} K L_{i,b}^B$$  \hspace{1cm} (5.4)

where the set of reference trajectories have been split up into $2k$ blocks ($k \in \mathbb{N}$) and $\overline{K L}_i^B$ is obtained by taking the average local KL divergence for residue $i$ measured between a set of half the blocks (with an index $b$) of the reference trajectories and the set of remaining blocks ($b^R$, the complement of the set of blocks $b$) over all combinations of blocks. We would like to note that our definition is a slight departure from the original as we have made $K L_{i,b}^B$ a true KL divergence between the sets of reference ensemble blocks (compare equations 5.3 & 5.4 with equation 26 in [233]). We can then obtain the probability, $p$, of obtaining a value greater than $K L_i$, calculated between ensembles, from the $K L_{i,b}^B$ values calculated within the reference ensemble, and reject the null hypothesis if $p < \alpha$ for some choice of $\alpha \in (0, 1)$ (in this study, we used $\alpha = 0.05$). If we are unable to reject the null hypothesis we set $\hat{K} L_i = 0$, where $\hat{K} L_i$ is the corrected KL divergence for
residue $i$. Otherwise, we reject the null hypothesis and set $\hat{KL}_i = KL_i - KL_i^B$. Additionally, we weighed each frame by the equilibrium MSM population of the corresponding state in creating the one-dimensional histograms in order to reduce bias introduced by initiating trajectories from non-Boltzmann distributions.

We have implemented the local KL divergence in Python and validated the code on a toy model (Code freely available at: https://github.com/ShuklaGroup/kl_divergence). In order to visualize population shifts in the dihedral angles of BAK1 with glutathionylation, we used the B-factor putty function in Open Source PyMOL 1.83 to show the local KL divergence for each residue (Figure 5.3a-c).

**pK$_a$ calculations**

We used PropKa 3.1 to predict the pK$_a$ of cysteine sulfur atoms in structures from simulation of BAK1 with no glutathionylation. For each BAK1 MSM state, we randomly chose 10 structures and calculated the pK$_a$ for the three solvent exposed cysteine residues (C353, C374, and C408) from each structure. The overall cysteine pK$_a$s were calculated by taking the average pK$_a$s within each state, multiplying by the MSM equilibrium populations and then creating kernel density estimates using the Scipy Python package.

**Bioinformatics analysis**

We created a multiple sequence alignment of the 940 known A. thaliana kinases using Clustal Omega. Using only the section of the alignment corresponding to the BAK1 kinase domain, we calculated frequencies of cysteine residues at each alignment position in order to gain a qualitative insight into how general the mechanisms of allostery through S-glutathionylation that we investigated in BAK1 are to the entire kinome. We expect that if the cysteine positions in BAK1 are shared with another kinase, it is more likely that the effects of S-glutathionylation on those cysteine residues will also be shared than if the cysteine positions are disparate.
5.5 Supplementary Information

**Supplementary Figure 5.1:** Scaling of Markov state model timescales for BAK1-SH. The 5 largest implied timescales for Markov state models built on molecular dynamics simulations of BAK1-SH with a range of number of clusters and lag times. The final model lag time was chosen such that the implied timescales as a function of lag time converge for all numbers of clusters.

**Supplementary Figure 5.2:** Scaling of Markov state model timescales for BAK1-C353SG. The 5 largest implied timescales for Markov state models built on molecular dynamics simulations of BAK1-C353SG with a range of number of clusters and lag times. The final model lag time was chosen such that the implied timescales as a function of lag time converge for all numbers of clusters.
Supplementary Figure 5.3: Scaling of Markov state model timescales for BAK1-C374SG. The 5 largest implied timescales for Markov state models built on molecular dynamics simulations of BAK1-C374SG with a range of number of clusters and lag times. The final model lag time was chosen such that the implied timescales as a function of lag time converge for all numbers of clusters.

Supplementary Figure 5.4: Scaling of Markov state model timescales for BAK1-C408SG. The 5 largest implied timescales for Markov state models built on molecular dynamics simulations of BAK1-C408SG with a range of number of clusters and lag times. The final model lag time was chosen such that the implied timescales as a function of lag time converge for all numbers of clusters.
Supplementary Figure 5.5: Allosteric effects of S-glutathionylation on the BAK1 ensemble. For each residue, all heavy atoms in each frame were aligned to the corresponding atoms in the crystal structure (PDBID: 3TL8, chain A) and the RMSD was measured, squared, and multiplied by the number of atoms in the RMSD measurement to give the Euclidean distance in residue-heavy-atom conformational space. The mean value of this squared distance is reported for each residue with the mean value for the corresponding residue in the non-glutathionylated simulation set subtracted.

Supplementary Table 5.1: AMD parameters for all systems.

<table>
<thead>
<tr>
<th>Name</th>
<th>$\alpha_D$</th>
<th>$E_{Threshold,D}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3TL8, A (BAK1-SH)</td>
<td>241</td>
<td>4123</td>
</tr>
<tr>
<td>3TL8, D (BAK1-SH)</td>
<td>241</td>
<td>4186</td>
</tr>
<tr>
<td>3TL8, G (BAK1-SH)</td>
<td>243</td>
<td>4154</td>
</tr>
<tr>
<td>3TL8, H (BAK1-SH)</td>
<td>248</td>
<td>4180</td>
</tr>
<tr>
<td>3UIM (BAK1-SH)</td>
<td>244</td>
<td>4114</td>
</tr>
<tr>
<td>3ULZ (BAK1-SH)</td>
<td>244</td>
<td>4117</td>
</tr>
<tr>
<td>3TL8, A (BAK1-C408SG)</td>
<td>242</td>
<td>4092</td>
</tr>
</tbody>
</table>
**Supplementary Table 5.2**: Lengths of simulation sets.

<table>
<thead>
<tr>
<th>Name</th>
<th>Adaptive AMD time (µs)</th>
<th>Serial AMD time (µs)</th>
<th>Unbiased MD time (µs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAK1</td>
<td>0.0</td>
<td>8.6</td>
<td>29.9</td>
</tr>
<tr>
<td>BAK1-C353-GSH</td>
<td>0.0</td>
<td>0.0</td>
<td>30.0</td>
</tr>
<tr>
<td>BAK1-C374-GSH</td>
<td>0.0</td>
<td>0.0</td>
<td>26.8</td>
</tr>
<tr>
<td>BAK1-C408-GSH</td>
<td>13.8</td>
<td>0.0</td>
<td>45.0</td>
</tr>
</tbody>
</table>

**Supplementary Table 5.3**: Parameters for MSM construction.

<table>
<thead>
<tr>
<th>Name</th>
<th>Lag time (ns)</th>
<th>Number of clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAK1</td>
<td>40</td>
<td>242</td>
</tr>
<tr>
<td>BAK1-C353-GSH</td>
<td>40</td>
<td>351</td>
</tr>
<tr>
<td>BAK1-C374-GSH</td>
<td>60</td>
<td>183</td>
</tr>
<tr>
<td>BAK1-C408-GSH</td>
<td>40</td>
<td>104</td>
</tr>
</tbody>
</table>
Chapter 6

A short outlook on the biophysics of brassinosteroid signaling

The work that I have presented in this thesis barely scratches the surface of all that is to be learned about the biophysics of brassinosteroid signaling. Construction of a structural model of the full BRI1 and BAK1 complex is an ambitious but attainable goal, and would be a useful tool for answering a number of fundamental questions about early events in brassinosteroid signaling. The general strategy and vision laid out below is strongly influenced by computational work on the human epidermal growth factor receptor [78, 235, 236, 237, 238, 239, 240, 241], which forms a homodimer upon epidermal growth factor binding.

The physical details of BRI1-BAK1 association remain largely unclear. The BRI1-BAK1 kinase domain complex or complexes are currently uncharacterized, though a mixture of experimental evidence and computational modeling holds promise for gaining insight into how the kinase domains interact. Knowledge of the kinase domain complex should help address a number of questions, including the role of phosphorylation in stabilizing a BRI1-BAK1 complex and the role of C-terminal tails and αC helix disorder in regulation of BRI1 and BAK1 kinase activity [106]. The fact that BRI1 and BAK1 can transphosphorylate one another helps narrow down a set of likely complexes where the active site of one kinase domain is close to a phosphorylation site of the other. The role of BRI1 and BAK1 transmembrane helices in the overall association process has been mentioned as another glaringly unknown component of brassinosteroid signal transduction [34], and careful consideration of the transmembrane helix association process is needed.

Once there exists a model for each section of the BRI1-BAK1 complex, simulations of the entire complex in a lipid bilayer should provide some insight into exactly how brassinosteroid binding activates the kinase domains of both proteins. A complete model of the BRI1-BAK1 complex could be used to reconcile experimental evidence for both apo BRI1-BAK1 pre-association and brassinolide-induced association and activation. In particular, a coarse-grained model derived from carefully modeled all-atom components of the BRI1-BAK1 complex could provide a picture of brassinolide-independent BRI1-BAK1 interactions, as coarse-grained modeling allows for simulation of much longer time scales than atomistic simulations do. Modeling the atomistic complex will likely be a challenge in and of itself, and protein-protein docking methods, together with biased molecular dynamics simulations and experimental data concerning the effects of...
mutations within predicted interfaces on association will be needed for efficient structure determination.

Another interesting question concerns understanding how BAK1 discriminates between its many co-receptors. The extracellular domain of BAK1 has been shown to interact with a number of other LRR-RLKs both in vitro [19] and in vivo [163]. It is surprising then that BAK1 can associate in a ligand-dependent manner with LRR-RLKs detecting a diverse set of ligands. Though compartmentalization within the membrane likely at least partly answers this question [57], it appears to be an interesting avenue for further investigation. Another question addressing LRR-RLKs more generally concerns the role of conformational changes in large LRR-RLK extracellular domains. Given the conformational change in BRI1 described in Chapter 2, it would be interesting to study the conformational dynamics of LRR-RLKs with known extracellular domain structures and examine how the free energy landscapes describing large-scale motion are altered by ligand binding and association of a smaller LRR-RLK. To our knowledge, our work is the first to focus on LRR-RLK extracellular domain conformational dynamics using molecular dynamics simulations, and with over 200 LRR-RLKs in A. thaliana [242], there are a large number of LRR-RLKs with interesting mechanisms of activation [7, 8] and with functions highly relevant to agriculture to be studied.

Perhaps less exciting but just as important is the job of further testing the results I have produced. I have received a healthy dose of skepticism concerning the disorder appearing in my simulations of the BAK1 kinase domain appearing in Chapter 3. Several people suggested that my choice of force field or starting structure may have produced the αC helix disorder as an artifact. Both of these explanations are certainly possible, while the question of the biological relevance of BAK1 T324 phosphorylation, just next to the αC helix, should be addressed. Biased molecular dynamics simulations using several force fields could be used to estimate the free energy of BAK1 αC helix unfolding and how it is affected by force field choice, while experimental evidence of αC helix disorder in the context of the whole kinase domain is needed. Another example of future work needed to verify my results could be biased sampling of the BRI1 extracellular domain with and without BAK1 association with the purpose of determining the stability of the secondary BRI1-BAK1 extracellular domain interface described in Chapter 2. The amount of sampling performed on the whole BRI1-BAK1-BL complex is entirely inadequate to make any certain statements about the stability of the secondary interface, and we cannot make any stronger conclusions beyond the ones we have made, given our current data. As always, the relevance of a secondary BRI1-BAK1 extracellular domain interface must be tested experimentally, where mutating key residues on the interface should decrease BRI1-BAK1 association free energy if the secondary interface really does stabilize the complex.

All of my work presented here has been on proteins from the model plant Arabidopsis thaliana. I emphasize that my results are on a model plant which is not directly relevant to agriculture, though I do
not believe this is in any way a weakness. Biology, and science in general, has been driven forward largely through the use of model systems that sacrifice detailed agreement with systems of more interest for ease of study. In my case, I had little choice but to use A. thaliana proteins, as crystal structures of BRI1 and BAK1 are exclusively from this plant. Integrating methods from computational biophysics into plant structural biology work will undoubtedly facilitate production of a more representative set of plant protein structures in the future, a goal of critical importance to the success of rational crop engineering.

As mentioned several times in the previous paragraph, all computational work on brassinosteroid signaling must be compared with and informed by experimental results, while simultaneously informing subsequent experimental work. Any mathematical model will always sacrifice some level of detail in order to facilitate proper analysis, and physics-based biomolecular simulations are no exception. Even with simplifying assumptions applied, it is impossible to tell with certainty whether most observables one would want to estimate from simulations, often corresponding to functions of conformational space integrals, have converged. Any biomolecular simulation can be used to produce experimentally testable predictions, whether the prediction is qualitative or quantitative in nature. Inversely, computational methods can provide nanoscopic interpretations of experimental results difficult to obtain through other experimental methods. While computational biophysics has not been widely applied to problems in plant molecular biology, I expect that in the future the combined efforts of structural biologists and computational biophysicists will prove to be highly fruitful.
References


Inkscape 0.91; http://inkscape.org.


Oh MH, Clouse SD, Huber SC. Tyrosine phosphorylation of the BRI1 receptor kinase occurs via a post-translational modification and is activated by the juxtamembrane domain. Front Plant Sci. 2012;3:175.


[185] Rambaut A. FigTree 1.4.2; 2014. Available from: http://tree.bio.ed.ac.uk/software/figtree/


