UNIVERSITY OF CALIFORNIA, SAN DIEGO

Protein Dynamics Through the Computational Nanoscope: An Atomic Perspective

A dissertation submitted in partial satisfaction of the
requirements for the degree of Doctor of Philosophy

in

Chemistry

by

Jamie Michelle Schiffer

Committee in charge:

Professor Rommie Amaro, Chair
Professor Gouri Ghosh
Professor Tracy Handel
Professor Elizabeth Komives
Professor Katja Lindenberg
Professor Susan Taylor

2016

The Dissertation of Jamie Michelle Schiffer is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2016

DEDICATION

I would like to dedicate this thesis to the memory of my father, Michael Charles Schiffer, and my cousin, Alyssa Abrego. It is for you both that I pursue scientific discovery, research, and education. You inspired me through the amazing lives you both chose to lead. You shaped the person I am today through your life and through your death. Love and miss you for the rest of my days.

i carry your heart with me (i carry it in
my heart) i am never without it (anywhere
i go you go, my dear; and whatever is done
by only me is your doing, my darling)
i fear
no fate (for you are my fate, my sweet) i want
no world (for beautiful you are my world, my true)
and it's you are whatever a moon has always meant
and whatever a sun will always sing is you

here is the deepest secret nobody knows
(here is the root of the root and the bud of the bud
and the sky of the sky of a tree called life; which grows
higher than the soul can hope or mind can hide)
and this is the wonder that's keeping the stars apart

i carry your heart (i carry it in my heart)

- e. e. cummings

TABLE OF CONTENTS

LIST OF FIGURES & TABLES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to acknowledge Professor Rommie Amaro for her support, encouragement, guidance, and patience. I am so thankful to Professor Amaro for taking me into the lab and showing me the world through the lens of the computational nanoscope. I never imagined I could feel so fulfilled with the work I am doing. I would also like to thank her for establishing BioChemCore and hiring me as the scientific director this past summer, it was an incredible opportunity and I am forever grateful!

I would also like to acknowledge Professor Elizabeth Komives for her support, mentorship, and collaboration. I would not have had half the opportunities afforded to me in graduate school if it were not for Professor Komives. I am indebted to her for so much, from the opportunity to teach a Biochemistry class of three hundred students, to learning about how the Waters Synapt works!

I would like to thank from the Komives lab my mentors Dr. Deepa Balasubramaniam and Dr. Holly Dembinski. Dr. Balasubramaniam was a tough mentor, and without her I probably would not have been as motivated to learn as much I could as quickly as I could. Dr. Dembinski was a patient mentor during the first few year of my graduate career and I am thankful for her guidance, friendship and for being one of the best shoulders to lean on.

I would also like to thank from the Amaro lab my mentors Dr. Robert Malmstrom and Dr. Victoria Feher. I do not think words could describe how much I have learned about myself and about science from these two mentors. Dr. Malmstrom and Dr. Feher have taught me such an incredible amount in every aspect of being a computational

chemist. I am also truly grateful for their friendship, and for constant sources of commiseration on the T4 Lysozyme L99A mutant, a system that perhaps no one else is as passionate about as Dr. Victoria Feher, and now because of her, myself.

I have to thank all of my coworkers and labmates from the Amaro lab, especially Tavina for being such an incredible friend and coworker. You have been honest with me when perhaps it was not in your best interest, a true sign of your giving nature. Thank you! I have to thank Bryn and Emilia for our work together in the MSM subgroup, and Sophia, Lane, Chris, Jeff, Sasha, Garrett, Sarah, and Ben. Special thank you to Ben, Chris, Jeff, Bryn, and Sarah for their outstanding help with BioChemCore 2016.

Thank you to all my friends that I have met through the graduate program in chemistry at UCSD for wonderful sources of stress relief and fun. Special thanks to Katherine, and Jason for living with me, to Naomi for just being amazing, Jessie, Clare, Lydia and Kara for excellent girls nights (wine galore), to Kendra, Hash, Gavin, and so many more.

Thank you to my three closest friends in all the world, Shaine, Courtney, and Lizzy for being constant sources of normalcy, support, laughter, and thoughtfulness. I could not have come out of these past few years without you three being just a phone call or drive away.

Thank you to my extended family. To my grandmother Pearl, my aunties Nancy, Anita, Mik, Deb and Mary-Ann, and my uncles Tom, Dennis, and Ed thank you. Thank you to my cousins Alyssa, Allison, Jaycob, Kenny, Eliot, Stephanie, Mandy, Jennifer, Kimberly, and Sabrina. My second family, Jacob, Rachel, Julie, Dave and Megan, thank

you for shaping who Justy is and for all of your continue support and kindess. I am fortunate to have the most supportive family.

Finally to my nuclear family, Mom, Dad, and Justin. Mom, thank you for reading to me when I was younger about famous women scientists and story books about other "young scientists" and then inserting my name in place of the character. Your subliminal messages worked, not so subuliminally though. Thank you for also being a badass example of how a woman can be brilliant, beautiful, and kind. Justin, thank you for being my rock, my best friend, the love of my life, and my sounding board for all big decisions and thoughts and ideas. I love you. Finally, thank you to my father, without whom I would not be here and who was my best friend, my greatest confidant, and my huge support system. I will miss you forever.

Chapter 2, in full, is a reprint of the material as it appears in Capturing Invisible Motions in the Transition from Ground to Rare Excited State of T4 Lysozyme L99A 2016. Schiffer, Jamie M., Feher, V. A., Malmstrom, R. D., Sida, R., and Amaro R. E., Biophysical Journal, 2016. The dissertation author was the co-primary investigator and author of this material.

Chapter 3, in full, is a reprint of the submitted material Benzene Dissociation through a Protein Excited State 2016. Schiffer, Jamie M., Feher, V. A., Mermelstein, D., Mih, N., Pierce, L. C. T., McCammon, A. J., Amaro, R. E. submitted. The dissertation author was the co-primary investigator and author of this material.

Chapter 4, in full, is a repreint of the material as it appears in How the Ankyrin and SOCS Box Protein, ASB9, Binds to Creatine Kinase 2016. Balasubramaniam, D.,

Schiffer, J., Parnell, J., Mir, S. P., Amaro, R. E., Komives, E. A., Biochemistry 2015. The dissertation author was a secondary investigator and author of this material.

Chapter 5, in full, is a reprint of the material as it appears in Model of the Ankryin and SOCS Box Protein, ASB9, E3 Ligase Reveals a Mechanism for Dynamic Ubiquitin Transfer 2016. Schiffer, Jamie. M., Malmstrom, R. D., Parnell, J., Ramirez-Sarmiento, C., Reyes, J., Amaro, R. E., Komives, E. A., Structure, 2016. The dissertation author was the primary investigator and author of this material.

Chapter 6, in part is currently being prepared for submission for publication of material. Schiffer, Jamie. M, Nguyen, T. T., Chan, G., Demir, O., Ghosh, G., Amaro, R. E., Structure-Based Dynamics of the Inhibitor of κB Kinase β Construe the Complexity at a Plausible NEMO Interface. The dissertation/thesis author in the primary investigator and author of this material.

VITA

2012	Bachelor of Arts, Boston University
2012-2016	Research Assistant, University of California, San Diego
2016	Doctor of Philosophy, University of California San Diego

ABSTRACT OF THE DISSERTATION

Protein Dynamics Through the Computational Nanoscope: An Atomic Perspective

by

Jamie Michelle Schiffer

Doctor of Philosophy in Chemistry

University of California, San Diego

Professor Rommie Amaro, Chair

Proteins dance; the atoms of a protein move through time in our cellular aqueous

environment, and each atom is important in this process. The dance of a protein is unique to that protein, and alone controls how that protein functions. In this way, our cells are a world unto their own, and as the proteins of a cell dance along, they orchestrate life. In this thesis, I discuss the dance, or dynamics, of five different proteins: the ankyrin repeat and SOCS box protein 9 (ASB9), creatine kinase (CK), the L99A mutant of T4 lysozyme, the Inhibitor of κB kinase 2 (IKK2), and the NFκB essential modulator (NEMO). Each of these proteins have unique sizes, shapes, dynamics, and roles in the cell. To study these proteins, I chose to use the computational microscope, performing molecular dynamics simulations at the atomic level. Through the solving of theoretical models of the bonds between atoms, iterated over and over again through time, a dance unfolded for each of these proteins, illuminating how these proteins perform their duties in our cells and in our test tubes. I demonstrate that ASB9 and CK interact through a highly disordered interface, uncover the dynamics which guide both excited state transitioning and benzene release in the L99A mutant of T4 lysozyme, and present a hypothesis for how NEMO appends itself to the IKK2 hexamer.

INTRODUCTION

Section 1: Synopsis

In this thesis concepts surrounding protein dynamics are explored with computational techniques, specifically molecular dynamics and network models. As an introduction into this thesis, a background on proteins in general and on the theory and application of various computational techniques are discussed. Following these topics, specific introductions on three different protein systems are reviewed.

For the background on proteins, in section 1.2 the timescale and dimensions of the world of proteins is noted. This is meant to introduce the reader to the world of proteins inside of our cells. In section 1.3, a brief background on protein structure is presented which reviews most of what is discussed in an undergraduate biochemistry course but diverts from this in exploring intrinsically disordered proteins. In section 1.4, the theory of how the small world of proteins sums up to thermodynamic quantities in the macroscopic world is reviewed through the standpoint of statistical mechanics.

This thesis focuses on how proteins move in solution at the atomic level and how proteins interact with one another, or protein dynamics and protein-protein interactions respectively. As a background on protein dynamics, section 1.5 provides a brief overview of the experimental techniques used and model systems for investigating protein dynamics. Section 1.6 provides a history of protein-protein interactions, and a brief overview of experimental and computational methods for studying such interactions.

To define protein motions, the work herein mostly concerns molecular dynamics simulations and aligning these simulations with experiment. Section 1.7 reviews the

theory of molecular dynamics, or MD. Section 1.8 explores the various methods for directly comparing simulation with experiment. Section 1.9 completes the introduction by providing a brief description of the three protein systems described in this these: the L99A mutant of T4 lysozyme, the ankyrin repeat and SOCS box protein 9 (ASB9) and its interaction with creatine kinase (CK), and the inhibitor of κB kinase beta (IKK2).

Section 2: A note on timescales and dimensions of the world inside our cells

Our cells are like little worlds. And in these little worlds, proteins are the worker bees; they make sure the work gets done. Just as each human being on earth has a job or a role in the world, each protein performs a different and critical function to the health of the whole cell. Each protein is a different shape and size, and finds home in different parts of the cell: some proteins reside in the cellular membranes, or the nuclear membranes, while others prefer to localize to cellular compartments in the cytoplasm or nucleus. The timescales of this world are much quicker that the timescales of our world. While it may take us thirty minutes just to get to work (depending on the traffic), most proteins are able to perform their functions on the microsecond timescale, or 0.000001 seconds. The spatial dimensions of this world are similarly much smaller than ours, with

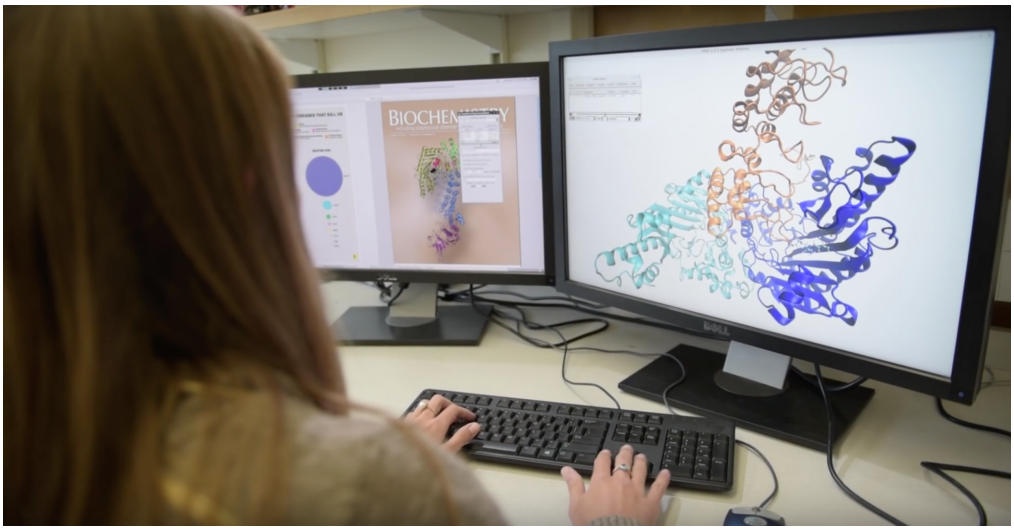


Figure 1.1: A dashing scientist and her computational nanoscope. Through the computational nanoscope it is now possible to watch the dynamics of proteins on the nanometer length scale and nanosecond to microsecond timescales.

the cell itself spanning on average only twenty-five micrometers (μm) and most proteins are within 10-30 nanometers (nm). Yet, perhaps the greatest distinction between this world and the one that we inhabit is that proteins are not surrounded by air as we are. Instead, proteins exist in an aqueous solution of salts and other proteins, much like triathletes at the beginning of a race. And once the race starts, so begins life, with proteins swimming somewhat aimlessly, or more accurately in an entropically favorable way, until they can achieve their purpose. Rather quickly compared to our timescales, the life of a protein ends, and proteins are recycled back into their building blocks so that they may be rebuilt into a new protein. Here we will learn about how different proteins achieve their purpose through the lens of the computation nanoscope.

Section 3: Protein Structure

Biochemistry courses introduce protein structure by breaking down the structure into four different categories: primary, secondary, tertiary, and quaternary. Here we review these principles and then introduce rationale to demonstrate how these principles are inherently flawed.

Primary structure is often described as a sequence of amino acids but it refers to much more than that: it is the basic atomic-level chemical structure of the amino acids that form a long peptide chain. The secondary structure of a protein purports how the amino acids of a protein fold locally. Tertiary structure describes the overall fold of a single protein peptide chain, and this fold is generally accomplished by burying hydrophobic side chains into the cavity of a protein, away from the aqueous salt solution that proteins float in. Quaternary structure designates how many proteins are found together in complexes so stable that the interactions are thought to be irreversible.

This protein structure code is an imperfect one. Secondary structures are more complex than to be easily categorized into bins of either alpha helices or beta sheets. The line between secondary structures and tertiary structures is often drawn in the sand, as is the division between tertiary and quaternary structures. Delineating quaternary structure can be confusing because most proteins associate with one either transiently or with some degree of separation. Moreover, expressing proteins as having an absolute primary, secondary, tertiary, or quaternary structure overlooks known scientific phenomenon such as mutants, protein quaking, protein dynamics, and protein signaling pathways. Here these oversights are explored.

Generally, in introductory biochemistry courses, two of the major secondary structural elements are introduced: alpha helices and beta sheets. But there are so many other structures that small sequences of amino acid within a large peptide chain can adopt, including but not limited to beta turns, coils, 3_{10} helices, and bends. What most obviously distinguishes these secondary structures from one another is how the backbones of the local peptide sequence form hydrogen bonds from one amino acid to another. Yet, these hydrogen bonds should not be confused as the force behind the folds of these secondary structures. Instead it is the hydrophobic-hydrophobic attraction and hydrophobic-water repulsion that induce the formation of these secondary structures. The

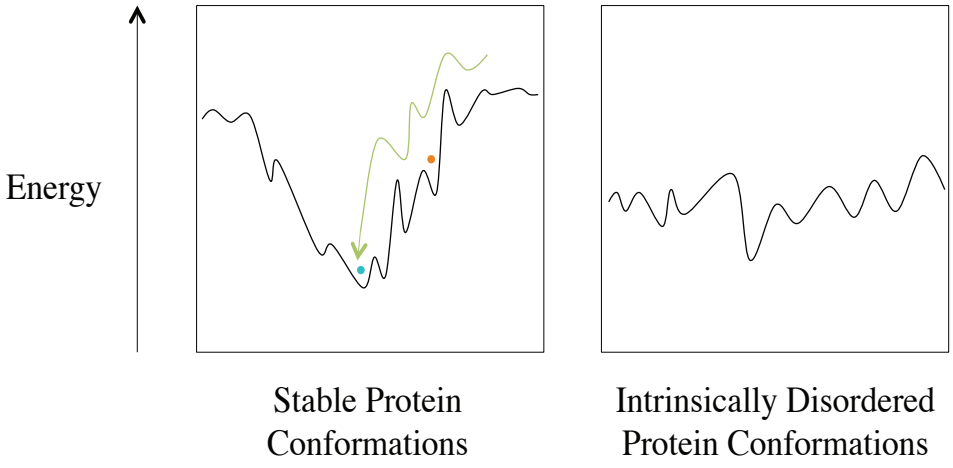


Figure 1.2: Two-dimensional depictions of free energy landscapes for stable, well folded proteins (left) and intrinsically disordered proteins (right). Ground states of a protein (blue dot) and excited states of a protein (orange dot) demonstrate how different protein conformations have different energies. A hypothetical folding pathway is also depicted to show how a protein samples multiple states as it folds (green arrow).

energy landscape (Figure 2) that is associated with these forces on the atoms and amino acids of a protein chain are typically depicted by a folding funnel, with the protein

dropping from one basin to another until it falls into the lowest energy structure, the folded structure. However, not all proteins fold or exist in this sort of energy landscape.

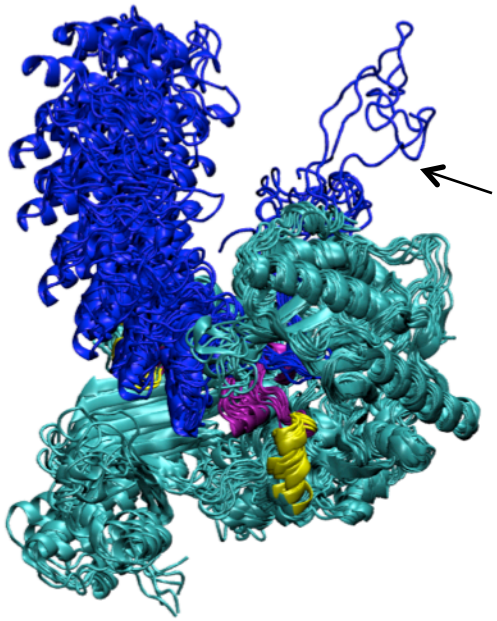


Figure 1.3: An image of the intrinsically disordered region of a protein in blue (arrow), which is important in protein interaction.

One such class of proteins whose energetic landscape cannot be explained through a folding funnel is the intrinsically disordered protein (IDP, see example in Figure 3 above). These proteins have flat energy landscapes with multiple small basins where different conformations with relatively equal energies are explored or sampled (Figure 2). In other words, these proteins have many different possible structures. This conformational heterogeneity (Figure 3) makes these proteins difficult to study with conventional structural biology methods, a point that is discussed later in the thesis. This characteristic also confers many unique characteristics to these proteins. For one, IDPs are highly gregarious proteins in that they interact with many other protein partners. They

are also more strictly regulated for this reason, because as can be imagined, this dynamic nature could result in misfolding and aggregation.[1] In fact, many misregulated IDPs are the root cause of malignancies and diseases.[2,3]

Most proteins do not perform their necessary cellular tasks in isolation. Instead, they interact through complexes of identical protein partners or through interactions with other proteins that have unique amino acids sequences, a point returned to more extensively in Section 1.7. While many higher order protein complexes do form stably in the cell and are energetically found at the bottom of a large energy funnel, other proteins interact transiently in a signaling pathway and only in certain chemical environments.

The energy of protein-protein interactions can be explain in terms of an enthalpy component (ΔH), the energy released due to stabilization of through space interactions of the electrostatic or van der Waals varieties, and an entropic component (ΔS), the randomness or heterogeneity of a protein's energy landscape. The total energy (ΔG) in a protein-protein interaction can be explain as a linear combination of these two terms: (1)

$$\Delta G = \Delta H - T\Delta S$$

However, the energy can also be written in terms of an equilibrium constant, which describes the concentrations of the proteins, say protein 1 and protein 2 (P_1, P_2), in complex versus the proteins alone in solution:

$$P_1 + P_2 \rightarrow P_1P_2 \quad (2)$$

$$\Delta G = -RT \bullet \ln(K_{eq}) \quad (3)$$

$$K_{eq} = \frac{[P_1P_2]}{[P_1][P_2]} \quad (4)$$

This equilibrium constant is also related to the ratio between the rates of the reverse and forward reaction in going from monomers of two proteins to a protein-protein complex:

$$K_{eq} = \frac{k_{off}}{k_{on}} \quad (5)$$

Determining these rates of binding and release can thus help us understand not just how quickly or slowly two proteins interact in solution but these rates can also be used to calculate the energies of interaction between two proteins. I will return to the theory and history of protein-protein interactions later in the introduction.

The last point that will be mentioned here is concerned with the difference between the cellular environment and the model shown in equation (2) above where one protein interacts with another protein. In the aqueous milieu of our cells, a single protein is not interacting with just another protein. Proteins are constantly bumping into one another due to the tight packing of proteins within our cells. Moreover, different types of proteins have been found to localize to different parts of the cell in a process called phase separation.[4] This processes has just recently begun to be attributed to IDPs in creation of phase-separated microenvironments within our cells. These microenvironments will ultimately also affect the rates and thus equilibriums of protein-protein interactions and should be considered when describing protein interactions in a cell versus protein interaction *in vitro* or *in silico*. In this thesis, the *in vitro* and *in silico* properties are studied in the hopes that these might provide a good starting point for understanding how these proteins move and function within our cells.

Section 4: Statistical Mechanics of Proteins

.

The connection between the atomic level motions of proteins and experimental observables can be described through a language, the language of statistical mechanics. Statistical mechanics is predicated on the notion that we can derive thermodynamic and kinetic quantities observed experimentally from probabilities and population averages of microscopic, or even nanoscopic, worlds, conditions, or events. In this section, this language is explored. First, terms ensemble, phase space, a priori probability, and Ergodic hypothesis are defined. Them from probability theory, the equation for entropy is derived. Finally, this section ends with discussion of the microcanonical ensemble (NVE), the canonical ensemble (NVT), and the constant pressure (isobaric-isothermal) ensemble (NPT).

Proteins do not exist alone in solution, and typically the properties of a protein that we measure from experiment consist of a population of an identical protein. The values obtained from these experiments are not absolutes, but rather averages. A group of conformational states, often termed microstates, with specific properties like number of particles N, volume V, temperature T, and energy E, is defined to be an ensemble.[5] Explained differently, an ensemble is a collection of a very large number of nanoscopic or microscopic things, with all replicas contributing to the thermodynamic properties of a macroscopic system.[6] To continue with the analogy begun in Chapter 1.2, the world is an ensemble of people at any given instance. At that instance, the world has a specific number of people, N, a specific volume, and a specific energy E. However, over time, that ensemble of people changes, which leads to the idea of phase space.

Phase space is an N-dimensional space, first coined and conceived by Gibbs, which encompasses all of the possible configurations of a given ensemble of N things. For the purposes of this thesis, these "things," are proteins. Thus, in the classical N-body or N-protein system at any time the ensemble can be completely describe by a single point in phase space.[6] Thus the state of people in the world right,… now, is only a single point in the phase space that is life on earth, assuming that the same number of people are always on planet earth. Yes, this is false assumption but it suits these purposes in explaining phase space here and is the same assumption made for describing proteins in a cell.

As will be explain in Chapter 1.7, investigations of the dynamics of proteins at the nanoscopic length scale can use techniques like molecular dynamics, which only investigate the motions of one protein through time. How then can this information be used to extrapolate thermodynamics properties of a protein ensemble? The answer comes in the form of the Ergodic hypothesis:

> "For a stationary random process, a large number of observations made on a single system at *M* arbitrary instants of time have the same statistical properties as observing *M* arbitrarily chosen systems at the same time from an ensemble of similar systems." – McQuarrie, D. *Statistical Mechanics*

Thus the snapshots of the world seen through the computational nanoscope are equivalent to observing an infinite amount of replicates of that world at a single point in time. These infinite replicates are what is observed through experimental techniques like X-ray crystallography, NMR spectroscopy, Isothermal titration calorimetry, HDXMS, and more. Thus it is possible to directly compare the results from the computational nanoscope with experimental observables.

There are a multitude of specific properties that can characterize an ensemble including by not limited to number of things or protein), volume of actual space, temperature, energy, pressure, and chemical potential. Here three different ensembles are addressed: the NVT ensemble or the canonical ensemble, the NVE ensemble or the microcanonical ensemble, and the NPT ensemble or the constant pressure ensemble, as these are most relatable to the world simulated in molecular dynamics simulations.

The canonical ensemble describes a set of systems that are not isolated, as the energies in the different systems can vary, but all have the same number of particles, volume, and temperature. The average energy and entropy of the canonical ensemble depends upon the temperature, number of particles, and volume by

$$E = E(N, V, \beta) = \frac{\sum_j E_j(N,V) e^{-\beta E_j(N,V)}}{\sum_j e^{-\beta E_j(N,V)}} \tag{6}$$

$$S = k \ln Q + kT \left(\frac{\partial ln Q}{dT}\right)_{N,V} \tag{7}$$

where β is 1/kT, where k is the Boltzman constant and T is the temperature. If we let the denominator be Q, then Q(N, V, β) can be used to describe the values for an array of other thermodynamic values including S and p. This Q is known as the canonical ensemble partition function.[6] Most molecular dynamics simulations are performed in this ensemble.

In the microcanonical ensemble, the energy, number of particles, and volume is held constant, and so is classically is known as an isolated system. Unlike in the canonical ensemble, the entropy of a microcanonical ensemble is unique in that it is proportional to the logarithm of the number of states that the system can occupy:

$$S = k\, ln\Omega(N, V, E) \tag{8}$$

This ensemble is rarely simulated as the conformational states of different energies in proteins are often important biologically and desirable for sampling.

Finally, the isobaric-isothermal ensemble, or constant pressure ensemble, is also commonly used to perform molecular dynamics simulations as it mirrors biologically relevant conditions rather well. In this ensemble, the partition function is written in terms of number of particles, temperature and pressure, and the entropy of the system is typically written as:

$$S = k\ ln\Delta + kT(\frac{\partial ln\Delta}{\partial T})_{N,p} \tag{9}$$

The partition functions for the canonical (Q), microcanonical (Ω), and isobaric-isothermal (Δ) ensembles are all logarithmically related to different thermodynamic variables. As can be seen in equation 8, the microcanonical partition function is logarithmic to the entropy of a microcanonical ensemble. The canonical partition function is logarithmic to the Helmholtz free energy, denoted A in most texts. Finally, the isobaric-isothermal partition function is logarithmic to the Gibbs free energy, and thus relates very nicely to many quantitatively measureable variables for an experimental ensemble. For a deeper look into statistical mechanics theory, please see McQuarrie's text "Statistcal Mechanics" referenced in section 1.8.

Section 5: Protein dynamics

The concept of proteins as dynamic, non-static systems was born in 1959 through hydrogen exchange experiments, not long after the discovery of secondary structural elements like α helices and β sheets,.[7,8] These results demonstrated that proteins "breathe" to allow solvent into buried protein interiors.[7] The conceptual leap from proteins as dynamic to conformational ensembles astoundingly followed shortly after in 1963.[9] The role of these conformational ensembles in ligand binding and other reaction mechanisms is a concept that the protein science community is still grappling with.[10,11]

What are conformational ensembles? Returning to the earth and people analogy, imagine that each person had identical genetic makeup, a world full of many copies of the same person. This earth then contains an ensemble of people, all moving differently, some with their hands raised, some walking, some speaking etc. Our cells, contain ensembles of proteins that are also in different physical states, or conformations. Experiments often deal with measuring averages over an ensemble of identical protein copies, where as simulations calculate the structures of a single protein copy through multiple time steps of a trajectory. It is possible to relate these two data sets because of the Ergodic hypothesis, a point discussed in chapter 1.4.

Historically, protein dynamics was a field that followed first from crystallography, which cannot resolve long-timescale dynamic motions and is limited mostly to solvent exposed side chain torsions or backbone coil and turn fluctuations. Today, protein dynamics is more commonly discerned experimentally with NMR and SAXS as well as fluorescence based techniques.[12,13] Complementing results from

experiment, the application of molecular dynamics and other simulations now allow scientists to align atomic level motions on the nanosecond to millisecond timescales with ensemble averages from experiment. In this thesis, the focus is on aligning results from molecular dynamics simulations with NMR and SAXS results and using these methods to reveal the dynamic personalities of proteins.[14]

Nuclear magnetic resonance (NMR) spectroscopy can measure both internal distances between atoms in a protein and the motions of atoms in a protein. These measurements are possible due the fact that some nuclei have nuclear spin magnetic moments, specifically the nuclei of ^{1}H, ^{13}C, and ^{15}N.

In a typical NMR experiment, these atomic nuclear spins in a protein align with an external magnetic field ($\mathbf{B}_1$) and then become disturbed away from their equilibrium aligned position through the application of a radio frequency pulse in a plane normal or perpendicular to the external magnetic field. After the pulse, the nuclei spins then precess about the axis of $\mathbf{B}_1$ at a very specific frequency. This frequency is dependent upon the strength of magnetic field witnessed by the nuclei. Each nuclei is affected by a different field strength because each nuclei exists in a unique chemical environment, with varying degrees of effective magnetic field shielding of the nuclei. This frequency is known as the Larmor precession frequency ($\mathbf{v}$).

The measured output from an NMR experiment without any mathematical treatment is free induction decay (FID), or the sums of all the induced nuclear moments over time along the axis of radio frequency pulse. This FID is then Fourier transformed to result in a plot in the frequency domain, so that each individual nuclei can be teased out

from the data. This mathematical treatment results in three primary measurements from NMR spectroscopy: chemical shift, intensity, and linewidth.

The three major methods for measuring protein dynamics through NMR spectroscopy are through the measurement of order parameters (ps-ns), chemical shift changes calculated from relaxation dispersion (μs-ms), and hydrogen-deuterium exchange chemical shift effects (seconds to days). Order parameters are calculated from orientation correlation functions C(t), or the dot product of a bond vector at some time t=0 and the bond vector at some future time, and these values quantify the amplitude of internal motions in a model free way. To study motions on the microsecond to millisecond regime, Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion is used, as are other relaxation techniques, which will not be reviewed here.

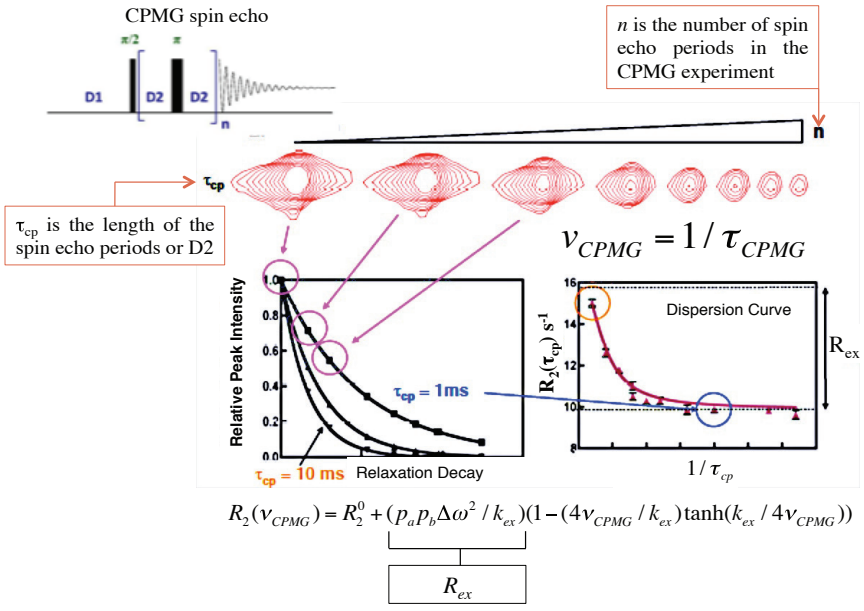


Figure 1.4: Schematic of how CPMG spin echo pulses can be used to derive R_{ex} between two protein states that exchange on slower timescales. Adapted from Palmer *et al.*[15]

With CPMG relaxation dispersion, a spin echo pulse is used to refocus magnetization vectors, however, if a nuclei samples two different chemical shifts in two consecutive spin echos, then line broadening will occur (Figure 4). These line-broadened spectra can be fit with various models to solve for the chemical shift difference, rates of interconversion, and populations of each chemical shift. Finally, hydrogen-deuterium exchange places a protein with ^{1}H into a deuterium solution (or vice versa), and measures the disappearance (or appearance) of chemical shifts over time due to local unfolding events, breathing and quaking that commonly happens in proteins on the millisecond to hours timescales.

In this thesis, we compare directly with the second kind of motions in the microsecond to millisecond regime through the use of back-calculated chemical shifts from simulations. See below for an explanation of how chemical shifts are back calculated from structures sampled in molecular dynamics simulations. In this thesis, structures from MD simulation are also compared to scattering experiments, a method for understanding the dynamic ensemble of a protein at a lower resolution than is achieved with NMR.

Small-angle X-ray scattering (SAXS) makes use of the scattering angle of protein or macromolecules in solution to understand the solution and thus dynamic properties of proteins that cannot be seen through in constrained crystal structures. The data reported from SAXS is the population of scattered X-rays of a given wavelength (λ) at a given angle (θ) from the sample based on the electron density of the protein itself as well as the first hydration shell. This shape of the curve at low angles q values, where $q = (4\pi \sin\theta/\lambda)$ can be fit to determine the radius of gyration through the Debye-Scattering curve. This

scattering curve can also be used to roughly back-calculate the electron density of a protein structure, a point to which we will return later in the introduction.

In the past twenty years it has become increasingly popular to study protein dynamics through the computational nanoscope, more specifically through the direct application of molecular dynamics simulations (see chapter 1.7) While an array of different techniques is available for performing conventional molecular dynamics simulations, it is generally difficult to achieve dynamics on the order of microseconds to milliseconds for an average sized protein (~ 30 kDa, or 300 residues). Thus many methodologies have arisen in the past few decades to gain access to these slower timescales. These include temperature enhanced molecular dynamics, replica exchange, metadynamics, accelerated molecular dynamics, Gaussian molecular dynamics and more. Additionally, advances in computational hardware have also enabled the accelerated sampling of longer timescales, through the advent of GPU computing, and in a very specific case, through the engineering of an application specific integrated circuit (ASIC) geared towards the calculations in a molecular dynamics trajectory. This later case was used in the construction of the Anton supercomputer, which is harnessed in this thesis for simulations of the L99A mutant of T4 lysozyme. Other techniques outside of molecular dynamics have also been applied to the study of long-timescale simulations, including anisotropic and elastic network models which is applied in this thesis for long-timescale dynamics of rather large protein complexes.

Computational biophysicists did not begin investigating protein dynamics by performing simulations on huge protein complexes. Instead initial study was focused on aligning simulations for smaller protein dynamic model systems to experimental results.

Myoglobin, bovine pancreatic trypsin inhibitor (BPTI), and T4 lysozyme are classic protein model systems for studying protein dynamics. Investigations of the timescale of fluctuations, or the kinetics, as well as the conformational pathways of these proteins have set the stage for studies of larger and more dynamic proteins today.[13] These proteins were the first to be examined because of their stabilities in solution and their ease of extraction in large quantities, as these investigations began before the advent of current molecular biology approaches.

Myoglobin is one of the most well-studied proteins in biology, and the first protein structure solved through X-ray crystallography.[16] It is well studied because it is plentiful and easily extracted from blood samples and it is important for oxygen transport in the body. The dynamic motions of myoglobin were initially probed through the method of flash photolysis to measure the rates of carbon monoxide and oxygen binding as a function of temperature and ligand concentration.[17]

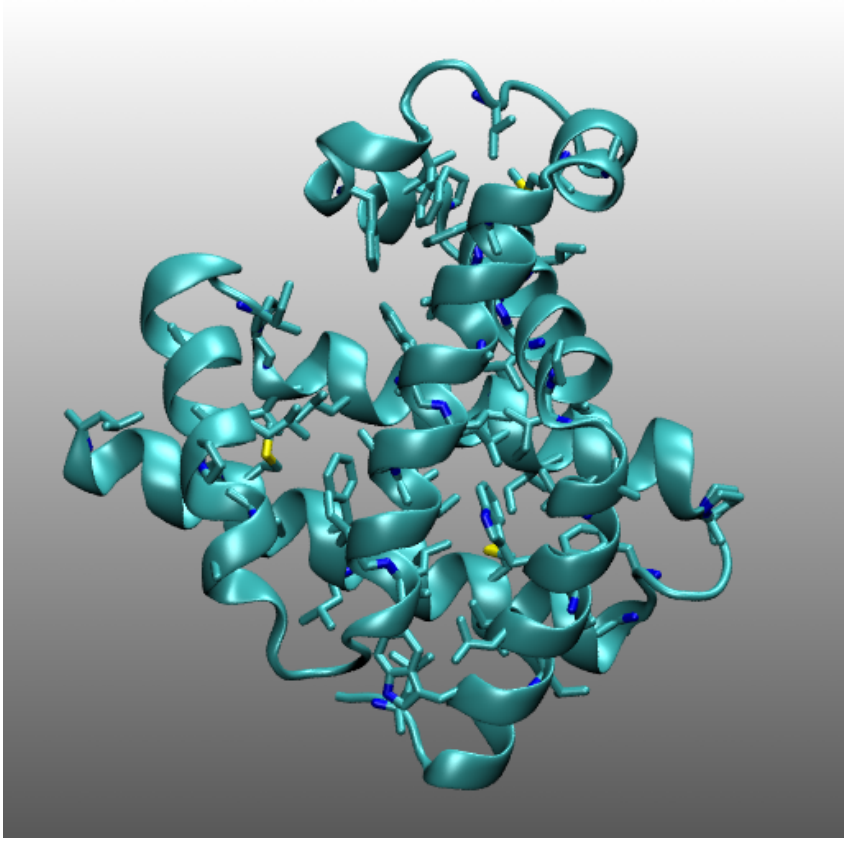


Figure 1.5: Myoglobin's structure and hydrophobic residues. This figure shows the backbone structure of myoglobin as well as the hydrophobic residues found in the backbone. The number of hydrophobic residues buried demonstrates how and why this protein was the first protein to be crystallized.

Bovine pancreatic trypsin inhibitor (BPTI) was the first protein system to be simulated with molecular dynamics.[18] Since then it has continued to be a model system in studying protein dynamics through the computational nanoscope. Simulations performed on the Anton supercomputer[19] recently explored the accuracy of accelerated molecular dynamics methodologies.[20] Aligning the results of long-timescale simulations of BPTI with experimental observables [21–25] continues to be an area of research and has revealed importantly that a single long trajectory predicts faster rates than seen experimentally.[22]

T4 lysozyme has become a model system for investigating the role of each amino acid in protein structure, with over 300 mutants studied for the effects on structure, function, stability, and dynamics.[26] These results mostly demonstrated that T4 lysozyme is resilient to amino acid changes, and provided the framework to explore these changes in other protein systems.[26] Additionally, cavity mutants in the T4 lysozyme provided the first example of engineering buried pockets capable of binding small molecules and conferring unique dynamics to a protein ensemble.[27–30] Here we investigate the L99A mutant of T4 lysozyme, a engineered protein than has come to be known as the model system for testing *in silico* ligand binding methods,[31–34] for investigating excited states in proteins,[35–42] and for determining how ligands bind to buried cavities.[27,43–45]

Following the motions of proteins at the atomic level is not a trivial task. Multiple different unique approaches to elucidating the "dynamic personality"[14] of proteins have been pursued in just the last decade, let alone the last century. The analysis of protein dynamics through the computational nanoscope presents an unparalleled avenue for discovery. Defining these motions is critical to our understanding of how proteins perform a range of biological functions, from catalysis to signal transduction through protein-protein interactions.

Section 6: Protein-Protein interactions

Here the importance of protein-protein interactions is reviewed in the context of signal transduction pathways. Once the function of protein-protein interactions is established, the methods for examining protein-protein interactions experimentally and computationally are considered.

Protein interactions are important in sending messages between the cell membrane to the nucleus through a domino chain of proteins: one protein tags another, which tags another, in signal transduction pathways which are like cellular equivalent of freeze tag. Sometimes, proteins interact with one another so that one may phosphorylate the other, as with kinases, and thereby alter the dynamics of the protein acted upon. Other times, proteins interact with one another tightly, with one inhibiting the other from performing its activity in the cell. Still further, some proteins interact with others to allow or block passage into different cellular compartments like the nucleus. In all of these processes, it is common for proteins to steer conformational changes by employing a protein artifice, allostery.

Allostery is a common feature of protein conformational changes, whereby an alteration at one surface dissipates to a different protein surface, and changes that protein's activity in some way. Thus, characterizing protein-protein interactions is of the utmost importance in understanding how our cells work. Scientists have developed a milieu of experimental and simulation techniques to investigate these processes, which are reviewed briefly in here. But first, presented here is a brief history of the study of protein-protein interactions.

The first demonstration of protein-protein complex interactions came from studies of antibodies and antigens in 1891, with the plant toxin ricin.[46] Paul Ehrlich injected non-toxic doses of ricin into mice, and then demonstrated that blood from the immunized mouse to a naïve mouse, which was then immune to ricin. The explanation for this phenomenon was that the mouse's body produced anti-ricin. However, at this time it was still not clear that ricin was a protein or that the antibodies were also proteins. Another early discovery of a protein-protein interaction, and certainly the first inhibitor complex, came in 1906 through a kinetic study of albumin in blood and its inhibition by trypsin.[47] It would take almost another 30 years until these enzymes were realized to be proteins.[48] In the 1920s, Svedberg with the ultracentrifuge demonstrated absolutely that proteins form complexes that can be separated out at low pH, and thus small subunits aggregate into larger complexes; the first example of quaternary structure.[48]

Today, protein-protein interactions are now studied by an array of techniques and can be isolated because of the major advancements during the molecular biology revolution of the 20th century. The optimization of cell culture, development of SDS gel electrophoresis, and discovery of polymerase chain reactions are just a few of the advancements that now allow scientists to isolate a single protein, purify it, and use an array of experimental techniques, from pull-down assays to surface Plasmon resonance spectroscopy to cell-based fluorescence assays, to characterize how, what, and when proteins interact with one another.

In this thesis, the focus will be on *in vitro* biophysical techniques for investigating protein-protein interactions and using the computational nanoscope to integrate this data into an atomic blueprint. The *in vitro* techniques reviewed include kinetic, pull-down, and

gel-based assays, isothermal titration calorimetry (ITC), surface plasmon resonance spectroscopy (SPR), nuclear magnetic resonance, and hydrogen deuterium exchange mass spectrometry. The *in silico* techniques used include protein-protein docking, molecular dynamics, Adaptive Poisson Boltzmann Solver,[49–51] protein-protein interaction predictors, and anisotropic network model analyses.

Some tools for investigating protein-protein interactions require very little instrumentation; simply a test tube and a column are enough to gain incredible insight. Upon purification, a single protein partner can be chemically linked and interact with a column much like the one it was purified from. The protein binding partner can then be run on the column and pulled-down. The resulting complex can be eluted and run on a gel compared to the two input purified proteins. From these sort of gel-based assays, qualification of the amount of protein interacting can be performed. Additionally, protein interactions can be measured through assays if one protein partner is known to inhibit the other. The effect of various concentrations of the inhibitor partner on an enzyme can provide give information on stoichiometry and relative binding rates. However, both the kinetic assays and pull-down methods are not exact and cannot report with absolute certainty what is happening at the molecular level, and so more advanced techniques like ITC, SPR and NMR are used to obtain this deeper level of insight. For an extensive look into these methodologies, please reference work by Phizicky and Fields.[52]

Isothermal titration calorimetry (ITC) is a technique that takes advantage of the fact that the heat capacity of a protein (ΔC_p) changes when complexed with another protein.[53] This makes sense in the context of the chemical environments of protein-protein interactions: often times a protein-protein interaction interface buries a significant

hydrophobic surface that would otherwise interact unfavorably with an aqueous environment.

In an ITC instrument there are two cells. One cell is a reference and one cell is the sample cell, with a volume V_o of one protein (Video 1: https://youtu.be/wHo_Zdnl3M8). The experiment consists of another protein, or small molecule ligand, injected periodically into this sample cell. The temperatures of the sample and reference cells are kept identical throughout the experiment, allowing calculation of heat emitted or absorbed upon complex formation between the sample protein and the injected protein. If the interaction is exothermic, the heat released by the interaction will increase the temperature of the sample cell relative to the reference cell. The instrument will sense the change in temperature, and reduces the input of heat to the sample cell relative to the reference cell to return the temperatures of the sample cell to the temperature of the reference cell.

As mentioned above, when the protein in the cell (P_c) interacts with the injected protein (P_i), heat may be dissipated into the solution. The heat (Q) evolved can be represented as:

$$Q = V_o \Delta H_b [P_c]_t K_a [P_i]/(1 + K_a[P_i] \quad (10)$$

Thus, if the enthalpy of binding (ΔH_b) is known, the association constant (K_a) can be calculated. After repetitive injections, a binding isotherm is measured from the changes in heat input to the sample cell relative to the reference cell. Injections continue well after saturation of P_c by P_i so that a baseline for the background heat of dissolving P_i can be measured. The integrated heats of these binding isotherms are the enthalpies of interaction, and this value can be incorporated into the equation above to solve for K_a.

Finally, this experiment can be performed over a range of temperatures to determine the temperature dependence of the enthalpy of interaction, which is simply the heat capacity (ΔC_p).

While ITC can measure the enthalpies, binding constants, and heat capacities for most protein-protein interactions, this technique is limited in its ability to measure protein-protein interaction kinetics when the timescales are too quick or the dissociation constant (K_D) is too small. This is where Surface Plasmon resonance spectroscopy is often adopted in biophysical chemistry. Surface Plasmon resonance, or SPR, is an optical technique that harnesses the ability of evanescent waves to report on the changes in refractive indices close to a sensor surface. In fact, evanescent waves are sensitive enough to measure changes in refractive indices due to an increase in molecular weight, which allows real-time measurements of protein-protein interactions at a Plasmon surface.[54–56] This relatively new technique came about in the early 1990s and has since revolutionized the measurements of kinetics of biomolecular interactions. [54,57] SPR harnesses the high affinity between the protein streptavidin and the biologically relevant small molecule biotin to link proteins of interest to a sensor surface.

There are three experimental steps of an SPR experiment: (1) coating the streptavidin embedded surface of a small gold chip with protein binding partner A that is biotinylated, (2) flowing free protein binding partner B in solution passed the linked protein binding partner A, (3) stripping the binding partner B from partner A with time or via chemical means like low pH solution. Step 2 and 3 result in association and dissociation curves respectively, and can be fit to Langmuir binding models to solve for the k_{on} and k_{off}. While SPR can be used for an array of other biophysical

characterizations, it is not within the scope of this thesis to review these here and the interested reader is referred the Biacore manual for a more extensive review.[57]

While ITC and SPR are able to provide quantitative measures of the entire protein ensembles interaction energies and kinetics, these methods are not capable of elucidating the interaction mechanism at the atomic level. For this type of investigation, biophysicists often turn to nuclear magnetic resonance (NMR) spectroscopy. For protein-protein interactions, it is common to label one of the protein interacting partners at a time, and obtain NMR spectra for the proteins alone in solution and then the proteins in complex. The differences in the spectra then allow the quantification of chemical environment changes for exact amino acids side chain and backbone atoms, as well as measure rates and thermodynamic quantities. This method was reviewed in greater detail in Section 1.6 and interested readers are referred to Cavanagh's text.[58]

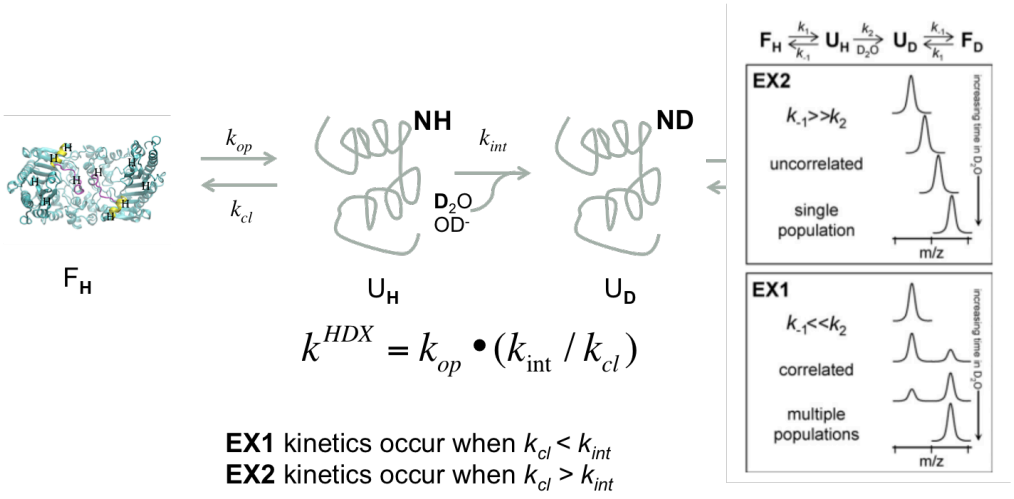


Figure 1.6: The Process of Hydrogen Deuterium Exchange. This schematic shows the different kinetic processes involved in hydrogen deuterium exchange including the role of protein unfolding, and the difference between EX1 and EX2 exchange regimes.

Finally, one of the best current tools for probing a protein-protein interface at the secondary structure level without also sampling local conformational changes comes from hydrogen deuterium exchange mass spectrometry (HDXMS).[59] This method works by measuring mass shifts due to incorporation of deuterium from solution into the backbone amide position of an amino acid. Amide protons are able to exchange with solution only when solvent exposed. Two different kinds of exchange processes are seen in these experiments. The first is known as EX1, and describes a hydrogen deuterium exchange process where the rate of local refolding (from an transiently unfolded structure) is less than the rate of hydrogen exchange. This exchange results in two mass peak populations in the spectrum (Figure 6). The second is known as EX2, and describes a hydrogen deuterium exchange process were the rate of local refolding is greater than the rate of hydrogen exchange, and so results in a single population shifting during exchange or incorporation of the deuterium. From HDXMS experiments, protection factors can be elucidated to determine which residue backbones are solvent exposed and on what timescale. Additionally, titration reactions can be used to elucidate equilibrium constants. For a more extensive review of HDXMS please see Percy *et al*.[60]

Computational methods for studying protein-protein interaction have become of great interest in recent year, especially with the rise of the CAPRI competition. Here we will discuss the methods used in this thesis to study protein-protein interactions computationally, which include protein-protein docking, protein interface prediction algorithms, and the Adaptive Poisson Boltzmann Solver (APBS). However, it is worth noting that there are many other methods including Monte-Carlo based[61] as well as other

docking methodologies[62,63] that will not be covered here and the interested reader is directed to other resources.

Rigid-body protein docking is limited by the dynamics of the proteins involved. However, protein docking requires relatively low computational power and so multiple conformations of a single protein structure can be docking iteratively. Here, this method was employed in both the ASB9-CK project and the IKK2-NEMO project. The ZDOCK and ClusPro2.0 protein-docking servers were used to dock an array of models of each protein set into one another. The ZDOCK algorithm uses shape-complementarity, electrostatics, and desolvation energies in the scoring function, and poses are created using a fast Fourier transform (FFT) correlation technqiue.[64] The ClusPro2.0 server makes use of the FFT correlation technique in the ZDOCK algorithm but then filters the output based only on a linear combination of desolvation and electrostatic energies. Then the ClusPro2.0 algorithm clusters the resulting docked complexes into various microstates, and based on statistical mechanic theory, presupposes that the most populated cluster is likely to be nearest to the native binding site. Input for both of these servers came from structures resulting from molecular dynamics simulations in the case of ASB9 and CK, and from anisotropic network models in the case of the IKK2-NEMO interaction.

Protein interface prediction algorithms make use of biophysical theories behind protein-protein interactions. For example, it is common to find large buried apolar surface areas in protein-protein interfaces, since the enthalpic contribution here to the protein complex is highly negative for proteins in aqueous environments. The methods used primarily in this thesis were server-based methods including metaPPISP (meta Protein-

protein Interaction Site Predictor) and SPPIDER (Solvent accessibility based Protein-Protein Interface iDEntification and Recognition).

Section 7: Molecular Dynamics simulations

Molecular dynamics simulations step the atoms of a protein through time through the use of empirical force fields, or force fields that are based on principles of molecular mechanics. In molecular mechanics, the atoms of a protein are treated as points in space, attached by strings to other points in space, with each point having an associated mass, charge, and volume. Several assumptions are implicated in the construction of molecular mechanics models. The first is known at the Born-Oppenheimer approximation, which states that, the motion of the nuclei and electrons of an atom can be separated and considered independently. This allows computational chemists to simplify the calculation of molecular motion from quantum mechanical, to a linear combination of spring forces through bonds and electrostatics or van der Waals forces through space.

The typical MD simulation utilizes a force field similar to the equation below as well as force field models for simulating water, the solvent for biological systems, where

$$V(\mathbf{r}) = \sum_{\text{bonds}} K_r(r - r_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{\text{eq}})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}\right] + \sum_{i<j} \frac{q_i q_j}{\epsilon R_{ij}}, \quad (11)$$

V(**r**) is the potential function for a molecular mechanic system. For the through bond summations including bonds, angles and dihedrals, K_r and $K\theta$ are spring constants, between bonds and angles respectively, and V_n is a weighting factor for dihedral angles (ϕ) away from an equilibrium value (γ). The last two summations in equation (7) are the

van der Waals potential, which is inversely proportional to the distance to the twelfth power between the atoms *i* and *j* and weighted by experimentally determined values A_{ij} and B_{ij}, and the electrostatic potential, which is proportional to the charges of the two atoms (q), and inversely proportional to the distance (R_{ij}).

While most force fields take the form of equation 7, there are a multitude of different available force fields with slightly different constants and weightings derived from different experimental values.[65,66] The main force fields currently under use are AMBER and CHARMM force fields.[67,68] Each has their own set of releases, starting in the 90s and continuing into the present day with Amberff12SB[69] and others. Details are still being worked out in the literature as to whether one force field biases towards a different secondary structure or which force fields are best for simulating a variety of proteins from stable proteins to intrinsically disordered proteins.[25,66,67,70–73] There are also many different water models afforded to each of these force fields, each representing the unique properties of water, especially the nonbonding orbitals of oxygen, in different ways. For example TIP3P waters create a pseudo bond between the two hydrogen atoms to keep the bond angle near experimentally determined 107.4^{o}.[74,75] Another force field model, TIP4P, adds a dummy atom at 120^{o} from the hydrogen atoms to maintain similarity to natural water.[76] Each water model comes with its own weaknesses and strengths, and in this thesis both the TIP3P and TIP4P water models are used.

Aside from explicit water models, there has been a recent increase in the use of implicit solvent methods for two reasons: decrease in computational load of the simulations and increase in dynamic conformational space sampled by a protein. Lower viscosity that is characteristic of implicit solvent models lends itself to accelerate protein

dynamics.[77,78] In the past few years, implicit solvent methods have improved and complementary force fields have been identified for performing implicit solvent MD.[79–81] Thus implicit solvent methods are a good option for exploring long-timescale dynamics of non-globular proteins and elongated structures; explicit simulations waste computing resources calculating the positions of water with protein models that do not accurately recapitulate experimental measurements of water anyways.

The greatest issue facing molecular dynamics simulations is recapitulating experimental observables, which is an issue due to the sampling limit, restricting conformational observations to a single part of phase space.[20,82] To overcome these challenges, scientists have adopted methods to speed up calculations and have designed computational hardware capable of increasing rates of simulation a thousand times over.

Two central approximations often used by computational chemists are the periodic boundary conditions and the particle mesh Ewald method. Periodic boundary conditions use the same principles that were found in the 1990's cell phone game, snake. When a molecule leaves through one surface, it enters directly back in through the surface 180^{o} in the opposite direction.[83] This approximation allows the reduction of total number of molecules simulations in an all-atom system. Particle mesh Ewald method speeds up the calculation by introducing an approximation in the through bond calculations, the most computationally expensive part of a force field calculation.[84] After a certain distance cutoff, approximately 10 Å, the electrostatics calculations are no longer explicitly calculated but instead are treated via a fast Fourier transform, which speeds up the calculation from an order of N^2 the calculation is an N•log(N) method.[84]

Multiple accelerated simulation methodologies also exist including accelerated molecular dynamics, metadynamics, temperature enhanced molecular dynamics, replica exchange molecular dynamics and more. Here we utilize the accelerated molecular dynamics method, and explain the details herein. For a more extensive review of accelerated simulation methodologies, please refer to Adcock *et al*.[23]

The design of computational hardware for speeding up simulations is perhaps the most effective means to date of increasing the sampling limit of MD simulations and consists of moving away from performing simulations on unparalleled central processing units (CPUs). The work in this thesis was performed both on Graphics Processing Units (GPUs), highly parallelized CPUs in supercomputers, and application specialized integrated circuit (ASIC) as part of the Anton supercomputer.

The computational speed-up from GPUs is a result of both higher computational power and high memory bandwidth needed for storing force, velocity and positive values for each atom in a trajectory. In contrast, the Anton supercomputer is made up of ASICs that were specifically designed to perform molecular dynamics simulations, with a torus

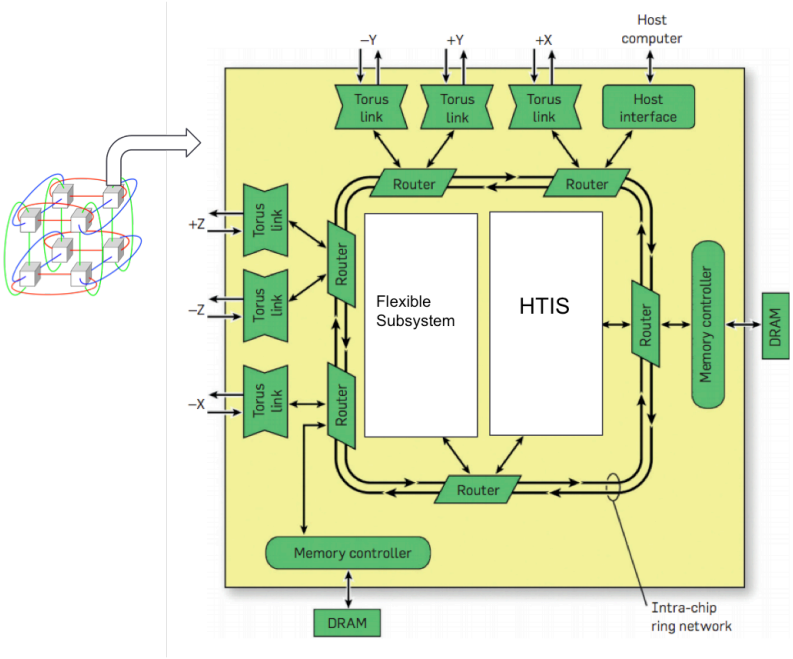


Figure 1.7: Anton's Application Specific Integrated Circuit (ASIC). This is a graphical representation of the ASIC in Anton with the flexible subsystem that performs the integration and through bond force calculations and the high throughput interaction subsystem (HTIS) that performs the through space force calculations. On the top left is showing the torus architecture, which the Anton supercomputer adopts for all 512 ASICs.

architecture and large local memory storage on each ASIC.[85,86] Though this machine is more closely related to a smart phone rather than to a typical computer, it is able to perform computational simulations that are more than two orders of magnitude faster than any other supercomputer composed of CPUs or GPUs.

Section 8: Methods for aligning simulation with experiment

Results from computational simulation are incredibly powerful and provide exquisitely detailed models for molecular processes. However, these results are just models and interpretation must be done in the context of how well these results align with experiment. In recent years, there has been a call-to-arms for scientists to keep creating tools for directly comparing experimental observables with all-atom details, and multiple tools have already been created.[87] Here, I review methods for aligning simulation with HDXMS, NMR, SAXS, and ITC or mutagenesis results.

Determining levels of solvent exposure in a protein backbone can reveal which domains of a protein are more prone to local unfolding. Uniting molecular dynamics simulations with HDXMS experiments can permit an understanding at the atomic level of how these solvent exposure events occur. Multiple methods have been employed in the literature for comparing results from HDXMS and computational simulations. Craig *et al* calculated predicted protection factors from protein folding simulation results using a coarse-grained Gö model.[88] These protection factors were calculated based on the number of native contacts and the distance between H-bonded residues found in the native state. Another approach for comparing simulation results with HDXMS protection factors is to compare solvent accessible surface areas (SASA) over the simulation for particular residues.[89] Others still have measured the internal distances between β sheet center of masses or other secondary structural changes on a molecular level.[90,91] In this thesis, instead of looking at the molecular level, hydrogen bonds with water are measured directly at the atomic level to understand the probabilities of various backbone amides for

interacting with water's nucleophile, oxygen.[92] While this analysis is the most detailed in the literature, it suffers from the lower timescales sampled compared to the other simulation methods used.

Compared to HDXMS, NMR observables provide multiple avenues for comparing directly to simulation results. NMR measurements like order parameters, RDCs and chemical shifts can be back calculated from simulation results.[89,93–98] Alternatively, kinetic and thermodynamic values calculated from NMR observables can also be compared to these same values calculated from MD simulations. In this thesis, the focus is on back calculating NMR chemical shifts and chemical shift differences that are realized from CPMG relaxation dispersion experiments, and so the focus here is on the multiple methodologies for back-calculating chemical shifts.[36]

Typical protein chemical shift prediction programs fall in to one of three categories: calculations based on sequence similarity through the use of chemical shift databases,[99,100] calculations based internal distances and atomic coordinates,[101] and a combination of the two.[99,102] Here, the combination method is used as this methodology has the best agreement with experiment to date.[99] However, calculation of chemical shifts from protein structures is still a heavy area of research and much improvement in these methodologies must occur before these practices become commonplace.[103–105]

Where as NMR has been utilized for decades to understand protein dynamical ensembles, SAXS is a relatively new method, and so fewer means exist for aligning experimental observables with computational models. However, an excellent tool from the Sali laboratory known as the fast open-source X-ray scattering (FOXS) allows the community to quickly calculate theoretical Debye scattering profiles based on atomic

level models and then compares these models to the actual SAXS profile.[106] This methodology has been integrated into other computational approaches such as in protein-protein docking methodologies[107] as well as metadynamics simulations.[108] Here we use the results to align solution structure scattering profiles of the ASB9-CK hetero-trimer with an ensemble of structures obtained from all-atom molecular dynamics simulations.

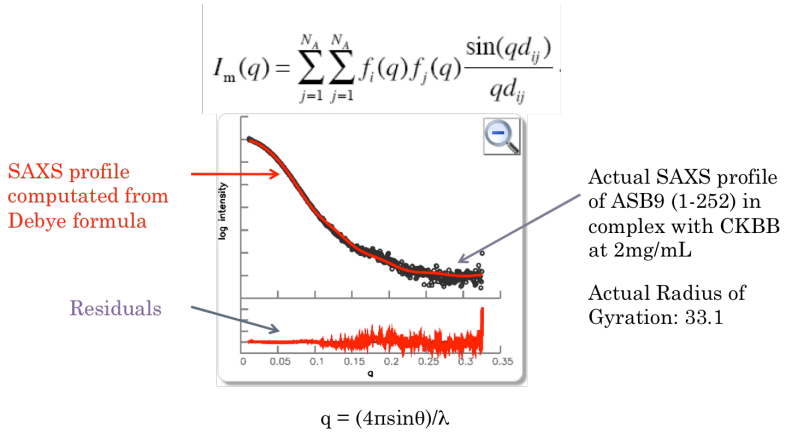


Figure 1.8: The Fast Open-Source X-ray Scattering (FOXS) server. This figure shows how the FOXS data is output upon solving the Debye scattering formula for a protein.

Finally, aligning simulation results with ITC and mutagenesis results can provide unique insight into the relationship between the thermodynamics and stabilities of a protein system and the conformational structure that underlie these energetic fluctuations. Dynamic effects of mutations on things like protein-protein interactions or enzymatic activity can be illuminated through computational simulations of wild-type proteins and mutants. Additionally, computational alanine scanning methods applied through Molecular Mechanics/Poisson-Boltzmann Surface Area (MM-PBSA) or through the computational more expensive thermodynamic integration (TI) methodologies can also be used to compare with experimental alanine scanning observables.[103] Similar

methodologies can also be used to estimate the binding energies from ITC results, however, these results are still plagued by systematic errors and future research must improve these methodologies.[104,105]

Section 9: A brief introduction to three very different protein systems

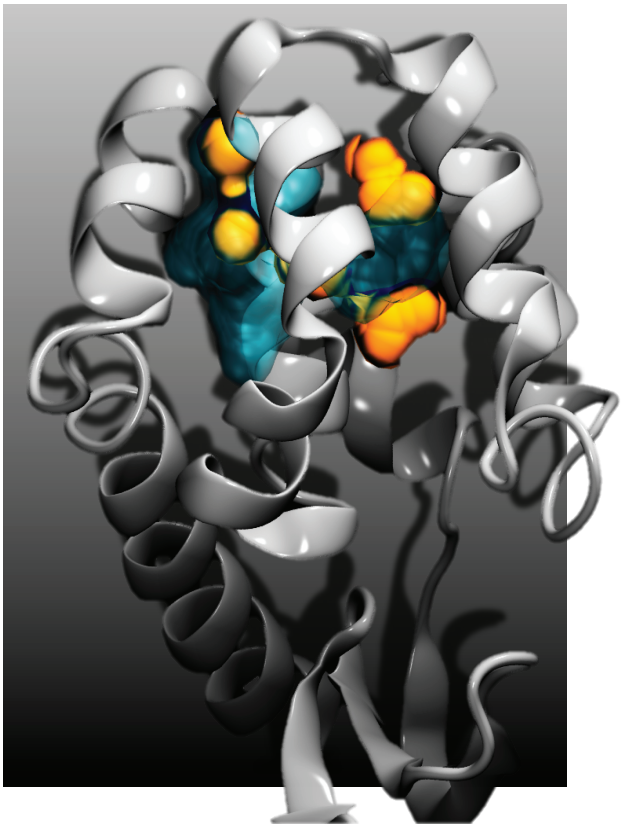


Figure 1.9: Mobile Defects in Engineered Protein Cavities. This image of the L99A mutant of T4 lysozyme shows the buried cavities that are present in the ground state (cyan) and the excited state (orange).

Often times, proteins perform biological or engineered functions on the microsecond timescale. When this is the case, these functions typically come about because of concerted backbone and side chain motions.[12,14,24] Sometimes, these concerted motions guide transitions from a ground state structure or ensemble, to an excited state structure.[22,35,36,40,109] These high-energy states are defined by low populations in the protein ensemble at standard temperature and pressure. These excited states are therefore difficult to resolve through experimental ensemble based methods. However, methods have been developed to determine populations, energy, and rates of exchange between

these high energy states.[97,110,111] These methods include NMR spectroscopy, and CPMG spin relaxation measurements, which have been previously discussed in section 5 of chapter 1.

Herein, results from computational simulations on the L99A mutant of T4 lysozyme, a model system for protein high-energy "invisible" states, are presented. The L99A mutant is a small protein composed of just one hundred and sixty-four residues.[28,35–38,42] This work reveals a previously unseen connection between excited state transitioning and egress of ligands bound to a buried cavity in the C-terminal domain of L99A. Additionally, these results add to a body of data which investigates the role of water in mobile defects in engineered protein cavities.[112,113] For further introduction into wild type T4 lysozyme structure, L99A structure and dynamics, and pioneering work in the fields of NMR spectroscopy and computational simulations please see chapters 2 and 3.

Proteins interact with one another on a variety of timescales, and these interactions can pass on critical information from one protein to the next. Here, work on a tight protein-protein interaction between ASB9 and CK is presented in chapters 4 and 5.[114]

The ankyrin repeat protein is a multi-domain protein composed of two hundred and ninety-four residues,[115–118] while creatine kinase forms a dimer in the cytoplasm,[119,120] and each monomer is composed of three hundred and eight-one residues. While ASB9 and CK interact on the microsecond timescale, their rates of dissociation from one another are too slow to measure experimentally,[114] suggesting that these proteins are found bonded inside our cells. The critical information in our cells in this

interaction is passed from CK through ASB9 to the rest of a huge protein complex that these two proteins are found in: E3 ligases.[121] E3 ligases are responsible for targeting proteins for

Figure 1.10: A Hetero-trimeric Protein Complex Critical for Energy Regulation in Our Cells. This image illuminates the structure of creatine kinase (CK, brown and green) in complex with the ankyrin repeat and SOCS box protein 9 (ASB9, blue). ASB9 has seven ankyrin repeat motifs and interacts with creatine kinase through its intrinsically disordered N-terminal domain and through ankyrin repeats 1 and 2.

degradation, as a control mechanism over cellular protein levels. As such, these protein complexes are incredibly important to cellular health,[122–125] and misregulation could lead to malignancies.[126–128] Work here is the first in the literature to provide a possible

mechanism for how Cullin-RING (CRL) E3 ligases are able to target proteins, like CK, for ubiquitin mediate degradation in our cells. To learn more about ASB9, CK, and E3 ligases, please see chapters 5 and 6.

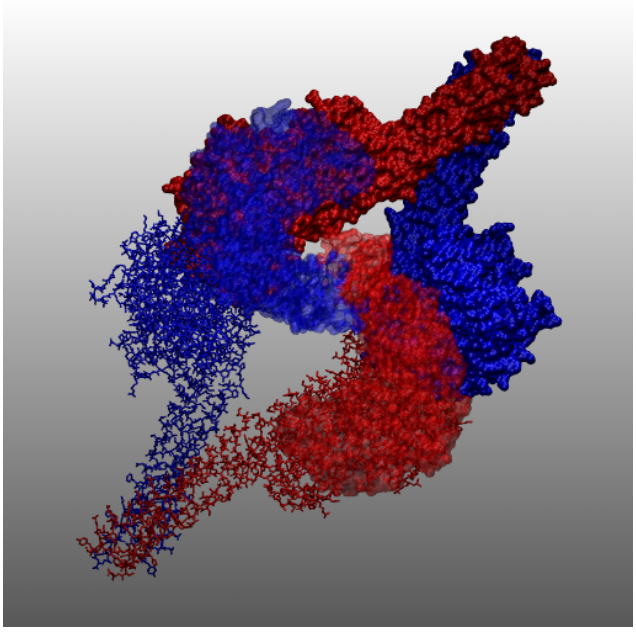


Figure 1.11: The Hexameric Structure of the IKK2/1 Signalasome. The red structures show the IKK1 monomers and the blue structures show the IKK2 monomers. These protein heterodimers interact with one another through the kinase domains

Finally, some proteins are quite large, and can serve as docking sites for other proteins to attract these proteins the same vicinity in the interior of a cell. For the IKK2 protein complex, or signalasome, these events can be difficult to study and occur transiently that measuring the timescales is still an area of investigation. IKK2 is a seven hundred and fifty-six residue protein that is found primarily as a dimer in cells.[129–134] However, recent unpublished work suggests that these proteins can transiently associate as hexamers in the cell to act like cellular aircraft carriers, with various other proteins docking onto to them and leaving when they have performed their tasks.

IKK2 is a kinase that can both phosphorylate other IKK2 molecules, in a process called autophosphorylation,[129] and phosphorylate proteins to activate them or target them for ubiquitin mediate degradation, like IκBs.[128,129,132,135] The ability of IKK2 to phosphorylate IκBs is dependent upon the interaction of IKK2 first with another protein, NEMO.[136,137] Here, protein-protein docking methods, anisotropic network models, and an array of other computational techniques are used to provide a preliminary model for the IKK2-NEMO interaction and the effect of this interaction on IKK2 dynamics. For more information on all proteins involved in this complex signal transduction pathway, please see chapter 7.

Section 10: References

(1) Gsponer, J.; Futschik, M. E.; Teichmann, S. A.; Babu, M. M. Tight Regulation of Unstructured Proteins: From Transcript Synthesis to Protein Degradation. *Science (80-.).* **2008**, No. 322, 1365–1368.

(2) Vavouri, T.; Semple, J. I.; Garcia-Verdugo, R.; Lehner, B. Intrinsic Protein Disorder and Interaction Promiscuity Are Widely Associated with Dosage Sensitivity. *Cell* **2009**, *138* (1), 198–208.

(3) Dyson, H. J. Making Sense of Intrinsically Disordered Proteins. *Biophysj* **2016**, *110* (5), 1013–1016.

(4) Bergeron-Sandoval, L.-P.; Safaee, N.; Michnick, S. W. Mechanisms and Consequences of Macromolecular Phase Separation. *Cell* **2016**, *165* (5), 1067–1079.

(5) Panagiotopoulos, A. Z. Statistical Mechanical Ensembles †. In *Statistical Thermodynamics*; 2014; pp 1–22.

(6) McQuarrie, D. A. *Statistical Mechanics*, 1st ed.; University Science Books: Mill Valley, 2000.

(7) Linderstrøm-Lang, K. U.; Schellman, J. A. The Enzymes. *Enzym.* **1959**, *1*, 443.

(8) Englander, S. W.; Mayne, L.; Bai, Y.; Sosnick, T. R. Hydrogen Exchange : The Modem Legacy of Linderstrom-Lang. *Protein Sci.* **1997**, *6*, 1101–1109.

(9) Straub, F. B. No Title. *Adv. Enzymol.* **1964**, *26* (89).

(10) Hammes, G. G.; Chang, Y.; Oas, T. G. Conformational Selection or Induced Fit : A Flux Description of Reaction Mechanism. *Proc Nat Acad Sci USA* **2009**, *106* (33), 13737–13741.

(11) Vogt, A. D.; Cera, E. Di. Conformational Selection or Induced Fit? A Critical Appraisal of the Kinetic Mechanism. *Biochemistry* **2012**, *51*, 5894–5902.

(12) Kharakoz, D. P. Protein Compressibility, Dynamics, and Pressure. *Biophys J* **2000**, *79* (1), 511–525.

(13) Karplus, M.; Kuriyan, J. Molecular Dynamics and Protein Function. **2005**, *102* (19).

(14) Henzler-Wildman, K.; Kern, D. Dynamic Personalities of Proteins. *Nature* **2007**, *450* (7172), 964–972.

(15) Palmer Kroenke, C. D., Loria, J. P., A. G. NMR Methods for Quantifying

Microsecond-to-Millisecond Motions in Biological Macromolecules. *Methods Enzym.* **2001**, *339*, 204–238.

(16) Kendrew, J. C.; Bodo, G.; DIntzis, H. M.; Parrish, R. G.; Wyckoff, H. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **1958**, *191*, 662–666.

(17) Austin, R. H.; Beeson, K. W.; Eisenstein, L.; Frauenfelder, H.; Gunsalus, I. C. Dynamics of Ligand Binding to Myoglobin. *Biochemistry* **1973**, *14* (24), 5355–5373.

(18) McCammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of Folded Proteins. *Nature* **1977**, *267*, 585–590.

(19) Shaw, D. E. DE; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. a; Jumper, J. M.; Salmon, J. K.; Shan, Y.; et al. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science (80-.).* **2010**, *330* (October), 341–346.

(20) Pierce, L. C. T.; Salomon-Ferrer, R.; Augusto F. De Oliveira, C.; McCammon, J. A.; Walker, R. C. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J. Chem. Theory Comput.* **2012**, *8* (9), 2997–3002.

(21) Weininger, U.; Modig, K.; Akke, M. Ring Flips Revisited: 13C Relaxation Dispersion Measurements of Aromatic Side Chain Dynamics and Activation Barriers in Basic Pancreatic Trypsin Inhibitor. *Biochemistry* **2014**, *53* (28), 4519–4525.

(22) Xue, Y.; Ward, J. M.; Yuwen, T.; Podkorytov, I. S.; Skrynnikov, N. R. Microsecond Time-Scale Conformational Exchange in Proteins: Using Long Molecular Dynamics Trajectory to Simulate NMR Relaxation Dispersion Data. *J. Am. Chem. Soc.* **2012**, *134* (5), 2555–2562.

(23) Adcock, S. a; Mccammon, J. A. Molecular Dynamics : Survey of Methods for Simulating the Activity of Proteins Molecular Dynamics : Survey of Methods for Simulating the Activity of Proteins. **2006**, *106* (February), 1589–1615.

(24) Karplus, M.; Gelin, B. R.; McCammon, J. a. Internal Dynamics of Proteins. Short Time and Long Time Motions of Aromatic Sidechains in PTI. *Biophys. J.* **1980**, *32* (1), 603–618.

(25) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins Struct. Funct. Bioinforma.* **2010**, *78* (8), 1950–1958.

(26) Baase, W. A.; Liu, L.; Tronrud, D. E.; Matthews, B. W. Lessons from the

Lysozyme of Phage T4. *Protein Sci.* **2010**, *19* (4), 631–641.

(27) Eriksson, A. E.; Baase, W. A.; Wozniak, J. A.; Matthews, B. W. A Cavity-Containing Mutant of T4 Lysozyme Is Stabilized by Buried Benzene. *Nature* **1992**, *355* (6358), 371–373.

(28) Maeno, A.; Sindhikara, D.; Hirata, F.; Otten, R.; Dahlquist, F. W.; Yokoyama, S.; Akasaka, K.; Mulder, F. A. A.; Kitahara, R. Cavity as a Source of Conformational Fluctuation and High-Energy State: High-Pressure NMR Study of a Cavity-Enlarged Mutant of T4Lysozyme. *Biophys. J.* **2015**, *108* (1), 133–145.

(29) Ando, N.; Barstow, B.; Baase, W. A.; Fields, A.; Matthews, B. W.; Gruner, S. M. Structural and Thermodynamic Characterization of T4 Lysozyme Mutants and the Contribution of Internal Cavities to Pressure Denaturation †. *Biochemistry* **2008**, *47* (42), 11097–11109.

(30) Eriksson, A. E.; Baase, W. A.; Zhang, X. J.; Heinz, D. W.; Blaber, M.; Baldwin, E. P.; Matthews, B. W. Response of a Protein Structure to Cavity-Creating Mutations and Its Relation to the Hydrophobic Effect. *Science (80-.).* **1992**, *255*, 178–183.

(31) Mann, G.; Hermans, J. Modeling Protein-Small Molecule Interactions: Structure and Thermodynamics of Noble Gases Binding in a Cavity in Mutant Phage T4 Lysozyme L99A. *J. Mol. Biol.* **2000**, *302* (4), 979–989.

(32) Deng, Y.; Roux, B. Calculation of Standard Binding Free Energies: Aromatic Molecules in the T4 Lysozyme L99A Mutant. *J. Chem. Theory Comput.* **2006**, *2* (5), 1255–1273.

(33) Purisima, E. O.; Hogues, H. Protein-Ligand Binding Free Energies from Exhaustive Docking. *J. Phys. Chem. B* **2012**, *116* (23), 6872–6879.

(34) Boyce, S. E.; Mobley, D. L.; Rocklin, G. J.; Graves, A. P.; Dill, K. A.; Shoichet, B. K. Predicting Ligand Binding Affinity with Alchemical Free Energy Methods in a Polar Model Binding Site. *J. Mol. Biol.* **2009**, *394* (4), 747–763.

(35) Mulder, F. A. A.; Mittermaier, A.; Hon, B.; Dahlquist, F. W.; Kay, L. E. Studying Excited States of Proteins by NMR Spectroscopy. *Nat. Struct. Biol.* **2001**, *8* (11), 932–935.

(36) Schiffer, J. M.; Feher, V. A.; Malmstrom, R. D.; Sida, R.; Amaro, R. E. Capturing Invisible Motions in the Transition from Ground to Rare Excited States of T4 Lysozyme L99A. *Biophys. J.* **2016**, *111*, 1–10.

(37) Korzhnev, D. M.; Orekhov, V. Y.; Dahlquist, F. W.; Kay, L. E. Off-Resonance R1rho Relaxation outside of the Fast Exchange Limit: An Experimental Study of a Cavity Mutant of T4 Lysozyme. *J Biomol NMR* **2003**, *26*, 39–48.

(38) Lerch, M. T.; López, C. J.; Yang, Z.; Kreitman, M. J.; Horwitz, J.; Hubbell, W. L. Structure-Relaxation Mechanism for the Response of T4 Lysozyme Cavity Mutants to Hydrostatic Pressure. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (19), E2437–E2446.

(39) Skrynnikov, N. R.; Dahlquist, F. W.; Kay, L. E. Reconstructing NMR Spectra Of "invisible" excited Protein States Using HSQC and HMQC Experiments. *J. Am. Chem. Soc.* **2002**, *124* (41), 12352–12360.

(40) Vallurupalli, P.; Hansen, D. F.; Kay, L. E. Structures of Invisible, Excited Protein States by Relaxation Dispersion NMR Spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (33), 11766–11771.

(41) Bouvignies, G.; Vallurupalli, P.; Hansen, D. F.; Correia, B. E.; Lange, O.; Bah, A.; Vernon, R. M.; Dahlquist, F. W.; Baker, D.; Kay, L. E. Solution Structure of a Minor and Transiently Formed State of a T4 Lysozyme Mutant. *Nature* **2011**, *477* (7362), 111–114.

(42) Vallurupalli, P.; Chakrabarti, N.; Pomes, R.; Kay, L. Atomistic Picture of Conformational Exchange in a T4 Lysozyme Cavity Mutant: An Experiment-Guided Molecular Dynamics Study. *Chem. Sci.* **2016**, *3*, 3602–3613.

(43) Mulder, F. A. A.; Hon, B.; Muhandiram, D. R.; Dahlquist, F. W.; Kay, L. E. Flexibility and Ligand Exchange in a Buried Cavity Mutant of T4 Lysozyme Studied by Multinuclear NMR. *Biochemistry* **2000**, *39* (41), 12614–12622.

(44) Merski, M.; Fischer, M.; Balius, T. E.; Eidam, O.; Shoichet, B. K. Homologous Ligands Accommodated by Discrete Conformations of a Buried Cavity. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (16), 5039–5044.

(45) Quillin, M. L.; Breyer, W. A.; Griswold, I. J.; Matthews, B. W. Size versus Polarizability in Protein-Ligand Interactions: Binding of Noble Gases within Engineered Cavities in Phage T4 Lysozyme. *J. Mol. Biol.* **2000**, *302* (4), 955–977.

(46) The Nobel Prize in Physiology or Medicine. **1908**.

(47) Hedin, S. G. Further Observations on the Time-Relations in the Action of Trypsin. *J. Physiol.* **1905**.

(48) Braun, P.; Gingras, A. History of Protein – Protein Interactions : From Egg-White. *Proteomics* **2012**, *12*, 1478–1498.

(49) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: An Automated Pipeline for the Setup of Poisson-Boltzmann Electrostatics Calculations. *Nucleic Acids Res.* **2004**, *32* (WEB SERVER ISS.), 665–667.

(50) Holst, M. J.; Saied, F. Numerical-Solution of the Nonlinear Poisson-Boltzmann

Equation: Developing More Robust and Efficient Methods. *J. Comput. Chem.* **1995**, *16* (3), 337–364.

(51) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (18), 10037–10041.

(52) Phizicky, E. M.; Fields, S. Protein-Protein Interactions : Methods for Detection and Analysis. **1995**, *59* (1), 94–123.

(53) Pierce, M. M.; Raman, C. S.; Nall, B. T. Isothermal Titration Calorimetry of Protein – Protein Interactions. **1999**, *221*, 213–221.

(54) Mcdonnell, J. M. Surface Plasmon Resonance : Towards an Understanding of the. **2001**, 572–577.

(55) Patching, S. G. Biochimica et Biophysica Acta Surface Plasmon Resonance Spectroscopy for Characterisation of Membrane Protein – Ligand Interactions and Its Potential for Drug Discovery ☆. *BBA - Biomembr.* **2014**, *1838* (1), 43–55.

(56) Torreri, P.; Ceccarini, M.; Macioce, P.; Petrucci, T. C.; Cellulare, B. Biomolecular Interactions by Surface Plasmon Resonance Technology. **2005**, *41* (4), 437–441.

(57) Merwe, P. A. Van Der. Surface Plasmon Resonance GENERAL PRINCIPLES OF BIACORE EXPERIMENTS. 1–50.

(58) Cavanagh, J.; Fairbrother, W. J.; Palmer III, A. G.; Rance, M.; Skelton, N. J. *Protein NMR Spectroscopy: Principles and Practice*; 2007.

(59) Mandell, J. G.; Falick, A. M.; Komives, E. A. Identification of Protein-Protein Interfaces by Decreased Amide Proton Solvent Accessibility. *Proc Nat Acad Sci USA* **1998**, *95*, 14705–14710.

(60) Percy, A. J.; Rey, M.; Burns, K. M.; Schriemer, D. C. Analytica Chimica Acta Probing Protein Interactions with Hydrogen / Deuterium Exchange and Mass Spectrometry — A Review. *Anal. Chim. Acta* **2012**, *721*, 7–21.

(61) Tremmel, I. G.; Weis, E.; Farquhar, G. D. The Influence of Protein-Protein Interactions on the Organization of Proteins within Thylakoid Membranes. *Biophys. J.* **2005**, *88* (4), 2650–2660.

(62) Kozakov, D.; Beglov, D.; Bohnuud, T.; Mottarella, S. E.; Xia, B.; Hall, D. R.; Vajda, S. How Good Is Automated Protein Docking? *Proteins Struct. Funct. Bioinforma.* **2013**, *81* (12), 2159–2166.

(63) Vajda, S.; Kozakov, D. Convergence and Combination of Methods in Protein – Protein Docking. **2009**, 164–170.

(64) Pierce, B. G.; Wiehe, K.; Hwang, H.; Kim, B. H.; Vreven, T.; Weng, Z. ZDOCK Server: Interactive Docking Prediction of Protein-Protein Complexes and Symmetric Multimers. *Bioinformatics* **2014**, *30* (12), 1771–1773.

(65) Ponder, B. J. A. Y. W.; À, D. A. C. FORCE FIELDS FOR PROTEIN SIMULATIONS I . Introduction Molecular Dynamics and Monte Carlo Simulations of Proteins , Which Directions the Field Is Taking in Developing New Models . Simple (Sometimes Called “‘ Class I ’”) Potential Energy Function : **2003**, *66*, 27–85.

(66) Martín-García, F.; Papaleo, E.; Gomez-Puertas, P.; Boomsma, W.; Lindorff-Larsen, K. Comparing Molecular Dynamics Force Fields in the Essential Subspace. *PLoS One* **2015**.

(67) Hornak, V. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. Proteins: Struct., Funct., Bioinf. 2006, pp 572–581.

(68) Vitalini, F.; Mey, A. S. J. S.; Noé, F.; Keller, B. G. Dynamic Properties of Force Fields. *J. Chem. Phys.* **2015**, *142* (8).

(69) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An Overview of the Amber Biomolecular Simulation Package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3* (2).

(70) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Systematic Validation of Protein Force Fields against Experimental Data. *PLoS One* **2012**, *7* (2), 1–6.

(71) Whitford, P. C.; Noel, J. K.; Gosavi, S.; Schug, A.; Sanbonmatsu, K. Y.; Onuchic, J. N. An All-Atom Structure-Based Potential for Proteins: Bridging Minimal Models with All-Atom Empirical Forcefields. *Proteins Struct. Funct. Bioinforma.* **2009**, *75* (2), 430–441.

(72) Henriques, J.; Cragnell, C.; Skepo, M. Molecular Dynamics Simulations of Intrinsically Disordered Proteins: Force Field Evaluation and Comparison with Experiment. *J. Chem. Theory Comput.* **2015**, *11* (7), 3420–3431.

(73) Rauscher, S.; Gapsys, V.; Gajda, M. J.; Zweckstetter, M.; De Groot, B. L.; Grubm??ller, H. Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field: A Comparison to Experiment. *J. Chem. Theory Comput.* **2015**, *11* (11), 5513–5524.

(74) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926.

(75) Smith, M. D.; Rao, J. S.; Segelken, E.; Cruz, L. Force-Field Induced Bias in the Structure of Aβ21-30: A Comparison of OPLS, AMBER, CHARMM, and GROMOS Force Fields. *J. Chem. Inf. Model.* **2015**, *55* (12), 2587–2595.

(76) Julin, J.; Shiraiwa, M.; Miles, R. E. H.; Reid, J. P.; P??schl, U.; Riipinen, I. Mass Accommodation of Water: Bridging the Gap between Molecular Dynamics Simulations and Kinetic Condensation Models. *J. Phys. Chem. A* **2013**, *117* (2), 410–420.

(77) Nguyen, H.; Maier, J.; Huang, H.; Perrone, V.; Simmerling, C. Folding Simulations for Proteins with Diverse Topologies Are Accessible in Days with a Physics-Based Force Field and Implicit Solvent. **2014**, 1–4.

(78) Nguyen, H.; Roe, D. R.; Simmerling, C. Improved Generalized Born Solvent Model Parameters for Protein Simulations. **2013**.

(79) Ma, I.; Contini, A.; Farmaceutiche, S.; Generale, C.; Marchesini, A. An Updated Test of AMBER Force Fields and Implicit Solvent Models in Predicting the Secondary Structure of Helical, β - Hairpin, and Intrinsically Disordered Peptides. **2016**.

(80) Anandakrishnan, R.; Drozdetski, A.; Walker, R. C.; Onufriev, A. V. Article Speed of Conformational Change : Comparing Explicit and Implicit Solvent Molecular Dynamics Simulations. *Biophysj* **2015**, *108* (5), 1153–1164.

(81) Perez, A.; Maccallum, J. L.; Brini, E.; Simmerling, C.; Dill, K. A. Grid-Based Backbone Correction to the Ff 12SB Protein Force Field for Implicit-Solvent Simulations. **2015**.

(82) Lane, T. J.; Shukla, D.; Beauchamp, K. A.; Pande, V. S. To Milliseconds and beyond: Challenges in the Simulation of Protein Folding. *Curr. Opin. Struct. Biol.* **2013**, *23* (1), 58–65.

(83) Roe, D. R.; Cheatham III, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Synamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084–3095.

(84) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089.

(85) Shaw, D. E.; Bowers, K. J.; Chow, E.; Eastwood, M. P.; Ierardi, D. J.; Klepeis, J. L.; Kuskin, J. S.; Larson, R. H.; Lindorff-Larsen, K.; Maragakis, P.; et al. Millisecond-Scale Molecular Dynamics Simulations on Anton. *Proc. Conf. High Perform. Comput. Netw. Storage Anal. SC 09* **2009**.

(86) Scarpazza, D. P.; Ierardi, D. J.; Lerer, A. K.; Mackenzie, K. M.; Pan, A. C.; Bank, J. A.; Chow, E.; Dror, R. O.; Grossman, J. P.; Killebrew, D.; et al. Extending the

Generality of Molecular Dynamics Simulations on a Special-Purpose Machine. *Proc. - IEEE 27th Int. Parallel Distrib. Process. Symp. IPDPS 2013* **2013**, 933–945.

(87) Papaleo, E. Integrating Atomistic Molecular Dynamics Simulations , Experiments , and Network Analysis to Study Protein Dynamics : Strength in Unity. **2015**, *2* (May), 1–6.

(88) Craig, P. O.; Joachim, L.; Weinkam, P.; Ho, R. M. B.; Ferreiro, D. U.; Komives, E. A.; Wolynes, P. G. Prediction of Native-State Hydrogen Exchange from Perfectly Funneled Energy Landscapes. **2011**, 17463–17472.

(89) Patel, S.; Vierling, E.; Tama, F. Replica Exchange Molecular Dynamics Simulations Provide Insight into Substrate Recognition by Small Heat Shock Proteins. *Biophysj* **2014**, *106* (12), 2644–2655.

(90) Petruk, A. A.; Defelipe, L. A.; Marti, M. A.; Turjanski, A. G. Molecular Dynamics Simulations Provide Atomistic Insight into Hydrogen Exchange Mass Spectrometry Experiments. **2013**, No. ii.

(91) Mao, Y.; Zlatic, C. O.; Gri, M. D. W.; Howlett, J.; Todorova, N.; Yarovsky, I.; Gooley, P. R. Hydrogen/Deuterium Exchange and Molecular Dynamics Analysis of Amyloid Fibrils Formed by a D69K Charge-Pair Mutant of Human Apolipoprotein C - II. **2015**.

(92) Schiffer, J. M.; Malmstrom, R. D.; Parnell, J.; Ramirez-Sarmiento, C.; Reyes, J.; Amaro, R. E.; Komives, E. A. Model of the Ankyrin and SOCS Box Protein, ASB9, E3 Ligase Reveals a Mechanism for Dynamic Ubiquitin Transfer. *Structure* **2016**, *24* (8), 1248–1256.

(93) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; Van Gunsteren, W. F.; Mark, A. E. Peptide Folding: When Simulation Meets Experiment. *Angew. Chemie Int. Ed.* **1999**, *38*, 236–240.

(94) Peter, C.; Daura, X.; Gunsteren, W. F. Van. Calculation of NMR-Relaxation Parameters for Flexible Molecules from Molecular Dynamics Simulations. **2001**, No. 1982, 297–310.

(95) Fuglestad, B.; Gasper, P. M.; Tonelli, M.; McCammon, J. A.; Markwick, P. R. L.; Komives, E. A. The Dynamic Structure of Thrombin in Solution. *Biophys. J.* **2012**, *103* (1), 79–88.

(96) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of Slow Molecular Order Parameters for Markov Model Construction. *J. Chem. Phys.* **2013**, *139* (1), 0–13.

(97) Skrynnikov, N. R.; Mulder, F. A. A.; Hon, B.; Dahlquist, F. W.; Kay, L. E.

Probing Slow Time Scale Dynamics at Methyl-Containing Side Chains in Proteins by Relaxation Dispersion NMR Measurements: Application to Methionine Residues in a Cavity Mutant of T4 Lysozyme. *J. Am. Chem. Soc.* **2001**, *123* (19), 4556–4566.

(98) Gu, Y.; Li, D.; Bru, R. NMR Order Parameter Determination from Long Molecular Dynamics Trajectories for Objective Comparison with Experiment. **2014**.

(99) Han, B.; Liu, Y.; Ginzinger, S. W.; Wishart, D. S. SHIFTX2: Significantly Improved Protein Chemical Shift Prediction. *J. Biomol. NMR* **2011**, *50* (1), 43–57.

(100) Shen, Y.; Bax, Æ. A. Protein Backbone Chemical Shifts Predicted from Searching a Database for Torsion Angle and Sequence Homology. **2007**, 289–302.

(101) Kohlhoff, K. J.; Robustelli, P.; Cavalli, A.; Salvatella, X.; Vendruscolo, M. Fast and Accurate Predictions of Protein NMR Chemical Shifts from Interatomic Distances. **2009**, 13894–13895.

(102) Wishart, D. S. Progress in Nuclear Magnetic Resonance Spectroscopy Interpreting Protein Chemical Shift Data. *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, *58* (1-2), 62–87.

(103) Martins, S. A.; Perez, M. A. S.; Moreira, I. S.; Sousa, F.; Ramos, M. J.; Fernandes, P. A. Computational Alanine Scanning Mutagenesis : MM-PBSA vs TI. **2013**.

(104) Genheden, S.; Ryde, U. The MM / PBSA and MM / GBSA Methods to Estimate Ligand-Binding Affinities. **2015**.

(105) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the Performance of MM/PBSA and MM/GBSA Methods: I. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2012**, *51* (1), 69–82.

(106) Schneidman-Duhovny, D.; Hammel, M.; Tainer, J. A.; Sali, A. Accurate SAXS Profile Computation and Its Assessment by Contrast Variation Experiments. *Biophys. J.* **2013**, *105* (4), 962–974.

(107) Schneidman-Duhovny, D.; Hammel, M.; Sali, A. Macromolecular Docking Restrained by a Small Angle X-Ray Scattering Profile. *J. Struct. Biol.* **2011**, *173* (3), 461–471.

(108) Kimanius, D.; Pettersson, I.; Schluckebier, G.; Lindahl, E.; Andersson, M. SAXS-Guided Metadynamics. *J. Chem. Theory Comput.* **2015**, *11* (7), 3491–3498.

(109) Masterson, L. R.; Yu, T.; Shi, L.; Wang, Y.; Gustavsson, M.; Mueller, M. M.; Veglia, G. CAMP-Dependent Protein Kinase A Selects the Excited State of the Membrane Substrate Phospholamban. *J. Mol. Biol.* **2011**, *412* (2), 155–164.

(110) Vallurupalli, P.; Hansen, D. F.; Lundstrom, P.; Kay, L. E. CPMG Relaxation Dispersion NMR Experiments Measuring Glycine 1H Alpha and 13C Alpha Chemical Shifts in the "Invisible" Excited States of Proteins. *J Biomol NMR* **2009**, *45*, 45–55.

(111) Mulder, F. A.; Skrynnikov, N. R.; Hon, B.; Dahlquist, F. W.; Kay, L. E. Measurement of Slow (Micros-Ms) Time Scale Dynamics in Protein Side Chains by (15)N Relaxation Dispersion NMR Spectroscopy: Application to Asn and Gln Residues in a Cavity Mutant of T4 Lysozyme. *J Am Chem Soc* **2001**, *123*, 967–975.

(112) Miyashita, O.; Onuchic, J. N.; Wolynes, P. G. Nonlinear Elasticity, Protein Quakes, and the Energy Landscapes of Functional Transitions in Proteins. *Proc. Natl. Acad. Sci.* **2003**, *100* (22), 12570–12575.

(113) Baron, R.; McCammon, J. A. Dynamics, Hydration, and Motional Averaging of a Loop-Gated Artificial Protein Cavity: The W191G Mutant of Cytochrome c Peroxidase in Water as Revealed by Molecular Dynamics Simulations. *Biochemistry* **2007**, *46* (37), 10629–10642.

(114) Balasubramaniam, D.; Schiffer, J.; Parnell, J.; Mir, S. P.; Amaro, R. E.; Komives, E. A. How the Ankyrin and SOCS Box Protein, ASB9, Binds to Creatine Kinase. *Biochemistry* **2015**, *54* (8), 1673–1680.

(115) Thomas, J. C.; Matak-Vinkovic, D.; Van Molle, I.; Ciulli, A. Multimeric Complexes among Ankyrin-Repeat and SOCS-Box Protein 9 (ASB9), ElonginBC, and Cullin 5: Insights into the Structure and Assembly of ECS-Type Cullin-RING E3 Ubiquitin Ligases. *Biochemistry* **2013**, *52* (31), 5236–5246.

(116) Muniz, J. R. C.; Guo, K.; Kershaw, N. J.; Ayinampudi, V.; Von Delft, F.; Babon, J. J.; Bullock, A. N. Molecular Architecture of the Ankyrin SOCS Box Family of Cul5-Dependent E3 Ubiquitin Ligases. *J. Mol. Biol.* **2013**, *425* (17), 3166–3177.

(117) Fei X Gu X, Qiu R, Mao Y, Ji C (2009) Protein Pept, Z. Y.; 16:333–335, L. Crystallization and Preliminary X-Ray Analysis of the Splice Variant of Human Ankyrin Repeat and Suppressor of Cytokine Signaling Box Protein 9 (hASB9-2). *Protein Pept Lett. 2009;16(3)333-5.* **2009**, *16*, 333–335.

(118) Debrincat, M. A.; Zhang, J. G.; Willson, T. A.; Silke, J.; Connolly, L. M.; Simpson, R. J.; Alexander, W. S.; Nicola, N. A.; Kile, B. T.; Hilton, D. J. Ankyrin Repeat and Suppressors of Cytokine Signaling Box Protein Asb-9 Targets Creatine Kinase B for Degradation. *J. Biol. Chem.* **2007**, *282* (7), 4728–4737.

(119) Bong, S. M.; Moon, J. H.; Nam, K. H.; Lee, K. S.; Chi, Y. M.; Hwang, K. Y. Structural Studies of Human Brain-Type Creatine Kinase Complexed with the ADP-Mg2+-NO3--Creatine Transition-State Analogue Complex. *FEBS Lett.* **2008**,

582 (28), 3959–3965.

(120) Shen, Y. Q.; Tang, L.; Zhou, H. M.; Lin, Z. J. Structure of Human Muscle Creatine Kinase. *Acta Crystallogr. D. Biol. Crystallogr.* **2001**, *57* (Pt 8), 1196–1200.

(121) Schiffer, J. M.; Malmstrom, R. D.; Parnell, J.; Ramirez-sarmiento, C.; Reyes, J.; Amaro, R. E.; Komives, E. A.; Schiffer, J. M.; Malmstrom, R. D.; Parnell, J.; et al. E3 Ligase Reveals a Mechanism for Dynamic Ubiquitin Transfer Article Model of the Ankyrin and SOCS Box Protein , ASB9 , E3 Ligase Reveals a Mechanism for Dynamic Ubiquitin Transfer. *Struct. Des.* **2016**, 1–9.

(122) Liu, J.; Nussinov, R. The Mechanism of Ubiquitination in the Cullin-RING E3 Ligase Machinery: Conformational Control of Substrate Orientation. *PLoS Comput. Biol.* **2009**, *5* (10).

(123) Duda, D. M.; Borg, L. A.; Scott, D. C.; Hunt, H. W.; Hammel, M.; Schulman, B. A. Structural Insights into NEDD8 Activation of Cullin-RING Ligases: Conformational Control of Conjugation. *Cell* **2008**, *134* (6), 995–1006.

(124) Brown, N. G.; VanderLinden, R.; Watson, E. R.; Qiao, R.; Grace, C. R. R.; Yamaguchi, M.; Weissmann, F.; Frye, J. J.; Dube, P.; Ei Cho, S.; et al. RING E3 Mechanism for Ubiquitin Ligation to a Disordered Substrate Visualized for Human Anaphase-Promoting Complex. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (17), 5272–5279.

(125) Liu, J.; Nussinov, R. The Role of Allostery in the Ubiquitin-Proteasome System. *Crit. Rev. Biochem. Mol. Biol.* **2013**, *48* (2), 89–97.

(126) Zhao, Y.; Sun, Y. Cullin-RING Ligases as Attractive Anti-Cancer Targets. *Curr. Pharm. Des.* **2013**, *19* (18), 3215–3225.

(127) Merlet, J.; Burger, J.; Gomes, J. E.; Pintard, L. Regulation of Cullin-RING E3 Ubiquitin-Ligases by Neddylation and Dimerization. *Cell. Mol. Life Sci.* **2009**, *66* (11-12), 1924–1938.

(128) Deshaies, R.; Joazeiro, C. RING Domain E3 Ubiquitin Ligases. *Annu. Rev. Biochem.* **2009**, *78*.

(129) Polley, S.; Huang, D. Bin; Hauenstein, A. V.; Fusco, A. J.; Zhong, X.; Vu, D.; Schröfelbauer, B.; Kim, Y.; Hoffmann, A.; Verma, I. M.; et al. A Structural Basis for IκB Kinase 2 Activation Via Oligomerization-Dependent Trans Auto-Phosphorylation. *PLoS Biol.* **2013**, *11* (6).

(130) Schmid, J. A.; Birbach, A. IκB Kinase β (IKKβ/IKK2/IKBKB)-A Key Molecule in Signaling to the Transcription Factor NF-κB. *Cytokine Growth Factor Rev.* **2008**, *19* (2), 157–165.

(131) Hauenstein, A. V.; Rogers, W. E.; Shaul, J. D.; Huang, D. Bin; Ghosh, G.; Huxford, T. Probing Kinase Activation and Substrate Specificity with an Engineered Monomeric IKK2. *Biochemistry* **2014**.

(132) Israël, A. The IKK Complex, a Central Regulator of NF-kappaB Activation. *Cold Spring Harb. Perspect. Biol.* **2010**, *2* (3), 1–14.

(133) Xu, G.; Lo, Y.-C.; Li, Q.; Napolitano, G.; Wu, X.; Jiang, X.; Dreano, M.; Karin, M.; Wu, H. Crystal Structure of Inhibitor of κB Kinase B. *Nature* **2011**, *472* (7343), 325–330.

(134) Liu, S.; Misquitta, Y. R.; Olland, A.; Johnson, M. A.; Kelleher, K. S.; Kriz, R.; Lin, L. L.; Stahl, M.; Mosyak, L. Crystal Structure of a Human IκB Kinase β Asymmetric Dimer. *J. Biol. Chem.* **2013**, *288* (31), 22758–22767.

(135) Patil, K. R.; Mohapatra, P.; Patel, H. M.; Goyal, S. N.; Ojha, S.; Kundu, C. N.; Patil, C. R. Pentacyclic Triterpenoids Inhibit Ikkβ-Mediated Activation of Nf-Kb Pathway: In Silico and in Vitro Evidences. *PLoS One* **2015**.

(136) May, M. J.; D'Acquisto, F.; Madge, L. A.; Glockner, J.; Pober, J. S.; Ghosh, S. Selective Inhibition of NF-kB Activation by a Peptide That Blocks the Interaction of NEMO with the IkB Kinase Complex. *Sci. Mag.* **2000**, *289*, 1550–1554.

(137) Rushe, M.; Silvian, L.; Bixler, S.; Chen, L. L.; Cheung, A.; Bowes, S.; Cuervo, H.; Berkowitz, S.; Zheng, T.; Guckian, K.; et al. Structure of a NEMO/IKK-Associating Domain Reveals Architecture of the Interaction Site. *Structure* **2008**, *16* (5), 798–808.

CHAPTER 2

Capturing Invisible Motions in the Transition from Ground to Excited States in T4 Lysozyme L99A

Please cite this article in press as: Schiffer et al., Capturing Invisible Motions in the Transition from Ground to Rare Excited States of T4 Lysozyme L99A, Biophysical Journal (2016), http://dx.doi.org/10.1016/j.bpj.2016.08.041

Biophysical *Journal*
Article

Biophysical Society

Capturing Invisible Motions in the Transition from Ground to Rare Excited States of T4 Lysozyme L99A

Jamie M. Schiffer,[1] Victoria A. Feher,[1,4,*] Robert D. Malmstrom,[1,2] Roxana Sida,[1,3] and Rommie E. Amaro[1,2,4,*]
[1]Department of Chemistry and Biochemistry and [2]National Biomedical Computation Resource, University of California, San Diego, La Jolla, California; [3]Centro de Enseñanza Técnica y Superior (CETYS) Campus Ensenada, Camino a Microondas Trinidad, Ensenada, Baja California, Mexico; and [4]Drug Design Data Resource, University of California, San Diego, La Jolla, California

ABSTRACT Proteins commonly sample a number of conformational states to carry out their biological function, often requiring transitions from the ground state to higher-energy states. Characterizing the mechanisms that guide these transitions at the atomic level promises to impact our understanding of functional protein dynamics and energy landscapes. The leucine-99-to-alanine (L99A) mutant of T4 lysozyme is a model system that has an experimentally well characterized excited sparsely populated state as well as a ground state. Despite the exhaustive study of L99A protein dynamics, the conformational changes that permit transitioning to the experimentally detected excited state (~3%, ΔG ~2 kcal/mol) remain unclear. Here, we describe the transitions from the ground state to this sparsely populated excited state of L99A as observed through a single molecular dynamics (MD) trajectory on the Anton supercomputer. Aside from detailing the ground-to-excited-state transition, the trajectory samples multiple metastates and an intermediate state en route to the excited state. Dynamic motions between these states enable cavity surface openings large enough to admit benzene on timescales congruent with known rates for benzene binding. Thus, these fluctuations between rare protein states provide an atomic description of the concerted motions that illuminate potential path(s) for ligand binding. These results reveal, to our knowledge, a new level of complexity in the dynamics of buried cavities and their role in creating mobile defects that affect protein dynamics and ligand binding.

INTRODUCTION

Proteins function through concerted atomic motions that guide transitions from highly populated ground-state ensembles to higher-energy states (1–4). Experimentally verified higher-energy states, or "excited states," typically constitute ~1% of the protein population, with exchange rates on the microsecond to millisecond timescale. To date, our understanding of these transient protein states has largely been achieved through interpretation of NMR relaxation dispersion spectroscopy (3,4). Relaxation dispersion measurements provide a high level of detail into the populations, exchange correlation times, and chemical shift differences between two states (3–5). Until recently, visualizing the specific structural implications of these dispersion measurements has been limited to comparisons with static crystal structures or models, because the slow timescale of these motions was beyond the reach of atomistic molecular simulation (6,7).

Long-timescale molecular dynamics (MD) simulations now make it possible to witness atomic motions on the nanosecond to microsecond timescale, with a suitable level of agreement with experiment (8–11). Additionally, with recent advances in specialized computational architectures like Anton, simulation timescales on the order of microseconds to milliseconds in a single trajectory are within reach (7,12–14).

Here, we report the transient atomic-level motions that govern the long-timescale dynamics of the L99A mutant of T4 lysozyme. L99A is a model system for studying ligand binding to buried protein cavities and protein excited states, and the conformational changes that govern these phenomena have been enigmatic despite nearly 25 years of experimental study (5,6,15–30). Experimental studies of the L99A mutant of T4 lysozyme have focused on defining the structures of and transition times between the ground state, the excited state, and ligand-bound states. Initial crystal structures of the L99A mutant indicate that the ~40 $Å^3$ cavity in the C-terminal domain of wild-type protein expands to an ~150 $Å^3$ cavity (26), a volume capable of accommodating substituted benzenes and noble gases (20,24,27–30). O_2 is simulated to bind and escape the buried cavity through

Submitted April 21, 2016, and accepted for publication August 16, 2016.
*Correspondence: vfeher@ucsd.edu or ramaro@ucsd.edu
Jamie M. Schiffer and Victoria A. Feher contributed equally to this work.
Editor: Amedeo Caflisch.
http://dx.doi.org/10.1016/j.bpj.2016.08.041

Please cite this article in press as: Schiffer et al., Capturing Invisible Motions in the Transition from Ground to Rare Excited States of T4 Lysozyme L99A, Biophysical Journal (2016), http://dx.doi.org/10.1016/j.bpj.2016.08.041

openings between the D, E, G, H, and J helices (31), whereas binding of substituted benzenes is accompanied by discrete rearrangements in the F and G helices (30). Substantial rearrangements also occur in these helices in an excited state that comprises ~3% of the protein ensemble at room temperature (5,25). NMR relaxation dispersion measurements and chemical-shift-based Rosetta-guided modeling were used previously to design a triple mutant of T4 lysozyme meant to mimic the excited state (5). This chemical-shift-based model places the F114 phenyl group into the center of the buried hydrophobic cavity in the excited state, blocking benzene binding (5). To date, the details of the conformational changes that guide transitions between ground, ligand-accessible, and excited states, at standard temperature and pressure conditions, remain unknown for the L99A mutant.

We performed conventional MD simulations on Anton starting from the ground-state apo crystal structure of L99A with the intention of witnessing the conformational transitions needed for ligand exchange (21). During a 30 μs MD simulation, L99A transitions from the ground state to a rare excited state of L99A. We compare the structures witnessed in the trajectory to chemical shifts and torsion angles determined experimentally. The simulation reveals multiple metastates and a previously unseen intermediate state that exists before the excited state in the trajectory. These transient conformational motions reveal, to our knowledge, a new level of intricacy in the concerted motions of protein interiors and suggest that short-lived metastates in protein ensembles may be critical to dynamic protein function.

MATERIALS AND METHODS

MD system preparation

We used Schrödinger PrepWizard and MD Desmond setup modules (32) to prepare T4 lysozyme structures for MD simulation. The T4 lysozyme L99A mutant crystal structure (PDB: 1L90) was prepared for the ground-state Anton simulations. The excited-state mimic was prepared using the coordinates from the solution structure of the L99A triple mutant (PDB 2LC9) and Schrödinger's Maestro (32) to mutate the residues A113 to glycine and P119 to arginine, the residues found in the T4 lysozyme L99A single mutant. We chose to simulate the triple mutant (PDB: 2LC9), ES-3mut, instead of the excited-state model (PDB: 2LCB), ES-Rosetta. However, given the low root mean-square (RMS) differences (~0.6 Å) between the models, the selection of either would have been appropriate. These systems were protonated at pH 5.5, with the pK_a-titratable residues determined using the Maestro integrated PROPKA (33). The pK_a of residue E11 is 4.8 for the ground-state crystal structure and >5.5 for the excited-state mimic. For consistency, we deprotonated E11 for both systems. We solvated each system in a TIP3P (32) cubic water box (~75.7 Å/side), with nine Cl^- ions to balance the charge of L99A and 150 mM NaCl. The total numbers of atoms in each system for the L99A crystal structure and each of the nine L99A excited-state mutant (PDB: 2LC9) structures were 43,507, 43,507, 43,540, 43,546, 43,498, 43,483, 43,504, 43,510, 43,504, and 43,525, respectively. Results from the first structure of the ES-3mut (PDB: 2LC9) are shown in the main body of the text, whereas measurements from all other simulations initiated from other 2LC9 structures are shown in Fig. S4 in the Supporting Material.

MD system parameterization

System coordinates from the preparation (above) were converted into an xleap-readable (34) format using in-house Python scripts. Each system was parameterized in xleap using the AMBER99SB (35) force field, and periodic boundary conditions were implemented.

Anton MD simulations

MD simulations on Anton were performed on the same parameterized, minimized, and equilibrated L99A crystal structure system described above. The Anton simulation was run in the NTP ensemble using Anton's multigrator (34) at 300 K, 1 bar, and with a 2-fs time step, and particle-mesh Ewald (PME) electrostatic approximations.

GPU-enabled MD simulations

Simulations not performed on Anton were performed with NVIDIA GK110 (GeForce GTX Titan, NVIDIA, Santa Clara, CA) graphics processing units (GPUs) using the CUDA version of PMEMD in AMBER12 (35). The L99A crystal structure system (derived from PDB: 1L90) was minimized and equilibrated using the GPU version of AMBER12 (36,37). We minimized in three stages, including a stage of heating: 1) 40 ps of hydrogen-only minimization with a restraint weight of 20 kcal/mol on the protein and solvent; 2) 4 ps of water minimization with a restraint weight of 20 kcal/mol on the protein and salt atoms; 3) 20 ps of water heating with a restraint weight of 20 kcal/mol on the protein and salt atoms; and 4) 40 ps of full minimization. We equilibrated the system using harmonic equilibration at 300 K over four sequential 500-ps runs, decreasing the restraint potential on the backbone at each step, starting at 4.0 kcal/mol and ending at 1.0 kcal/mol. GPU-enabled AMBER12 production runs were carried out as an NTP ensemble at 300 K and 1 bar with a 2-fs time step and PME (38) electrostatic approximation and nonbonding cutoff of 10.0 Å. MD input files are provided as part of the data-sharing files.

Excited-state mimic MD simulations

The ES-3mut system was also minimized and equilibrated in two stages using the GPU version of AMBER12 (36,37): 1) 6 ps of hydrogen-only minimization with a restraint weight of 100 kcal/mol were performed, with the 500 steps of steepest-descent algorithm minimization and 2500 steps of conjugate-gradient algorithm minimization; and 2) 12 ps of solvent-only minimization with a restraint weight of 100 kcal/mol, with 500 steps of steepest-descent algorithm minimization and 5500 steps of conjugate-gradient algorithm minimization (39). This protocol was chosen because we did not want to minimize into an arbitrary and nonnative energy minimum. Instead, we performed extensive equilibration so that the remutated L99A protein could more readily search the energetic landscape of L99A. We heated the system to 300 K for 100 ps with a restraint on the protein backbone of 4.0 kcal/mol and equilibrated the system using harmonic equilibration at 300 K over three sequential 100-ps runs, decreasing the restraint potential on the backbone of L99A on each step, starting at 3.0 kcal/mol and ending at 1.0 kcal/mol. GPU-enabled AMBER12 production runs were carried out as an NTP ensemble at 300 K and 1 bar, with Langevin dynamics (40) with a collision frequency of 1.0 ps^{-1}, a pressure relaxation time of 2.0 ps, with SHAKE bond constraints on hydrogen bonds (41), and with a 2-fs time step and PME (38) electrostatic approximation and nonbonding interaction cutoff of 9.0 Å. MD input files are provided as part of the data sharing.

Please cite this article in press as: Schiffer et al., Capturing Invisible Motions in the Transition from Ground to Rare Excited States of T4 Lysozyme L99A, Biophysical Journal (2016), http://dx.doi.org/10.1016/j.bpj.2016.08.041

Simulation analysis

MD trajectories were processed using CPPTRAJ (42) and VMD (43). All frames were aligned using the backbone atoms of residues 75–155, which encompasses the C-terminal domain of T4 lysozyme L99A. CPPTRAJ was used to perform dihedral-, distance-, angle-, and hydrogen-bond-based measurements. VMD was used to perform RMS-deviation (RMSD)-based analysis of the F/G helix (residues 108–124), the I helix (residues 135–142), the C/D helix (residues 75–90), and the buried hydrophobic residues (residues 84, 99, 102, 114, 133, 138, and 153).

SHIFTX2 and SHIFTX+ calculations

Two NMR chemical-shift prediction methods were utilized for our studies. SHIFTX+ is an algorithm that takes into account only structural coordinates for chemical-shift prediction. The SHIFTX2 algorithm incorporates a number of sequence-based features from a training set as well as structural coordinates (44). Han et al. (44) report that SHIFTX2 has improved correlation coefficients between observed and predicted backbone chemical shifts compared to the SHIFTX+, so SHIFTX2 was used to predict chemical shifts where direct comparisons were made between MD simulation results and reported experimental values (see Fig. 2). Here, $^{13}C\alpha$ chemical shift predictions were calculated for each frame, then averaged over frames 1–12,569, representing the ground state, and frames 112,746–125,691, representing ES-Anton, of the Anton trajectory. A pH of 5.5 was input for the SHIFTX2 algorithm.

Chemical-shift predictions reported in Fig. 4 *d* are based on chemical shifts calculated for each frame of the Anton trajectory with the SHIFTX+ algorithm.

POVME calculations

A VMD plug-in for POVME 2.0 software (44) was used to calculate pocket volumes and buried surfaces in the L99A C-terminal-domain Anton trajectory. A sphere with 8 Å radius was centered near b-carbons for residues L84, A99, and M102 (sphere-center coordinates 26.04, 21.48, 25.47). Volume calculations were performed on each cluster centroid, determined by RMSD-based clustering (see below), as well as on every 10th frame of the Anton trajectory. Fig. 4 *a* depicts cluster centroids from each of the most populated RMSD-based clusters.

RMSD-based clustering

The Anton trajectory structures were clustered using the GROMOS algorithm (45) in GROMACS version 4.5.5 (46) with heavy-atom trace RMSD cutoff of 1.2 Å for T4 lysozyme L99A mutant.

Principal component analysis

The Anton trajectory was stripped of solvent and aligned against the backbone of residues 75–155 in the C-terminal domain. Principal component analysis was performed for all atoms for each frame of the trajectory using CPPTRAJ. Three principal components were specified for calculation.

RESULTS AND DISCUSSION

A single Anton trajectory simulates a ground-to-excited-state transition

The single Anton trajectory captures a significant conformational change from the ground state to the sparsely populated excited state of the T4 lysozyme L99A mutant. The Anton excited state (ES-Anton) is defined by key structural features that have been experimentally characterized (5,25). These features are the refolding of the F and G helices into a single helix and changes in the F114 χ_1 and ψ torsion angles that position the phenyl ring of F114 into the buried cavity (Figs. 1, S1, and S2). Structural changes from the ground state to the ES-Anton state were followed through hydrogen-bond distances between residues in the F and G helices, changes in dihedral angles of F114 (Fig. 1, *b–e*), and internal distances between the F114 phenyl group and hydrophobic residues lining the L99A cavity (Fig. 4 *c*). These internal distances define the same metastates as can be seen with the first principal component of the Anton trajectory (Fig. S6). Folding of the F and G helices into a single helix, as measured by backbone $i \rightarrow i + 4$ hydrogen-bond distances (Fig. 1, *b* and *c*), starts at 18.6 μs in the Anton trajectory with the G113 to S117 hydrogen-bond formation and completes by 27 μs with the G110-to-F114 hydrogen-bond formation, recapitulating experimental distances (5). Additionally, the F114 ψ and χ_1 torsion angles relax to experimentally determined values for the excited state of $-60°$ and $-164°$ at 18.6 μs and 27 μs, respectively (5). These measurements indicate that starting from the ground state, the simulated L99A mutant of T4 lysozyme reaches the sparsely populated excited state within 27 μs of simulation.

Verifying the simulated excited state by comparison with experiment

Calculated chemical-shift differences ($\Delta\varpi$) between the simulated ground-state (1–4 μs) and ES-Anton (27–30 μs) structures agree with experimental $\Delta\varpi$ values measured for the full-length L99A lysozyme protein and for those residues in the F/G helix that have the largest measured $\Delta\varpi$ (Fig. 2).

However, some differences are observed between ES-Anton and what has been previously modeled with the Rosetta mutagenesis approach (5). Specifically, the pitch of the F/G helices and A helix (Fig. S1) and the ϕ, ψ, χ_1 torsion angles of residues in the C, D, E, H, I, and J helices (Fig. S3).

For a direct comparison between previous models and our simulated excited state, we performed separate MD simulations initiated on the published excited-state model of L99A, specifically the triple mutant L99A/G113/R119P (PDB: 2LC9) (5), hereafter referred to as ES-3mut.

In simulations initiated from ES-3mut, we observe that the majority of the C-terminal domain backbone relaxes to structures more similar to that of ES-Anton than to the starting structure (Figs. 3, *a–h*, S4, and S5). The pitch of the F/G helices relaxes to the pitch of ES-Anton within a few hundred nanoseconds, as evidenced by backbone RMSD of the F/G helices (Fig. 3, *a–c*). Additionally, RMSD measurements of residues in the C and D helices (residues 75–90, atoms N, C, C_α, and O) and buried hydrophobic residues

Please cite this article in press as: Schiffer et al., Capturing Invisible Motions in the Transition from Ground to Rare Excited States of T4 Lysozyme L99A, Biophysical Journal (2016), http://dx.doi.org/10.1016/j.bpj.2016.08.041

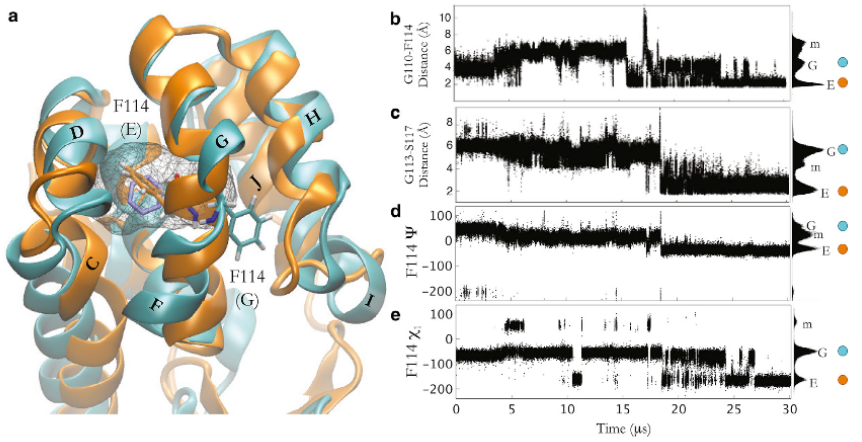


FIGURE 1 The Anton trajectory samples the transition from ground state to excited and cavity-accessible states of T4 lysozyme L99A. (*a*) Superposition of the C-terminal domain L99A ground-state crystal structure (PDB: 3DMV) and the Anton-simulated average excited-state structure. The gray mesh outlines the buried cavity in the ground state. The positions of F114 in the ground state (G) and the excited state (E) are shown. (*b–e*) Structural measures for conformational differences over the Anton trajectory. Hydrogen-bond distances between the backbone of helical residues G110 and F114 (*b*) and G113 and S117 (*c*) and the F114 ψ (*d*) and χ_1 (*e*) torsion angle values are shown. The ground (G), meta (m), and excited (E) states are shown in the various populations in the histograms to the right of the plots. The circles indicate the experimentally determined values of the ground and excited states. To see this figure in color, go online.

(residues 84, 99, 102, 114, 133, 138, and 153, trace atoms) relax to structures more similar to ES-Anton than to ES-3mut (Fig. 3, *d* and *h*). Residues in the C, D, F/G, and I helices sample the same ϕ, ψ, χ_1 torsion angles as sampled in the ES-Anton model (Figs. 3, S4, and S5), and in one of the triplicate simulations from the ES-3mut model, residues in the I helix also sample ψ torsion angles sampled in the ES-Anton trajectory (Figs. 3, *e–g*, and S4).

Some differences exist in the torsion angles of the residues in the ES-Anton structures and ES-3mut simulations. For example, there are multiple possible explanations for the differences in the ϕ, ψ, χ_1 torsion angles of these residues in the ES-Anton model and ES-3mut simulations. The ψ torsion angles sampled in the I helix in both models are outside of the helical regime and occur before the excited-state transition (Fig. S3). It is possible that the structure guidance required in interpretation of relaxation dispersion measurements for the L99A mutant may have confined the conformational ensemble of this excited state. It is also possible that known errors introduced by the force field (47) could have contributed to the local unfolding of the I helix in the ES-Anton and ES-3mut models (5). However, force field bias is also known to increase helical content (48), which could contribute to the increase in timescales we see in the refolding of the F/G helix into a single helix (see below in Discrepancies in Timescales of Excited-State Transitions

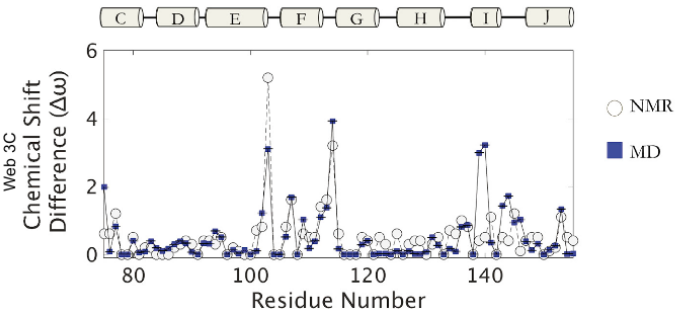


FIGURE 2 The Anton-simulated excited state agrees with relaxation dispersion data. Absolute values of the $^{13}C_\alpha$ chemical-shift differences, $|\Delta\varpi^{13C\alpha}|$, between the ground and excited states back-calculated with SHIFTX2 from structures from the Anton trajectory (*squares*) compared to NMR-relaxation-derived $|\Delta\varpi^{13C\alpha}|$ values (*circles*) for the C-terminal domain (5). To see this figure in color, go online.

Please cite this article in press as: Schiffer et al., Capturing Invisible Motions in the Transition from Ground to Rare Excited States of T4 Lysozyme L99A, Biophysical Journal (2016), http://dx.doi.org/10.1016/j.bpj.2016.08.041

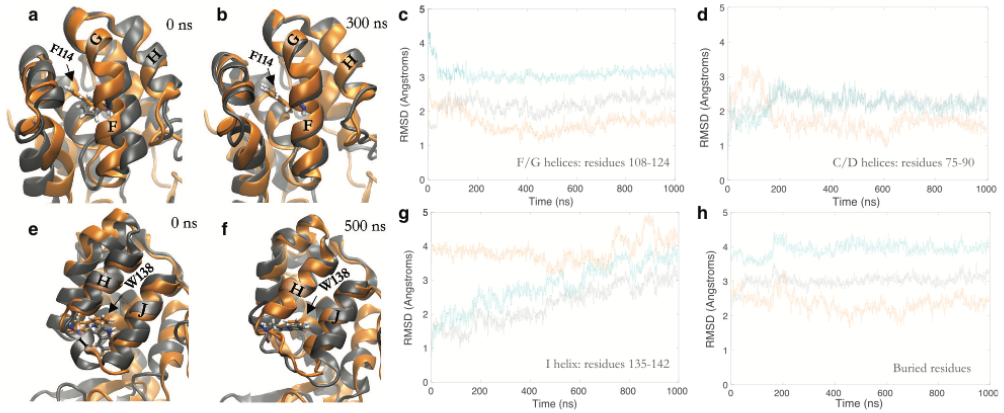


FIGURE 3 Simulations from a previous model of the excited state relax to the excited state simulated with the Anton trajectory. The FG helices (*a* and *b*) and the I helix (*e* and *f*) of the previous excited-state model starting structure (*gray*; PDB: 2LC9) relax to the excited state from the Anton trajectory (*orange*) after 300 ns and 500 ns, respectively (*b* and *f*, *gray*). (*c*, *d*, *g*, and *h*) RMSD time plots of the simulations initiated from coordinates of the previous excited-state model to the MD Anton excited-state average structure (*orange*), to the excited-state model starting structure (*gray*), and to the ground-state crystal structure (*teal*) are reported as an average over three parallel simulations for the residues of the F/G helices, C/D helices (residues 75–90), I helix (residues 135–142), and buried residues (trace of residues 84, 99, 102, 114, 133, 138, and 153).

in Simulation and Experiment). Alternatively, it is also possible that the excited state is composed of multiple states that vary in the conformation of the residues in the C, D, E, H, I, and J helices but are consistent in the position of F114, with a χ_1 of ~160°, and the refolding of the F/G helices into a single helix.

Defining the metastates and intermediate state in the Anton trajectory

To define conformational changes that occur in the Anton trajectory further, we measured the buried pocket volumes (Fig. 4, *a* and *b*), the internal distances between the F114 phenyl ring and hydrophobic residues lining the L99A cavity (Fig. 4 *c*), and back-calculated chemical shifts from every 10th frame of the Anton trajectory using the SHIFTX+ algorithm (Fig. 4 *d*). Additionally, we performed PCA over the entire Anton trajectory. Taken together, these measurements indicate four metastates and an intermediate state that are sampled in addition to the ground state and the excited state of the L99A mutant of T4 lysozyme (Figs. 1 *a* and S6; Movies S1–S4). Conformational states defined along both the first and second principal components (PC1/2) align well with the metastates witnessed from the F114-V103 internal distance and buried volume measurements (Fig. S6; Movies S1–S4). The differences between the principal components are motions of the N-terminal domain relative to the C-terminal domain.

As described above for the Anton simulation (Fig. 2), inspection of the back-calculated SHIFTX+ values for F114 $^{13}C_\alpha$ shows that they recapitulate the experimentally derived differences between the ground and excited states over the entire Anton trajectory. Additionally, the back-calculated chemical shifts of the intermediate state and metastates are within error of the chemical shifts for the excited state and ground state, respectively.

The four metastates (m1–m4) witnessed in the single Anton trajectory have chemical shifts similar to those of the ground state, maintain the F helix as a 3_{10} helix separate from the G helix, and position F114 just outside of the buried cavity despite fluctuation in cavity volume and changes in F114-V103 internal distances throughout these metastates (Fig. 1). The internal distance between F114 and V103 was chosen as a metric here because V103 neighbors the buried L99A cavity, and has the largest chemical shift difference ($\Delta\varpi$) between the ground and the excited state in both ^{15}N and $^{13}C_\alpha$ spectra. In the following paragraphs, we describe the conformational differences that define each of these states as unique from the ground and excited states and show that these differences are manifested in the back-calculated chemical shifts, calculated using the SHIFTX+ algorithm (49). Interestingly, the back-calculated $^{13}C_\alpha$ chemical shifts of F114 in these metastates are within error of the back-calculated $^{13}C_\alpha$ chemical shifts in the ground state despite the wide range of conformations sampled in the four metastates.

The metastates are unique from the ground state, as well as one another, primarily in the angle between the D helix and the G helix, as well as the position of the F114 side chain (Fig. 4). In the m1 state, the angle between the D and G helices is more acute than in the ground state, which manifests in a decreased V103-F114 distances and

Please cite this article in press as: Schiffer et al., Capturing Invisible Motions in the Transition from Ground to Rare Excited States of T4 Lysozyme L99A, Biophysical Journal (2016), http://dx.doi.org/10.1016/j.bpj.2016.08.041

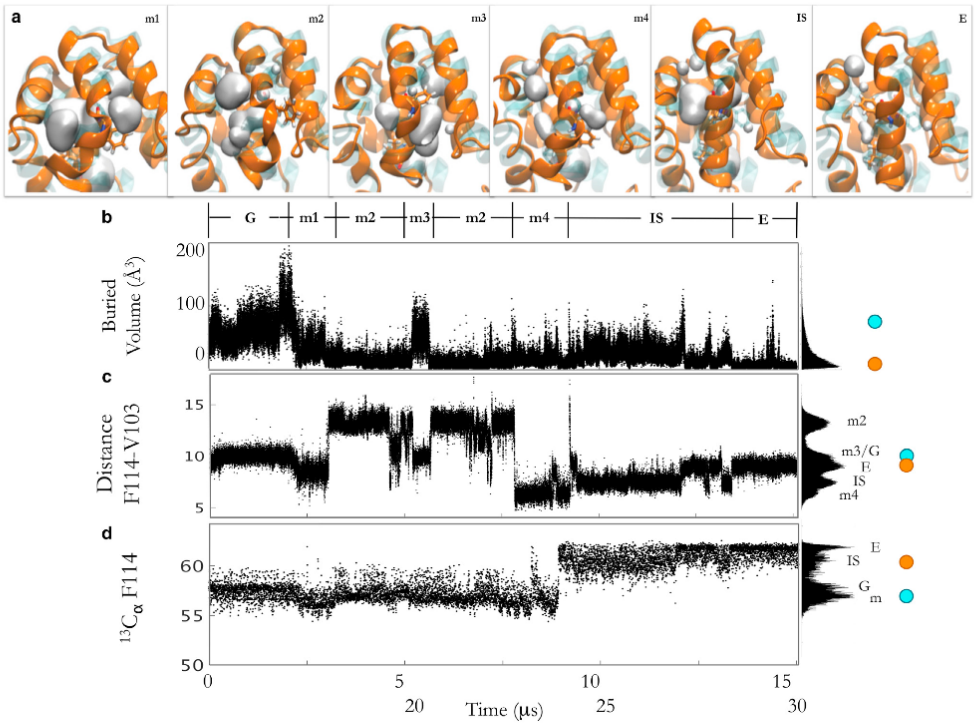


FIGURE 4 The Anton trajectory samples the transition from ground state to excited state. (*a*) Representative structures of each of the metastates (m1–m4), the intermediate state (IS), and the excited state (E). The Anton trajectory, the ground state (PDB: 3DMV) as a shadow, and the cavity volumes are depicted. F114 is shown for reference in each representative structure. (*b–d*) The buried volume, distance between F114 and V103, and back-calculated chemical shifts of the $^{13}C_{\alpha}$ for F114 using SHIFTX+ are plotted over time in the Anton trajectory. To the right of each plot is the population histogram for the buried volumes, distances, and back-calculated chemical shifts in the entire trajectory. The circles to the right show the values from the ground state and the excited state from experiment. To see this figure in color, go online.

decreased F114 ψ angle. The m2 state is characterized by an increased distance between F114 and V103, due to D helix cracking that flips the phenyl ring of F114 outside of the cavity between the G and H helices. The m3 state has the F114 χ_1 angle similar to that of the excited state, resulting in the phenyl ring of F114 flipped into the buried cavity, but the D and G helices are still two separate helices with a wide angle between the helices. Finally, the m4 state is similar to the m3 state in that F114 is still flipped into the buried cavity, but with a χ_1 angle similar to that of the ground state, shortening the distance between F114 and V103.

An intermediate state (IS) (Fig. 4) with distinct conformational features from the excited state (ES-Anton), occurs in the Anton trajectory between 18 and 27 μs. The IS has the same F114 ψ angle as the excited state but differs by the F114 χ_1 angle, as it flips between the ground and excited state values (Fig. 1 *e*). It also contains a larger buried cavity volume (Fig. 4 *b*). The fluctuation in the F114 χ_1 angle was seen in high-temperature simulations of the ES-3mut (15), suggesting that these two states are similar in energy (see Verifying the Simulated Excited State by Comparison with Experiment). Despite the F114 χ_1 and cavity-volume differences between the IS and the ES-Anton, the back-calculated chemical shifts from these structures are not distinct and are equal to chemical shifts determined experimentally (Fig. 4 *d*, *orange circle*) (5). Thus, the intermediate state, although distinctly observed in the simulation, may represent an equilibration event along the course of the trajectory.

Dynamic fluctuation of the buried cavity of L99A

The C-terminal domain of apo L99A experiences significant core mobility and deformation (Fig. 5; Movie S1). These

Please cite this article in press as: Schiffer et al., Capturing Invisible Motions in the Transition from Ground to Rare Excited States of T4 Lysozyme L99A, Biophysical Journal (2016), http://dx.doi.org/10.1016/j.bpj.2016.08.041

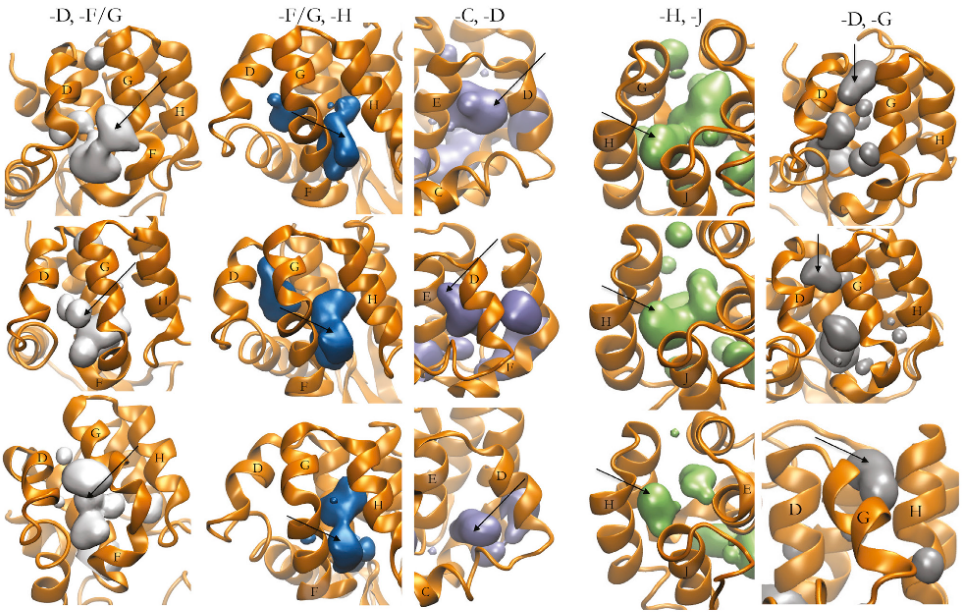


FIGURE 5 Five mobile defects through the protein surface are seen in the Anton trajectory. Three examples of mobile defects through the protein surface occur between the D and F/G helices (*solid surfaces, far left*) at 5.38 μs, 6.87 μs, and 8.37 μs (*top to bottom*), between the F/G and H helices (*third from left*) at 2.41 μs, 19.22 μs, and 24.21 μs (*top to bottom*), between the C and D helices (*third from left*) at 340 ns, 550 ns, and 6.78 μs (*top to bottom*), between the H and J helices (*fourth from left*) at 630 ns, 4.95 μs, and 5.14 μs (*top to bottom*), and between the D and G helices (*right*) at 7.77 μs, 8.27 μs, 11.66 μs (*top to bottom*). To see this figure in color, go online.

internal conformational rearrangements result in the overall reduction in the volume of the buried cavity from ~70 Å^3 in the equilibrated ground state to <10 Å^3 in the final simulated excited state (Fig. 4 *b*), as measured with POVME 2.0 (44). These rearrangements are accompanied by mobile defects, or cavity migrations, to other regions of the C-terminal domain. As previously mentioned, the buried cavity in the apo ground-state crystal structure (PDB: 1L90) was initially measured to be ~150 Å^3. However, with the POVME method (44) we used, the volume of the buried cavity in this ground-state crystal structure (PDB: 1L90) was ~65.5 Å^3. We attribute this discrepancy to variability in methods of measuring cavities in proteins (50). The volume reduction seen in the Anton trajectory aligns with the previously described rationale for using high pressure to shift protein conformational equilibria to alternative packing of the hydrophobic core by filling void volumes (51), and has been seen experimentally in L99A for F114 or water filling the void volume (16–18,52).

Occasionally, these internal-cavity deformations reach the protein surface, and over the course of the trajectory, five distinct openings occur from the buried pocket of L99A to solution (Fig. 5). Four of these openings from the buried cavity to solution have been witnessed through enhanced simulation techniques (data not shown). The D/G, D/F/G, and H/J openings (Fig. 5) have also been previously simulated in the egress and ingress of molecular oxygen (31). Although the openings through the D/F/G pocket are larger and able to accommodate benzene, the H/J and D/G pockets form narrower surface openings that cannot, suggesting that different ligands may bind to the buried cavity through different pathways.

The openings of the buried cavity to solvent through the D and F/G helices occur during transitions between the various metastates. Recall that these metastates are defined by the location of the phenyl group of F114 in the buried pocket and the angle between the F and G helices. This suggests that the movement of the F114 side chain and bending of the F and G helices is correlated with the transient sliding of helices that permits the opening of the buried cavity to solution. In other words, transitioning between these metastates may play a role in ligand exchange. Additionally, the D/F/G surface has previously been identified by simulation to both bind benzene and be the opening through which benzene binds to the buried cavity (53,54).

Please cite this article in press as: Schiffer et al., Capturing Invisible Motions in the Transition from Ground to Rare Excited States of T4 Lysozyme L99A, Biophysical Journal (2016), http://dx.doi.org/10.1016/j.bpj.2016.08.041

Discrepancies in timescales of excited-state transitions in simulation and experiment

The opening at the interface of the D and F/G helices, where benzene has been previously simulated to bind and pass through to the buried cavity (54), occurs repeatedly throughout the simulation and on timescales commensurate with experimental rates of benzene binding (μs). Experimentally, the rate of benzene binding is rapid (μs) (27) compared to the transition from the ground state of L99A to the excited state (~ms) (20). However, in the Anton simulation, we observed the transition to the excited state between 18.6 μs and 27 μs, suggesting a transition timescale two orders of magnitude faster than found in experiment. Although only a single transition in one single long-timescale simulation was observed, and therefore no determination of the timescale for the transition can be made, the order-of-magnitude difference between our observed transition and the experimental rate is consistent with previously reported discrepancies between simulation and experiment for slow ring flips (on the order of microseconds) in proteins (55,56). The fortuitous observation of this transition is consistent with force-field-induced biases toward greater helical content (48) that could have led to the faster transition timescale, since refolding of the F and G helices into a single helix is a key structural change on the path from the ground state to the excited state of L99A.

CONCLUSION

Despite the intense study of T4 lysozyme L99A as a model protein system for ligand binding, the Anton supercomputer simulation reported here represents the longest single-trajectory MD simulation for this protein. What emerges are atomistic details for the previously enigmatic conformational changes known to occur for L99A mutant ligand exchange and the ground-to-excited-state transition.

Remarkably, there is a high degree of cavity plasticity, exceeding what might be expected from previous crystallographic studies (18,19), that involves residues, metastates, and conformations beyond the discrete crystallographic states observed for L99A bound to a congeneric series of benzene-related compounds (24,30). Further, the number of rotamers observed for buried L99A residues exceeds that seen in the several-hundred-microsecond simulations of other protein cores, such as ubiquitin, RNaseH, and β-lactamase (56). Although dramatic, we find that the torsional, hydrogen-bond, secondary-structure, and back-calculated chemical shifts from the L99A simulation are in agreement with previously measured experimental data. And the extent of the core plasticity is consistent with this protein's ability to undergo the conformational change that buries the F114 side chain into the engineered C-terminal cavity in the excited state without extensive backbone conformational change.

Our studies demonstrate that the engineered binding pocket of L99A has a much higher degree of plasticity when simulated beyond a few hundred nanoseconds and that the plasticity is not limited to a discrete number of buried side chains. This perhaps presents a challenging dilemma for those utilizing this protein as a model system for free-energy perturbation studies, as a number of different core side-chain rotamers may define the protein ground state and/or ligand-bound states. Furthermore, the residues involved in the plasticity are not limited to a particular region of the binding pocket, but rather are found in all regions surrounding the binding pocket, a fact that is somewhat dissimilar to what is observed for native-protein enzyme pockets where multiple conformations are typically limited to a particular loop or pocket subregion.

Finally, the structural details presented here are consistent with a growing body of data implicating activated volumes and mobile defects of internal cavities in protein conformational change (16,17,51). It is possible that mobile defects may not be unique to protein transitions related to excited states, and that fluctuations in small, buried volumes could contribute to conformational change in many protein mechanisms.

SUPPORTING MATERIAL

Six figures and four movies are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(16)30764-0.

AUTHOR CONTRIBUTIONS

J.M.S. performed simulations and analyses; R.S. performed analyses in Fig. S4 and RMSD calculations; J.M.S. and V.A.F. wrote the article, with significant revisions from R.D.M.; and V.A.F., J.M.S., R.D.M., and R.E.A. planned the study and outlined simulations to perform. R.E.A. was the leading principal investigator, and V.A.F. initiated and directed the study.

ACKNOWLEDGMENTS

R.E.A. is a co-founder of Actavalon, Inc., and V.A.F. has financial interest in Actavalon, Inc. Readers are welcome to comment on the online version of this article. Correspondence and requests for materials should be addressed to VAF (vfeher@ucsd.edu) or REA (ramaro@ucsd.edu).

This work was funded in part by the Director's New Innovator Award Program (NIH DP2 OD007237 to R.E.A.). Funding and support from the National Biomedical Computation Resource (NBCR) was provided through NIH P41 GM103426. J.M.S. was supported by National Institutes of Health Molecular Biophysics Training Grant T32 GM008326. Anton computer time was provided by the Pittsburgh Supercomputing Center (PSC) and the National Center for Multiscale Modeling of Biological Systems (MM Bios) through grant P41GM103712-S1 from the National Institutes of Health. D.E. Shaw Research generously made the Anton machine at PSC available.

Please cite this article in press as: Schiffer et al., Capturing Invisible Motions in the Transition from Ground to Rare Excited States of T4 Lysozyme L99A, Biophysical Journal (2016), http://dx.doi.org/10.1016/j.bpj.2016.08.041

REFERENCES

1. Boehr, D. D., D. McElheny, …, P. E. Wright. 2006. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science.* 313: 1638–1642.
2. Masterson, L. R., T. Yu, …, G. Veglia. 2011. cAMP-dependent protein kinase A selects the excited state of the membrane substrate phospholamban. *J. Mol. Biol.* 412:155–164.
3. Baldwin, A. J., and L. E. Kay. 2009. NMR spectroscopy brings invisible protein states into focus. *Nat. Chem. Biol.* 5:808–814.
4. Henzler-Wildman, K., and D. Kern. 2007. Dynamic personalities of proteins. *Nature.* 450:964–972.
5. Bouvignies, G., P. Vallurupalli, …, L. E. Kay. 2011. Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature.* 477:111–114.
6. Xue, Y., J. M. Ward, …, N. R. Skrynnikov. 2012. Microsecond timescale conformational exchange in proteins: using long molecular dynamics trajectory to simulate NMR relaxation dispersion data. *J. Am. Chem. Soc.* 134:2555–2562.
7. Shaw, D. E., R. O. Dror, …, B. Towles. 2009. Millisecond-scale molecular dynamics simulations on Anton. *Proc. Conf. High Perform. Comput. Netw. Storage Anal. SC 09.* 39.
8. Sgourakis, N. G., Y. Yan, …, A. E. Garcia. 2007. The Alzheimer's peptides Aβ40 and 42 adopt distinct conformations in water: a combined MD/NMR study. *J. Mol. Biol.* 368:1448–1457.
9. Sgourakis, N. G., M. Merced-Serrano, …, A. E. Garcia. 2011. Atomic-level characterization of the ensemble of the Aβ(1–42) monomer in water using unbiased molecular dynamics simulations and spectral algorithms. *J. Mol. Biol.* 405:570–583.
10. Jensen, L. B., K. Mortensen, …, C. Foged. 2010. Molecular characterization of the interaction between siRNA and PAMAM G7 dendrimers by SAXS, ITC, and molecular dynamics simulations. *Biomacromolecules.* 11:3571–3577.
11. Fuglestad, B., P. M. Gasper, …, E. A. Komives. 2012. The dynamic structure of thrombin in solution. *Biophys. J.* 103:79–88.
12. Shaw, D. E., P. Maragakis, …, W. Wriggers. 2010. Atomic-level characterization of the structural dynamics of proteins. *Science.* 330: 341–346.
13. Dror, R. O., T. J. Mildorf, …, D. E. Shaw. 2015. SIGNAL TRANSDUCTION. Structural basis for nucleotide exchange in heterotrimeric G proteins. *Science.* 348:1361–1365.
14. Vitalini, F., A. S. Mey, …, B. G. Keller. 2015. Dynamic properties of force fields. *J. Chem. Phys.* 142:084101.
15. Vallurupalli, P., N. Chakrabarti, …, L. Kay. 2016. Atomistic picture of conformational exchange in a T4 lysozyme cavity mutant: an experiment-guided molecular dynamics study. *Chem. Sci.* 7:3602–3613.
16. Nucci, N. V., B. Fuglestad, …, A. J. Wand. 2014. Role of cavities and hydration in the pressure unfolding of T4 lysozyme. *Proc. Natl. Acad. Sci. USA.* 111:13846–13851.
17. Lerch, M. T., C. J. López, …, W. L. Hubbell. 2015. Structure-relaxation mechanism for the response of T4 lysozyme cavity mutants to hydrostatic pressure. *Proc. Natl. Acad. Sci. USA.* 112:E2437–E2446.
18. Collins, M. D., M. L. Quillin, …, S. M. Gruner. 2007. Structural rigidity of a large cavity-containing protein revealed by high-pressure crystallography. *J. Mol. Biol.* 367:752–763.
19. Eriksson, A. E., W. A. Baase, …, B. W. Matthews. 1992. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science.* 255:178–183.
20. Mulder, F. A. A., B. Hon, …, L. E. Kay. 2000. Flexibility and ligand exchange in a buried cavity mutant of T4 lysozyme studied by multinuclear NMR. *Biochemistry.* 39:12614–12622.
21. Mulder, F. A. A., B. Hon, …, L. E. Kay. 2002. Slow internal dynamics in proteins: application of NMR relaxation dispersion spectroscopy to methyl groups in a cavity mutant of T4 lysozyme. *J. Am. Chem. Soc.* 124:1443–1451.
22. Korzhnev, D. M., V. Y. Orekhov, …, L. E. Kay. 2003. Off-resonance R1ρ relaxation outside of the fast exchange limit: an experimental study of a cavity mutant of T4 lysozyme. *J. Biomol. NMR.* 26:39–48.
23. Skrynnikov, N. R., F. W. Dahlquist, and L. E. Kay. 2002. Reconstructing NMR spectra of "invisible" excited protein states using HSQC and HMQC experiments. *J. Am. Chem. Soc.* 124:12352–12360.
24. Morton, A., and B. W. Matthews. 1995. Specificity of ligand binding in a buried nonpolar cavity of T4 lysozyme: linkage of dynamics and structural plasticity. *Biochemistry.* 34:8576–8588.
25. Mulder, F. A. A., A. Mittermaier, …, L. E. Kay. 2001. Studying excited states of proteins by NMR spectroscopy. *Nat. Struct. Biol.* 8:932–935.
26. Eriksson, A. E., W. A. Baase, …, B. W. Matthews. 1992. A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. *Nature.* 355:371–373.
27. Feher, V. A., E. P. Baldwin, and F. W. Dahlquist. 1996. Access of ligands to cavities within the core of a protein is rapid. *Nat. Struct. Biol.* 3:516–521.
28. Morton, A., W. A. Baase, and B. W. Matthews. 1995. Energetic origins of specificity of ligand binding in an interior nonpolar cavity of T4 lysozyme. *Biochem.* 34:8564–8575.
29. Baldwin, E., W. A. Baase, …, B. W. Matthews. 1998. Generation of ligand binding sites in T4 lysozyme by deficiency-creating substitutions. *J. Mol. Biol.* 277:467–485.
30. Merski, M., M. Fischer, …, B. K. Shoichet. 2015. Homologous ligands accommodated by discrete conformations of a buried cavity. *Proc. Natl. Acad. Sci. USA.* 112:5039–5044.
31. Kitahara, R., Y. Yoshimura, …, F. A. A. Mulder. 2016. Detecting O_2 binding sites in protein cavities. *Sci. Rep.* 6:20534.
32. Jorgensen, W. L., J. Chandrasekhar, …, M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
33. Rostkowski, M., M. H. M. Olsson, …, J. H. Jensen. 2011. Graphical analysis of pH-dependent properties of proteins predicted using PROPKA. *BMC Struct. Biol.* 11:6.
34. Lippert, R. A., C. Predescu, …, D. E. Shaw. 2013. Accurate and efficient integration for molecular dynamics simulations at constant temperature and pressure. *J. Chem. Phys.* 139:164106.
35. Hornak, V., R. Abel, …, C. Simmerling. 2006. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins.* 65:712–725.
36. Salomon-Ferrer, R., A. W. Götz, …, R. C. Walker. 2012. Routine microsecond molecular dynamics simulations with AMBER. 2. Explicit-solvent particle mesh Ewald. *J. Chem. Theory Comput.* 9:3878–3888.
37. Case, D. A., T. E. Cheatham, 3rd, …, R. J. Woods. 2005. The Amber biomolecular simulation programs. *J. Comput. Chem.* 26:1668–1688.
38. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an $N\cdot\log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.
39. Press, W., B. Flannery, S. Teukolsky, and W. Vetterling. 1986. Numerical Recipes: The Art of Scientific Computing, 3rd ed. Cambridge University Press, Cambridge, United Kingdom.
40. Lamm, G., and A. Szabo. 1986. Langevin modes of macromolecules. *J. Chem. Phys.* 85:7334–7348.
41. Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *J. Comput. Phys.* 23: 327–341.
42. Roe, D. R., and T. E. Cheatham, 3rd. 2013. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* 9:3084–3095.
43. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38, 27–28.

Please cite this article in press as: Schiffer et al., Capturing Invisible Motions in the Transition from Ground to Rare Excited States of T4 Lysozyme L99A, Biophysical Journal (2016), http://dx.doi.org/10.1016/j.bpj.2016.08.041

44. Durrant, J. D., L. Votapka, …, R. E. Amaro. 2014. POVME 2.0: An enhanced tool for determining pocket shape and volume characteristics. *J. Chem. Theory Comput.* 10:5047–5056.

45. Baron, R., and J. A. McCammon. 2007. Dynamics, hydration, and motional averaging of a loop-gated artificial protein cavity: the W191G mutant of cytochrome *c* peroxidase in water as revealed by molecular dynamics simulations. *Biochemistry.* 46:10629–10642.

46. Hess, B., D. van der Spoel, and E. Lindahl. 2008. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. 4:435–447.

47. Thompson, E. J., A. J. DePaul, …, E. J. Sorin. 2010. Evaluating molecular mechanical potentials for helical peptides and proteins. *PLoS One.* 5:e10056.

48. Smith, M. D., J. S. Rao, …, L. Cruz. 2015. Force-field induced bias in the structure of Aβ21–30: a comparison of OPLS, AMBER, CHARMM, and GROMOS force fields. *J. Chem. Inf. Model.* 55:2587–2595.

49. Han, B., Y. Liu, …, D. S. Wishart. 2011. SHIFTX2: significantly improved protein chemical shift prediction. *J. Biomol. NMR.* 50:43–57.

50. Gellatly, B. J., and J. L. Finney. 1982. Calculation of protein volumes: an alternative to the Voronoi procedure. *J. Mol. Biol.* 161:305–322.

51. Roche, J., J. A. Caro, …, C. A. Royer. 2012. Cavities determine the pressure unfolding of proteins. *Proc. Natl. Acad. Sci. USA.* 109: 6945–6950.

52. Maeno, A., D. Sindhikara, …, R. Kitahara. 2015. Cavity as a source of conformational fluctuation and high-energy state: high-pressure NMR study of a cavity-enlarged mutant of T4 lysozyme. *Biophys. J.* 108:133–145.

53. Wang, K., J. D. Chodera, …, M. R. Shirts. 2013. Identifying ligand binding sites and poses using GPU-accelerated Hamiltonian replica exchange molecular dynamics. *J. Comput. Aided Mol. Des.* 27:989–1007.

54. Miao, Y., V. A. Feher, and J. A. McCammon. 2015. Gaussian accelerated molecular dynamics: unconstrained enhanced sampling and free energy calculation. *J. Chem. Theory Comput.* 11:3584–3595.

55. Weininger, U., K. Modig, and M. Akke. 2014. Ring flips revisited: ^{13}C relaxation dispersion measurements of aromatic side chain dynamics and activation barriers in basic pancreatic trypsin inhibitor. *Biochemistry.* 53:4519–4525.

56. Bowman, G. R., and P. L. Geissler. 2014. Extensive conformational heterogeneity within protein cores. *J. Phys. Chem. B.* 118:6417–6423.

Biophysical Journal, Volume *111*

Supplemental Information

Capturing Invisible Motions in the Transition from Ground to Rare Excited States of T4 Lysozyme L99A

Jamie M. Schiffer, Victoria A. Feher, Robert D. Malmstrom, Roxana Sida, and Rommie E. Amaro

Supplementary Figures

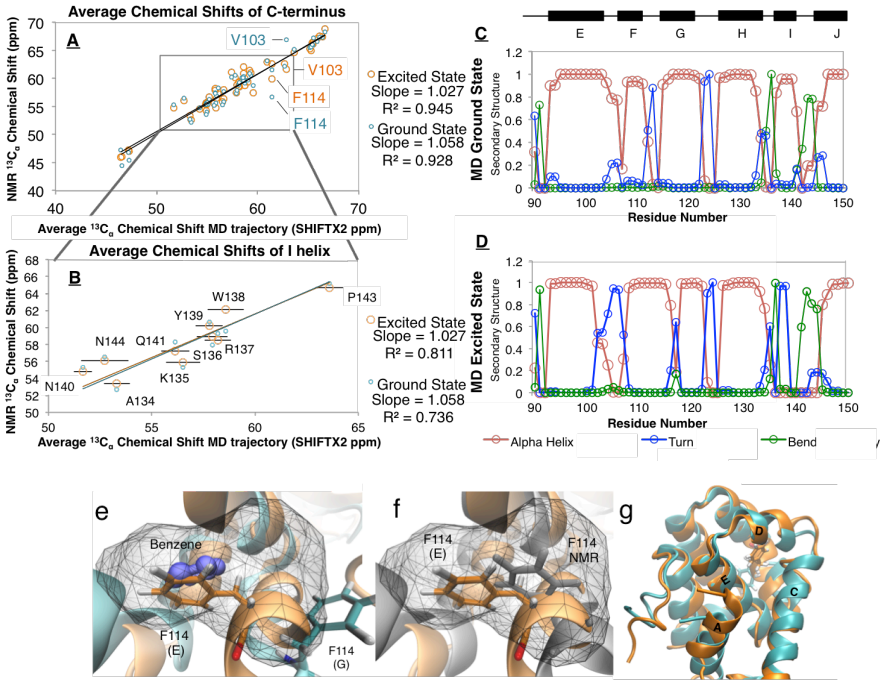


Figure S1 | Characterizing the structures of the T4 lysozyme L99A excited state from MD simulations with SHIFTX2 and secondary structure calculations. (a) The SHIFTX2 average back-calculated chemical shifts of residues in the C-terminal domain (residues 75-155) from the excited state simulated with Anton (x axis) are compared to the chemical shifts determined by relaxation dispersion NMR spectroscopy (y axis) of the excited state (orange). The fits to the experimental ground state chemical shifts are shown for comparison (cyan). (b) Plot that is the same as (a) but looking closer at residues in the I helix. Error bars are showing the standard deviation of the back-calculated chemical shifts of the excited state from the simulation. (c/d) Secondary structures calculated from frames of the ground state and excited states from MD simulation including the alpha helices (red), turns (blue), and bended structures (green). (e) Location of F114 in the buried pocket in the ground (G) and excited (E) states shown in the context of the buried ground state cavity (grey mesh). (f) Same as (e) but showing the location of F114 in the previously published Rosetta model (grey). (g) View of the ground and the excited state structures of the L99A mutant from simulation from the back face of the C-terminal domain to highlight the change in pitch of the A helix in the excited state (orange).

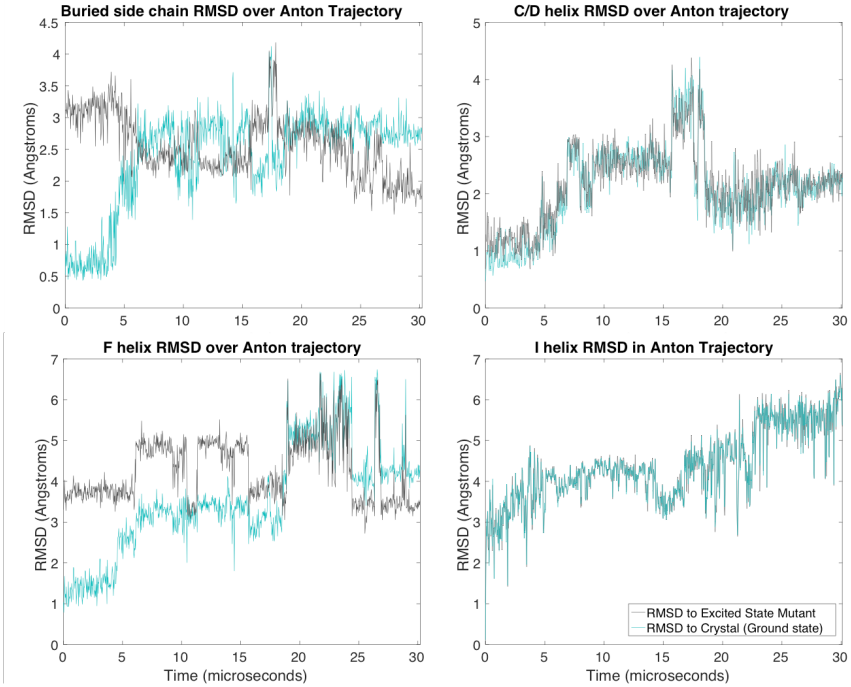


Figure S2 | Root-mean-squared deviation (RMSD) of the backbone atoms in the C-D, F-G, and I helices and in the side chain atoms (residues M102, F114, L133, W138, F153) of the buried hydrophobic residues from the Anton trajectory compared to the excited state mimic mutant and to the ground state crystal structure.

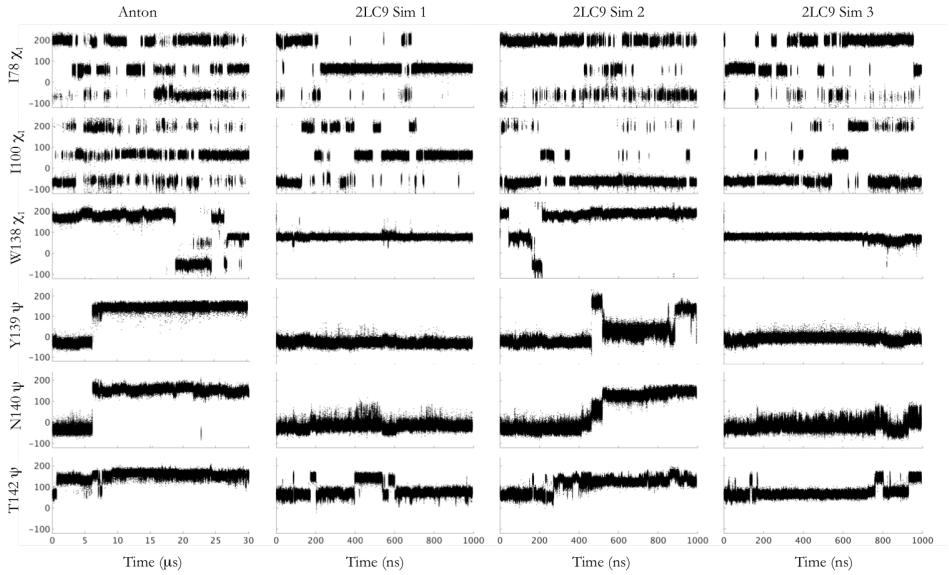


Figure S3 | The χ and ψ angle changes in the Anton trajectory of buried hydrophobic and I helix residues are also witnessed in simulations initiated from the previously published Rosetta model. Residues I100, I78, and W138 all neighbor the buried cavity of the L99A mutant. In the Anton trajectory, these residues sample different **χ** angles in the excited state in the ground state, which were not previously seen in the Rosetta model. However, simulations initiated from this Rosetta model also sample these same excited state χ_1 angles in at least one of the triplicate simulations (2LC9_1, 2LC9_2, 2LC9_3). Residues Y139, N140, and T142 also sample unique **ψ** angles early on in the Anton trajectory. In at least one, if not more, of the triplicate Rosetta model simulations, these same **ψ** angles are also sampled.

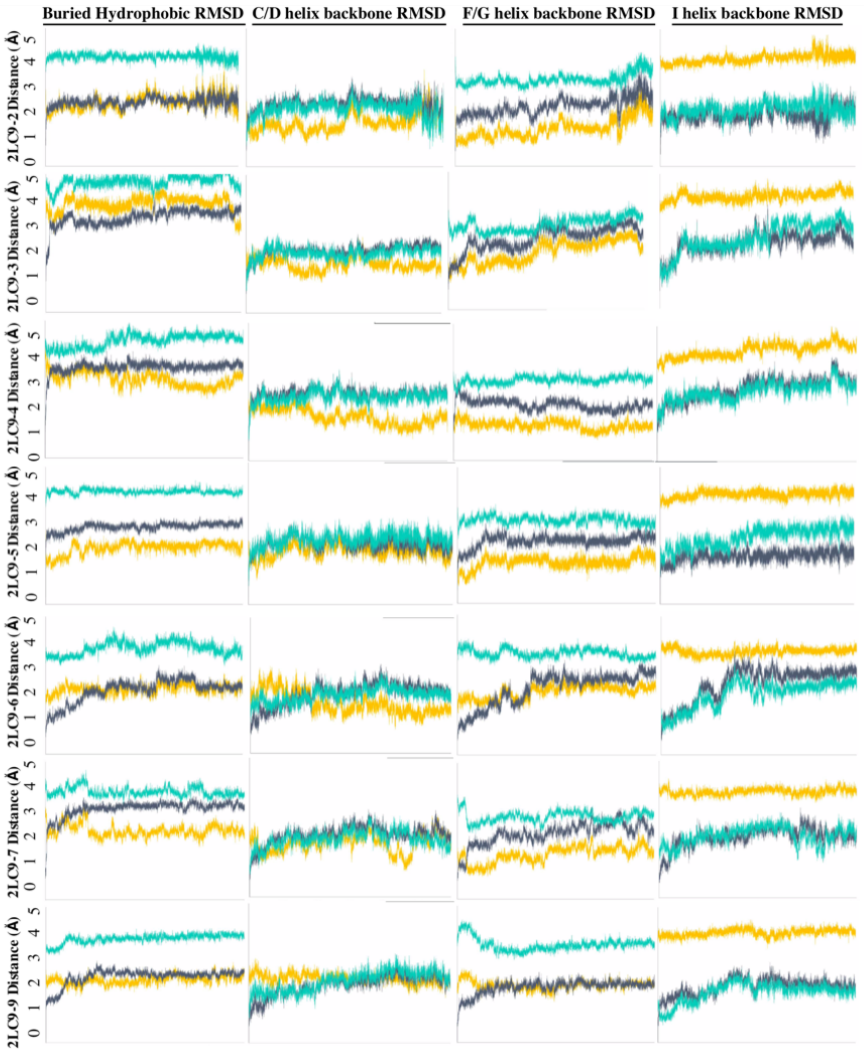


Figure S4 | Average RMSD measurements from simulations of in-silico re-mutated (A113G/P119R) triple mutant model of the L99A excited state (PDB: 2LC9). Each row represents RMSD measurements made during the simulations initiated from a different starting structure, but each

were a member of the 2LC9 PDB (10 structures in total). RMSD time plots are showing the RMSD of the simulated structures to the MD Anton excited state average structure (orange), to the excited state model starting structure (gray, 2LC9), and to the ground state crystal structure of L99A (teal) and are reported as an average over three parallel simulations. The first column is showing the RMSD time plots of buried residues (trace of V78, L87, M102, V103, V111, F114, L133, W138, F153). The second column is showing the RMSD time plots of the backbone of the C/D helix residues (residue 75-90). The third column is showing the RMSD time plots of backbone of the F/G helix residues (residue 108-122). The fourth column is showing the RMSD time plots of the backbone of the I helix residues (residue 135-145).

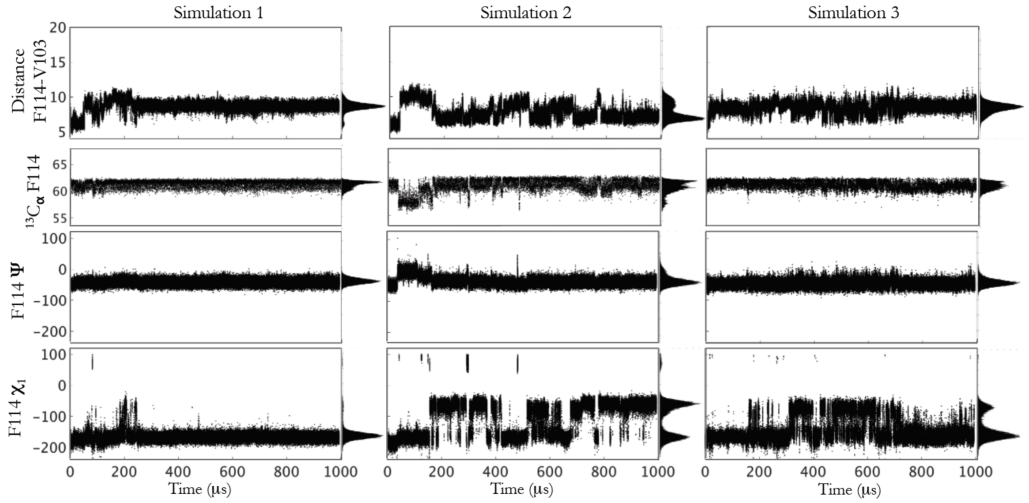


Figure S5 | Analysis of simulations initiated from previous model of the excited state. (A) Distance between F114 phenyl ring and amide nitrogen of V103, (B) back-calculated chemical shift of the $^{13}C_\alpha$ of F114, (C/D) and ψ and χ_1 angle plots for F114 during the three simulations initiated from the first structure of the previous published NMR excited state model (PDB 2LC9). The population histograms are also shown on the right of each plot.

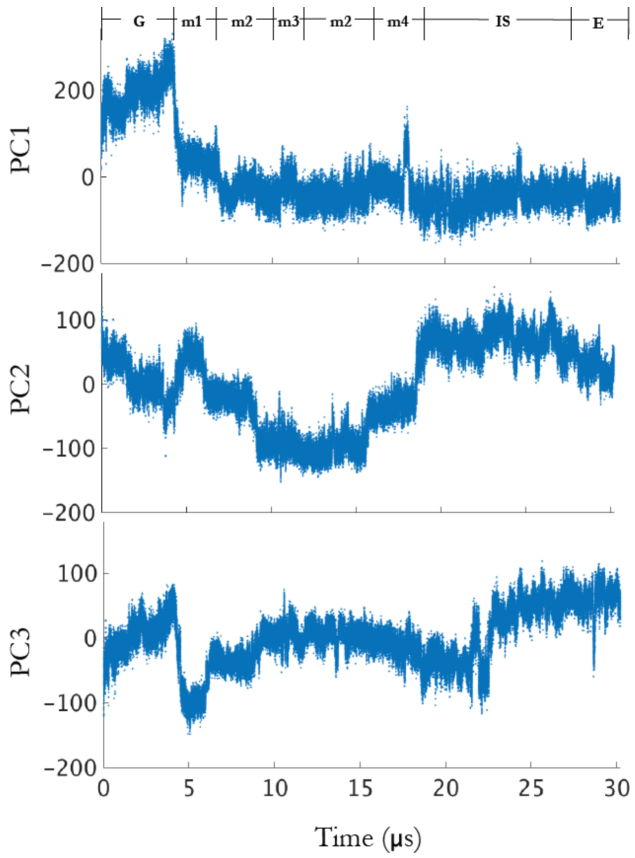


Figure S6 | Principal component (PC) analysis of Anton Trajectory from ground to excited state. Principal component changes over the simulation times for PC1, PC2, and PC3. PC1 identifies all of the metastates and has a slight difference between the IS and E states. PC2 and PC3 identify more clearly the difference between intermediate and excited states. PC1, PC2, and PC3 are accompanied with movies S2-S4 respectively.

ACKNOWLEDGEMENTS

Chapter 2, in full, is a reprint of the material as it appears in Capturing Invisible Motions in the Transition from Ground to Rare Excited State of T4 Lysozyme L99A 2016. Schiffer, Jamie M., Feher, V. A., Malmstrom, R. D., Sida, R., and Amaro R. E., Biophysical Journal, 2016. The dissertation author was the co-primary investigator and author of this material.

CHAPTER 3

Benzene Dissociation through a Protein Excited State

Feher, V. A. [†,1,4,*], Schiffer, J. M. [†,1], Mermelstein, D.[1], Mih, N.[5], Pierce, L.C.T.[2], McCammon, A. J.[1], Amaro, R. E.[1,3,4,*]

Author Affiliations

1. Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0378, United States.
2. Relay Therapeutics, 215 First Street, 3rd Floor, Cambridge, Massachusetts 02142, United States.
3. National Biomedical Computation Resource, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0608, United States.
4. Drug Design Data Resource, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0608, United States.

1. Department of Bioinformatics and Systems Biology, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0378, United States.

Correspondence to: vfeher@ucsd.edu or ramaro@ucsd.edu

Keywords ligand dissociation, L99A, ligand exchange, high-energy state, mobile defects

Classification BIOLOGICAL SCIENCES: Biophysics and Computational Biology

Author Contributions L.P. and V.F. performed aMD simulations; J.S. performed and analyzed APBS, POVME, internal distances, and SHIFTX2 (back-calculated chemical shift) calculations. D.M., N.M., J.S. and V.F. analyzed the aMD simulations. J.S. and V. F. wrote the paper with input from all other authors. R. A. and AJM were the leading principal investigator. V. F. initiated and directed the study.

Main Points

- **Benzene exits the buried cavity of L99A through the excited state.**
- Benzene exits and enters through different surface openings in L99A.
- Benzene egress and ground to excited state transitioning pass through the same intermediate state. .
- The L99A mutation exaggerates dynamic motions witnessed in aMD simulations of the wild type T4 lysozyme implicating the existing WT* cavity in lysozyme functionally-related mobile defects.

Abstract

The atomic level elucidation of mechanisms for protein-ligand dissociation is the next frontier in biophysical characterization of enzyme catalysis and pharmaceutical drug design. Although experimental kinetics methods can suggest that a "conformational change" is required for ligand release, the precise motions that ultimately coordinate release are only known for a handful of proteins. Even for the historically well-studied T4 lysozyme L99A cavity mutant, atomic-level mechanisms of ligand release are just beginning to surface. Here, we add to this growing body of data, and report the mechanism of benzene dissociation from L99A through the application of accelerated molecular dynamics simulations. We observe benzene dissociation in a multistep process through cooperativity of core residue side-chain rotamers. These concerted motions also occur in transition to the experimentally characterized high-energy state. Thus, benzene dissociates through the high energy state of L99A, providing an explanation for the similarities in experimentally determined rates of these two dynamic phenomonen.

Introduction

While receptor flexibility has been acknowledged as important for ligand binding and molecular recognition,[1–3] its importance is less emphasized in studies of ligand release, largely due to the difficulty in studying ligand release and the lack of atomic-level information on these processes. Currently, only a few examples in the literature exist of such simulation studies, and typically ligand egress is simulated either through temperature acceleration or through steered or biased simulations.[4–9] There are concerns with both of these methods.[10,11] With temperature accelerated MD, there is the concern

that the protein of interest may be sampling along the unfolding pathway and be illuminating a dynamical pathway not observed at standard temperatures. With steered or biased MD, MD is setup along a predetermined reaction pathway, and thus the quality of the simulated dynamic model is limited by the quality of information available about a given reaction pathway. Thus, there remains a need to establish an atomic framework for the mechanism of small molecule release in proteins and enzymes through simulation while still preserving the underlying free energy landscape.

L99A has been an attractive model system for studying ligand exchange with buried protein cavities because the cavity is hydrophobic and seemingly empty under standard temperatures and pressures. [12–21] L99A contains an engineered >100 $Å^3$ cavity that is buried >5Å from the protein surface, and has an enigmatic ligand exchange path.[22] Despite this large vacant cavity, the crystal structure of L99A is identical to the wild-type protein (Figure 1). This similarity proved to be misleading, as this buried cavity is associated with two experimentally-detected dynamic motions not witnessed in the wild-type protein: the formation of a sparsely populated excited state[23–25] and the binding of ligands to the buried cavity despite the absence of a clear path from solvent in the crystal structure.[12,13,15,21,26–32] Both of these findings pointed to unique dynamics that could not be realized through crystallography or NMR spectroscopy alone. As in other non-rigid condensed system, stochastic fluctuations elicit small structural changes in proteins while large structural changes occur through of concerted fluctuations.[2] Over the past two decades experimentalists have sought to discover which types of structural changes sampled by the L99A mutant; large and concerted or small and stochastic.

Pioneering NMR studies revealed initial insight into the dynamics of this protein model system. In studies published over two decades ago demonstrated that ligands enter into the cavity rapidly with bimolecular rate constants of ~10^6-10^7 $M^{-1}s^{-1}$, suggestive of a low activation barrier to binding. These studies also calculated a much slower egress rate of ~ms^{-1} at 30 ºC.[27] Seminal work with Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion NMR spectroscopy revealed that a low populated (3%, 25ºC) excited state was just 2.0 kcal/mol higher in energy than the ground state. This work also pointed to residues in the E, F, and I helices as undergoing the largest structural changes, with other residues surrounding the buried cavity also seeing shifts in their chemical environments over a broad timescale range.[24,33,34] More recently, through the use of a chemical shift-based triple mutant model for the excited state, it was shown that benzene does not bind to the excited state.[25] This result was further confirmed through Gaussian-accelerated molecular dynamics studies, showing benzene egress through the ground state between the D, F, and G helices.[35]

All atom simulations have continued to lend perspective on the underlying free energy landscape and conformational dynamics of L99A. Our most recent work describes a long-timescale trajectory performed on the Anton supercomputer with conventional unbiased MD, which captured a transition from ground to the "invisible" excited state of L99A.[36] This simulation also shows that a pocket opens up between the F, H and I helices large enough to permit benzene release only once the transition to the excited state has occurred. Mobile defects create pockets between an array of other helices occur in a state independent manner, and were also observed in simulations of O_2 ingress and egress.[37] Pocket openings between the D, F and G helices occur in the ground state structure,

which further supports previous findings that benzene binds to the ground state ensemble through this surface (Figure 1).

Recent metadynamics simulations of the L99A mutant have witnessed the transition from ground to excited state as well as the path of benzene egress.[7] These ground to excited state simulations were biased towards sampling between the crystallographic ground state and a previously published Rosetta model.[25] However, as we demonstrated in our most recent results,[36] this Rosetta model is limited by the constraints set in the model building process, and MD simulations from these models relaxed into the excited conformational state we sampled through long-timescale simulation on the Anton supercomputer.[36] Additionally, the adiabatic biased MD (ABMD) simulations of benzene egress witnessed benzene egress through a 'ratcheting potential' that selects for fluctuations to a ligand-free state. This method therefore biases along a pathway that may or may not recapitulate the underlying conformational states of the free energy landscape of L99A.

Here we perform accelerated MD simulations of the L99A mutant T4 lysozyme with benzene bound and through an atomic-level mechanistic approach we witness benzene release through the same surface as the ABMD simulations.[7] Additionally, we show that benzene exits through the high-energy state of L99A, uniting decades of disparate data on this historically well-studied protein into an atomic framework. In accelerated molecular dynamics (aMD) a biasing potential accelerates the crossing of high-energy barriers without modifying the underlying conformational landscape, and thus does not require any prior knowledge of the landscape.[38] This method has been shown to recapitulate relaxation dispersion NMR spectroscopy results, providing us with

confidence in application of this method towards to study of benzene egress.[39] Results from these aMD simulations also agree with unbiased adiabatic MD of L99A performed on the Anton Supercomputer.[36] However, as with all accelerated or exchange simulation methods, there is always concern about how well the perturbed landscaped may recapitulate the conformational states of the native free energy landscape. For this reason, here we compare the results from aMD with results from conventional molecular dynamics and relaxation-dispersion NMR spectroscopy and show that aMD is able to recapitulate results from both computational and experimental methods without any preset bias towards these results. We witness the same transient packing defects and ring rotations during the egress of benzene in our aMD simulations were also seen in the excited state transitioning with conventional MD.[36] Thus, as chemical shift calculations and rotamer measurements confirm,[40] benzene ultimately dissociates from L99A through the NMR-characterized excited state. Finally, our simulations demonstrate that the motions associated with excited state transitioning and benzene release are similar to those seen in wild type. These results provide an answer to a decades old question of whether cavity-creating mutants significantly alter the dynamic nature of the wild-type core,[27] demonstrating here that dynamics associated with cavity mutations are simply exaggerating wild type motions.

Results

Benzene leaves through a transient surface opening in the L99A excited state

The time constant for benzene exchange with the L99A cavity mutant is ~1 millisecond,[27,28,33,41] a timescale beyond the reach of conventional MD simulation. In

order to sample slower conformational changes of L99A we used accelerated molecular dynamics (aMD); a methodology that has recently been shown to accurately recapitulate experimentally measured long-lived conformational phenomena such as torsional populations and water-protein exchange.[39,42] Here we perform aMD simulations of the L99A cavity mutant (apo),[43] of benzene bound to the L99A cavity mutant (holo),[16] and of WT* T4 lysozyme as a control simulation.[44]

In one of three triplicate simulations initially set up to explore benzene – L99A lysozyme conformational states we observed that benzene unbinds from the protein through a transient surface opening between the F/G, H, and I helices, (FGH surface, Movie S1). The distance between the location of benzene at FGH surface and of benzene in the L99A crystallized is on average 9.3 Å throughout the aMD simulation. In the L99A-benzene bound crystal, benzene is less than 6 Å from the solvent through the D and G helices (DG surface). How and why does benzene move from its ground state position in its path of egress?

Further analysis of the protein conformation from which benzene leaves reveals a structure more consistent with the experimentally characterized excited state conformation (refs) than the ground state used to initiate the simulations. The excited state of the L99A cavity mutant has a number of structural hallmarks determined experimentally and confirmed through previous simulations.[7,25,36,45] Most notably, the helices F and G refold into a single helix and sidechain of F114 locates from the ground state surface exposed position to one with the side-chain buried in the cavity formed by the L99A mutation.[7,25,36,45] Benzene egress coincides with the transition of the L99A protein from the ground to excited state, and leaves the protein in a series of two steps.

The transition from S1 to S2 occurs at 2.2 x 10^5 aMD timesteps, and the transition from S2 to S3 occurs at 2.5 x 10^5 aMD timesteps. The latter transition coexists with the transition of the two helices F and G to a single helix, specifically a decrease in i to i+4 aminde to carbonyl distances for residues G113 – S117, V111 - T115 and the F114 torsional angle changes. For comparison, back-calculated chemical shifts from the aMD trajectories S3 state recapitulate chemical shift differences from the Anton-simulated excited state and chemical shift based model of the excited state[25,36] (Figure 2). Thus benzene leaves through the NMR-characterized high energy state.

Benzene passes through the same states that the aromatic ring of F114 travels in the transition from ground to excited state, albeit in opposite directions (Figure 3). Internal distances between F114 or benzene and residues that surround the buried cavity demonstrate the similarity in the paths of egress by benzene and of excited state transitioning by the aromatic ring of F114 (Figure 3). The crystal structure of benzene in the S1 state is identical to the position of the F114 aromatic ring in the excited state. The position of benzene in the S2 state is the same position of the F114 aromatic ring in the intermediate state.[36] Finally, the position of benzene in the S3 state is the same position of the F114 aromatic ring in the ground state, crystal apo structure of L99A.

Benzene is simulated to leave only once through the FHI surface, yet multiple other pocket openings occur in our simulations, as we similarly seen in the transition to the excited state.[36] For example, mobile defects between the DG, HJ, and at the helix termini of the D, E, G, H and J helices from the buried cavity occur. All of these surfaces have been previously simulated to allow other ligands to ingress and egress into the buried cavity, but these events are not witnessed here for benzene.

To understanding the preferences of cavity openings for benzene release, we performed pocket volume and shape measurements with POVME2.0.[46,47] These results illuminate pocket openings at the DG surface large enough to allow benzene egress (Supplementary figure #). However, benzene does not leave from this surface even when completely solvent exposed. The pocket opening at the FHI pocket is narrower than the opening at the DG surface, and yet benzene leaves through the FHI pocket.

Cavity volume and mobile defects alone cannot account for the preference of benzene to egress through the FHI pocket. We next turned to the Adaptive Poisson-Boltzman Solver (APBS)[48,49] to calculate the electrostatics surrounding the cavity openings simulated to both bind benzene (DG) and release it (FHI) in the hopes that these calculations might reveal a role of electrostatics in benzene release. Indeed, we find that the FHI cavity is more polar than the DG cavity. The polarity of the FHI pocket is due to local helix unwinding that exposes backbone amides at the pocket surface. It is of interest therefore, that the mere opening of a path between the core and surface is not sufficient for egress, contrary to what is observed for ingress.[35]

Dynamic motions surrounding the buried cavity result in productive and non-productive excursions from the ground state

Of interest is that the mere opening of a path between the core and surface are not sufficient for egress, contrary to what is observed for ingress. Both productive excursions, in which benzene egresses, and non-productive excursions, in which benzene remains in the buried L99A cavity, are seen in the aMD simulations of holo L99A (Figure 2). Throughout the simulation for apo and benzene bound L99A, N-terminal

residues of the F-helix are observed to unwind and rewind. This unwinding is consistent with an array of experimental observables: low NMR order parameters,[50] lower hydrogen exchange protection factors relative to WT[51,52] high crystallographic B-factors,[53] propensity to deform with local point mutations and deformability upon ligand binding to the nearby cavity,[15,16,54] especially most recently to a congeneric series of ligands.[21] The difference between these two conformational shifts lies in the degree of concerted motions for both backbone and side chain torsion angles.

There are cases where benzene moves around in the buried cavity but does not leave during our aMD trajectories. In these non-productive excursions (column 1, Figure 2), two instances of rotamer shifts are witness, at 3,000 aMD timesteps and at 13,000 aMD timsteps. The first non-productive excursion results from shifting of both F114 and F153 χ_1 rotamer flips, but are unaccompanied by any backbone conformational changes or motion of benzene. The second non-productive excursion results from F114 χ_1 rotamer flips as well as folding of the F and G helices into a single helix, as witnessed in the reduction in distance between backbone nitrogen of G113 and the backbone carbonyl oxygen of S117, and is accompanied by benzene shifting to an S2-like state. However, in this second excursion, F153 and L133 χ_1 rotamer flips do not accompany the transition, and benzene returns back to its crystallographic position. Interestingly, for the second excursion, when F114 χ_1 and ψ rotamer flips occur in unison, it also results in shifts in the back-calculated chemical shifts[40] of F114 and V103.

The productive dissociation of benzene from the L99A cavity occurs via a series of concerted side-chain and backbone rotations for the F and G helices that were also

seen in the transition to the excited state (Figure 2). In the simulated productive excursion that permits benzene release, unwinding of the F-helix leads to N-terminal extension of the G-helix through hydrogen bonds between G110-F114, A112-N116, and G113-S117 (Figure 2, Supplementary Figure 2) as well as χ_1 rotamer flips for hydrophobic core residues V103, V111, F114, L121, L133, and F153 (Figure 2, Supplementary Figure 3). The concerted motions that allow benzene to shift out of its ground state are identical to those seen in the excited state transition.

Snapshots of the benzene locations in the buried cavity demonstrate that in our simulation, benzene leaves through a stochastic series of three states (S1, S2, and S3 in Figure 5A). Transitions between these states occur due to the concerted motions described above. As the F/G helix rewinds, L121, L133 and F153 undergo rotamer changes opening up an adjacent cavity, the second state benzene samples in its exit pathway (S2 in Figure 5). Concurrently with benzene movement, the F114 ring relaxes into a more buried position in the L99A pocket, just as it does in the excited state. Benzene remains in this pocket for ~25 ns before the F153 side chain rotates back to its original angle, and benzene moves towards the protein surface to a position occupied by the F114 phenyl group in the ground state, what we refer to here as the S3 state (Figure 5).

Incredibly, prior to benzene passage from S1 to S2, an enormous and hydrophobic cavity opens to solution between the D, F, and G helices. This opening completely exposes benzene and the surrounding hydrophobic residues to aqueous solution (Figure 6A). This same cavity opening occurs preceding F114 side chain flipping into the buried cavity on the path to the excited state in the Anton trajectory (Figure 6B). This identical

exposure of a large surface area of hydrophobic residues (Supplementary Fig. 1) thus occurs prior to concerted motions in benzene leaving or excited state transitioning, implying that this energetically unfavorable mobile defect could play a role in both processes.

Wild-type T4 lysozyme samples similar yet attenuated plasticity relative to the cavity-expanded mutant

Acclerated MD simulations on WT* lysozyme (PDB = 1L63) were conducted for comparison to long timescale simulation on apo L99A and benzene bound L99A. Concerted motions seen in WT T4 lysozyme simulations are reminiscent of those seen for the benzene-L99A complex (Figure 7A/B). The difference between the fluctuations seen with WT and those seen in both the apo and holo simulations of L99A, are that these motions do not result in a stable conformation, and relaxation back into a "ground" state proceeds within a few 100 ns of aMD simulation. In contrast, once these concerted motions occur in the cavity mutant in both the holo and apo states the protein moves into the excited state because there is available space in the buried cavity for F114 to reside. In wild-type, these concerted motions of internal and backbone dihedral angles serve to open up the groove between the F-G, H, and I helices, the groove which binds the peptide portion of the peptidoglycan substrate of T4 lysozyme. (Figure 7C/D). In fact, this motion is important for the threading of the peptide portion of peptidoglycan out of the active site cleft as part of the mechanism of processivity of T4 lysozyme.

Discussion

The rates of benzene leaving and excited state transitioning are experimentally identical (~1 ms^{-1}). Here we show that these phenomenon occur through the same atomic level mechanism, explaining their similarities in rates for the first time.[27,55] Interestingly, previous simulations have not identified that benzene leaves through the excited state of L99A.[7] Interestingly, the concerted rotations of L133 and F114 have been previously implicated as important and assigned as collective variables in the conformational transition from ground to excited state,[7] but F153 and L121 rotations have not. This observation suggests that our aMD simulation reveals a greater level of cooperativity between more residues neighboring the buried cavity than previously realized.[45] Additionally, Metadynamics simulations did not account for other collective variables, like the G113-S117 backbone distances, that we show here to be of importance to the concerted motions leading to the excited state, and thus benzene release. Our results also suggest that the off pathway requires a greater degree of cooperativity than the on pathway as witnessed from Gaussian accelerated MD (gaMD).[35]

Taken together, the results from this aMD simulation, the Anton trajectory,[36] and the Markov State Model (MSM)[45] of the high temperature triple mutant simulations reveal that these rates are so similar because these events pass through the same transition state. The MSM of the triple mutant[45] identified an intermediate state that was both the most populated as well as the state through which the highest flux was measured from ground to excited state (state 12). Both the S2 state of benzene in our aMD and the intermediate state from the Anton trajectory recapitulate this state's backbone structure and chi angles of buried F114. Thus, here, we confidently rename this state as a

transition state in the path from ground to excited state, and thus we sample this transition state in both benzene egress and excited state transitioning. These results imply that this transition state is the rate-limiting step resulting from identical concerted motions, mobile defects, and ring flips in the buried cavity in both ligand egress and excited state transitioning.

Cavity volume fluctuation and concerted motions between states and as part of ground, transition, and excited states demonstrate that the idea of rigid and discrete "sites" of buried cavities[37] is not reality. Even in the ground state ensemble of our aMD and long-timescale cMD simulations the buried cavity samples a range of cavity volumes and shapes, opening up paths between the buried cavity and solvent through an array of different helices (Figure 3, 4). Not only the dynamics or the kinetics, but also the shape, size, and electrostatics of buried cavities in proteins cannot exclusively be defined through crystallography alone, as crystallography is selecting for a single ground state conformation. These fluctuations in cavity volumes and electrostatics have vast implications for the role of dynamics of buried cavities in drug binding and release, and thus is an important consideration in drug discovery efforts.

The location of mobile defects and the surrounding surface electrostatics and sizes may guide different ligands to bind through different conformations of the same protein state(i.e. ground, excited, intermediate). For example, in the ground state ensemble from our simulations, the FHI surface is highly charged, perhaps explaining why this surface is not the location for benzene ingress but a strong candidate for benzene egress (Figure 4B, Supplemental figure 1). A pocket through this surface to the buried cavity has been previously observed in a several instances: in the crystal structure of the L99G mutant,[62]

in crystal structure of L99A bound to a congeneric series of ligands,[21] in the L99A excited state from long-timescale simulations performed on the Anton supercomputer,[36] and in adiabatic-biased MD simulations of benzene dissociation.[7] Interestingly, the L99G mutant shifts the F and G helices closer to the D helix. The shifted F/G helices open up a pocket between the F, H and I helices which accommodates five water molecules in the crystal structure, further suggesting that this pocket is polar.

Initial crystal structures of the L99A mutant revealed that this mutation expands the 40 $Å^3$ cavity to ~150 $Å^3$. Without this knowledge, it is likely that this mutant would not continue to be studied today. However, the exact effects of these expansions on the dynamic atomic-level motions are just beginning to be realized through *in silico* exploration.[7,35–37,45,56,57] Here, we contribute to this body of *in silico* work, and answer the decades old question of whether a single mutant can manifest an entirely new dynamic motion.[27] Our results show that the L99A mutant does not illicit a completely unique set of dynamics, but rather exaggerates the breathing and quaking motions[58] intrinsic to wild-type T4 lysozyme. These results have broad implication for how to design new proteins from wild-type proteins, to engineer exaggerated motions in protein interiors.

Over the past few years, a wealth of simulation data has arisen for the L99A mutant dynamics, each providing another piece in solving the dynamic puzzle. The results herein could not stand without the multitude of other simulation techniques used, and even then caution should be taken in analyzing the results of any of these methods in isolation. For our results presented here, as well as other recent work,[35] aMD and gaMD alter the underlying energy landscape and thus may produce unnatural conformation states. For our recently published simulations on Anton, as well as simulations of O_2

ingress and egress,[37] long timescale cMD has been shown to sample unnatural states in a force-field dependent manner.[59–61] In Hamiltonian replica exchange MD,[26] sampling along predefined Hamiltonians can miss underlying conformational changes in the structure of the protein. Temperature enhanced MD and metadynamics also suffer from their own biases and inaccuracies.[7,45] However, when results from all six of these simulation methods are taken together and agreements between the simulations are seen, we can be more confident about each individual result. Collectively, these simulations have reveled paths of ligand ingress and egress, key transition states in the pathway from ground to excited state, and the conformational changes, kinetics, thermodynamics, and electrostatics of the underlying energy landscape of this once enigmatic protein. What remains to be answered are questions of the role of water in dynamics and further exploration and characterization of the free energy landscape in terms of unique collective variables like buried cavity shapes and electrostatics.

Author Information REA is a co-founder of Actavalon, Inc. VAF has an equity interest in Actavalon, Inc. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to V. F. (vfeher@ucsd.edu)

Footnotes

† co-authors

Methods

For all POVME, SHIFTX2 chemical shift measurements, dihedral angle measurements,

internal distance measurements, and simulation details of the Anton trajectory please reference our previous work.[36]

aMD simulation setup

Three systems were set up for accelerated molecular dynamics simulations using identical methods described previously[42] with the following exceptions; PDB codes 1L63.pdb (pH = 5.3), 1L90.pdb, and 181L.pdb were used for the WT*,[44] L99A apo mutant[63] and L99A benzene bound mutant lysozyme,[54] respectively, 30 Cl^- an 21 Na^+ ions. A 10-Å cutoff radius was used for range-limited interactions, with Particle Mesh Ewald electrostatics[64] for long-range interactions, simulations run at 300K. Run on GPUs using AMBER 11 SANDER 2010.[65,66] 181L.pdb system run on AMBER 12 SANDER 2012. ~1.5us of simulation for each system (WT*, L99A, L99ABNZ).

The benzene for 181L.pdb was parameterized using Jaguar v. X, Schrodinger, Inc., for geometry optimization and charge calculations, then incorporated into the starting structure MD parameter file using GAFF and XLeap.

To calculate boost potentials, 50 ns of conventional molecular dynamics was used as a reference and as the starting point for the accelerated simulation, as described previously.[42] Boost potentials for the three systems were calculated as follows:

164 amino acids,

alpha (dihedral) = (1/5) (4kcal/mol/aminoacid * 164) = 132

4 kcal/mol/aminoacid * 164 = 656

E(dih-boost) = E(dih) + (4*164)

1L63:

alpha(dih) = 132

E(dih-boost) = 1879 +656 = 2535 kcal/mol

Alpha(PE) = 33543*0.175 = 5870

E(PE-boost) = -95797 + 5970 = -89887 kcal/mol

1L90:

alpha(dih) = 132

E(dih-boost) = 1884 kcal/mol * 656 = 2540 kcal/mol

33512 atoms, 164 amino acids, EPtot = -95693

alpha(PE) = 33512 * 0.175 = 5865

E(PE-boost) = -95693+5865 = -89829 kcal/mol

181L:

alpha(dih) = same

E(dih-boost) = 1775 +656 = 2431 kcal/mol

Alpha(PE) = 34738*0.175 = 6079

E(PE-boost) = -96676 + 6079 = -90597 kcal/mol

APBS calculations

The PDB2PQR[49] automated pipeline was used to investigate the electrostatics of buried cavities and pocket openings from the aMD trajectories. A pH of 5.5 was chosen to align with the pH from multiple NMR studies[28,34] and our simulations.[36] The PARSE force field was chosen for the calculation.[67]

Acknowledgements

This work was funded in part by the Director's New Innovator Award Program NIH DP2 OD007237 to REA. Funding and support from the National Biomedical Computation Resource (NBCR) is provided through NIH P41 GM103426. JMS was supported by the NIH Molecular Biophysics Training Grant T32 GM008326. Anton computer time was provided by the Pittsburgh Supercomputing Center (PSC) and the National Center for Multiscale Modeling of Biological Systems (MM Bios) through Grant P41GM103712-S1 from the National Institutes of Health. D.E. Shaw Research generously made the Anton machine at PSC available.

References

(1) Baron, R.; McCammon, J. A. Dynamics, Hydration, and Motional Averaging of a Loop-Gated Artificial Protein Cavity: The W191G Mutant of Cytochrome C Peroxidase in Water as Revealed by Molecular Dynamics Simulations. *Biochemistry* **2007**, *46* (37), 10629–10642.

(2) Karplus, M.; Gelin, B. R.; McCammon, J. a. Internal Dynamics of Proteins. Short Time and Long Time Motions of Aromatic Sidechains in PTI. *Biophys. J.* **1980**, *32* (1), 603–618.

(3) Wüthrich, K.; Wagner, G. Internal Dynamics of Proteins. *Trends Biochem. Sci.* **1984**, *9* (4), 152–154.

(4) Kuczera, K.; Lambry, J. C.; Martin, J. L.; Karplus, M. Nonexponential Relaxation after Ligand Dissociation from Myoglobin: A Molecular Dynamics Simulation. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90* (12), 5805–5807.

(5) Long, D.; Mu, Y.; Yang, D. Molecular Dynamics Simulation of Ligand Dissociation from Liver Fatty Acid Binding Protein. *PLoS One* **2009**, *4* (6), 1–10.

(6) Hu, Y.; Liu, H. A Case Study on Temperature Accelerated Molecular Dynamics Simulation of Ligand Dissociation: Inducer Dissociation From the Lac Repressor

Protein. *J. Phys. Chem. A* **2014**.

(7) Wang, Y.; Papaleo, E.; Lindorff-larsen, K. Mapping Transiently Formed and Sparsely Populated Conformations on a Complex Energy Landscape. **2016**.

(8) Swegat, W.; Schlitter, J.; Krüger, P.; Wollmer, A. MD Simulation of Protein-Ligand Interaction: Formation and Dissociation of an Insulin-Phenol Complex. *Biophys. J.* **2003**, *84* (3), 1493–1506.

(9) Sonoda, M. T.; Martínez, L.; Webb, P.; Skaf, M. S.; Polikarpov, I. Ligand Dissociation from Estrogen Receptor Is Mediated by Receptor Dimerization: Evidence from Molecular Dynamics Simulations. *Mol. Endocrinol.* **2008**, *22* (7), 1565–1578.

(10) Abrams, C.; Bussi, G. Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration. *Entropy* **2014**, *16* (1), 163–199.

(11) Adcock, S. a; Mccammon, J. A. Molecular Dynamics : Survey of Methods for Simulating the Activity of Proteins Molecular Dynamics : Survey of Methods for Simulating the Activity of Proteins. **2006**, *106* (February), 1589–1615.

(12) Lopez, C. J.; Yang, Z.; Altenbach, C.; Hubbell, W. L. Conformational Selection and Adaptation to Ligand Binding in T4 Lysozyme Cavity Mutants. *Proc Natl Acad Sci U S A* **2013**, *110*, E4306–E4315.

(13) Baldwin, E.; Baase, W. A.; Zhang, X.; Feher, V.; Matthews, B. W. Generation of Ligand Binding Sites in T4 Lysozyme by Deficiency-Creating Substitutions. *J Mol Biol* **1998**, *277*, 467–485.

(14) Mann, G.; Hermans, J. Modeling Protein-Small Molecule Interactions: Structure and Thermodynamics of Noble Gases Binding in a Cavity in Mutant Phage T4 Lysozyme L99A. *J Mol Biol* **2000**, *302*, 979–989.

(15) Morton, A.; Baase, W. A.; Matthews, B. W. Energetic Origins of Specificity of Ligand Binding in an Interior Nonpolar Cavity of T4 Lysozyme. CC:See Also the next Paper: A. Morton and B.W. Matthews, Biochem 34, 8576-8588 (1995), Specificity of Ligand Binding in a Buried Nonpolar Cavity of T4 Lysozy. *Biochem* **1995**, *34* (27), 8564–8575.

(16) Eriksson, A. E.; Baase, W. A.; Wozniak, J. A.; Matthews, B. W. A Cavity-Containing Mutant of T4 Lysozyme Is Stabilized by Buried Benzene. *Nature*

1992, *355* (6358), 371–373.

(17) Quillin, M. L.; Breyer, W. A.; Griswold, I. J.; Matthews, B. W. Size versus Polarizability in Protein-Ligand Interactions: Binding of Noble Gases within Engineered Cavities in Phage T4 Lysozyme. *J. Mol. Biol.* **2000**, *302* (4), 955–977.

(18) Purisima, E. O.; Hogues, H. Protein-Ligand Binding Free Energies from Exhaustive Docking. *J. Phys. Chem. B* **2012**, *116* (23), 6872–6879.

(19) Malmstrom, R. D.; Watowich, S. J. Using Free Energy of Binding Calculations to Improve the Accuracy of Virtual Screening Predictions. *J. Chem. Inf. Model.* **2011**, *51* (7), 1648–1655.

(20) Ucisik, M. N.; Zheng, Z.; Faver, J. C.; Merz, K. M. Bringing Clarity to the Prediction of Protein-Ligand Binding Free Energies via "Blurring." *J Chem Theory Comput* **2014**, *10*, 1314–1325.

(21) Merski, M.; Fischer, M.; Balius, T. E.; Eidam, O.; Shoichet, B. K. Homologous Ligands Accommodated by Discrete Conformations of a Buried Cavity. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (16), 5039–5044.

(22) Eriksson, A. E.; Baase, W. A.; Zhang, X.-J.; Heinz, D. W.; Blaber, M.; Baldwin, E. P.; Matthews, B. W. Response of a Protein Structure to Cavity-Creating Mutations and Its Relation to the Hydrophobic Effect. *Science (80-.).* **1992**, *255* (5041), 178–183.

(23) Korzhnev, D. M.; Orekhov, V. Y.; Dahlquist, F. W.; Kay, L. E. Off-Resonance R1?? Relaxation outside of the Fast Exchange Limit: An Experimental Study of a Cavity Mutant of T4 Lysozyme. *J. Biomol. NMR* **2003**, *26* (1), 39–48.

(24) Mulder, F. A. A.; Mittermaier, A.; Hon, B.; Dahlquist, F. W.; Kay, L. E. Studying Excited States of Proteins by NMR Spectroscopy. *Nat. Struct. Biol.* **2001**, *8* (11), 932–935.

(25) Bouvignies, G.; Vallurupalli, P.; Hansen, D. F.; Correia, B. E.; Lange, O.; Bah, A.; Vernon, R. M.; Dahlquist, F. W.; Baker, D.; Kay, L. E. Solution Structure of a Minor and Transiently Formed State of a T4 Lysozyme Mutant. *Nature* **2011**, *477* (7362), 111–114.

(26) Wang, K.; Chodera, J. D.; Yang, Y.; Shirts, M. R. Identifying Ligand Binding Sites and Poses Using GPU-Accelerated Hamiltonian Replica Exchange Molecular Dynamics. *J. Comput. Aided. Mol. Des.* **2013**, *27* (12), 989–1007.

(27) Feher, V. a; Baldwin, E. P.; Dahlquist, F. W. Access of Ligands to Cavities within the Core of a Protein Is Rapid. *Nat. Struct. Biol.* **1996**, *3* (6), 516–521.

(28) Mulder, F. A.; Skrynnikov, N. R.; Hon, B.; Dahlquist, F. W.; Kay, L. E. Measurement of Slow (Micros-Ms) Time Scale Dynamics in Protein Side Chains by (15)N Relaxation Dispersion NMR Spectroscopy: Application to Asn and Gln Residues in a Cavity Mutant of T4 Lysozyme. *J Am Chem Soc* **2001**, *123*, 967–975.

(29) Vallurupalli, P.; Hansen, D. F.; Lundstrom, P.; Kay, L. E. CPMG Relaxation Dispersion NMR Experiments Measuring Glycine 1H Alpha and 13C Alpha Chemical Shifts in the "Invisible" Excited States of Proteins. *J Biomol NMR* **2009**, *45*, 45–55.

(30) Lerch, M. T.; López, C. J.; Yang, Z.; Kreitman, M. J.; Horwitz, J.; Hubbell, W. L. Structure-Relaxation Mechanism for the Response of T4 Lysozyme Cavity Mutants to Hydrostatic Pressure. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (19), E2437–E2446.

(31) Quillin, M. L.; Breyer, W. A.; Griswold, I. J.; Matthews, B. W. Size versus Polarizability in Protein-Ligand Interactions: Binding of Noble Gases within Engineered Cavities in Phage T4 Lysozyme. *J Mol Biol* **2000**, *302*, 955–977.

(32) Deng, Y.; Roux, B. Calculation of Standard Binding Free Energies: Aromatic Molecules in the T4 Lysozyme L99A Mutant. *J. Chem. Theory Comput.* **2006**, *2* (5), 1255–1273.

(33) Mulder, F. A. A.; Hon, B.; Muhandiram, D. R.; Dahlquist, F. W.; Kay, L. E. Flexibility and Ligand Exchange in a Buried Cavity Mutant of T4 Lysozyme Studied by Multinuclear NMR. *Biochemistry* **2000**, *39* (41), 12614–12622.

(34) Skrynnikov, N. R.; Dahlquist, F. W.; Kay, L. E. Reconstructing NMR Spectra Of "invisible" excited Protein States Using HSQC and HMQC Experiments. *J. Am. Chem. Soc.* **2002**, *124* (41), 12352–12360.

(35) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **2015**, *11* (8), 3584–3595.

(36) Schiffer, J. M.; Feher, V. A.; Malmstrom, R. D.; Sida, R.; Amaro, R. E. Capturing Invisible Motions in the Transition from Ground to Rare Excited States of T4 Lysozyme L99A. *Biophys. J.*

(37) Kitahara, R.; Yoshimura, Y.; Xue, M.; Kameda, T.; Mulder, F. A. A. Detecting O2 Binding Sites in Protein Cavities. *Sci. Rep.* **2016**, *6* (October 2015), 20534.

(38) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120* (24), 11919–11929.

(39) Fuglestad, B.; Gasper, P. M.; Tonelli, M.; McCammon, J. A.; Markwick, P. R. L.; Komives, E. A. The Dynamic Structure of Thrombin in Solution. *Biophys. J.* **2012**, *103* (1), 79–88.

(40) Han, B.; Liu, Y.; Ginzinger, S. W.; Wishart, D. S. SHIFTX2: Significantly Improved Protein Chemical Shift Prediction. *J. Biomol. NMR* **2011**, *50* (1), 43–57.

(41) Skrynnikov, N. R.; Mulder, F. A. A.; Hon, B.; Dahlquist, F. W.; Kay, L. E. Probing Slow Time Scale Dynamics at Methyl-Containing Side Chains in Proteins by Relaxation Dispersion NMR Measurements: Application to Methionine Residues in a Cavity Mutant of T4 Lysozyme. *J. Am. Chem. Soc.* **2001**, *123* (19), 4556–4566.

(42) Pierce, L. C. T.; Salomon-Ferrer, R.; Augusto F. De Oliveira, C.; McCammon, J. A.; Walker, R. C. Routine Access to Millisecond Time Scale Events with Accelerated Molecular Dynamics. *J. Chem. Theory Comput.* **2012**, *8* (9), 2997–3002.

(43) Eriksson, A. E.; Baase, W. A.; Zhang, X. J.; Heinz, D. W.; Blaber, M.; Baldwin, E. P.; Matthews, B. W. Response of a Protein Structure to Cavity-Creating Mutations and Its Relation to the Hydrophobic Effect. *Science (80-.).* **1992**, *255*, 178–183.

(44) Nicholson, H.; Anderson, D. E.; Dao Pin, S.; Matthews, B. W. Analysis of the Interaction between Charged Side Chains and the α-Helix Dipole Using Designed Thermostable Mutants of Phage T4 Lysozyme. *Biochemistry* **1991**, *30* (41), 9816–9828.

(45) Vallurupalli, P.; Chakrabarti, N.; Pomes, R.; Kay, L. Atomistic Picture of Conformational Exchange in a T4 Lysozyme Cavity Mutant: An Experiment-Guided Molecular Dynamics Study. *Chem. Sci.* **2016**, *3*, 3602–3613.

(46) Durrant, J. D.; Votapka, L.; S??rensen, J.; Amaro, R. E. POVME 2.0: An Enhanced Tool for Determining Pocket Shape and Volume Characteristics. *J. Chem. Theory Comput.* **2014**, *10* (11), 5047–5056.

(47) Durrant, J. D.; de Oliveira Cé, A. F.; McCammon, J. A. POVME: An Algorithm for Measuring Binding-Pocket Volumes. *J Mol Graph Model* **2011**, *29*, 773–776.

(48) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (18), 10037–10041.

(49) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: An Automated Pipeline for the Setup of Poisson-Boltzmann Electrostatics Calculations. *Nucleic Acids Res.* **2004**, *32* (WEB SERVER ISS.), 665–667.

(50) Xu, J.; Baase, W. A.; Baldwin, E.; Matthews, B. W. The Response of T4 Lysozyme to Large-to-Small Substitutions within the Core and Its Relation to the Hydrophobic Effect. *Protein Sci.* **1998**, *7* (1), 158–177.

(51) McIntosh, L. P.; Griffey, R. H.; Muchmore, D. C.; Nielson, C. P.; Redfield, A. G.; Dahlquist, F. W. Proton NMR Measurements of Bacteriophage T4 Lysozyme Aided by 15N Isotopic Labeling: Structural and Dynamic Studies of Larger Proteins. *Proc. Natl. Acad. Sci.* **1987**, *84* (5), 1244–1248.

(52) Xue, M.; Kitahara, R.; Yoshimura, Y.; Mulder, F. A. A. Aberrant Increase of NMR Signal in Hydrogen Exchange Experiments. Observation and Explanation. *Biochem. Biophys. Res. Commun.* **2016**, 8–11.

(53) Baase, W. A.; Liu, L.; Tronrud, D. E.; Matthews, B. W. Lessons from the Lysozyme of Phage T4. *Protein Sci.* **2010**, *19* (4), 631–641.

(54) Morton, A.; Matthews, B. W. Specificity of Ligand Binding in a Buried Nonpolar Cavity of T4 Lysozyme: Linkage of Dynamics and Structural Plasticity. *Biochemistry* **1995**, *34* (27), 8576–8588.

(55) Mulder MIttermaier, A., Hon, B., Dahlquist, F. W., Kay, L. E., F. A. A. Studying Excited States of Proteins by NMR Spectroscopy. *Nat. Struct. Biol.* **2001**, *8*.

(56) Sakuraba, S.; Kono, H. Spotting the Difference in Molecular Dynamics Simulations of Biomolecules. *J. Chem. Phys.* **2016**, *145* (7), 074116.

(57) Xu, M.; Caflisch, A.; Hamm, P. Protein Structural Memory Influences Ligand Binding Mode(s) and Unbinding Rates. *J. Chem. Theory Comput.* **2016**, *12* (3).

(58) Miyashita, O.; Onuchic, J. N.; Wolynes, P. G. Nonlinear Elasticity, Protein Quakes, and the Energy Landscapes of Functional Transitions in Proteins. *Proc.*

Natl. Acad. Sci. **2003**, *100* (22), 12570–12575.

(59) Xue, Y.; Ward, J. M.; Yuwen, T.; Podkorytov, I. S.; Skrynnikov, N. R. Microsecond Time-Scale Conformational Exchange in Proteins: Using Long Molecular Dynamics Trajectory to Simulate NMR Relaxation Dispersion Data. *J. Am. Chem. Soc.* **2012**, *134* (5), 2555–2562.

(60) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Systematic Validation of Protein Force Fields against Experimental Data. *PLoS One* **2012**, *7* (2), 1–6.

(61) Martín-García, F.; Papaleo, E.; Gomez-Puertas, P.; Boomsma, W.; Lindorff-Larsen, K. Comparing Molecular Dynamics Force Fields in the Essential Subspace. *PLoS One* **2015**.

(62) Wray, J. W.; Baase, W. A.; Lindstrom, J. D.; Weaver, L. H.; Poteete, A. R.; Matthews, B. W. Structural Analysis of a Non-Contiguous Second-Site Revertant in T4 Lysozyme Shows That Increasing the Rigidity of a Protein Can Enhance Its Stability. *J Mol Biol* **1999**, *292*, 1111–1120.

(63) Eriksson, A. E.; Baase, W. A.; Matthews, B. W. Similar Hydrophobic Replacements of Leu99 and Phe153 within the Core of T4 Lysozyme Have Different Structural and Thermodynamic Consequences. *J. Mol. Bio.* **1993**, *229*, 747–769.

(64) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089.

(65) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26* (16), 1668–1688.

(66) Salomon-Ferrer, Romelia Goetz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8* (5), 1542–1555.

(67) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **1994**, *98* (7), 1978–1988.

(68) Desvaux, H.; Dubois, L.; Huber, G.; Quillin, M. L.; Berthault, P.; Matthews, B. W.

Dynamics of Xenon Binding inside the Hydrophobic Cavity of Pseudo-Wild-Type Bacteriophage T4 Lysozyme Explored through Xenon-Based NMR Spectroscopy. *J. Am. Chem. Soc.* **2005**, *127* (33), 11676–11683.

Figures and Figure legends

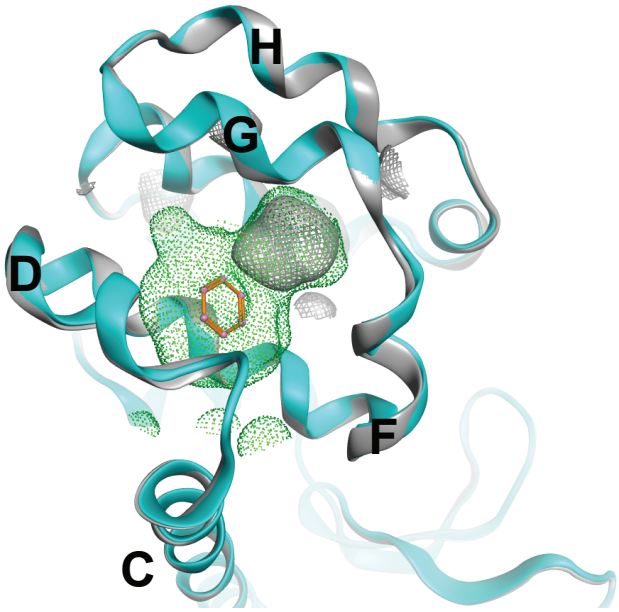


Figure 3.1: **Crystallographic comparison of T4 lysozyme WT* and L99A benzene bound mutant illustrate few differences apart from C-terminal domain cavity expansion.** The backbone atom superposition of wild-type T4 lysozyme (PDB: 1L63, gray) and the backbone structure of the L99A cavity forming mutant (PDB: 181L, cyan) C-terminal domain have RMSD less than crystallographic error. The cavity formed by the L99A mutant expands the buried internal wild type ~ 30 Å^3 cavity (grey mesh) to ~ 150 Å^3 cavity (green mesh) with no obvious solvent to internal cavity pathways. The L99A mutant cavity can accommodate substituted benzenes[16,21] (benzene depicted as orange ball and stick) and multiple atoms of noble gases.[31,68] The WT* binds 1 atom of noble gas.

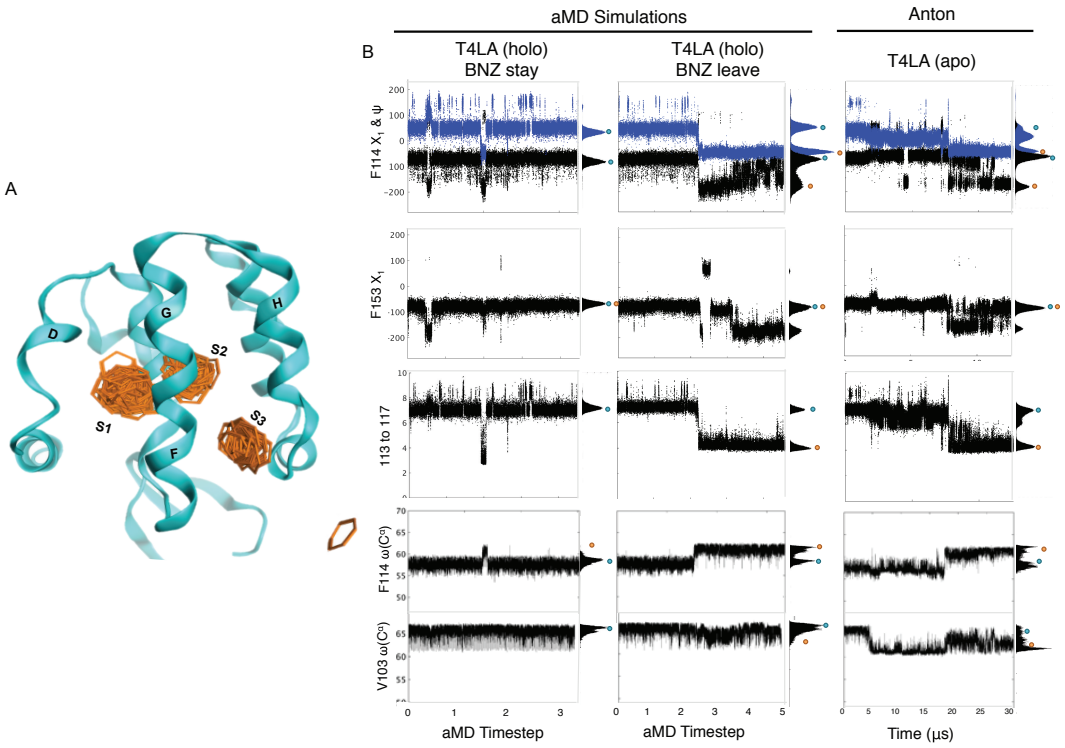


Figure 3.2: Productive and non-productive excursions in aMD simulations of the L99A cavity mutant of T4 Lysozyme. (a) Benzene samples three different states in the aMD trajectory where it escapes from the buried cavity. The S1 state is the location of benzene in the ground state crystal structure. Following the S1 state, benzene travels to S2 and then to S3 befor being released between the F, H and I helices. (b) Rotamer changes, hydrogen bond distances, and chemical shifts are shown over aMD simulation trajectories for apo L99A simulation where no excursions are seen from the ground state, holo L99A simulation where non-productive excursions result in benzene staying in the buried cavity, and holo L99A simulation where benzene egresses at ~ 22,000 aMD timesteps (or ~150 ns). Additionally, the long-timescale cMD Anton trajectory is showing the transition from ground to excited state which takes place at ~ 18 μs as we previously described. For all plots, histograms on the side of each figure are showing the relative populations of each state over the given trajectory. The circles above each of the populations identify the experimentally-determined excited state (orange) and the experimentally-determined ground state (cyan).

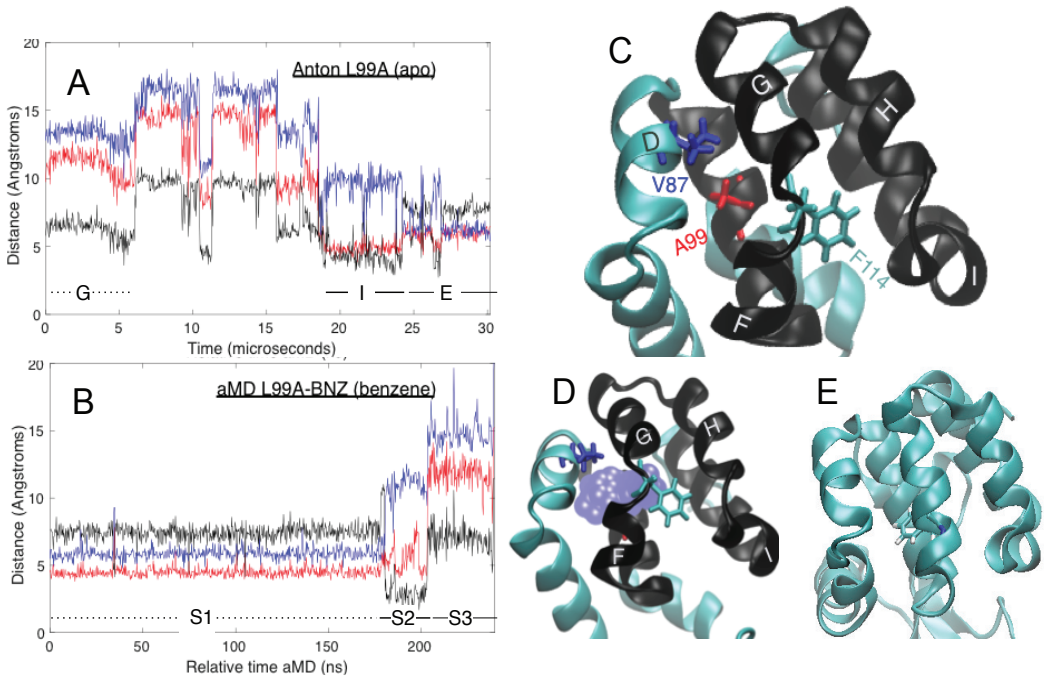


Figure 3.3: **The concerted motions of benzene leaving and excited state transitioning.** (b,c) Internal distances between the center of mass between the aromatic ring of F114 (b) or the aromatic ring of benzene (c) and the beta carbon of A99 (red), the beta carbon of V87 (blue), and the center of mass of the four helix bundle (black). (d,e) The locations of A99 (red), V87 (blue), F114 (cyan) and the four helix bundle (black) are shown for reference. The location of the buried cavity is also shown in purple (e) with the C-terminal domain of L99A from ground state apo crystal structure (PDB: 4W51). (f) The conformation of L99A when benzene samples the S2 state for comparison with the highest populated intermediate structure sampled with high temperature MSMs.[45]

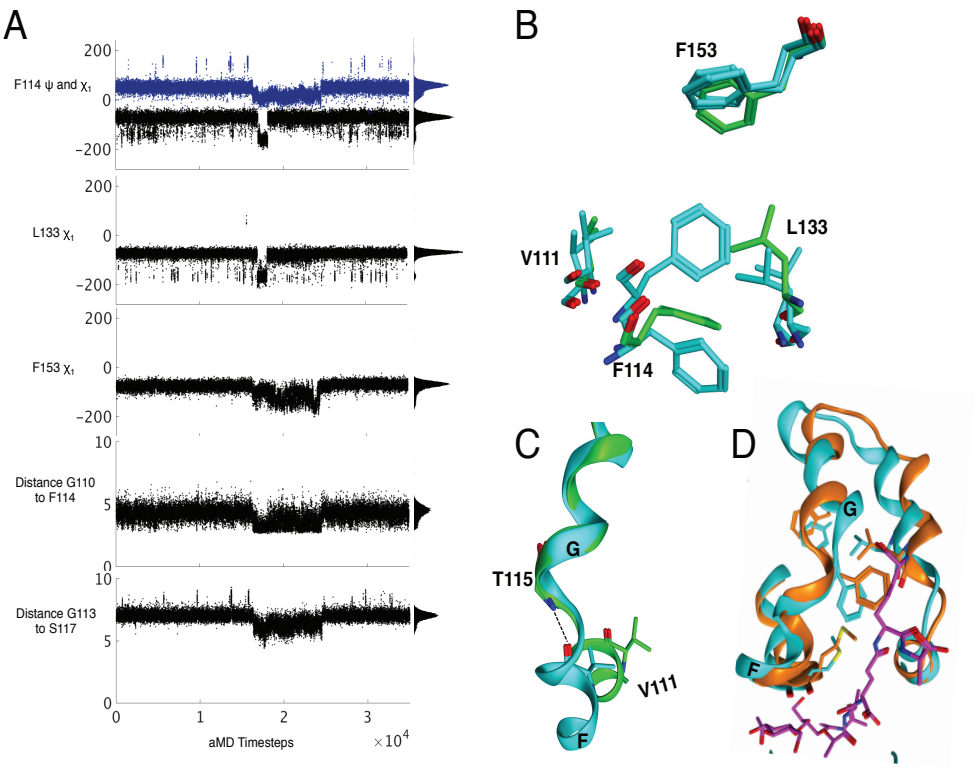


Figure 3.4: **Wild type T4 lysozyme transition from ground state to a high energy-like state.** (a) Conformational changes that occur in the wild-type aMD simulations mirror what occurs in the L99A (holo) simulations, with concerted motions and backbone breakages occuring at around 17000 aMD timesteps or ~75 ns. For all plots, histograms on the side of each figure are showing the relative populations of each state over the given trajectory. (b) Showing the relative motions possible from the crystal structure (green) to the high energy-like state (cyan) for F114, L133, F153, and V111. (c) Helices F and G transiently form a more continuous single helix with the G110 carbonyl - F114 amine and V111 carbonyl – T115 amine distances averaging 2.8 – 3.2 Å. (d) Helices F, G, H and I depicted in the ground state (cyan) and high energy-like state (orange) with the peptidoglycan substrate superimposed to demonstrate the importance of this cleft between the F/G and H/I helices for processivity of T4 lysozyme.

Supplementary Figures

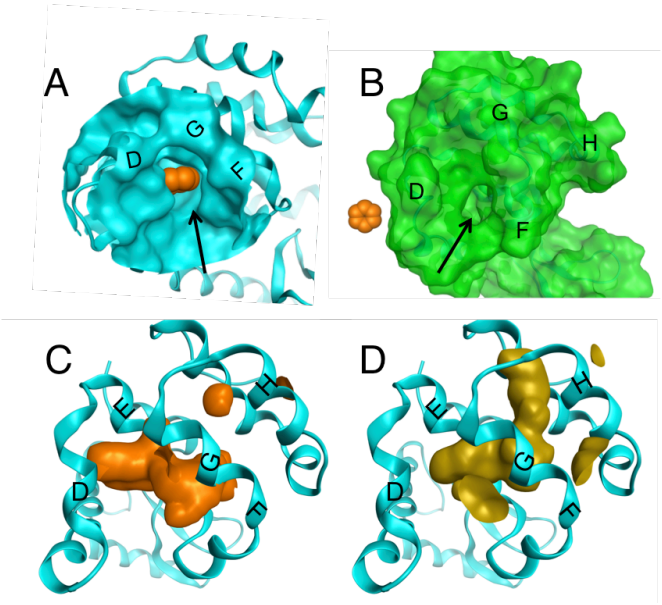


Figure 3.S1: Transient pocket openings where benzene is known to bind in the L99A cavity mutant of T4 lysozyme. (A/B) the pocket between the D, F, and G helices opens to solution throughout the simulation revealing a large enough defect to allow benzene release. (C/D) Other mobile defects leave to solvent exposure of the buried cavity through the C and D helices (C) and through the H and J helices (D).

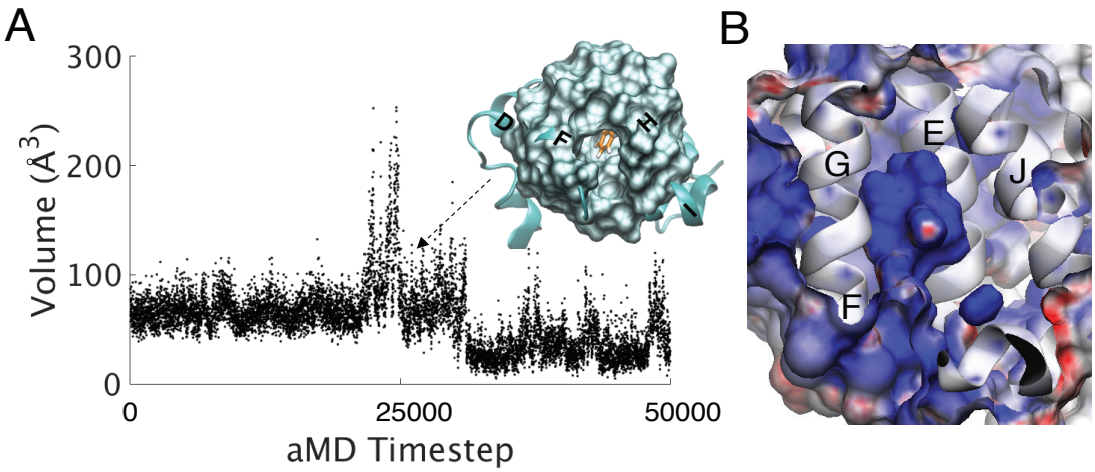


Figure 3.S2: A buried cavity between the F and H helices becomes solvent exposed upon transitioning to the L99A excited state. (A) the buried volume changes throughout the trajectory but it also unstable within each state varying on the order of 50 Å within a single state. Upon transitioning to the excited state at ~ 25,000 aMD timesteps, an enormous cavity opens exposing benzene in S2 to solvent. (B) This excited state cavity defect is positively charged as determined through APBS calculations.

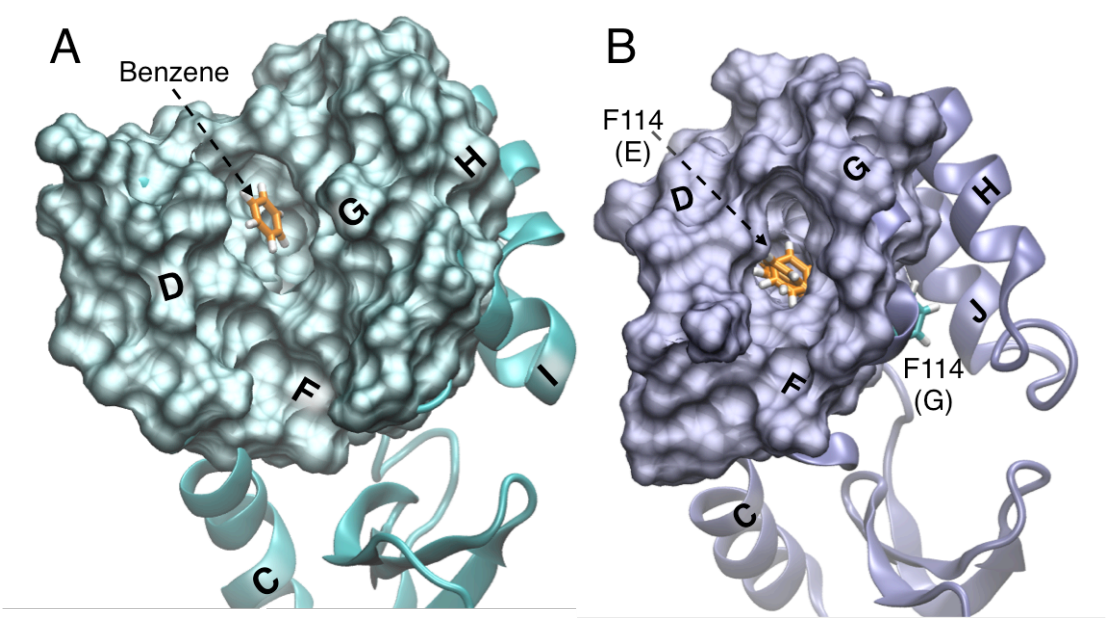


Figure 3.S3: A large mobile defect occurs prior to excited state transitioning in both the benzene bound aMD simulations (A) and the excited state Anton trajectory (B). This mobile defect is so large that it completely exposes benzene in the aMD trajectory and F114 in the Anton trajectory.

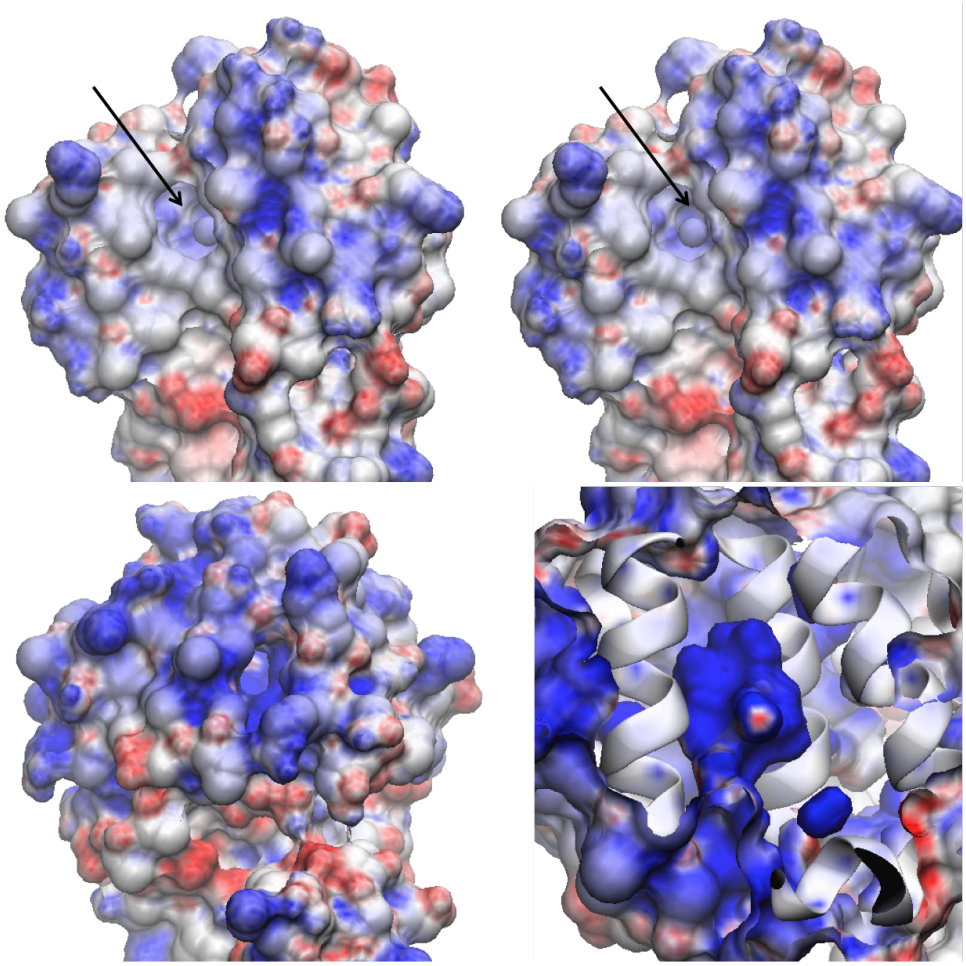
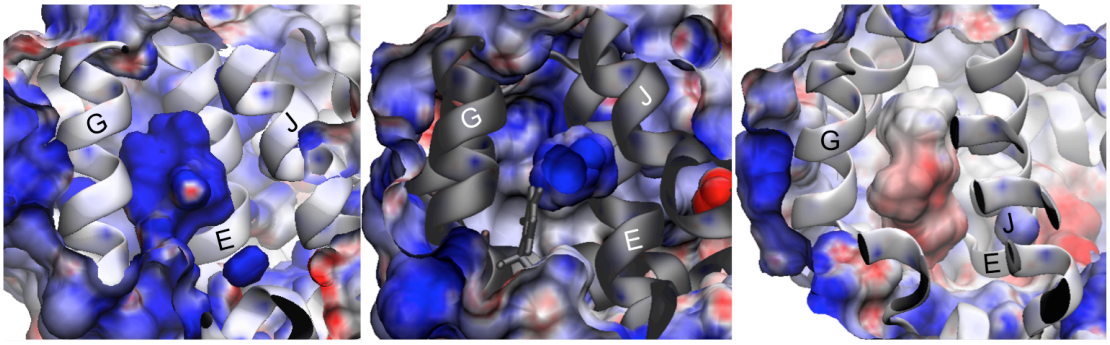


Figure 3.S4: Electrostatics of mobile defects in ground and excited states of the L99A mutant from simulation. (top) Mobile defect through the D, F and G helices in the ground state forms a cavity opening with hydrophobic surfaces. (bottom) Contrastingly, the mobile defect through the F and H helices in the excited state is highly polar.



L99A BNZ leaving state aMD

WT T4 lysozyme aMD processive state

L99G T4 lysozyme Crystal

Figure 3.S5: Polarity of buried cavities present in excited-like states of T4 lysozyme. (A) The L99A Benzene leaving state contains a mobile defect with a positively charge cavity surface. (B) The processive state seen in our aMD trajectories of the WT T4 lysozyme reveal a similarly positively charged buried cavity. (C) The cavity opening between the F and H helices seen in the L99G T4 lysozyme mutant is negatively charged and accomodated 5 water molecules in the crystal structure.

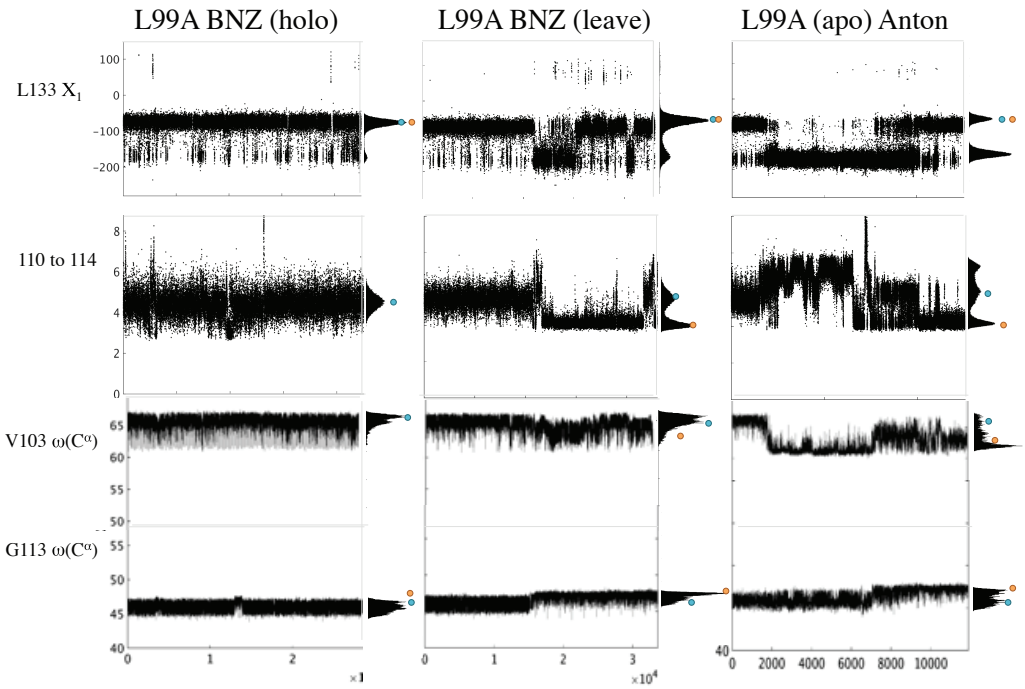


Figure 3.S6: Dihedral angle, backbone distances, and back-calculated chemical shift differences measured over three different condition trajectories of the L99A cavity mutant. The column to the left is showing these values for L99A bound to benzene where benzene is maintained in the buried cavity. The middle column is showing the values for the trajectory where benzene leaves. The last column is showing the results from the Anton trajecotry of apo L99A in the transition from ground to excited state.

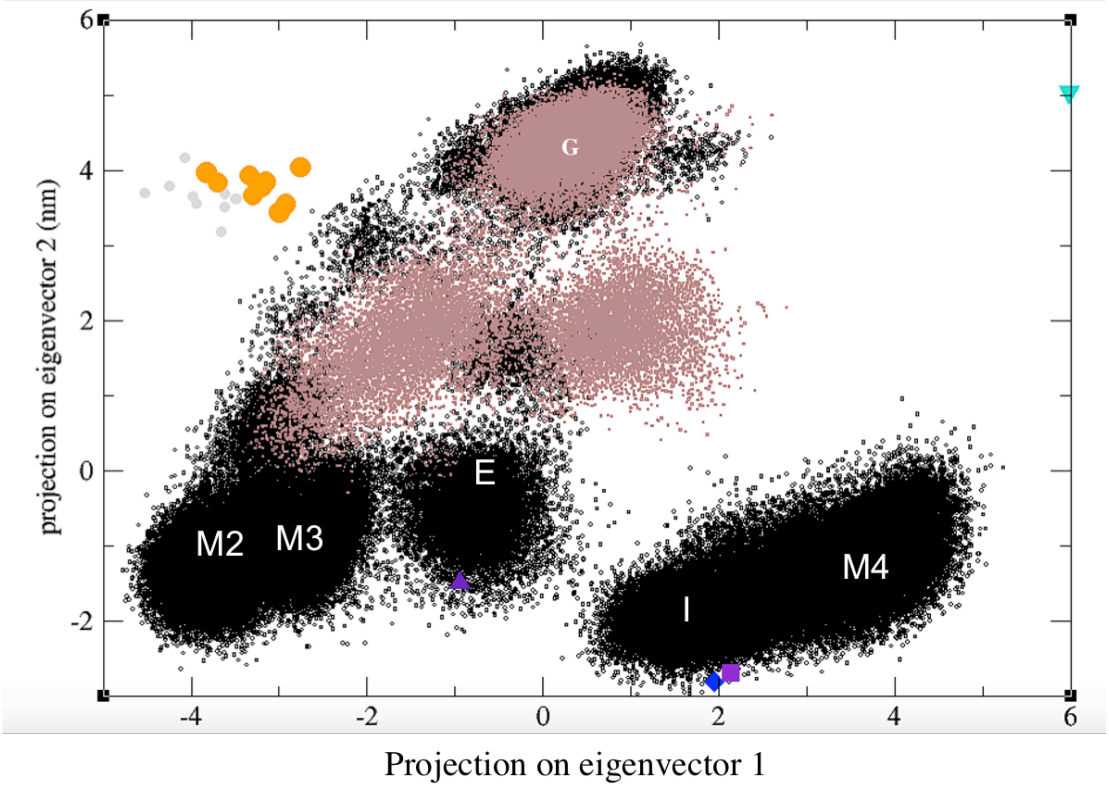


Figure 3.S7: Principal component analysis results for both the Anton trajectory (black) and the aMD Benzene leaving trajectory (mauve). The first principal componenet is on the x axis and the second is on the y axis. The location in phase space of the excited state triple mutant model and Rosetta model are shown in orange and grey circles at the top left corner respectively. The crystal structures of WT and L99A ground state are shwon in purple triangle (PDB: 1L91), blue diamond (PDB: 1L90), and puple squares (PDB: 3DMV) respectively. The most recent crystal structure of the L99A mutant is shown in cyan triangle (PDB: 4W51).

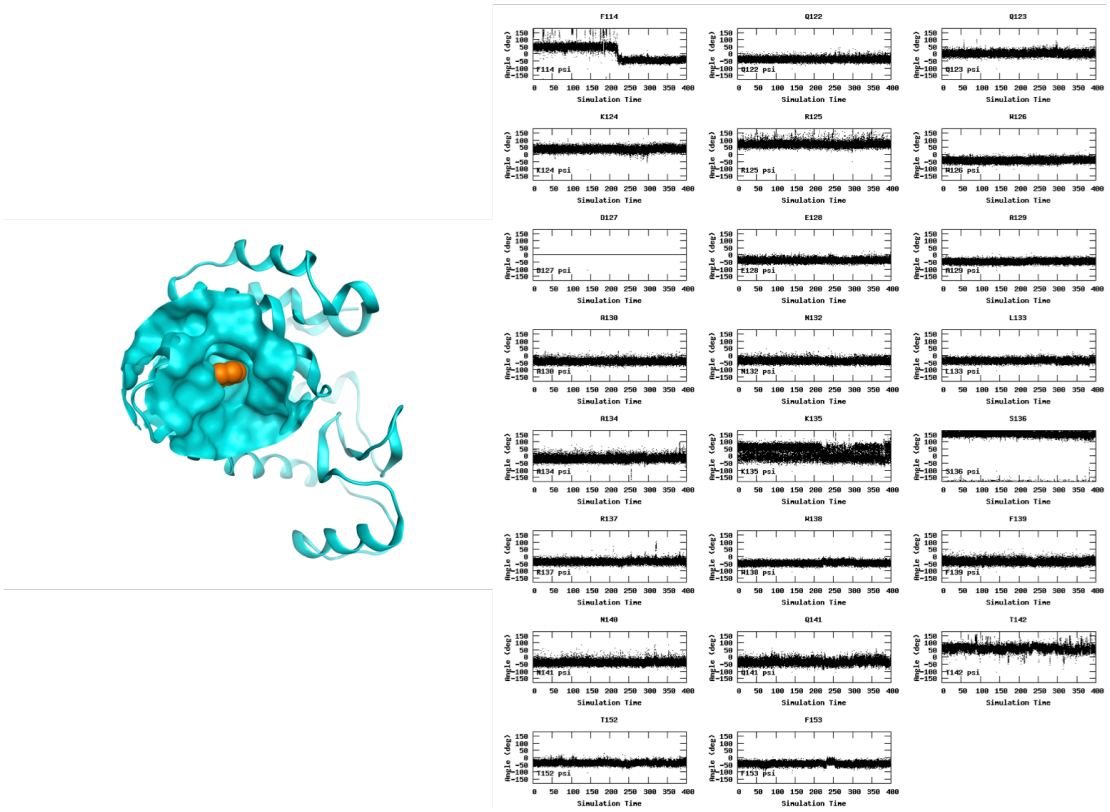
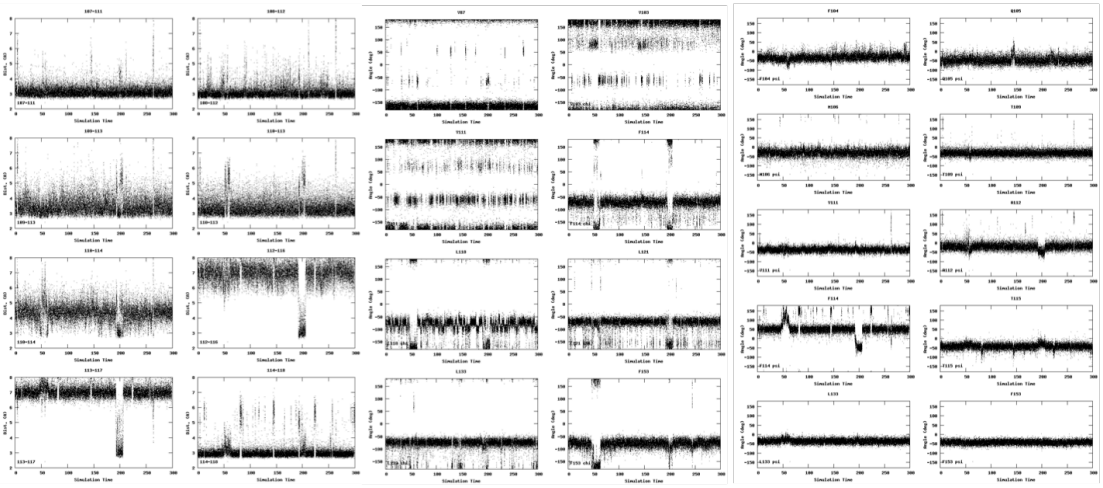


Figure 3.S8: Dihedral angle changes for multiple side chains of the L99A mutant during the holo (Benzene leaving) simulation.



dist(i, i+4) carbonyl to amide residue chi1 angle residue psi angle

Figure 3.S9: Internal distances and dihedral changes in non-productive excursions of the L99A holo simulations.

ACKNOWLEDGMENTS

Chapter 3, in full, is a reprint of the submitted material Benzene Dissociation through a Protein Excited State 2016. Schiffer, Jamie M., Feher, V. A., Mermelstein, D., Mih, N., Pierce, L. C. T., McCammon, A. J., Amaro, R. E. (submitted). The dissertation author was the co-primary investigator and author of this material.

CHAPTER 4

How the Ankyrin and SOCS box protein 9, ASB9, binds to Creatine Kinase

BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

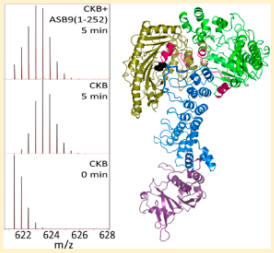
How the Ankyrin and SOCS Box Protein, ASB9, Binds to Creatine Kinase

Deepa Balasubramaniam, Jamie Schiffer, Jonathan Parnell, Stephan P. Mir, Rommie E. Amaro, and Elizabeth A. Komives*

Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0378, United States

Supporting Information

ABSTRACT: The ankyrin repeat and SOCS box (ASB) family is composed of 18 proteins and belongs to the suppressor of cytokine signaling (SOCS) box protein superfamily. The ASB proteins function as the substrate-recognition subunits of ECS-type (ElonginBC-Cullin-SOCS-box) Cullin RING E3 ubiquitin ligase (CRL) complexes that specifically transfer ubiquitin to cellular proteins targeting them for degradation by the proteasome. ASB9 binds to creatine kinase (CK) and targets it for degradation; however, the way in which ASB9 interacts with CK is not yet known. We present a complete characterization of the binding of ASB9 to CK. One ASB9 molecule binds to a dimer of CK. The binding affinity of ASB9(1−252) was extremely tight, and no dissociation could be observed. Deletion of the 34 N-terminal amino acids forming ASB9(35−252) resulted in weakening of the binding, so that a binding affinity of 2.6 nM could be measured. Amide hydrogen−deuterium exchange (HDXMS) experiments showed that both ASB9(1−252) and ASB9(35−252) protected the same region of CK, residues 182−203, which forms one side of the active site. The HDXMS experiments indicated that the N-terminal disordered region and first ankyrin repeat of ASB9 are protected from exchange in the complex. Molecular docking yielded a structural model consistent with all of the data that suggested the N-terminal residues of ASB9(1−252) may lie in one CK active site. This model was corroborated by enzymatic activity assays and mutational analysis.



Proteasome-dependent protein degradation occurs when ubiquitin is transferred to the ε-amine of lysine residues within the doomed protein.[1,2] The transfer of ubiquitin requires a three-enzyme system comprised of an E1 ubiquitin-activating enzyme, an E2 ubiquitin-conjugating enzyme, and an E3 ubiquitin ligase. It is the E3 ligase that binds the doomed protein, brings it together with the E2 enzyme, and catalyzes the transfer of ubiquitin. More than 600 human E3 ligases have been identified. The most well-characterized is the SCF family, for which structures have been available for more than 10 years.[3] The Cullin RING E3 ligases (CRL) make up the largest family of E3 ligases in eukaryotes and have separate substrate-binding and catalytic subunits.[4,5] The substrate-recognition protein binds to the N-terminal domain of the Cullin subunit (Cul1−5 or Cul7), and a RING protein (Rbx1 or Rbx2), which in turn recruits the E2 ubiquitin conjugate, binds to the C-terminal domain. Neddylation of the Cullin C-terminal domain is thought to alter the conformation, so that the substrate and ubiquitin are brought into the proximity of each other.[6]

The ankyrin repeat and SOCS box (ASB) family is composed of 18 proteins and belongs to the suppressor of cytokine signaling (SOCS) box protein superfamily. The ASB proteins interact with Cul5-Rbx2 to form a functional E3 ubiquitin ligase.[7] ASB family members function as the substrate-recognition subunits of ECS-type (ElonginBC-Cullin-SOCS-box) Cullin RING E3 ubiquitin ligase (CRL) complexes that specifically transfer ubiquitin to cellular proteins targeting them for degradation by the proteasome. The quaternary multi-subunit complex formed by ASB9, Elongin B, Elongin C (EloBC), and Cullin 5 was recently characterized,[8] and the structure of ASB9 bound to Elongins B and C has also been reported.[9] However, the interactions of the ASB subunits with their substrates and the assembly of the ASB subunits within ECS-type ubiquitin ligases remain poorly understood.

The Elongin BC, Cullin 5, and Rbx2 proteins are the same for each ASB-containing CRL, so it is thought that the ASB protein is responsible for binding the target protein and that each ASB binds to a different target protein. The ASBs are composed of an N-terminal ankyrin repeat domain (ARD) with different numbers of ankyrin repeats, and a C-terminal SOCS box domain.[7] ASB proteins bind and recognize their specific substrate through the ARD. The CRL containing ASB9 binds to creatine kinase (CK) and targets it for degradation;[10,11] however, the way in which ASB9 interacts with CK is not yet known. The structure of ASB9 (residues 37−294) bound to Elongin BC[9] and the structure of ASB9 (residues 19−252)[12,13] both show similar architectures of the well-folded ARD of ASB9. Interestingly, a sequence C-terminal to the ARD appears

Received: November 16, 2014
Revised: February 2, 2015
Published: February 5, 2015

to form part of an additional ankyrin repeat. A structural model for the ASB9−CK interaction was proposed on the basis of the way in which other ankyrin repeat proteins bind their targets; however, the experimental evidence for this model would also have been consistent with a variety of interaction modes.[13]

ASB9 is predominantly expressed in the kidney and testes, and it has been shown to bind to and ubiquitinate brain-type cytosolic creatine kinase (CKB)[10] and ubiquitous mitochondrial creatine kinase (uMtCK).[11] Recent evidence suggests that ASB9 could be a biomarker for human breast cancer,[14] and it has also been linked to colorectal cancer.[15] The reduction of the level of uMtCK expression by siRNA led to an increased rate of cell death, and reduced rates of proliferation, migration, and invasion in HCC cell lines. Transient overexpression of ASB9 reduced uMtCK protein levels in HCC cells, suggesting that the increased uMtCK levels that correlate with poor prognosis may be due to a reduced level of ASB9 expression.[16]

We present the complete characterization of the binding of ASB9 to CK, including measurement of the binding affinity and stoichiometry. In addition, we performed amide hydrogen−deuterium exchange (HDXMS) to determine the binding site,[17] and we present a structural model consistent with all of the data.

■ EXPERIMENTAL PROCEDURES

Cloning, Expression, and Purification. ASB9(1−252) and ASB9(35−252) were cloned into the pHis8 vector with an N-terminal His8 tag. ASB9(1−252)D32A was generated by site-directed mutagenesis from the ASB9(1−252) construct. All ASB9 proteins were expressed in BL21(DE3) cells, grown to an OD_{600} of 0.7, and induced with 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) overnight at 18 °C in M9ZN medium (M9 medium supplemented with 10 g/L NZ amine). The cell pellet was resuspended in phosphate buffer [50 mM sodium phosphate (pH 8.0) and 300 mM NaCl] containing 10 mM imidazole and 0.5 mM phenylmethanesulfonyl fluoride (PMSF), sonicated on ice, and centrifuged for 30 min at 12000 rpm. The supernatant was loaded over a 10 mL Ni-NTA (His-Pur, Pierce) column at 4 °C, washed with phosphate buffer containing 50 mM imidazole, and eluted with phosphate buffer containing 250 mM imidazole. Fractions containing ASB9 were pooled and dialyzed into 20 mM Tris (pH 8.5), 200 mM NaCl, 0.5 mM EDTA, and 1 mM DTT. Aliquots (2 mL) of the dialyzed protein were flash-frozen in liquid nitrogen, stored at −80 °C, and further purified on a Superdex-200 (S-200) gel filtration column before each experiment (GE Healthcare).

CKB was cloned in pET11a and expressed in BL21(DE3) cells, grown to an OD_{600} of 0.7, and induced with 0.4 mM IPTG overnight at 18 °C in M9ZN medium (M9 medium supplemented with 10 g/L NZ amine). The protein was purified over a blue Sepharose column as described previously.[18] The cells were resuspended in buffer A [10 mM MES (pH 6.0), 20 mM KCl, and 1 mM DTT] containing 0.5 mM PMSF, sonicated on ice, and centrifuged for 30 min at 12000 rpm. The supernatant was loaded onto a 50 mL blue Sepharose column at 4 °C and washed with 5 column volumes of buffer A, and CKB was eluted with 2 column volumes of buffer B [10 mM TES (pH 8.0), 20 mM KCl, and 1 mM DTT]. Fractions containing the enzyme were pooled and dialyzed into 50 mM HEPES, 0.1 mM EDTA, and 1 mM DTT (pH 7.0), and aliquots were stored at −80 °C. CKB fractions were also purified on a S-200 gel filtration column (GE Healthcare), and the eluted protein was concentrated in a 10 kDa molecular mass cutoff Amicon Ultra centrifugal filter (Millipore).

The complex (CKB+ASB9) was formed by mixing a 10 μM solution of ASB9 with a 15 μM (monomer concentration) solution of CKB to achieve a 1.14 ASB9:CKB monomer excess. The mixture was incubated for 2 h at 4 °C and purified by S200 size exclusion chromatography.

Activity Assay. Creatine kinase activity was determined by using a coupled enzyme assay as described previously.[19] The creatine kinase reaction is coupled to pyruvate kinase and lactate dehydrogenase reactions, and the conversion of NADH to NAD is observed at 340 nm to monitor the amount of creatine phosphate formed. Assays were performed at 30 °C in pH 9.0 triethanolamine buffer.[20] The assay mixture contained 0.2 mM NADH, 0.4 mM phosphoenolpyruvate, 5 mM ATP, 6 mM magnesium acetate, 13 mM potassium acetate, 28−56 units/mL pyruvate kinase, and 54−108 units/mL lactate dehydrogenase. The creatine kinase concentration was 3 nM (monomer), and the creatine concentration was varied from 10 to 100 mM. The assay mixture was incubated at 30 °C for 3 min, and the reaction was initiated by the addition of enzyme. The velocity of the reaction is calculated using an extinction coefficient of 6290 M^{-1} cm^{-1} for NADH. For assays testing the role of ASB9 binding in CKB activity, the complex (CKB+ASB9) was added to achieve a CKB concentration of 3 nM. V_{max} and K_M values were determined by fitting the data to the Michaelis−Menten kinetic model in Kaleidagraph (Synergy, Inc.).

Stoichiometry of the ASB9−CK Complex. Multiangle light scattering (MALS) analysis was performed to determine the stoichiometry of the ASB9−CK complex. In sequential runs, ASB9 alone, CK alone, and the ASB9−CK complex were analyzed. The proteins were buffer exchanged into 20 mM Tris (pH 8.0) and 100 mM NaCl, and 100 μL of a protein sample was analyzed via analytical size exclusion chromatography using a GE Healthcare Superdex 200 10/300 GL column with a flow rate of 0.4 mL/min. The column flowed directly into a miniDAWN TREOS MALS instrument (Wyatt Technology) and an Optilab UT-rEX instrument (Wyatt Technology). The Astra 6 instrument (Wyatt Technologies) was used to determine the weight-average molar mass of the eluting peaks using the intensity of the Rayleigh scattering as a function of scattering angle (LS) along with the buffer-subtracted refractive index (dRI).[21]

Measurement of Binding Thermodynamics. All isothermal titration calorimetry (ITC) experiments were performed on a VP-ITC calorimeter (MicroCal, Inc.). ASB9 constructs and CKB were purified by size exclusion chromatography on an S200 column in 50 mM bistrispropane (pH 8.5) and 200 mM sodium chloride, immediately prior to use. Protein concentrations were determined by measuring the absorbance at 280 nM and calculated using extinction coefficients as determined by amino acid analysis (for ASB9 constructs, ε = 24400 M^{-1} cm^{-1}; for CKB, ε = 33700 M^{-1} cm^{-1}). Experiments were performed at 30 °C; around 25 injections of 10 μL of 50 μM CKB dimer were made into 5 μM ASB9 in the cell. Isotherms were analyzed using Origin software (OriginLab) as described previously[22] and fit to a single-binding site model.

Measurement of Binding Kinetics. For surface plasmon resonance (SPR) experiments, CKB was cloned in pMCSG51 to add a biotinylation tag (Midwest Center for Structural Genomics). This plasmid has an N-terminal His tag to aid

purification, followed by an AVI tag that can be biotinylated on a specific lysine using *Escherichia coli* biotin ligase (BirA), which is co-expressed with the gene of interest. Biotinylated CKB was purified using a combination of Ni-NTA (using the phosphate buffers as the ASB9 constructs described above) and size exclusion chromatography over a S200 column. SPR experiments were performed on a Biacore 3000 instrument (GE Healthcare). Biotinylated CKB was immobilized on streptavidin (SA) chips in a high-salt buffer [50 mM HEPES (pH 8.0), 500 mM NaCl, 0.5 mM sodium azide, and 0.005% Tween 20]. Biotinylated CKB (300 RU) was immobilized on FC3 and data from FC1 (no CK immobilized) was automatically subtracted from FC3. Data collection rate set to high. The running buffer used for the binding experiments consisted of 50 mM HEPES (pH 8.0), 200 mM NaCl, 0.5 mM sodium azide, and 0.005% Tween 20. It is to be noted that ASB9 constructs have to be freshly purified by size exclusion chromatography and maintained at 4 °C for binding to be observed. Injections were made using the kinject mode at a rate of 20 μL/min, with a 240 s association time and a 300 s dissociation time for ASB9(35−252). The ASB9(35−252) concentration was varied from 0 to 20 nM. The data were analyzed using BiaEvaluation version 4.1 with a simple 1:1 Langmuir binding model.

Hydrogen−Deuterium Exchange Mass Spectrometry (HDXMS). HDXMS was performed on CKB, ASB9(1−252), ASB9(35−252), and CKB in complex with ASB9(1−252) and ASB9(35−252). Complexes were made by adding an excess of ASB9 to CKB and vice versa and incubating the complexes for 30 min at 4 °C. HDXMS experiments were conducted at Lilly, Inc., in San Diego, CA, using a Waters Synapt G2S system with H/DX technology as described previously,[23,24] with the exception that the quench buffer contained 100 mM phosphate, 2 M guanidine hydrochloride, and 320 mM TCEP (pH 2.4). Each sample (5 μL) was mixed with 55 μL of D_2O buffer (containing 0.1× PBS) for several deuteration times (10 s to 10 min) at 15 °C. The exchange was quenched for 2 min at 1 °C with an equal volume of quench buffer. A portion of the quenched sample (50 μL) was injected onto an online pepsin column (Applied Biosystems, Poroszyme Immobilized Pepsin cartridge). The resulting peptic peptides were then separated on a C18 column (Waters, Acquity UPLC BEH C18, 1.7 μm, 1.0 mm × 50 mm) fit with a Vanguard trap column using a 3 to 85% acetonitrile (containing 0.1% formic acid) gradient over 12 min at a flow rate of 40 μL/min. The separated peptides were directed into a Waters SYNAPT G2s quadrupole time-of-flight (qTOF) mass spectrometer. The mass spectrometer was set to collect data in the MS^E, ESI^+ mode; in a mass acquisition range of m/z 255.00−1950.00; with a scan time of 0.4 s. Continuous lock mass correction was accomplished with infusion of the LeuEnk peptide every 30 s (mass accuracy of 1 ppm for the calibration standard). The peptides were identified using PLGS version 2.5 (Waters, Inc.). The relative deuterium uptake for each peptide was calculated by comparing the centroids of the mass envelopes of the deuterated samples with the undeuterated controls using DynamX version 2.0 (Waters Corp.).

Molecular Modeling. The extended structure of the ASB9 intrinsically disordered N-terminal domain, ASB9(1−35), was generated on the basis of the amino acid sequence in Maestro. Eleven parallel simulations were performed using the AMBER ff99SB force field[25] in a TIP3P water box.[26] The simulated system contained a single copy of the 35-residue ASB9 N-terminus in an orthorhombic water box with dimensions of 67 Å × 63.63 Å × 64.36 Å, with periodic boundary conditions. The N-terminus and C-terminus of ASB9 were capped with acetyl and methyl groups, respectively. The total system contained 61016 atoms, including a low (0.15 M) concentration of sodium chloride in the water solvent. Simulations were performed with NVIDIA GK110 (GeForce GTX Titan) GPUs using the CUDA version of PMEMD in AMBER12.[27] The ASB9(1−35) system was relaxed using 86000 steps of energy minimization followed by one restrained 250 ps heating step up to 303 K. Three consecutive restrained 250 ps equilibration steps were performed. The restraint force was applied to the backbone atoms in these equilibration steps starting at 3.0 kcal $Å^{-2}$ mol^{-1} for the first step, then 2.0 kcal $Å^{-2}$ mol^{-1} for the second step, and 1.0 kcal $Å^{-2}$ mol^{-1} for the last step before the performance run. Long-range electrostatics were calculated using the particle mesh Ewald (PME) method,[28] with a 10 Å cutoff and a 1.2 Å grid spacing. Ten simulations were performed at 303 K for 25 ns each, for a total of 250 ns of sampling time. The trajectory structures were clustered using the GROMOS algorithm[29] in GROMACS version 4.5.5,[30] with a backbone root-mean-square deviation (rmsd) cutoff of 3.5 Å for ASB9(1−35). The top eight (90% of the total ensemble) cluster centroids were analyzed, and the most elongated and compact cluster centroids were appended onto the crystal structure of the ARD of ASB9.

The molecular model of ASB9(1−252) was built from the 2.2 Å crystal structure of splice variant 2 of the ARD of ASB9 from hASB9-2 [Protein Data Bank (PDB) entry 3D9H]. The top eight cluster centroids of the ASB9(1−35) simulation (representing 90% of the total ensemble) were analyzed, and the most elongated and compact cluster centroids were appended onto the crystal structure of the ARD of ASB9 in Maestro. The two models of ASB9(1−252), one with an elongated N-terminus and one with a compact N-terminus, were docked into two models of brain-type CK (CKB): the crystal structure of CKB (PDB entry 3DRB, which has one open monomer and one closed monomer)[31] and a model of the open−open CKB crystal structure generated in VMD by overlaying two crystal structures over one another. In VMD, the rmsd trajectory tool extension was used to align two CKB crystal structures.

The ZDOCK server was used to dock both models of ASB9 into both models of CKB (http://zdock.umassmed.edu/).[32] The top 10 structures of each of the four dockings were compared on the basis of their ZDOCK scores (1103−1778 points), the consistency of the binding interface determined from the rmsd of the docked orientation from the top 10 docked complexes, and the steric overlap of the SOCS box domain with CKBB when the ASB9−Elongin BC crystal structure (PDB entry 3ZNG)[8] is overlaid with the docked complexes.

RESULTS

The Affinity of the ASB9−CK Complex Requires Two Regions of ASB9. The affinity of the ASB9−CK complex was measured by both surface SPR and ITC. For these experiments, the binding affinity of the ARD of ASB9 [ASB9(35−252)] was compared to the affinity of a construct containing both the ARD and the 34 N-terminal residues. Both SPR and ITC revealed that the binding affinity between ASB9(1−252) and CK is extremely high. For ASB9(1−252), no dissociation was observed, even when 3 M guanidine hydrochloride was included in the dissociation buffer in the SPR experiments (Figure 1A). ITC experiments corroborated this observation.

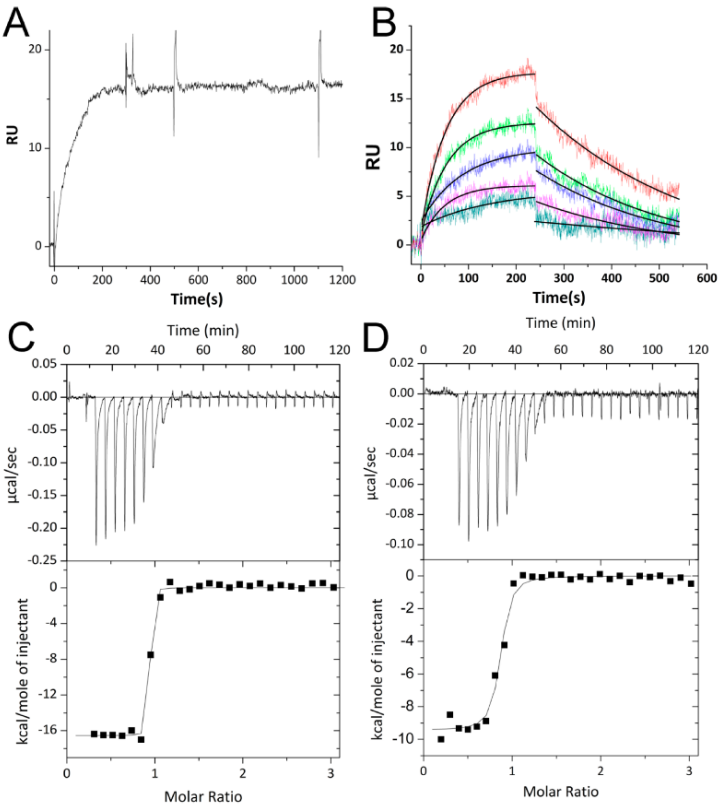


Figure 1. Measurement of binding kinetics and thermodynamics. (A) SPR experiment in which 30 nM ASB9(1−252) was passed over immobilized CKB at a rate of 20 μL/min. Regeneration was attempted by adding 3 M guanidine hydrochloride to the regeneration solution at 500 s, but no dissociation was observed. (B) SPR experiment in which 8, 9, 10, 12, and 14 nM ASB9(35−252) was passed over immobilized CKB at a rate of 20 μL/min. Regeneration was conducted with buffer containing 0.5 M guanidine hydrochloride and 10 mM ATP. The curves were fit globally using a 1:1 Langmuir binding model to yield the following binding constants: $k_d = 3.5 \times 10^{-3}$ s^{-1}, $k_a = 1.36 \times 10^6$ M^{-1} s^{-1}, and $K_D = 2.6 \times 10^{-9}$ M. (C) ITC thermogram and fit for 50 μM CKB (dimer concentration) (in the syringe) binding to 5 μM ASB9(1−252) (in the cell). Because the binding was so tight, only the ΔH could be determined from the data [−16.5 ± 0.2 kcal/mol (N = 0.9)]. (D) ITC thermogram and fit for 50 μM CKB (dimer concentration) (in the syringe) binding to 5 μM ASB9(35−252). For this truncated version of ASB9, ΔH = −9.9 ± 0.2 kcal/mol (N = 0.78).

Although the c value was too high to measure the K_d, the ΔH of the interaction was large, −16.5 kcal/mol, and could be determined with a high degree of precision (Figure 1C). Sensorgrams collected with immobilized CKB and ASB9(35−252) from 0 to 14 nM revealed a k_d of 3.5×10^{-3} s^{-1} and a k_a of 1.36×10^6 M^{-1} s^{-1}, yielding a K_D of 2.6×10^{-9} M (Figure 1B). Although the c value for this interaction was still high, the ΔH of the interaction was only −9.8 kcal/mol for the truncated version (Figure 1D). These results strongly suggested a bipartite interaction of ASB9 with CK.

Stoichiometry of the ASB9−CK Complex. CK is a dimeric protein, and therefore it was necessary to determine whether one molecule of ASB9 bound to a dimer of CK or whether each monomer of CK could bind an ASB9 molecule. Size exclusion chromatography and analytical ultracentrifugation experiments both suggested a stoichiometry of one ASB9 per CK dimer, but because of the elongated shape of ASB9, the results were inconclusive. MALS analysis of a mixture of CK with an excess of ASB9 showed conclusively that the stoichiometry was one ASB9 bound to one dimer of CK (Figure 2). The experiment was performed with an excess of ASB9, which eluted at the expected volume for monomeric ASB9.

Partial Inhibition of CK Activity by ASB9. The enzymatic activity of CK to form creatine phosphate was measured in a coupled assay.[19] In the absence of ASB9, the V_{max} was $(3.36 \pm 0.07) \times 10^{-7}$ M^{-1} s^{-1} (for comparison to literature values, this translated to a k_{cat} of 6.7×10^3 min^{-1}) and the K_M

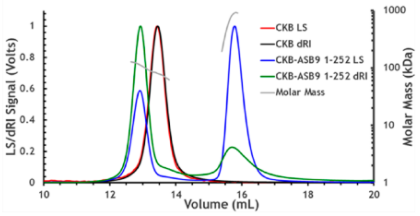


Figure 2. Multiangle light scattering (MALS) and refractive index curves for CKB and the CKB–ASB9 complex. The calculated molecular masses were 85.3 kDa for CKB, 29 kDa for ASB9(1–252), and 114 kDa for CKB and ASB9(1–252) (one CKB dimer and one ASB molecule). Both light scattering (LS) and refractive index with buffer subtracted (dRI) were measured and are plotted for each protein sample.

was 2.1 ± 0.2 mM. When the complex of ASB9(1–252) with CK was assayed, partial inhibition was observed so that the V_{max} of the complex was (2.1 ± 0.06) × 10^{-7} M^{-1} s^{-1} and the K_M was 3.4 ± 0.3 mM (Figure 3). This result was consistent with

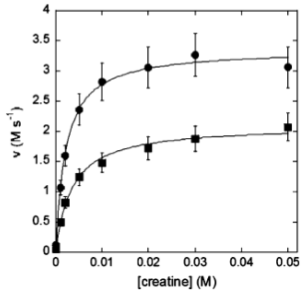


Figure 3. Enzyme activity of (■) CKB and (●) CKB with ASB9(1–252) at 3 nM enzyme and 30 °C. The K_M was 2.1 ± 0.2 mM, and the V_{max} was (3.36 ± 0.07) × 10^{-7} M^{-1} s^{-1} (for comparison to literature values, this translated to a k_{cat} of 6.7 × 10^3 min^{-1}). Binding of ASB9(1–252) to CKB partially inhibits the activity of CKB, increasing the K_M slightly to 3.4 ± 0.3 mM and decreasing the V_{max} to (2.1 ± 0.06) × 10^{-7} M^{-1} s^{-1}.

the MALS result that showed one ASB9 bound to a dimer of CK, likely leaving one active site available in the CK dimer. The complex of ASB9(35–252) with CK was also assayed, and this complex had full CK activity hinting that the N-terminus of ASB9 may occlude the active site of one monomer of CK in the ASB9(1–252)–CK complex.

HDXMS Mapping of the Interface between ASB9 and CK. To obtain information about where ASB9 and CK come in contact, amide hydrogen–deuterium exchange experiments (HDXMS) were performed in which the proteins were exposed to deuterated buffer for short periods of time to map the surface-exposed amides in the individual proteins and in the complex.[17] A Synapt G2S mass spectrometer was used for the study, yielding amide exchange measurements across nearly the entire sequence of both proteins (Figure 1 of the Supporting Information). These experiments revealed only one region of CKB, residues 182–203, was protected upon binding to ASB9(1–252) and ASB9(35–252). This region was covered by two peptides, 182–192 and 193–203, both of which showed a decreased level of exchange in the complex (Figure 4A–C).

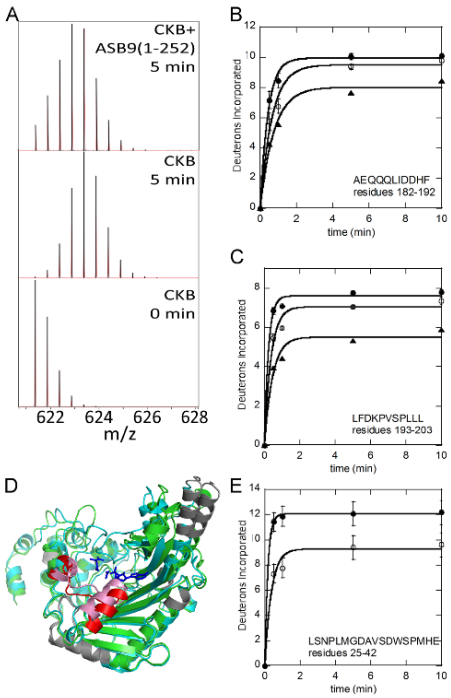


Figure 4. Amide hydrogen–deuterium exchange (HDXMS) results. (A) Peptide mass envelope for the CKB peptide LFDKPVSPLLL from CKB alone or in complex with ASB9(1–252) after exchange for 5 min showing a clear decrease in the level of deuterium uptake upon binding to ASB9. (B and C) D_2O uptake curves for CKB peptides AEQQQLIDDHF and LFDKPVSPLLL for (●) CKB, (○) CKB with ASB9(35–252), and (▲) CKB with ASB9(1–252). (D) Overlay of structures of human apo CKB (open, green) and CKB (closed, cyan) bound to Mg-ADP, NO_3 (shown as blue sticks), and creatine (shown as blue sticks). The region of CKB that shows a difference in deuterium incorporation upon binding to ASB9 is colored red in the open conformer and pink in the closed conformer. CKB regions that had no coverage are colored gray. (E) Deuterium uptake plot for the ASB9 peptide LSNPLMGDAVSDWSPMHE for (●) ASB9(1–252) and (○) ASB9(1–252) with CKB.

The protected residues 182–203 mapped to a helix–coil–helix region that is located right in front of the substrate-binding pocket of the CK active site (Figure 4D). The helix, residues 182–189, undergoes a conformational change and closes upon substrate binding as shown by the overlay of the two monomers from the dimeric CK crystal structure (PDB entry 3DRB).

The HDXMS results for ASB9 showed subtle changes throughout the protein; however, the main site of protection appeared to be N-terminal residues 25–42 (Figure 4E). A full

analysis of the HDXMS data on ASB9 will be presented in a separate study.

Molecular Modeling. Although a model of the ASB9−CK interaction has already been proposed,[13] this model was created from the crystal structure of ASB9 in which the first 35 residues were not observed, likely because of the N-terminus being disordered. As our binding studies implicated the N-terminus of ASB9 as a major contributor to binding affinity, we sought to model the disordered N-terminus and predict a model that included these residues. Multiple simulations of the N-terminal residues, ASB9(1−35), were performed, and the structures were clustered. Two clusters emerged from these simulations, one in which the N-terminus was extended and another in which it was folded back on itself. Representative models from these two cluster families were then attached to ASB9 to form a model of the full-length protein to create two different models of full-length ASB9. To explore whether it is the open state of CK or the closed state, two different models of the CK dimer were used. One corresponded to the CK crystal structure (PDB entry 3DRB) in which the A chain has an open active site and the B chain has the active site closed (and an ADP bound, which was removed for the docking). A second model in which both active sites were open was created starting again from the 3DRB structure but replacing the B chain (closed structure) with a replica of the A chain (open structure). The two different ASB9 models were docked against these two CK models, and the results of each trial were scored using a combination of the actual docking score from ZDOCK and the consistency of the docked orientation from the top 10 docked complexes. Only when the CK dimer in which both active sites were open was docked with ASB9 that had an extended N-terminus were consistent solutions with good ZDOCK scores obtained. The consistent solution from this docking trial mainly had the N-terminus and first ankyrin repeat of ASB9 docked between the two monomers of CK with the extended N-terminus lying in one of the CK active sites as was shown by a second scoring described in Experimental Procedures (Figure 5).

Role of Asp32 of ASB9. The molecular docking revealed a possible role for Asp32 of ASB9 in inhibition of CK activity. The docked structure showed Asp32 taking the place of one of the phosphates of ATP in the active site (Figure 5C). To test the importance of this interaction, we prepared the D32A mutant of ASB9 and compared its activity to that of the wild type. ITC experiments revealed the mutant protein interacted with CK with a ΔH of −3.7 kcal/mol, even lower than the binding enthalpy for the N-terminal truncation ASB9(35−252) (Figure 6). Activity assays showed that the D32A mutant partially inhibited the activity of CK. Whereas ASB9(1−252) inhibited the V_{max} of CK by 38 ± 4%, ASB9(D32A) inhibited the V_{max} by 22 ± 3% and ASB9(35−252) did not inhibit the activity at all.

■ DISCUSSION

In order to build up a functional model of how ASB9 interacts with CK, we performed a number of experiments and interpreted these in light of a docked molecular model. First, we found that one molecule of ASB9 interacts with a dimer of CK in an asymmetric manner. The N-terminus of ASB9 and the first ankyrin repeat nestle into the cleft between the two monomers of CK. The HDXMS data reveal a single region of CK that is protected from exchange in the complex, residues 182−203. This helix−strand−helix structure forms one lip of the CK active site and is known to undergo a conformational change, being open in the apoenzyme and closed in the substrate-bound form.[31]

A

	CKB (open/open)	CKB (open/closed)
Elongated N-term ASB9	**Score range 1230-1778** RMSD: 3.4 +/- 1.6 100% without steric overlap	**Score range 1113-1405** RMSD: 9.8 +/- 6.2 83% without steric overlap
Compact N-term ASB9	**Score range 1200-1516** RMSD: 12.9 +/- 1.7 44% without steric overlap	**Score range 1108-1325** RMSD: 11.3 +/- 2.1 57% without steric overlap

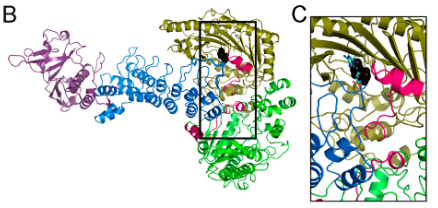


Figure 5. Best docked model of the ASB9−CK complex. (A) Table of results from docking of two different structural models of ASB9. One with an extended and the other with a collapsed N-terminus were docked with CKB structures that had either open or closed active sites. The solutions with the best ZDOCK score were obtained from the docking of the ASB9 with an extended N-terminus and an open CKB. The solutions were also overlapped with the ASB9−EloBC crystal structure (PDB entry 3ZKJ), and the percentage of structures that did not have steric clashes between the SOCS−EloBC and CKB structures is also reported in the table. RMSD is reported in Å. (B) Docked model of the ASB9(1−252)−CKB complex. The regions of CK (green and olive) that were shown to be protected upon ASB9 (blue) binding by HDXMS are colored bright pink. Aspartate 32 within the disordered N-terminus in ASB9 that binds in the ATP-binding pocket of CK is shown as black spheres. Elongin BC, which binds at the other end of the ASB9 ARD, is colored purple. (C) Close-up view of the active site of the CK monomer in which the N-terminus of ASB9 docks. The pose is the same as in panel B, but the ATP has been added back into the structure to show the superposition of Asp32 with the ATP.

Second, comparison of full-length ASB9(1−252) with an N-terminally truncated mutant, ASB9(35−252), revealed that both bound tightly, but the N-terminal residues made the affinity even higher. Interestingly, the N-terminal truncation, although it binds CK tightly, does not inhibit enzyme activity, and even full-length ASB9 only partially inhibits CK activity. Because of the stoichiometry of the complex (one ASB9 per dimer of CK), one active site remains open in the complex, and therefore, only partial inhibition is seen. Some studies have suggested that only one active site of CK functions at a time,[31] and our studies reveal that the ASB9−CK complex retains much of the CK activity. The partial inhibition seen is consistent with the docked model, which shows the disordered N-terminus of ASB9 lying in one active site with Asp32 interacting with the positively charged residues that normally interact with ATP. The D32A mutation shows partial inhibition consistent with weakening of this probably disordered interaction that would inhibit enzyme activity. Thus, it is

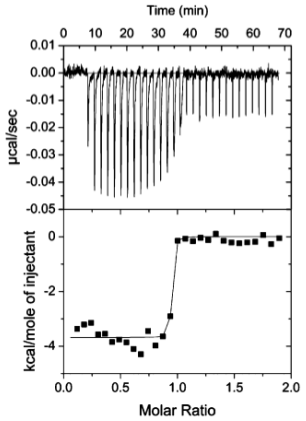


Figure 6. ITC thermogram and fit for 50 μM CKB (dimer concentration) (in the syringe) binding to 5 μM ASB9(1−252) D32A. For this mutant version of ASB9, $\Delta H = -3.7 \pm 0.08$ kcal/mol.

possible that CK could exist in cells as the ASB9 complex and still retain significant activity.

Third, we found that ASB9(1−252) interacts with CK so tightly that dissociation was not observed even in the presence of denaturant. This result is reminiscent of the interaction of IκBα with NFκB, for which dissociation is also not observed.[22,33] As with IκBα, this result suggests that proteasomal degradation is the only way to dissociate the complex.[34] The result also suggests that the ASB9−CK complex may be a stable subcomponent of the complete ECS-type CR E3 ligase. Studies of the other protein−protein interactions within the complex will reveal whether any of the other components weaken this interaction or whether the complex remains intact until proteasomal degradation occurs.

Our model of the ASB9−CK complex differs significantly from that predicted previously when only the crystallographically ordered portion of ASB9 was used in the docking.[13] We observed that the N-terminal residues contributed substantially to the binding based on ITC, SPR, HDXMS, and enzymatic assays. Therefore, we sought a way to predict some possible structures of the disordered N-terminus of ASB9 to generate a more realistic docked model.[13] Remarkably, the docked model we obtained, with the N-terminus included, was completely consistent with HDXMS data showing where the two proteins made surface contact. The C-terminus of ASB9, which must interact with Elongin BC and the rest of the ECS-type CR E3 ligase, protrudes from the center of the CK dimer placing the CK at the tip of the long ARD, almost like a lollypop on the end of a stick. Our model suggests that a simple hinging motion in the ECS-type CR E3 ligase would allow contact of the CK with the ubiquitin at the other end of the E3 ligase complex. Such a model is only one of several possible mechanisms for ubiquitin transfer requiring further study.

■ ASSOCIATED CONTENT

Supporting Information

Figure 1 showing all of the deuterium uptake curves for CK alone and in complex with ASB9(1−252) and ASB9(35−252). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: ekomives@ucsd.edu. Phone: (858) 534-3058.

Funding

D.B. is supported by American Heart Association Postdoctoral Award 14POST18970079. J.S. is supported by National Institutes of Health (NIH) Molecular Biophysics Training Grant T32 GM008326-23. This work was funded in part through a NIH Director's New Innovator Award (DP2-OD007237) to R.E.A. and by NIH Grant P01 GM071862 to E.A.K. Support from the National Biomedical Computation Resource (NBCR, P41 GM103426) to R.E.A. is also gratefully acknowledged.

Notes

The authors declare no competing financial interest.

■ ABBREVIATIONS

CKB, brain-type creatine kinase; uMtCK, ubiquitous mitochondrial creatine kinase.

■ REFERENCES

(1) Hershko, A., and Ciechanover, A. (1998) The ubiquitin system. *Annu. Rev. Biochem. 67*, 425−479.

(2) Schulman, B. A., and Harper, J. W. (2009) Ubiquitin-like protein activation by E1 enzymes: The apex for downstream signalling pathways. *Nat. Rev. Mol. Cell Biol. 10*, 319−331.

(3) Zheng, N., Schulman, B. A., Song, L., Miller, J. J., Jeffrey, P. D., Wang, P., Chu, C., Koepp, D. M., Elledge, S. J., Pagano, M., Conaway, R. C., Conaway, J. W., Harper, J. W., and Pavletich, N. P. (2002) Structure of the Cul1-Rbx1-Skp1-F boxSkp2 SCF ubiquitin ligase complex. *Nature 416*, 703−709.

(4) Petroski, M. D., and Deshaies, R. J. (2005) Function and regulation of cullin-RING ubiquitin ligases. *Nat. Rev. Mol. Cell Biol. 6*, 9−20.

(5) Zimmerman, E. S., Schulman, B. A., and Zheng, N. (2010) Structural assembly of cullin-RING ubiquitin ligase complexes. *Curr. Opin. Struct. Biol. 20*, 714−721.

(6) Duda, D. M., Borg, L. A., Scott, D. C., Hunt, H. W., Hammel, M., and Schulman, B. A. (2008) Structural insights into NEDD8 activation of cullin-RING ligases: Conformational control of conjugation. *Cell 134*, 995−1006.

(7) Kohroki, J., Nishiyama, T., Nakamura, T., and Masuho, Y. (2005) ASB proteins interact with Cullin5 and Rbx2 to form E3 ubiquitin ligase complexes. *FEBS Lett. 579*, 6796−6802.

(8) Thomas, J. C., Matak-Vinkovic, D., Van Molle, I., and Ciulli, A. (2013) Multimeric Complexes among Ankyrin-Repeat and SOCS-box Protein 9 (ASB9), ElonginBC, and Cullin 5: Insights into the Structure and Assembly of ECS-type Cullin-RING E3 Ubiquitin Ligases. *Biochemistry 52*, 5236−5246.

(9) Muniz, J. R., Guo, K., Kershaw, N. J., Ayinampudi, V., von Delft, F., Babon, J. J., and Bullock, A. N. (2013) Molecular architecture of the ankyrin SOCS box family of Cul5-dependent E3 ubiquitin ligases. *J. Mol. Biol. 425*, 3166−3177.

(10) Debrincat, M. A., Zhang, J. G., Willson, T. A., Silke, J., Connolly, L. M., Simpson, R. J., Alexander, W. S., Nicola, N. A., Kile, B. T., and Hilton, D. J. (2007) Ankyrin repeat and suppressors of cytokine signaling box protein asb-9 targets creatine kinase B for degradation. *J. Biol. Chem. 282*, 4728−4737.

(11) Kwon, S., Kim, D., Rhee, J. W., Park, J.-A., Kim, D.-W., Kim, D.-S., Lee, Y., and Kwon, H.-J. (2010) ASB9 interacts with ubiquitous mitochondrial creatine kinase and inhibits mitochondrial function. *BMC Biol. 8*, 1−23.

(12) Fei, X., Zhang, Y., Gu, X., Qiu, R., Mao, Y., and Ji, C. (2009) Crystallization and preliminary X-ray analysis of the splice variant of human ankyrin repeat and suppressor of cytokine signaling box protein 9 (hASB9-2). *Protein Pept. Lett. 16* (3), 333−335.
(13) Fei, X., Gu, X., Fan, S., Yang, Z., Li, F., Zhang, C., Gong, W., Mao, Y., and Ji, C. (2012) Crystal Structure of Human ASB9-2 and Substrate-Recognitionof CKB. *Protein J. 31*, 275−284.
(14) Zhong, L., Ge, K., Zu, J. C., Zhao, L. H., Shen, W. K., Wang, J. F., Zhang, X. G., Gao, X., Hu, W., Yen, Y., and Kernstine, K. H. (2008) Autoantibodies as potential biomarkers for breast cancer. *Breast Cancer Res. 10*, R40-41−R40-48.
(15) Tokuoka, M., Miyoshi, N., Hitora, T., Mimori, K., Tanaka, F., Shibata, K., Ishii, H., Sekimoto, M., Doki, Y., and Mori, M. (2010) Clinical significance of ASB9 in human colorectal cancer. *Int. J. Oncol. 37*, 1105−1111.
(16) Uranbileg, B., Enooku, K., Soroida, Y., Ohkawa, R., Kudo, Y., Nakagawa, H., Tateishi, R., Yoshida, H., Shinzawa, S., Moriya, K., Ohtomo, N., Nishikawa, T., Inoue, Y., Tomiya, T., Kojima, S., Matsuura, T., Koike, K., Yatomi, Y., and Ikeda, H. (2014) High ubiquitous mitochondrial creatine kinase expression in hepatocellular carcinoma denotes a poor prognosis with highly malignant potential. *Int. J. Cancer 134*, 2189−2198.
(17) Mandell, J. G., Falick, A. M., and Komives, E. A. (1998) Identification of protein-protein interfaces by decreased amide proton solvent accessibility. *Proc. Natl. Acad. Sci. U.S.A. 95*, 14705−14710.
(18) Chen, L. H., White, C. B., Babbitt, P. C., McLeish, M. J., and Kenyon, G. L. (2000) A comparative study of human muscle and brain creatine kinases expressed in *Escherichia coli*. *J. Protein Chem. 19*, 59−66.
(19) Tanzer, M. L., and Gilvarg, C. (1959) Creatine and creatine kinase measurement. *J. Biol. Chem. 234*, 3201−3204.
(20) Wang, P. F., McLeish, M. J., Kneen, M. M., Lee, G., and Kenyon, G. L. (2001) An unusually low pK_a for Cys282 in the active site of human muscle creatine kinase. *Biochemistry 40*, 11698−11705.
(21) Trathnigg, B. (1995) Determination of MWD and Chemical Composition of Polymers by Chromatographic Techniques. *Prog. Polym. Sci. 20*, 615−650.
(22) Bergqvist, S., Croy, C. H., Kjaergaard, M., Huxford, T., Ghosh, G., and Komives, E. A. (2006) Thermodynamics reveal that helix four in the NLS of NF-κB p65 anchors IκBα, forming a very stable complex. *J. Mol. Biol. 360*, 421−434.
(23) Dembinski, H., Wismer, K., Balasubramaniam, D., Gonzalez, H. A., Alverdi, V., Iakoucheva, L. M., and Komives, E. A. (2014) Predicted disorder-to-order transition mutations in IκBα disrupt function. *Phys. Chem. Chem. Phys. 16*, 6480−6485.
(24) Fang, J., Nevin, P., Kairys, V., Venclovas, C., Engen, J. R., and Beuning, P. J. (2014) Conformational analysis of processivity clamps in solution demonstrates that tertiary structure does not correlate with protein dynamics. *Structure 22*, 572−581.
(25) Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Bioinf. 65*, 712−725.
(26) Jorgensen, W. L. (1982) Revised TIPS for simulations of liquid water and aqueous solutions. *J. Chem. Phys. 77*, 4156−4163.
(27) Salomon-Ferrer, R., Goetz, A. W., Poole, D., Le Grand, S., and Walker, R. C. (2013) Routine microsecond molecular dynamics simulations with AMBER: Part II: Particle Mesh Ewald. *J. Chem. Theory Comput. 9*, 3878−3888.
(28) Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems. *J. Chem. Phys. 98*, 10089−10092.
(29) Baron, R., and McCammon, J. A. (2007) Dynamics, hydration, and motional averaging of a loop-gated artificial protein cavity: The W191G mutant of cytochrome c peroxidase in water as revealed by molecular dynamics simulations. *Biochemistry 46*, 10629−10642.
(30) Hess, B., Kutzner, C., Van Der Spoel, D., and Lindahl, E. (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput. 4*, 435−447.
(31) Eder, M., Schlattner, U., Becker, A., Wallimann, T., Kabsch, W., and Fritz-Wolf, K. (1999) Crystal structure of brain-type creatine kinase at 1.41 Å resolution. *Protein Sci. 8*, 2258−2269.
(32) Pierce, B. G., Wiehe, K., Hwang, H., Kim, B. H., Vreven, T., and Weng, Z. (2014) ZDOCK Server: Interactive Docking Prediction of Protein-Protein Complexes and Symmetric Multimers. *Bioinformatics 30*, 1771−1773.
(33) Alverdi, V., Hetrick, B., Joseph, S., and Komives, E. (2014) Direct observation of a transient ternary complex during IκBα-mediated dissociation of NF-κB from DNA. *Proc. Natl. Acad. Sci. U.S.A. 111*, 225−230.
(34) Ferreiro, D. U., and Komives, E. A. (2010) Molecular Mechanisms of System Control of NF-κB Signaling by IκBα. *Biochemistry 49*, 1560−1567.

ACKNOWLEDGMENT

Chapter 4, in full, is a repreint of the material as it appears in How the Ankyrin and SOCS Box Protein, ASB9, Binds to Creatine Kinase 2016. Balasubramaniam, D., Schiffer, J., Parnell, J., Mir, S. P., Amaro, R. E., Komives, E. A., Biochemistry 2015. The dissertation author was a secondary investigator and author of this material.

CHAPTER 5

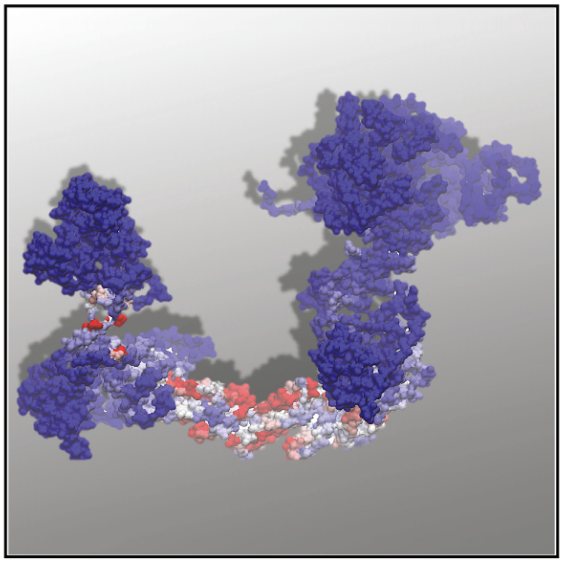
Model of the Ankyrin and SOCS Box Protein, ASB9, E3 Ligase Reveals a Mechanism for Dynamic Ubiquitin Transfer

Article

Structure

Model of the Ankyrin and SOCS Box Protein, ASB9, E3 Ligase Reveals a Mechanism for Dynamic Ubiquitin Transfer

Graphical Abstract



Authors

Jamie M. Schiffer, Robert D. Malmstrom, Jonathan Parnell, Cesar Ramirez-Sarmiento, Javiera Reyes, Rommie E. Amaro, Elizabeth A. Komives

Correspondence

ramaro@ucsd.edu (R.E.A.), ekomives@ucsd.edu (E.A.K.)

In Brief

Cullin-RING E3 ligases (CRLs) are multi-protein complexes that transfer ubiquitin over 60 Å to a protein doomed for degradation. Schiffer et al. construct an atomic model for a CRL that ubiquitinates creatine kinase (CK) and uncover dominant modes of motion that could permit ubiquitin transfer.

Highlights

- Protein-protein docked models of ASB9-CK align with new ITC and SAXS data
- Simulations from these models converge to a similar complex interface
- Key residues in this interface are validated with mutagenesis and pull-down assays
- A dominant mode of motion of the CK-targeting E3 ligase is revealed

Schiffer et al., 2016, Structure *24*, 1248–1256
August 2, 2016
http://dx.doi.org/10.1016/j.str.2016.05.016

CellPress

Structure
Article

Model of the Ankyrin and SOCS Box Protein, ASB9, E3 Ligase Reveals a Mechanism for Dynamic Ubiquitin Transfer

Jamie M. Schiffer,[1] Robert D. Malmstrom,[1,2] Jonathan Parnell,[1] Cesar Ramirez-Sarmiento,[3,4] Javiera Reyes,[3] Rommie E. Amaro,[1,2,*] and Elizabeth A. Komives[1,*]
[1]Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0378, USA
[2]National Biomedical Computation Resource, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0608, USA
[3]Departamento de Biologia, Facultad de Ciencias, Universidad de Chile, Las Palmeras 3425, Casilla 653, Santiago 7800003, Chile
[4]Present address: Institute for Biological and Medical Engineering Pontificia Universidad Católica de Chile Avda. Vicuña Mackenna 4860, Macul - Santiago, Chile
*Correspondence: ramaro@ucsd.edu (R.E.A.), ekomives@ucsd.edu (E.A.K.)
http://dx.doi.org/10.1016/j.str.2016.05.016

SUMMARY

Cullin-RING E3 ligases (CRLs) are elongated and bowed protein complexes that transfer ubiquitin over 60 Å to proteins targeted for proteasome degradation. One such CRL contains the ankyrin repeat and SOCS box protein 9 (ASB9), which binds to and partially inhibits creatine kinase (CK). While current models for the ASB9-CK complex contain some known interface residues, the overall structure and precise interface of the ASB9-CK complex remains unknown. Through an integrative modeling approach, we report a third-generation model that reveals precisely the interface interactions and also fits the shape of the ASB9-CK complex as determined by small-angle X-ray scattering. We constructed an atomic model for the entire CK-targeting CRL to uncover dominant modes of motion that could permit ubiquitin transfer. Remarkably, only the correctly docked CK-containing E3 ligase and not incorrectly docked structures permitted close approach of ubiquitin to the CK substrate.

INTRODUCTION

Ubiquitin-mediated degradation regulates cellular concentrations of proteins critical to cellular function, including metabolic enzymes and signaling proteins. Cells have evolved to conjugate ubiquitin to a substrate protein by a cascade of three enzymes: E1 ubiquitin-activating enzyme, E2 ubiquitin-conjugating enzyme, and E3 ubiquitin ligase. E3 ligases catalyze the transfer of ubiquitin from E2 enzymes to protein substrates (Deshaies and Joazeiro, 2009) and, after multiple rounds, the poly-ubiquitination of a protein substrate flags it for destruction by the proteasome (Petroski and Deshaies, 2005). Although over 600 human E3 ligases have been identified, only a small percentage of E3 ligase structures have been solved, and even less is known about the conformational dynamics of these multi-protein complexes, despite being ubiquitous and attractive therapeutic targets against tumor malignancies (Deshaies and Joazeiro, 2009; Kerscher et al., 2006; Zhao and Sun, 2013). The Cullin-RING E3 ligases (CRLs) are the most abundant type of E3 ligases and form an extended multi-protein complex. Previous models of CRLs show a distance of over 60 Å between ubiquitin, positioned on the C-terminal end of the Cullin protein, and the target protein, positioned at the N-terminal end of the Cullin protein. Thus, substantial conformational changes must be required for ubiquitin transfer (Thomas et al., 2013; Duda et al., 2008).

One such CRL contains the ankyrin repeat and suppressor of cytokine signaling (SOCS) box protein 9 (ASB9), which forms the substrate recognition domain in an Elongin-Cullin-SOCS box-type (ECS-type) E3 ligase. ASB9 interacts directly with the E3 ligase protein substrate, creatine kinase (CK), targeting it for ubiquitin-mediated degradation (Debrincat et al., 2007; Kwon et al., 2010). As one of 18 members of the ASB family, the largest family of SOCS-box-containing proteins (Kohroki et al., 2005), ASB9 is composed of an intrinsically disordered N terminus, an ankyrin repeat domain, and the SOCS box domain. While the SOCS box interacts with Elongin B, Elongin C, and Cullin 5 in the E3 ligase complex (Thomas et al., 2013; Muniz et al., 2013), the ankyrin repeat domain and N-terminal disordered segment interact with CK (Fei et al., 2012; Balasubramaniam et al., 2015). CKs play a critical role in the buffering of ATP/ADP ratios in excitable cells such as in cardiomyocytes and neural cells, and as such are crucial for cellular homeostasis in the vertebrate heart and brain (Wang et al., 2011; Bush et al., 2011).

Based on a crystal structure of the ASB9/Elongin B, C (Figure 1), an initial model was proposed for the interaction of ASB9 with a dimer of CK (Muniz et al., 2013). Our group then used an array of solution biophysical studies to obtain experimental evidence of where CK docks on ASB9 and presented a revised model that aligned with our experimental data (Balasubramaniam et al., 2015). Enzymatic assays indicated that the ASB9-CK interaction retained half-the-sites reactivity, and our second-generation model therefore had the intrinsically disordered N terminus of ASB9 occluding the active site of one CK monomer. Isothermal titration calorimetry (ITC) and surface plasmon resonance data revealed that the ankyrin repeat domain of ASB9 binds to CK with nanomolar affinity, and revealed that the N-terminal disordered residues contribute to even tighter affinity (Balasubramaniam et al., 2015). Hydrogen-deuterium exchange

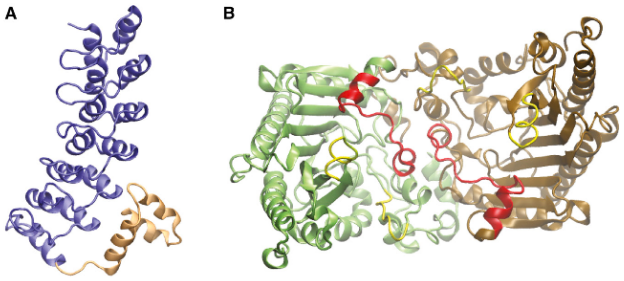


Figure 1. Structures of ASB9 and CK
(A and B) Structures of (A) the ankyrin repeat domain (purple) and SOCS box domain (orange) of ASB9 (PDB: 3ZKJ) and (B) the CK homo-dimer made up of two open monomers (PDB: 3DRB). The active site loops (yellow) and hydrogen deuterium exchange protected residues (red) are shown for reference.

mass spectrometry (HDXMS) showed that when ASB9 binds CK residues 182–203 are protected (Balasubramaniam et al., 2015). The second-generation model captured these biophysical results including the interaction of the N-terminal repeats of ASB9 with the HDX-protected residues in one of the CK monomers.

Here, we use small-angle X-ray scattering (SAXS), ITC, pull-down assays, computational modeling, and simulation to create and validate a third-generation model of the ASB9-CK interaction that not only captures the interaction interface, but also the overall shape of the complex. We disrupt binding of ASB9 with CK through mutations of residues predicted to be in the binding interface by the model. Finally, we use this model to provide a description of the dynamic motions of ECS-type E3 ligase complexes that enable ubiquitin transfer to the target protein. Our results are, to our knowledge, the first to report an atomic-level mechanism that permits ubiquitin transfer in CRLs, which are the largest family of E3 ligases, ubiquitinate almost 20% of proteins degraded through the ubiquitin-mediate degradative pathway, and are an attractive therapeutic target in cancer (Zhao and Sun, 2013).

RESULTS AND DISCUSSION

Isothermal Titration Calorimetry Refines which Regions of the N-Terminal Disordered Segment Contribute to Binding

Our second-generation model predicted that residues 1–18 were not in contact with CK, whereas residues 19–35 of the disordered N terminus of ASB9 lay in one active site of CK. To test whether this prediction was correct, we performed ITC experiments, titrating CK into a truncated mutant of ASB9 that only contains residues predicted to be in the complex interface from the docked poses, ASB9(19–252). Although the *c* value was too high to measure the K_d, the ΔH of the interaction was large, −19.2 kcal/mol, and could be determined with a high degree of precision (Figure 2A). This enthalpic contribution fully accounted for the CK interaction with ASB9, suggesting that residues 1–19, do not interact with CK in the binding interface.

Third-Generation Docked Model of the ASB9-CK Complex

Based on our ITC results, the ASB9 structure (PDB: 3D9H) was modified to include structures of the intrinsically disordered residues 19–35 that were not present in the crystal structure. Structures of ASB9(19–35) were generated with molecular dynamics (MD) simulations and these trajectories showed that this region sampled elongated structures similar to what was observed for the intrinsically disordered N-terminal 35 residues of ASB9 as we have previously reported (Balasubramaniam et al., 2015).

The crystal structure of CK (PDB: 3B6R) has one monomer of the dimer in an open conformation and one in a closed conformation. As previously described in the second-generation docking, we generated an open-open dimer from the crystal structure coordinates (Balasubramaniam et al., 2015). Based on the ITC results presented here, our current docked model started from further minimized and equilibrated structures of the ASB9(19–252) and of CK (open-open dimer).

Using the ZDOCK 3.0.2 server (Pierce et al., 2014), we obtained the top ten best-docked poses (Table S1). The top scoring model, which we refer to as ZDOCK-1, had a score of 1,778 and the second highest scoring model, ZDOCK-2, had a score of 1,425 indicating a strong discrimination between the models. However, all of the ten poses appeared very similar in the orientation of the ASB9 ARD with respect to the CK active site (Figure 2B).

Analysis of ASB9-CK Complex Scattering Profiles

We collected SAXS data for both ASB9(19–252)-CK and ASB9(1–252)-CK, and both yielded scattering profiles with monotonically decaying scattering as a function of the scattering vector q and calculated R_g of 33.1 and 32.4 Å, respectively (Figure 3A). We used the Debye formula for spherical scatters (Schneidman-Duhovny et al., 2011, 2013) to compute the SAXS profile from the ZDOCK-1 and ZDOCK-2 poses. The profile calculated from the ZDOCK-1 pose agreed with both the ASB9(19–252)-CK and the ASB9(1–252)-CK complexes, with χ^2 values of 1.79 and 1.21, and R_g values of 31.84 and 32.38 Å, respectively. The profile calculated from the ZDOCK-2 pose showed poor agreement with the ASB9(19–252)-CK complex, having a χ^2 value of 7.34 and an R_g value of 31.90, but agreed with the ASB9(1–252)-CK complex, with a χ^2 value of 1.21 and an R_g value of 32.15 Å. Other protein-protein docking servers, such as RosettaDock (Lyskov and Gray, 2008), FOXSDock (Schneidman-Duhovny et al., 2011), and ClusPro2.0 (Kozakov et al., 2013; Comeau et al., 2004) were also used and attained poses similar to the ZDOCK-1 and ZDOCK-2 poses, but agreement as assessed by χ^2 was not as good as for the ZDOCK-1 pose (Figure S1 and Tables S1–S4). We chose to further explore the ZDOCK-1, which agreed best with the ASB9(19–252)-CK complex and the ZDOCK-2 pose, which agreed best with the ASB9(1–252)-CK complexes.

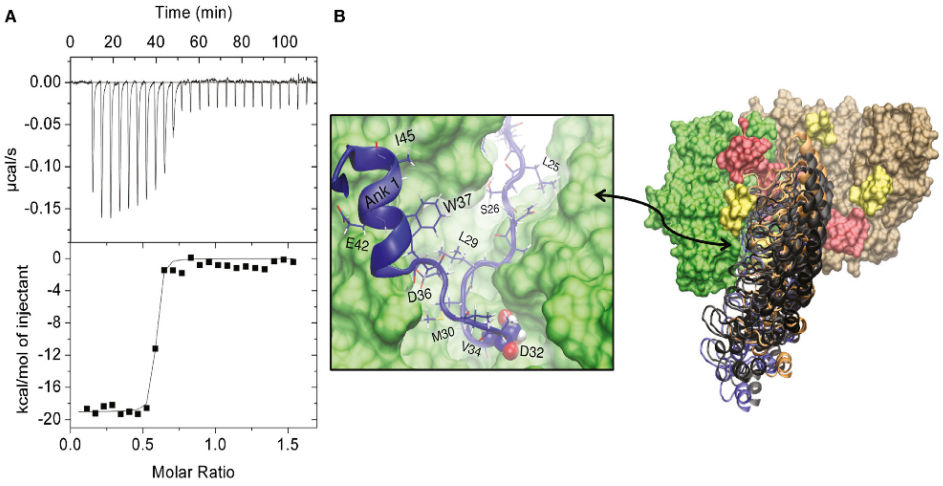


Figure 2. N-Terminal Residues of ASB9 Form Half of the Binding Interface between ASB9 and CK

(A) ITC thermogram and fit for 37 μM CK (dimer concentration) (in syringe) binding to 5 μM ASB9(19–252) (in cell).

(B) An overlay of six of our best docked models from the third-generation docking demonstrate the consistency in docking. ZDOCK1 (blue) and ZDOCK2 (orange) poses of ASB9 (blue) are shown for reference. CK dimer (green/brown) is depicted in a surface model, with active site loops (yellow) and HDXMS-protected residues (red) similarly depicted. Residues of the intrinsically disordered N terminus of ASB9 loop through the active site of CK in our third-generation docked model.

Because the ZDOCK-1 pose agreed better with the SAXS profiles than poses from other docking servers, these were not pursued further.

The ZDOCK-1 and ZDOCK-2 poses appeared highly similar, with both poses showing the intrinsically disordered residues 23–35 interacting with one open active site in CK. However, the strong variation in χ^2 agreement with the SAXS data for the ASB9(19–252)-CK complex suggested significant differences. Indeed, when further analyzed, the total buried surface area in the ZDOCK-1 structure was 3,970 Å^2, whereas only 2,901 Å^2 was buried in the ZDOCK-2 pose. The buried surface area in the ZDOCK-1 pose also involved many more interacting side chains including ASB9 residues D36, W37, E42, I45, and H46 of ank1 and L78 of ank2, which interact with the active site loop of the occluded CK monomer. Residues S35 and P39 interact with the HDX-protected residues of that same CK monomer while residues H48, Q49, L50, N54, L55, and Q59 of ank1 and H81 and S83 of ank2 interact with the HDX-protected residues of the open CK monomer. In addition, ASB9 residues A67, D68, H69, A100, D101 W102, E131, D133, L134, I164, S165, and H166 interact with active site loops of the occluded CK monomer (Figure 4A). The large amount of buried surface area is consistent with the very tight binding interaction (Balasubramaniam et al., 2015; Chothia and Janin, 1975; Chakravarty

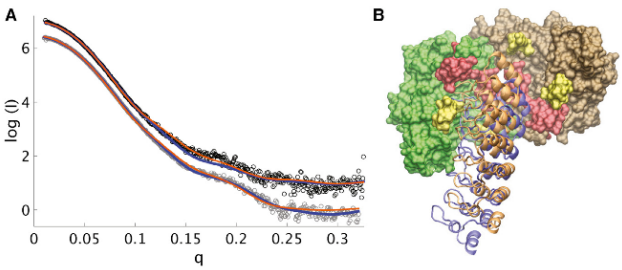


Figure 3. Docked Structures Align with SAXS Profiles for the ASB9(1–252) Complex with CK and for the ASB9(19–252) Complex with CK

(A) SAXS profiles for ASB9(1–252)-CK complex (black) and for ASB9(19–252)-CK complex (gray) are shown as a scatterplot. Theoretical scattering profiles for the ZDOCK1 pose, with or without residues 1–18, are shown in blue. Theoretical scattering profiles for the ZDOCK2 pose, with or without residues 1–18, are shown in orange.

(B) Relative orientation of the ankyrin repeat domain of ASB9 in the ZDOCK1 (blue) or ZDOCK2 (orange) pose, respectively. HDXMS-protected residues (red) and active site loops (yellow) of CK are shown for reference.

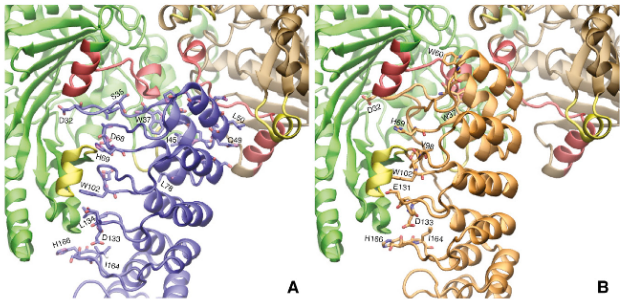


Figure 4. Residues from ASB9 Involved in the Binding Interface for the Third-Generation Model

(A) Residues from ASB9 in the ZDOCK1 pose include D32, S35, W37, I45, Q49, L50, D68, H69, L78, W102, D133, L134, I164, and H166.

(B) Residues for ASB9 in the ZDOCK2 pose include D32, W37, W60, H69, V98, W102, E131, D133, I164, and H166.

et al., 2013). In the ZDOCK-2 pose, only residues W37 and W60 of ASB9 ank 1 interact with HDX-protected residues of the occluded CK active site. In addition, residues H69, V98, T99, W102, E131, L134, D133, I164, S165, and H166 interact with the CK active site loops (Figure 4B).

ZDOCK-1 and ZDOCK-2 Poses Converge to Similar Binding Interface in MD Simulations

To test the stability and understand the structural dynamics of the ASB9-CK complex, we performed all-atom MD simulations initiated from the atomic structures of the ZDOCK-1 and ZDOCK-2 poses (Table 1). We simulated ASB9(23–252) and not ASB9(19–252), because in both the ZDOCK-1 and ZDOCK-2 poses, ASB9 residue R23 is the last N-terminal residue to lie within the CK open cleft, and we wished to avoid issues with force fields in oversampling collapsed structures of intrinsically disordered proteins (Henriques et al., 2015). Simulations were performed in triplicate for both the ZDOCK-1 and ZDOCK-2 poses.

Within tens of nanoseconds, the structures from the ZDOCK-2 simulations relaxed to an ASB9-CK interface that mimicked the ZDOCK-1 pose, while the interface of the ZDOCK-1 pose was maintained for over 300 ns of simulation (Figure 5, Movie S1). The relaxation of the ZDOCK-2 structures was measured with all-atom root-mean-square deviation (rmsd) of the most critical interface residues from ASB9 (see below, residues D32, W37, I45, L50, L78, W102) and internal distances between residues implicated in the complex interface in the ZDOCK-1 pose (Figure 5A, Movie S1, and Figure S1). The internal distance between ASB9 residues I45 and L78 and CK active site loop residues P67 and Y68 relaxes to and is maintained at between 4 and 6 Å in all three parallel simulations from the ASB9(23–252)-CK complex (Figure 5A). A similar relaxation occurs for the internal distances between the center of mass of ank1 and CK residue V198, which is a residue in one of the HDXMS-protected peptides (Movie S1, Figure S1).

Simulations provide new insight into the dynamics of CK that explain the protection of CK residues A182 to F192 (peptide 1 from HDXMS) and residues L193 to L203 (peptide 2 from HDXMS) in the helix-coil-helix region, located in front of the substrate-binding pocket, when in complex with ASB9. In the ZDOCK1 pose, ank1 of ASB9 binds directly to peptide 2 in both CK monomers. However, ASB9 does not bind directly to the CK peptide 1 in either CK monomer. Our simulations demonstrate that nanosecond motions lead to helix cracking causing (Kreuzer and Elber, 2013; Miyashita et al., 2003) exposure of residues E183 to L187 to exchange. This cracking is significantly dampened when ASB9 is bound to CK, and likely accounts for the protection of these residues previously reported with HDXMS (Balasubramaniam et al., 2015).

Residues Critical to the ASB9-CK Interaction Verified with Mutagenesis

Structures from the ZDOCK1 simulation were clustered based on backbone rmsd (Figure 6A), and the cluster centroids were analyzed using the Robetta Alanine Scan server (Kortemme and Baker, 2002; Kortemme et al., 2004) to predict the contributions of individual residues to the binding free energy of the ASB9-CK complex (Figure 6B). Residues that were previously found to be important in the ASB9-CK interaction, including D32, I45, and L78, were predicted to have a high ΔΔG between the alanine mutant and wild-type residue. In addition, residues W37, L50, and W102 were also predicted to have a high ΔΔG and are highly conserved among ASB9 variants. Pull-down assays on these mutants revealed that, indeed, residues W37, I45, L50, L78, and W102 are critical to the interaction of ASB9 with CK (Figure 6C). Thus, theoretical alanine scanning (Kortemme and Baker, 2002; Kortemme et al., 2004) of the MD ensemble of ASB9-CK accurately predicted the key and necessary residues in the binding interface of the ASB9-CK complex.

Mechanism of Ubiquitin Transfer to CK

With our experimentally validated model for the ASB9-CK complex, we constructed a complete model of the CK-targeting ECS-type E3 ligase, to uncover the conformational change necessary for ubiquitin transfer to CK. The E3 ligase model containing full-length ASB9, Elongins B and C, Cullin 5, Nedd8, Rbx1, E2, and ubiquitin, but without the CK substrate, was built as previously described (Muniz et al., 2013). The CK dimer was then added to this model based on our validated ZDOCK1 pose. To model the distance between the active thioester bond of E2-ubiquitin and lysines on CK, the ankyrin repeat domain of ASB9 from the E3 ligase model was aligned with the ankyrin repeat domain of ASB9 from the ZDOCK1 model. A distance of 82 Å was measured between the active thioester bond and the closest lysine on CK (Figure 7A).

The large distance between ubiquitin and CK in our E3 ligase model suggests the need for a large conformational change

Table 1. ZDOCK1 and ZDOCK2 System Setup

	ASB9(1–252)-CK ZDOCK1	ASB9(1–252)-CK ZDOCK2	ASB9(23–252)-CK ZDOCK1	ASB9(23–252)-CK ZDOCK2
Box dimensions (Å)	123.73 × 99.8 × 125.96	119.7 × 128.4 × 97.2	129.35 × 97.6 × 91.2	88.5 × 126.3 × 108.8
Total number of atoms	131,749	125,446	96,116	102,915
Neutralizing salt	37 NA^+	36 NA^+	38 NA^+	37 NA^+

The box dimensions, number of atoms, and neutralizing salt atoms added to the four systems simulated: ZDOCK1 simulations, with and without the 22 N-terminal residues of ASB9, and ZDOCK2 simulations, with and without the 22 N-terminal residues of ASB9.

in the transfer of ubiquitin, and we found that calculating the theoretical motions of this E3 ligase model based on the structure alone provided incredible insight. We performed elastic network model analysis of our E3 ligase model to decipher if any slow motions of this multi-domain protein complex could be predicted (Zheng and Doniach, 2003). Indeed, the slowest and most dominant normal mode observed in our E3 ligase model involves the coordinated folding of the Rbx1 linker domain with the bending of Cullin 5, and to a lesser extent Elongin C and ankyrin repeats 4–6 of ASB9 (Figures S2 and S3), bringing the active thioester of E2-ubiquitin within 5.2 Å of the nearest lysine in CK (Figure 7B).

Remarkably, the dominant normal modes that bring the ubiquitin close to its CK substrate were not observed in elastic network models of the CK-E3 ligase models constructed with the ZDOCK-2 structure. Instead, the normal mode that achieved any proximity was a twisting motion; however the ubiquitin and CK remained over 50 Å apart (Figure 7C). In addition, models that did not include the substrate, CK, were also incapable of motions that brought the ubiquitin close to ASB9. Thus, not only did our latest model fit with all of the experimental data, it was the only one that moved in a manner consistent with the function of the entire E3 ligase, ubiquitin transfer. To our knowledge, this is the first reported model that suggests how CRLs transfer ubiquitin over 60 Å. Our model strongly suggests that correct substrate docking is required to initiate the large-scale motions necessary for ubiquitin transfer in ASB-containing E3 ligases.

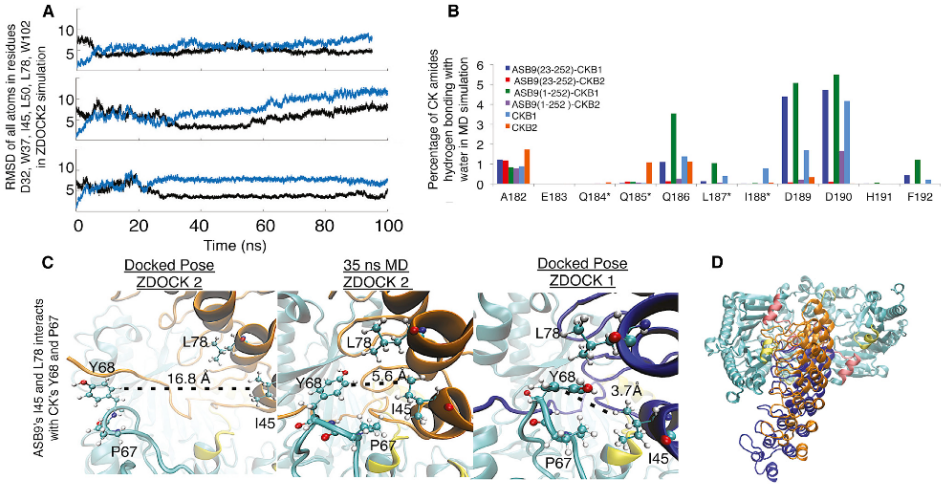


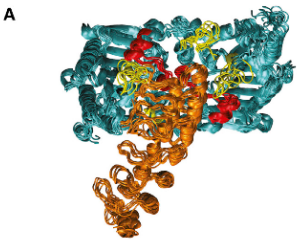
Figure 5. Simulations from the Best Docked Poses (ZDOCK1, ZDOCK2) Converge to a Similar Binding Interface and Provide an Explanation for HDX Protection of CK Residues by ASB9

(A) Root-mean-square deviation (rmsd) of all atoms in residues in the ZDOCK1 interface, D32, W37, I45, L50, L78, and W102, measured over three separate MD trajectories starting from the ZDOCK2 pose. Within a few nanoseconds, the rmsd in each trajectory decreases to <5 Å, and is maintained under 5 Å in two of the three simulations.

(B) Percentages of amide hydrogen bonding with water are reported for residues of peptide 1 in CK from simulation of ASB9(1–252) in complex with CK, simulations of ASB9(23–252) in complex with CK, and simulations of CK alone in solution.

(C) Figures showing the initial and final distances between ASB9 and CK residues in the ZDOCK2 simulation compared with the distance of the ZDOCK1 pose. This interface eventually converges to a very similar interface.

(D) For reference, the structures of the ZDOCK1 pose and ZDOCK2 pose.



B

ASB9 residue	Average ΔΔG	Sequence Conserve	Sequence Variety
D32	1.61 ± 0.25	15/18	G
V34	1.11 ± 0.14	16/18	E
D36	2.59 ± 0.84	16/18	Y
W37	1.17 ± 0.24	17/18	-
I45	0.70 ± 0.10	14/18	L, V
Q49	1.78 ± 0.68	1/18	L, V
L50	1.46 ± 0.12	18/18	-
L78	0.70 ± 0.37	18/18	-
W102	1.43 ± 0.86	17/18	R

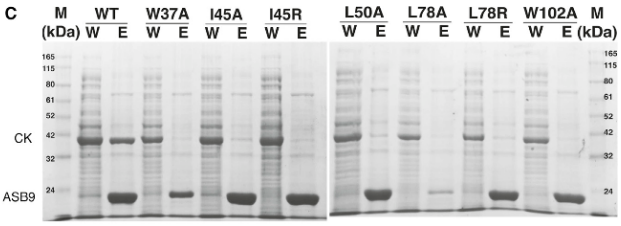


Figure 6. MD Simulations Reveal Which Residues Are Critical in the Binding Interface and Show Mobility of Active Site Residues

(A) Cluster centroids of the ASB9(23–252)-CK simulations.

(B) Robetta Alanine Scan results reported as an average over all cluster centroids.

(C) Pull-down assays of indicated alanine mutants reveals that CK is not pulled down sufficiently by any of these mutants. The wash (W) and elution (E) contents are shown for wild type and for each mutant.

EXPERIMENTAL PROCEDURES

Experimental Procedures used in this study for mutagenesis, protein purification, ITC, CK assay, and protein docking were previously reported (Balasubramaniam et al., 2015). Here, the same method was used for performing protein docking experiments as previously reported, except that the starting structures of ASB9(1–252) and CK(open-open) were minimized and equilibrated to sample more favorable side-chain and hydrogen-bond orientations using Maestro (Schrodinger Release 2015-4). Additionally, ASB9(19–35) was simulated to help build an ASB9(19–252) model for docking experiments.

Small-Angle X-Ray Scattering

Proteins were purified as previously described (Balasubramaniam et al., 2015), concentrated to 2 mg/ml for the ASB9(1–252)-CK complex, and 3 mg/ml for the ASB9(19–252)-CK complex, and sent out to the SIBYLS beamline for analysis.

Molecular Dynamics Simulations

Three parallel ASB9(1–252)-CK (CK in complex with ASB9) and ASB9(23–252)-CK simulations starting from the ZDOCK-1 and ZDOCK-2 poses were performed using the AMBER ff99SB force field (Hornak et al., 2006) in a TIP3P (Jorgensen et al., 1983) water box. The orthorhombic water box around the heterotrimer was built with the dimensions listed in Table 1, with periodic boundary conditions prepared in Maestro (Schrodinger Release 2015-4). The N and C termini were capped with acetyl and methyl groups, respectively. The total system included a low (0.15 M) concentration of sodium chloride in water and 36 Na^+ to neutralize. Analysis of the protonation states of the system at pH 7.0 was performed with PROPKA (Li et al., 2005). Simulations were performed with the NVIDIA GK110 (GeForce GTX Titan) GPU using the CUDA version of PMEMD in AMBER12 (Hornak et al., 2006; Case et al., 2005). The simulations were relaxed using 119,000 steps of energy minimization. For 3,000 steps, only hydrogens were minimized with 100.0 kcal/Å^2 × mol of restraint, 500 steps of steepest descent algorithm-based minimization, and 2,500 steps of conjugate-gradient algorithm-based minimization. Then, all hydrogens and waters were minimized in 6,000 steps. All waters, hydrogens, and protein side chains were minimized during the next 40,000 steps of minimization. During the last 70,000 steps of minimization all atoms were minimized. Following minimization, the system was heated to 303.15 K in 4,000 steps with a 4.0 kcal/Å^2 × mol restraint on the protein backbone atoms. Three equilibration steps were then performed, each totaling 250 ps of simulation in which all of the backbone atoms except ASB9(1–35) were restrained with 3.0 kcal/Å^2 × mol, 2.0 kcal/Å^2 × mol, and finally 1.0 kcal/Å^2 × mol of force. Long-range electrostatics were calculated using the particle mesh Ewald method (Case et al., 2005), with a 10 Å cutoff and a 1.2 Å grid spacing. Three simulations initiated from the CK homodimer (PDB: 3DRB), and five simulations initiated from the Maestro (Schrodinger Release 2015-4)-built ASB9(19–35) were performed with the same force field, water model, salt concentration, and simulation conditions as described for the ASB9-CK simulations above. The net charge of the solute was counter-balanced with the addition of 30 Na^+ ions for the CK homo-dimer and no counterbalance was needed for the ASB9(19–35) system. The orthorhombic boundary conditions for the CK homodimer were 89.57 × 87.0 × 117.2 Å, with a total of 86,108 atoms in the system. The cubic boundary conditions for ASB9(19–35) were 73.4 × 73.5 × 73.4 Å per side, with a total of 32,623 atoms in the system. The CK homo-dimer structure was taken from the PDB: 3DRB (Bong et al., 2008). The ligands were removed from the structure since all of the experiments were performed on apo CK homo-dimer. Residues 1–5 were not resolved in the crystal structure and were not built into the CK homo-dimer since the location of these residues varies among CK analogs (Lahiri et al., 2002; Rao et al., 1998; Shen et al., 2001; Tisi et al., 2001) and, while important for interactions with other proteins they are not important for CK's interaction with ASB9.

MD Analysis

Backbone rmsd-based conformational clustering was performed using the gromos algorithm (Daura et al., 1999) with GROMACS 4.0.5 (Hess et al., 2008) software. For the ASB9-CK systems, atomic coordinates were extracted at every 6 ps and aligned to the backbone atoms of the CK monomer with ASB9's D32 docked into its active site. Various rmsd cutoffs were used for the entire backbone (1.25, 1.55, 1.7, and 2.1 Å), with 2.1 Å used for consistency as this produced the best clustering for the 25 ns simulations. After calculating the rmsd distance matrix, the gromos algorithm clusters the structures and determines the cluster centroid as previously described (Demir et al., 2011).

The number of hydrogen bonds formed between each residue's backbone amide and oxygen water in the simulation was performed using in-house python scripts and cpptraj (Roe and Cheatham, 2013). The standard cpptraj settings for angle and distance definitions for a hydrogen bond were used. Each frame of the trajectory was considered and the total number of hydrogen bond contacts was tallied and reported as a percentage of the total frames in the simulation.

Robetta Analysis

Robetta analysis (Kim et al., 2004) was performed on each cluster centroid of the above-described 2.1 Å rmsd-based clustering without specifying any mutation list.

Mutational Analysis of Robetta Results

CKB, cloned into a pET11a vector, and N-terminal His-tagged ASB9(35–252) mutants, cloned into a pHis8 vector, were co-expressed in *Escherichia coli*

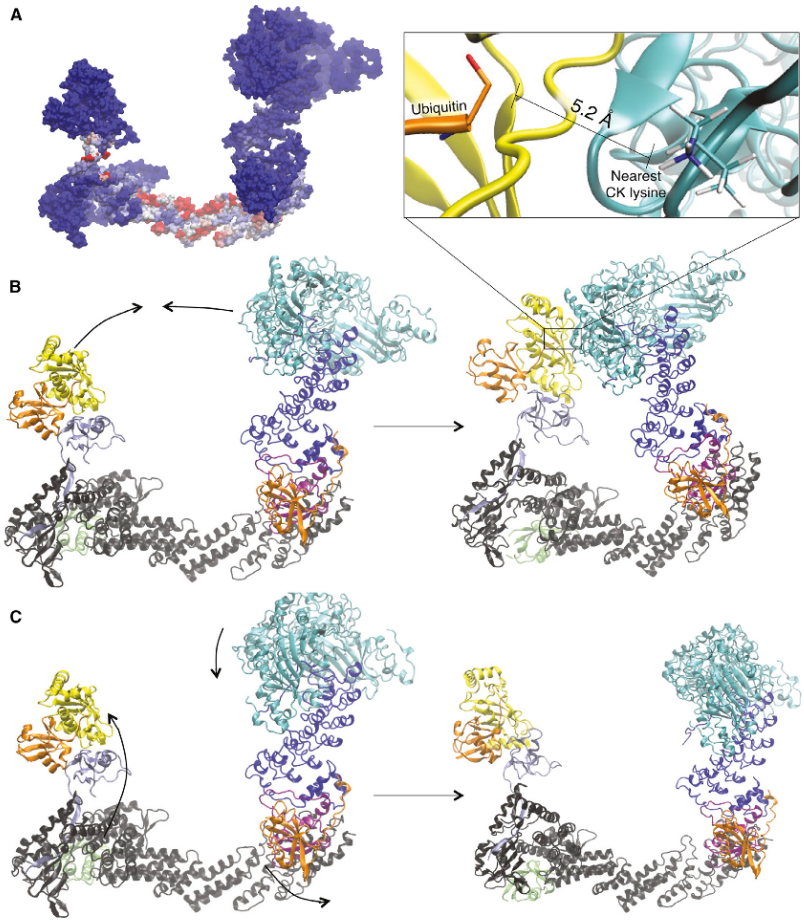


Figure 7. The CK-Targeting ECS-type E3 Ligase Motion Revealed by the Elastic Network Model

(A) Surface representation of E3 ligase colored based on the mobility of residues in the elastic network model.

(B) The lowest-frequency normal mode of E3 ligase model, built with the ZDOCK-1 pose for the ASB9-CK complex, brings active thioester of E2-ubiquitin within 5.2 Å of a CK lysine.

(C) The normal mode of the E3 ligase model, built with the ZDOCK-2 pose for the ASB9-CK complex, brings the active thioester of E2-ubiquitin closest to a CK lysine, within ~50 Å.

BL21(DE3) cells. ASB9 mutants W37A, I45A, I45R, L50A, L78A, L78R, and W102A were generated by site-directed mutagenesis from the ASB9(35–252) construct. Cells were grown at 37°C to an optical density of 0.7, and induced with 0.4 mM isopropyl β-D-1-thiogalactopyranoside overnight at 18°C in 50 ml Luria-Bertani medium. The cell pellet was resuspended in 2 ml of buffer A (50 mM Tris-HCl [pH 8.0], 50 mM NaCl, 20 mM imidazole, 1 mM DTT, 0.5 mM phenylmethanesulfonylfluoride), sonicated on ice, and centrifuged for 20 min at 4°C and 14,000 rpm. The supernatant was loaded onto a 1 ml HisTrap HP (GE Healthcare) column at 4°C equilibrated with buffer A, then washed with 2 ml of buffer A and eluted with 3 ml of buffer A supplemented with 500 mM imidazole. Then, 20 μl aliquots of each purification step (supernatant, SN; flow-through, FT; wash, W; elution, E) were analyzed by SDS-PAGE with 15% polyacrylamide gels.

Elastic Network Model

The CK Cullin-5 E3 ubiquitin ligase complex was built as previously described (Thomas et al., 2013), with the ASB9(19–252)-CK complex of the equilibrated ZDOCK-1 pose, the equilibrated ZDOCK-2 pose, and cluster

centroids from both simulations. The various E3 ligases were uploaded to the AD-ENM (Elastic Network Model) server (Zheng and Brooks, 2005). For the output specification, ten of the lowest modes were output, with an amplitude of displacement per mode of 24 Å, and 30 frames were generated. For the parameters settings, a 0 Å distance for coarse graining was chosen as well as a 15 Å distance cutoff for elastic interaction between carbon atoms.

SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures, four tables, and one movie and can be found with this article online at http://dx.doi.org/10.1016/j.str.2016.05.016.

AUTHOR CONTRIBUTIONS

J.S., R.D.M., C.S., R.E.A., and E.A.K. collaborated in planning of the project. J.S. executed the majority of the experiments and all the computational work, performed the analysis, and wrote the paper with E.A.K. R.D.M. provided mentorship and in-house scripts, and assisted in building the computational protein systems. J.P. purified the protein for SAXS analysis. C.S. and J.R. performed the mutagenesis experiments. R.E.A. and E.A.K. were leading investigators and edited the paper.

ACKNOWLEDGMENTS

This work was funded in part by the Director's New Innovator Award Program NIH DP2 OD007237 and an NSF XSEDE award RAC CHE060073N to R.E.A. Funding and support from the National Biomedical Computation Resource is provided through NIH P41 GM103426. J.M.S. was supported by NIH Molecular Biophysics Training Grant T32 GM008326. Additional resources from the Keck II Computing and Visualization facility are gratefully acknowledged. R.E.A. is a co-founder of Actavalon, Inc.

Received: February 24, 2016
Revised: April 25, 2016
Accepted: May 12, 2016
Published: July 7, 2016

REFERENCES

Balasubramaniam, D., Schiffer, J., Parnell, J., Mir, S.P., Amaro, R.E., and Komives, E.A. (2015). How the ankyrin and SOCS box protein, ASB9, binds to creatine kinase. Biochemistry *54*, 1673–1680. http://pubs.acs.org/doi/abs/10.1021/bi501420n.

Bong, S.M., Moon, J.H., Nam, K.H., Lee, K.S., Chi, Y.M., and Hwang, K.Y. (2008). Structural studies of human brain-type creatine kinase complexed with the ADP-Mg2+-NO3–creatine transition-state analogue complex. FEBS Lett. *582*, 3959–3965. http://dx.doi.org/10.1016/j.febslet.2008.10.039.

Bush, D.J., Kirillova, O., Clark, S.A., Davulcu, O., Fabiola, F., Xie, Q., Somasundaram, T., Ellington, W.R., and Chapman, M.S. (2011). The structure of lombricine kinase: implications for phosphagen kinase conformational changes. J. Biol. Chem. *286*, 9338–9350.

Case, D.A., Cheatham, T.E., 3rd, Darden, T., Gohlke, H., Luo, R., Merz, K.M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R.J. (2005). The Amber biomolecular simulation programs. J. Comput. Chem. *26*, 1668–1688.

Chakravarty, D., Guharoy, M., and Robert, C.H. (2013). Reassessing buried surface areas in protein – protein complexes. Protein Sci. *22*, 1453–1457.

Chothia, C., and Janin, J. (1975). Principles of protein-protein recognition. Nature *256*, 705–708.

Comeau, S.R., Gatchell, D.W., Vajda, S., and Camacho, C.J. (2004). ClusPro: a fully automated algorithm for protein-protein docking. Nucleic Acids Res. *32*, 96–99.

Daura, X., Gademann, K., Jaun, B., Seebach, D., van Gunsteren, W.F., and Mark, A.E. (1999). Peptide folding: when simulation meets experiment. Angewandte Chem. Int. Ed. *38*, 236–240. http://doi.wiley.com/10.1002/(SICI)1521-3773(19990115)38:1/2<236::AID-ANIE236>3.0.CO;2-M.

Debrincat, M.A., Zhang, J.G., Willson, T.A., Silke, J., Connolly, L.M., Simpson, R.J., Alexander, W.S., Nicola, N.A., Kile, B.T., and Hilton, D.J. (2007). Ankyrin repeat and suppressors of cytokine signaling box protein Asb-9 targets creatine kinase B for degradation. J. Biol. Chem. *282*, 4728–4737.

Demir, Ö., Baronio, R., Salehi, F., Wassman, C.D., Hall, L., Hatfield, G.W., Chamberlin, R., Kaiser, P., Lathrop, R.H., and Amaro, R.E. (2011). Ensemble-based computational approach discriminates functional activity of p53 cancer and rescue mutants. PLoS Comput. Biol. *7*, e1002238.

Deshaies, R., and Joazeiro, C. (2009). RING domain E3 ubiquitin ligases. Annu. Rev. Biochem. *78*, 399–434.

Duda, D.M., Borg, L.A., Scott, D.C., Hunt, H.W., Hammel, M., and Schulman, B.A. (2008). Structural insights into NEDD8 activation of cullin-RING Ligases: conformational control of conjugation. Cell *134*, 995–1006.

Fei, X., Gu, X., Fan, S., Yang, Z., Li, F., Zhang, C., Gong, W., Mao, Y., and Ji, C. (2012). Crystal structure of Human ASB9-2 and substrate-recognition of CKB. Protein J. *31*, 275–284.

Henriques, J., Cragnell, C., and Skepo, M. (2015). Molecular dynamics simulations of intrinsically disordered proteins: force field evaluation and comparison with experiment. J. Chem. Theor. Comput. *11*, 3420–3431.

Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E. (2008). GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. J. Chem. Theor. Comput. *4*, 435–447.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of multiple amber force fields and development of improved protein backbone parameters. Proteins *65*, 572–581.

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., and Klein, M.L. (1983). Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. *79*, 926. http://scitation.aip.org/content/aip/journal/jcp/79/2/10.1063/1.445869.

Kerscher, O., Felberbaum, R., and Hochstrasser, M. (2006). Modification of proteins by ubiquitin and ubiquitin-like proteins. Annu. Rev. Cell Dev. Biol. *22*, 159–180.

Kim, D.E., Chivian, D., and Baker, D. (2004). Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res. *32*, 526–531.

Kohroki, J., Nishiyama, T., Nakamura, T., and Masuho, Y. (2005). ASB proteins interact with Cullin5 and Rbx2 to form E3 ubiquitin ligase complexes. FEBS Lett. *579*, 6796–6802. http://linkinghub.elsevier.com/retrieve/pii/S0014579305013761.

Kortemme, T., and Baker, D. (2002). A simple physical model for binding energy hot spots in protein-protein complexes. Proc. Natl. Acad. Sci. USA *99*, 14116–14121.

Kortemme, T., Kim, D.E., and Baker, D. (2004). Computational alanine scanning of protein-protein interfaces. Sci. STKE *2004*, l2. http://www.ncbi.nlm.nih.gov/pubmed/14872095.

Kozakov, D., Beglov, D., Bohnuud, T., Mottarella, S.E., Xia, B., Hall, D.R., and Vajda, S. (2013). How good is automated protein docking? Proteins *81*, 2159–2166.

Kreuzer, S.M., and Elber, R. (2013). Catch bond-like kinetics of helix cracking: Network analysis by molecular dynamics and Milestoning. J. Chem. Phys. *139*, 121902, http://dx.doi.org/10.1063/1.4811366.

Kwon, S., Kim, D., Rhee, J.W., Park, J.A., Kim, D.W., Kim, D.S., Lee, Y., and Kwon, H.J. (2010). ASB9 interacts with ubiquitous mitochondrial creatine kinase and inhibits mitochondrial function. BMC Biol. *8*, 23.

Lahiri, S.D., Wang, P.F., Babbitt, P.C., McLeish, M.J., Kenyon, G.L., and Allen, K.N. (2002). The 2.1 Å structure of Torpedo californica creatine kinase complexed with the ADP-Mg2+-NO3 –creatine transition-state analogue complex. Biochemistry *41*, 13861–13867.

Li, H., Robertson, A.D., and Jensen, J.H. (2005). Very fast empirical prediction and rationalization of protein pK a values. Proteins *61*, 704–721.

Lyskov, S., and Gray, J.J. (2008). The RosettaDock server for local protein-protein docking. Nucleic Acids Res. *36*, 233–238.

Miyashita, O., Onuchic, J.N., and Wolynes, P.G. (2003). Nonlinear elasticity, protein quakes, and the energy landscapes of functional transitions in proteins.

Proc. Natl. Acad. Sci. USA *100*, 12570–12575, http://dx.doi.org/10.1073/pnas.2135471100.

Muniz, J.R., Guo, K., Kershaw, N.J., Ayinampudi, V., von Delft, F., Babon, J.J., and Bullock, A.N. (2013). Molecular architecture of the ankyrin SOCS box family of Cul5-dependent E3 ubiquitin ligases. J. Mol. Biol. *425*, 3166–3177.

Petroski, M.D., and Deshaies, R.J. (2005). Function and regulation of cullin-RING ubiquitin ligases. Nat. Rev. Mol. Cell Biol. *6*, 9–20.

Pierce, B.G., Wiehe, K., Hwang, H., Kim, B.H., Vreven, T., and Weng, Z. (2014). ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. Bioinformatics *30*, 1771–1773.

Rao, J.K., Bujacz, G., and Wlodawer, A. (1998). Crystal structure of rabbit muscle creatine kinase. FEBS Lett. *439*, 133–137. http://www.ncbi.nlm.nih.gov/pubmed/9849893.

Roe, D.R., and Cheatham, T.E., III (2013). PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. J. Chem. Theor. Comput. *9*, 3084–3095. http://www.chemeurope.com/en/publications/580240/ptraj-and-cpptraj-software-for-processing-and-analysis-of-molecular-dynamics-trajectory-data.html.

Schneidman-Duhovny, D., Hammel, M., and Sali, A. (2011). Macromolecular docking restrained by a small angle X-ray scattering profile. J. Struct. Biol. *173*, 461–471. http://dx.doi.org/10.1016/j.jsb.2010.09.023.

Schneidman-Duhovny, D., Hammel, M., Tainer, J.A., and Sali, A. (2013). Accurate SAXS profile computation and its assessment by contrast variation experiments. Biophys. J. *105*, 962–974. http://dx.doi.org/10.1016/j.bpj.2013.07.020.

Shen, Y.Q., Tang, L., Zhou, H.M., and Lin, Z.J. (2001). Structure of human muscle creatine kinase. Acta Crystallogr. D Biol. Crystallogr. *57*, 1196–1200.

Thomas, J.C., Matak-Vinkovic, D., Van Molle, I., and Ciulli, A. (2013). Multimeric complexes among ankyrin-repeat and SOCS-box protein 9 (ASB9), ElonginBC, and cullin 5: insights into the structure and assembly of ECS-type Cullin-RING E3 ubiquitin ligases. Biochemistry *52*, 5236–5246.

Tisi, D., Bax, B., and Loew, A. (2001). The three-dimensional structure of cytosolic bovine retinal creatine kinase. Acta Crystallogr. D Biol. Crystallogr. *57*, 187–193.

Wang, Y., Wang, S., Gao, Y.S., Chen, Z., Zhou, H.M., and Yan, Y.B. (2011). Dissimilarity in the folding of human cytosolic creatine kinase isoenzymes. PLoS One *6*, 28–34.

Zhao, Y., and Sun, Y. (2013). Cullin-RING Ligases as attractive anti-cancer targets. Curr. Pharm. Des. *19*, 3215–3225. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4034125&tool=pmcentrez&rendertype=abstract.

Zheng, W., and Brooks, B.R. (2005). Normal modes based prediction of protein conformational changes guided by distance constraints. Biophys. J. *88*, 3109–3117.

Zheng, W., and Doniach, S. (2003). A comparative study of motor-protein motions by using a simple elastic-network model. Proc. Natl. Acad. Sci. USA *100*, 13253–13258.

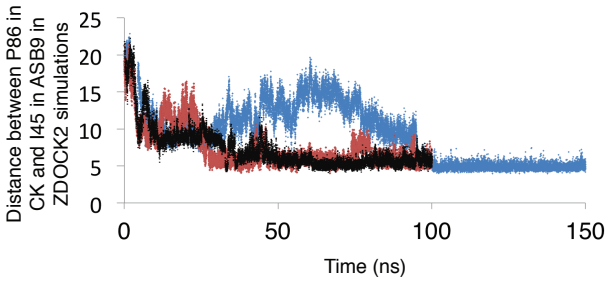
Structure, Volume *24*

Supplemental Information

Model of the Ankyrin and SOCS Box Protein, ASB9, E3 Ligase Reveals a Mechanism for Dynamic Ubiquitin Transfer

Jamie M. Schiffer, Robert D. Malmstrom, Jonathan Parnell, Cesar Ramirez-Sarmiento, Javiera Reyes, Rommie E. Amaro, and Elizabeth A. Komives

Supplemental Material



Supplementary Figure 1, related to Figure 5 | The distance between I45 in ASB9 and P86 in CK over three separate simulations starting from the ZDOCK2 structure. These residues interact very tightly in the ZDOCK1 pose.

Movie 1, related to Figure 5 | The ZDOCK2 pose for ASB9 is shown in purple, the two creatine kinase dimers are depicted in cyan with the active site loops in yellow and the HDX-MS residues is red. Within a few nanoseconds, the interfaces of the ZDOCK2 pose moves towards the ZDOCK1 pose (orange).

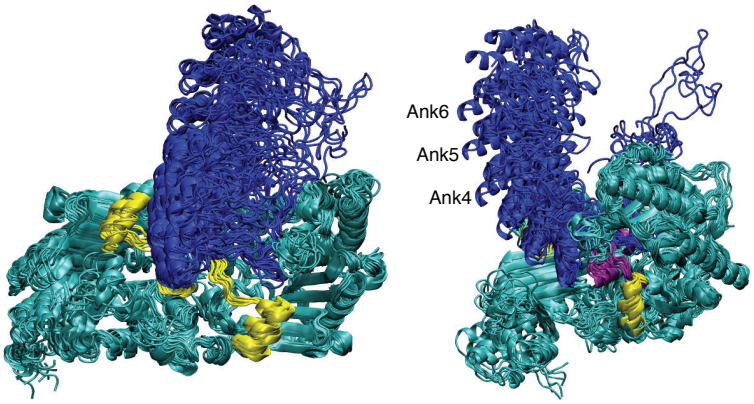


Figure S2, related to Figure 7 | Dynamic motion of ankyrin repeats 4-6 in simulations of ASB9-CK shown with cluster centroids from ASB9(1-252)-CK simulations.

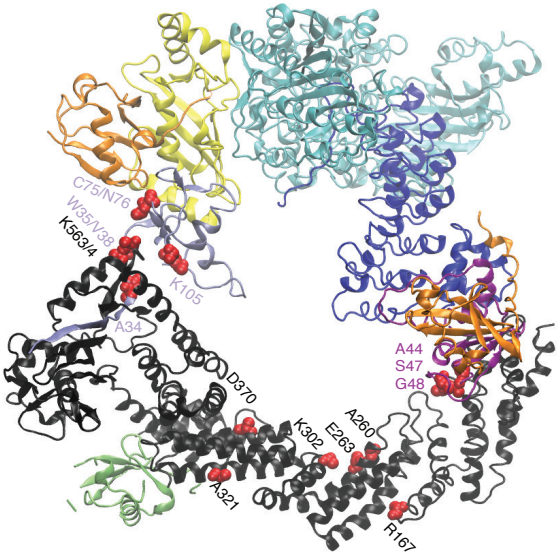


Figure S3 related to Figure 7 | Key fluctuating residues in the slow motions of the full E3 Ligase model that permit ubiquitin (orange) transfer from E2 (yellow) to creatine kinase (cyan).

	CKBB (open/closed)					CKBB (open/open)				
	Docking Rank and Score	χ^2 1-252	χ^2 19-252	R_g(Å)	Binding to CK cleft with D32 in active site?	Docking Rank and Score	χ^2 1-252	χ^2 19-252	R_g(Å)	Binding to CK cleft with D32 in active site?
(ZDOCK)	1 - 1405	2.40	7.78	31.61	yes	1 - 1778	1.21	1.79	32.38	yes
	2 - 1370	2.95	11.70	32.39	yes	2 - 1425	1.21	7.34	31.9	yes
	3 - 1290	4.08	63.68	31.81	NO	3 - 1415	1.58	5.86	32.68	yes
	4 - 1191	3.92	23.91	31.17	NO	4 - 1409	17.22	140.89	32.37	NO
	5 - 1168	2.85	11.09	32.39	yes	5 - 1387	6.86	84.09	33.65	NO
	6 - 1163	2.53	9.42	31.54	yes	6 - 1361	4.49	53.14	31.94	NO
	7 - 1137	5.43	10.82	31.22	NO	7 - 1350	1.69	6.81	32.44	yes
	8 - 1123	12.74	59.90	30.57	NO	8 - 1317	1.66	5.24	32.75	yes
	9 - 1114	2.95	25.40	31.86	yes	9 - 1274	2.66	4.16	31.72	yes
	10 - 1113	5.95	49.98	35.09	NO	10 - 1230	1.77	6.76	32.62	yes

Table S1, related to Figure 3: This table summarizes docking scores, χ^2 values to the ASB9(1-252)-CK complex and ASB9(19-252)-CK complex, the Debye scattering calculated radius of gyration (R_g(Å)), and whether ASB9 docks into the cleft between the two CK monomers with the residue D32 of ASB9 in the active site of one CK monomer for each of the four docking protocols that have been previously described (*12*) with the ZDOCK 3.0.2 algorithm (*15*). The χ^2 values and the R_g values were determined with the FOXS SAXS server (*20*).

	Docking Rank and Score	χ^2 19-252	R_g(Å)	Binding to CK cleft with D32 in active site?
ASB9 Elongated N-Terminus	-936.852	20.3401	32.67	yes
	-936.102	121.4404	35.3	NO
	-935.34	129.2769	36.78	NO
	-935.254	113.6356	36.78	NO
	-935.226	68.3929	37.59	NO
	-934.934	95.4529	36.07	NO
	-934.835	163.0729	36.93	NO
	-934.712	134.7921	36.61	NO
	-934.664	124.5456	37.63	NO
	-934.662	64.9636	37.2	NO

Table S2, related to Figure 3: This table summarizes docking scores, χ^2 values to the ASB9(1-252)-CK complex and ASB9(19-252)-CK complex, the Debye scattering calculated radius of gyration (R_g(Å)), and whether ASB9 docks into the cleft between the two CK monomers with the residue D32 of ASB9 in the active site of one CK monomer for ASB9 with a compact and elongated N-terminus docked into the CK open/open dimer with the RosettaDOCK algorithm (*16*). Both docking protocols were initiated from the top ranked docked poses from the ZDOCK protocol above. The χ^2 values and the R_g values were determined with the FOXS SAXS server (*20*).

	Energy 19-252	χ2 19-252	R_g(Å)	Binding to CK cleft with D32 in active site?
FOXSDock	-1962.2	2.65	32.88	yes
	-1864.1	2.09	32.51	yes
	-1802.1	4.08	32.56	yes
	-1739.7	4.71	31.69	yes
	-1687.6	5.00	32.33	yes
	-1784.7	7.39	32.41	yes
	-1555.9	4.01	32.41	yes
	-1504.5	4.07	32.49	no - D32
	-1407.8	2.63	31.89	no - D32
	-1499.7	5.18	32.4	yes

Table S3, related to Figure 3: This table summarizes docking energy scores, χ^2 values to the ASB9(1-252)-CK complex and ASB9(19-252)-CK complex, the Debye scattering calculated radius of gyration (R_g(Å)), and whether ASB9 docks into the cleft between the two CK monomers with the residue D32 of ASB9 in the active site of one CK monomer for ASB9 with a elongated N-terminus docked into the CK open/open dimer with the FOXSdock algorithm (*16*).

	χ2 19-252	R_g(Å)	Binding to CK cleft with D32 in active site?
ClusPro2.0	18.49	30.97	yes
	4.33	32.09	yes
	10.43	31.18	yes
	23.91	31.63	NO
	36.60	30.86	yes
	4.04	32	yes
	3.31	31.58	yes
	13.40	30.99	yes
	47.75	31.27	yes
	62.88	32.55	NO

Table S4, related to Figure 3: This table summarizes the χ^2 values to the ASB9(1-252)-CK complex and ASB9(19-252)-CK complex, the Debye scattering calculated radius of gyration (R_g(Å)), and whether ASB9 docks into the cleft between the two CK monomers with the residue D32 of ASB9 in the active site of one CK monomer for ASB9 with a elongated N-terminus docked into the CK open/open dimer with the ClusPro2.0 algorithm (*16*).

ACKNOWLEDGEMENTS

Chapter 5, in full, is a reprint of the material as it appears in Model of the Ankryin and SOCS Box Protein, ASB9, E3 Ligase Reveals a Mechanism for Dynamic Ubiquitin Transfer 2016. Schiffer, Jamie. M., Malmstrom, R. D., Parnell, J., Ramirez-Sarmiento, C., Reyes, J., Amaro, R. E., Komives, E. A., Structure, 2016. The dissertation author was the primary investigator and author of this material.

CHAPTER 6

Uncovering Dynamic Motions at a Plausible NFκB Signalosome Interface

Introduction

The NFκB pathway

The Nuclear factor κB (NF-κB) pathway is central to the regulation of gene expression in cells, especially in pathways responsible for inflammatory response, proliferation, and differentiation. Protein phosphorylation events propagate the signals in this pathway from cell surface receptors to nuclear transcription. NF-κB proteins are the transcription factors of this pathway and are inhibited by inhibitors of κB kinases (IκBs). These IκBs are in turn regulated through an array of other kinases, the most important of which is the focus of this research.

There are two major channels through the NF-κB pathway: the canonical and non-canonical pathways. The canonical pathway depends on the interaction between the NF-κB essential modulator (NEMO) and the inhibitor of κB kinase (IKK) for propagation of the signal from IKK to IκBs. The non-canonical pathway depends on phosphorylation by the NF-κB interacting kinase (NIK). Here we will focus our attention on the canonical pathway and the atomic-level details associated with the IKK-NEMO interaction.[1,2]

The canonical signaling pathway arises due to inflammatory cytokines, radiation, stress signals and pathogenic assaults (Figure 1). Misregulation of this pathway can lead to disease states including cancer, chronic inflammation, and heart disease.[1] In the canonical signaling pathway, IKK complex phosphorylates IκBs. IκBs are ankyrin repeat proteins with an intrinsically disordered C-termini.

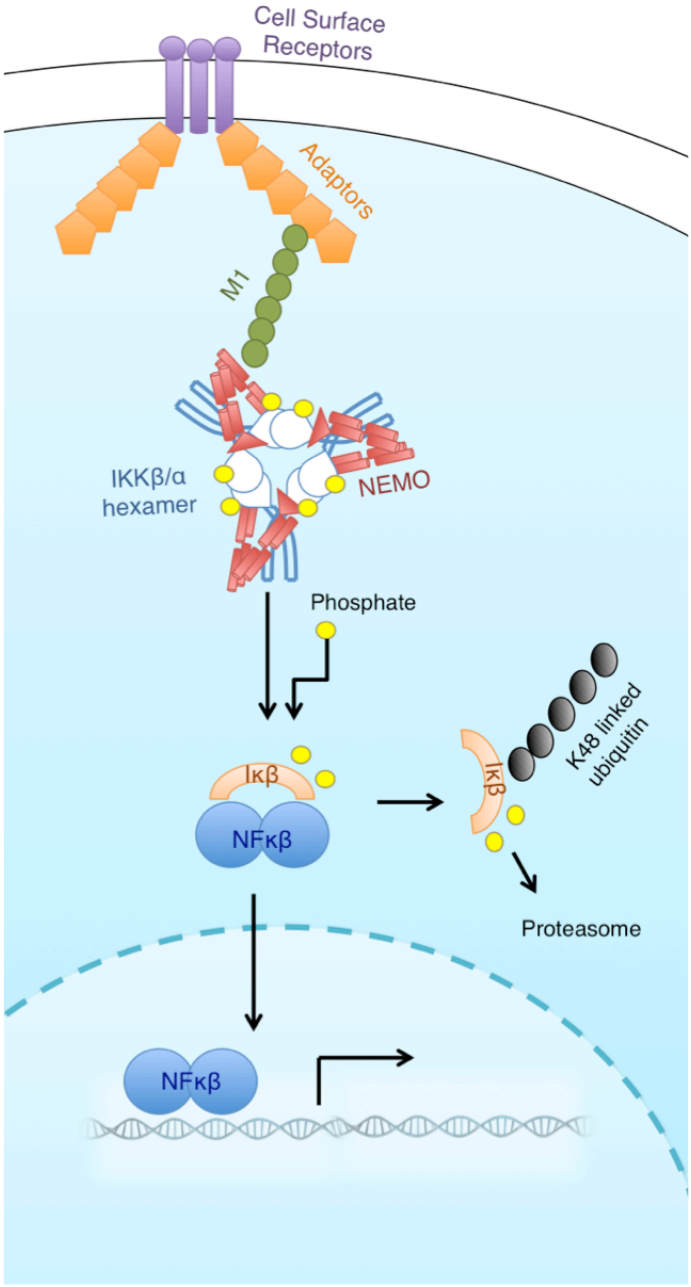


Figure 6.1: The NFκB pathway connects signals at the cell surface to the transcription factor NFκB. The central signaling hub is the IKK1/2 or IKKα/β signalasome complex (blue). The signals are relayed from one kinase to another by phosphorylation (yellow).

Phosphorylation of IκBs strips it from NF-κBs (Figure 1).[3,4] Once dissociated, NF-κB translocates to the nucleus, where it mediates transcription of genes in a context-dependent manner. Thus, the IKK complex is a central regulatory complex in the NF-κB pathway.

Well-resolved crystal structures the of α-helical scaffold/dimerization domain (SDD), ubiquitin-like domain (ULD) and kinase domain (KD) of IKKβ provide 2.8 Å resolution of the quaternary structure of the active IKKβ dimers.[5–7] These structures also reveal the importance of phosphorylation on the activation of the kinase loops, a mechanism that is reminiscent of the phosphorylation dependent activity of the IKK-related kinase TBK-1.[6,8,9] However, these crystal structures are difficult to reconcile with initial size exclusion chromatography studies which identified a high molecular weigh kinase complex as responsible for IκB phosphorylation, approximately 700 kDa.[10] Thus what is the quaternary structure of IKK2 in the full signalasome and what other proteins are involved?

There are four solved structures of NEMO subdomain in complex with a variety of other proteins. One crystal structure consists of a dimer of the NEMO-binding domain (NBD) or IKK2 in complex with two molecules of residues 49 to 101 of NEMO.[11] NEMO also interacts with the viral protein vFLIP through NEMO residues 150 to 272.[12] Another NEMO structure includes residues 267 to 329 bound as a dimer to two molecules of diubiquitin.[11,13–15] This structure implicates NEMO in the localization of signalasomes to ubiquitin chains near cellular membranes.

The IKK signalasome complex is likely composed of the catalytic subunits IKKα, IKKβ, and the regulatory subunit NEMO or IKKγ. NEMO has been shown to act as an IκB scaffold, directly recruiting IκB to IKKβ in the fully competent signalasome and thus permitting stimulus-specific signal transduction through the ubiquitious signaling complex of IKKβ.[11] Studying these proteins alone as well as in complex has proven to be challenging due in large part to issues with purification of large quantities. However, the complex of IKKs/NEMO could account for the high molecular weight complex witnessed in initial biochemistry assays of the fully competent signalasome.

The stoichiometry, precise complex structure, and interfaces of the full signalosome (IKK complex), which contains IKKβ, IKKα, and NEMO in a 1:1:1 ratio, are still a matter of debate. Recent work from the Ghosh lab has just uncovered that IKKα dimers arrange in a trimeric arrangement, or essentially for a hexamer of IKKα, through cryo-electron microscopy at a 4.5 Å resolution (Polley *et al submitted)*. Here we aim to define the structure of the minimally competent IKK complex with the ultimate goal of providing key therapeutic insight for the treatments of arthritis, autoimmune disorders, and lymphoid malignancies. The new hexameric structure of IKKα and structures of the IKKβ and NEMO subdomains were used to create a model of the competent signalasome. Moreover, the dynamic composition and large size of this complex (~700 to 900 kDa) call for a novel and interdisciplinary approach in determining the stoichiometry, complex interfaces, and dynamic ensembles. Our integrative modeling approach combines information from crystallographic data, mutagenesis, and cryoEM data with computational modeling approaches like protein interface prediction and

protein-protein docking servers, anisotropic network models, and molecular dynamics simulations to uncover the dynamic motions of IKK2 on relevant timescales and produce a dynamic picture of experimentally validated interaction surfaces between IKK2 and NEMO.

Materials and Methods

Computational protein structure preparation

We used Schrödinger PrepWizard modules[16] to prepare the IKK2 structure for MD simulation. The IKK2 S117E/S181E crystal structures (PDB: 4E3C)[7] with chains AF, chains BC, and chains DE forming the three dimers. Residues S177E and S181E were mutated back to serines using Schrödinger's maestro. These systems were protonated at pH 7.4 with the pKa titratable residues determined using the Maestro integrated PROPKA.[17] While E16 was protonated in some of the crystal structures, for consistency we deprotonated E16 for all systems and all monomers. Missing residues and side chains were added with PRIME.[18] The total numbers of atoms in all three dimer systems were 21160 atoms, and an unperturbed charge of the unit of -10.00. The box dimensions for the AF, BC, and DE dimers were (141.87, 141.0, 112.6), (136.4, 146.2, 96.96), and (148.4, 141.9, and 75.51) respectively, demonstrating the differences in volume occupied by these three dimers.

The hexameric model of the IKK2/IKK1 signalasome were built starting from the IKK1 hexameric model from cryoelectron microscopy.[19] The IKK2 sequence was thread through the various monomers of the IKK1 hexamer using SWISS-MODEL server.[20] The various monomer structures had different scores for the IKK2 threaded sequences (Table 1). The IKK1 hexameric model as well as the various NEMO structures, described below in *Protein-Protein docking*, were the same way as the IKK2 dimers with the same use of PROPKA and PRIME at the same pH value.

Table 6.1: Scores from SWISS-MODEL for threading the IKK2 sequence through the IKK1 monomers solved in the IKK1 hexameric structure.

Protein-protein interaction prediction

	Chain A	Chain B	Chain C	Chain D	Chain E	Chain F
Qmean4	-4.14	-5.42	-4.26	-3.95	-4.14	-4.43
Qmean6	-3.07	-3.80	-3.07	-2.83	-3.07	-3.27
All_atom	-1.065	-1.35	-0.35	-0.64	-0.71	-0.78
Cbeta	-0.717	-0.84	0.007	0.10	-0.56	-0.96
Solvation	0.60	0.12	0.55	0.46	0.56	0.29
Torsion	-3.91	-5.03	-4.16	-3.84	-3.96	-4.12

Protein-protein interactions were predicted using the metaPPISP[21] and SPPIDER[22] servers on the maestro fixed structures of the dimer and hexameric IKK2 models. Results were converted using in-house python scripts into the B-factor columns of IKK2 PDB files for visualization with VMD. These predictions were also performed for every tenth frame of the fifty frame PDB file created as output from anisotropic network model (ANM) performed with ProDy (see below).[23]

Anisotropic Network Models

The ProDy[23] integrated VMD interface was used to determine long-range motions of the various IKK2 dimers from PDB 4E3C. All atoms were used in the anisotropic network model calculation, with 10 modes output, a distance cutoff of 15 Å and a force constant of 1.0. These same parameters were used to model the dynamics of the full IKK1/2 hexameric structure, except that all-atom models were not possible for this larger structure. Instead, all C_{alpha} , C_{beta}, and N atoms were used to construct the model.

APBS electrostatics

Surface electrostatics were calculated using the PDB2PQR[24] and the APBS[25,26] webservers. These calculations were performed on the maestro prepared IKK2 dimer chains (PDB: 4E3C), on the IKK1/2 signalasome complex modeled as described above, and on intermediate states witnessed through anisotropic network models. These calculations were performed at pH 7.4 using the PARSE force field, with an internal naming scheme, optimized hydrogen bonds, inserting white spaces between coordinates, and using PROPKA[17] to assign protonation states.

Protein-protein docking

The ClusPro2.0 server[27] was used to dock models of the dimeric and hexameric IKK2 models with various NEMO structures.[11,13–15] These structures included the complex between NEMO and the NEMO-binding domain (NBD) or IKK2 (PDB: 3BRV),[11] the complex between NEMO and diubiquitin (PDB: 2ZVO),[13] and just the NEMO CC2-LZ domain in complex with itself (PDB: 4BWN, 3FX0).[14] All of these NEMO structures were prepared with Maestro as described above.

Principal Component Analysis

The PCA method included in Gromacs was used to perform PCA analysis on the C_{alpha} and C_{beta} atoms of the IKK2 dimers from the implicit solvent simulations. The various structures obtained from ANM were projected onto the first and second principal components to show the space simulated.

Solvent accessible surface area

Solvent accessible surface area was calculated from crystal, cryoEM, ANM, and MD implicit solvent simulations for the entire protein, and for residues L456, V542, and Q545. These calculations were performed through VMD using the Tk console with the following script:

```
> set outfile [open my_protein.txt w]
> set nf [molinfo top get numframes]
> set all [atomselect top "resid 54 to 103"]
> for {set i 0} {$i<$nf} {incr i} {
> $all frame $i
> $all update
> set sasa [measure sasa 1.4 $all]
> puts $outfile "Frame $i, SASA $sasa"
> }
> close $outfile
```

Molecular dynamics minimization and equilibration

Implicit solvent molecular dynamics simulations were performed for all three dimers found in the PDB 4E3C. Structure from maestro were used to build prmtop and inpcrd files using Leap module of AmberTools17. The structures were built using ff14SBonlysc force field with mbondii3 bond lengths, and implicit generalized born solvent models,[28] as was prescribed in Nguyen *et al*.[29] The protein systems were then minimized and equilibrated in four stages: one heating from 0 K to 100 K for 4 ps followed by three 250 ps stages. The first 250 ps stage was heating from 100 K to 310 K with heavy atom positional restraints of 10 kcal/mol$Å^2$. The second 250 ps equilibration stage was at 310 K with heavy atom restraints reduced down to 1 kcal/mol$Å^2$. The final 250 ps equilibration stage was at 310 K with heavy atom restrained reduced down to 0.1 kcal/mol$Å^2$. Production runs were performed in triplicate using constant temperature,

with an implicit salt concentration of 150 mM using a modified generalized Born theory based on the Debye Hückel limiting law for ion screening of interactions, and a cutoff distance of 16 for the effective Born radii. Bonds involving hydrogens were constrained by the SHAKE algorithm[30] with a tolerance of 0.00001. Temperature was controlled with a Langevin thermostat[31] with a collision frequency of $\gamma = 1.0\ ps^{-1}$.

Rosetta Model building

The NEMO structures were built using the Robetta model generation based on the sequence of NEMO. Two domains were identified and models were built for both domains based on sequence homology.

Results

The experimentally measured molecular weight for the IKK signalasome suggests that the competent complex consists of a trimer of IKK1/2 dimers with three NEMO dimers. Based on this, here we integrate computational protein-protein docking of crystallized subdomain of IKK proteins and anisotropic and Gaussian network models with experimental validation of our models.

We used both dimer IKK2 models from crystallography and a model structure of the IKK1/IKK2 hexameric complex to understand potential NEMO binding sites, long-timescale dynamics, and the effect of dynamics on atomic level properties of the complexes. The model of the IKK1/IKK2 hexameric model was created from cryoEM structures of IKK1 hexamer (unpublished) and homology models of IKK2 sequence into this hexamer.

Two crystal structures of IKK2 dimer were used, one is the dimer structure of an inhibited IKK2 structure (PDB: 4KIK) and the other are three dimer structures of the IKK2 S177E/S181E constitutively active dimers (chain BC) (PDB 4E3C). These constitutively active dimers are unique from one other by the angle in the SDD domains of the dimer. Chains AF of this crystal structure have the largest angle between the SDD domains (open) and chains BC of this crystal structure have the smallest angle between the SDD domains (closed) and thus the closest distance between the kinase domains (KD) of the dimer.

Protein-Protein Docking to the IKK2 dimers and IKK1/IKK2 hexamer

We have used the MetaPPISP protein interface prediction server and the ClusPro2.0 docking server to propose a first-generation model (Figure 2) for the interaction between NEMO (PDB 3BRV, 4BWN, 3CL3) and the IKK2 dimers (PDB 4KIK). The MetaPPISP method predicted that residues on the IKK2 dimer protein surface in the kinase domain (KD) are likely areas of protein-protein interaction (Figure 2). As a positive control for the MetaPPISP method, these are the same residues in the trimeric interface of the IKK1 hexamer determined from cryoEM.

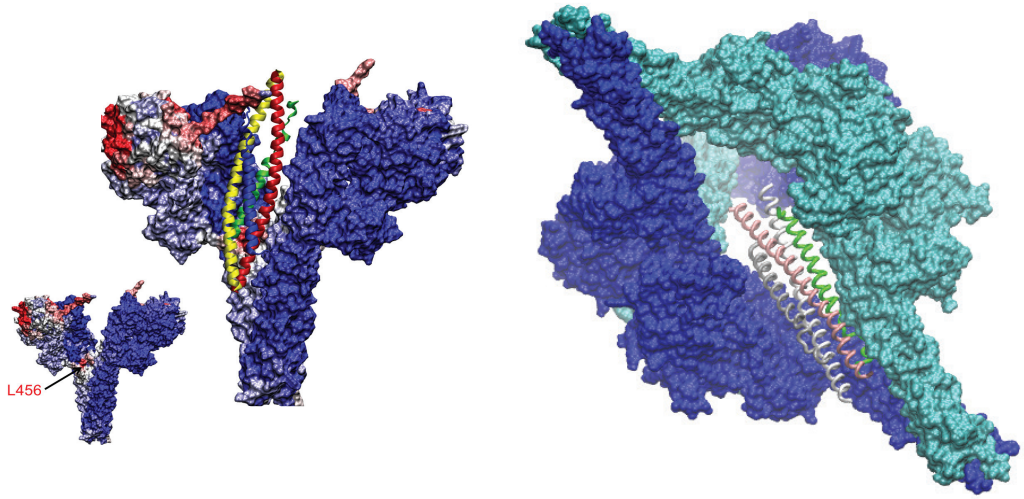


Figure 6.2: Protein-protein interface prediction and protein-protein docking point to a similar interface for the interaction between IKK2 and NEMO in both the IKK2 dimer (left) and IKK1(cyan)/IKK2(blue) hexamer (right). On the left, the IKK2 dimer is colored based on the score from the metaPPISP server, with red highlighting regions likely to be involved in a protein-protein interface from PDB:4KIK, and blue highlighting regions less likely to be in a protein-protein interface. The NEMO structure in complex with the NEMO binding domain (NBD) of IKK2 are shown in cartoon in both images.

The MetaPPISP and SSPIDER method, another protein-protein interface prediction server, both predicted that residues L456, V542, and Q545 of the SDD domain

(from here referred to as the LVQ patch) of the IKK2 dimer are likely to be involved in a protein-protein interaction (Figure 2). Other surfaces that were highlighted as likely to form protein-protein interactions included the active loop and areas around the active site, as well as the intrinsically disordered N-terminus of both dimers. This disordered N-terminus was modeled into the structures of IKK2 dimer using maestro.

Additionally, these same IKK2 residues from the SDD domain stated above are also predicted to interact with the N-terminal domain of NEMO (PDB 3BRV) in eight of the top ten binding protein-protein docking poses determined with the ClusPro2.0 server (Figure 2). These IKK2 residues were also predicted to be in the interface between IKK2 and the CC1-LZ domain of NEMO (PDB 4BWN) from docking with ClusPro2.0. However, the $IKK2_{NBD}$ sequence, which the N-terminus of NEMO interacts with, is within 40 amino acids of the end of the SDD domain and this tethering may lend itself to facilitate further protein-protein interactions with the SDD of IKK2 as predicted here.

NEMO has over 400 residues and has not been crystalized in full. Instead, small segments of NEMO have been crystallized, like NEMO bound to the NBD of IKK2 (PDB: 3BRV, $IKK2_{NBD}$) or NEMO bound to di-ubiquitin (PDB: 4BWN). Size exclusion chromatography results measure a 700 kDa complex for NEMO alone, suggesting that the shape of NEMO alone is likely an extended. Here we use the crystallized structures of NEMO bound to the NBD of IKK2 as well as an extended model of full length IKK2 built using the Rosetta model-building server.[32]

The docked structures of the NEMO-$IKK2_{NBD}$ with the IKK2 dimer structures solved in the presence of an inhibitor Staurosporine analog K252a (PDB: 4KIK),[33] were filtered based on where the known hexameric interface of IKK1 is, and the knowledge

that NEMO is known to recruit IκBs to the kinase domain of IKKs. The docking results from ClusPro2.0 are filtered into four major docking categories: balanced, electrostatics, hydrophobic, and van der Waals + electrostatics. Given than few hydrophobic residues are found to be solvent exposed in the NEMO structures, the hydrophobic method seemed less apropos, and indeed there was less consistency in the docking results from this method. We therefore inspected the balanced, electrostatic and van der Waals docking results. From these docking results, a small set of NEMO residues consistently docked at the complex interfaces between NEMO and the SDD domain of IKK2: R66, Q67, N69, Q70, R73, E74, E77, H81, and E88.

Docking performed on the NEMO-$IKK2_{NBD}$ with the IKK1/IKK2 hexameric model resulted in similar binding interfaces as seen for the electrostatic and van der Waals docking results from the IKK2 dimer. However, unlike when docking was performed on the IKK2 dimer, the docking results with the IKK1/IKK2 hexamer resulted in consistent structures for all four docking methods afforded through ClusPro2.0: balanced, hydrophobic, Vdw/Elec and electrostatic dockings all had the NEMO-$IKK2_{NBD}$ docking into the SDD domain of one dimer and then threading through the center of the IKK1/IKK2 hexamer (Figure 2B). Additionally, in this docking experiment, the NEMO-$IKK2_{NBD}$ preferentially docked into the IKK2 structures, with the NEMO residues noted above burying significant surface area in the SDD domain of an IKK molecule and interacting with the LVQ patch described above.

APBS of the IKK2 dimer models reveals different electrostatic characteristics based on KD distances.

Despite the fact that the dimers from the IKK2(S177E/S181E) contain the same residues, the electrostatics at the surfaces of each of the three dimer chains are drastically different (Figure 3). While the kinase domain of chains A and F are hydrophobic or positively charged (blue), the kinase domain of chains B and C are mostly positively charged in the active site and negatively charged on the outer face. The kinase domain of chains D and E are even more drastically different from the AF dimers, with very little positive charge in the active site but negative charged at the outer face.

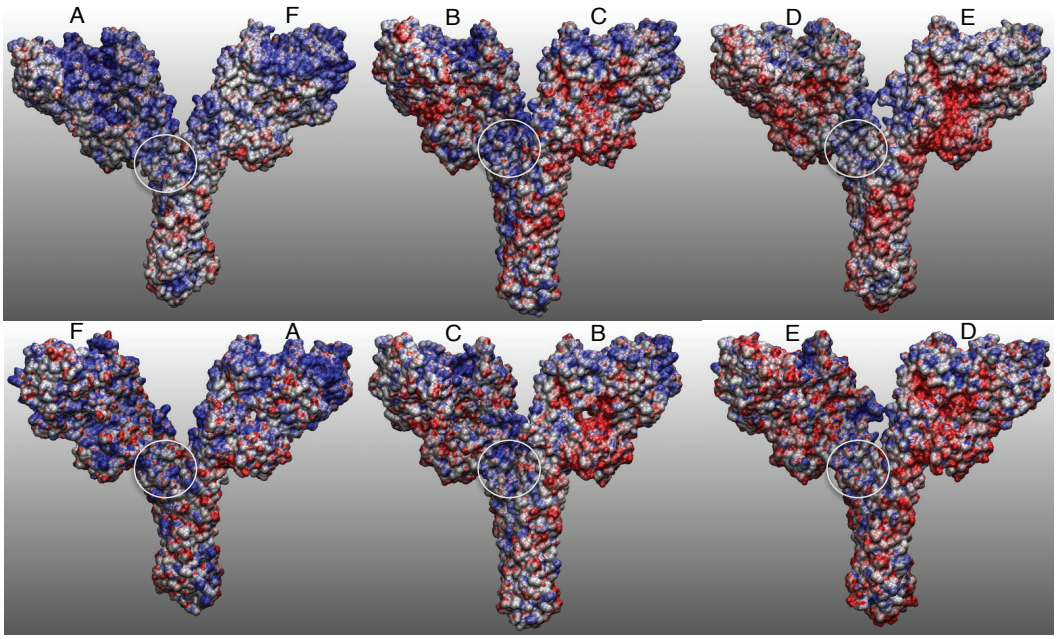


Figure 6.3: APBS electrostatics are mapped onto the dimer models from PDB: 4E3C with the AF chains having the greatest distances between the kinase domains and the BC chains having the least. The circle on each structure shows the plausible NEMO interaction interface and it drawing attention to the polarity of this interface in the BC chains in the center versus the more open IKK2 dimers.

Perhaps more interesting however in the context of our NEMO docking experiments, the interface of the BC dimer is highly positively charged along the SDD domain, especially where NEMO docks. This positively charged surface could explain

the cluster of three glutamic acid residues in NEMO often found in the docked interfaces, and the preference of the docked NEMO structures for one chain over the other.

Effects of dynamics on electrostatics, solvent accessible surface area, and protein-protein

To investigate the role of dynamics in NEMO and IKK2 binding, we performed anisotropic network model (ANM) analysis on the four IKK2 dimer models that docking was performed on. Anisotropic network models calculate the long-timescale dynamics of proteins by calculating springs between one atom of a protein and any atom that is within one cutoff distance. We then analyzed the resulting structures from ANM by measuring solvent accessible surface area (SASA) changes and performing metaPPISP measurements on the intermediate structures in the ANM pathways.

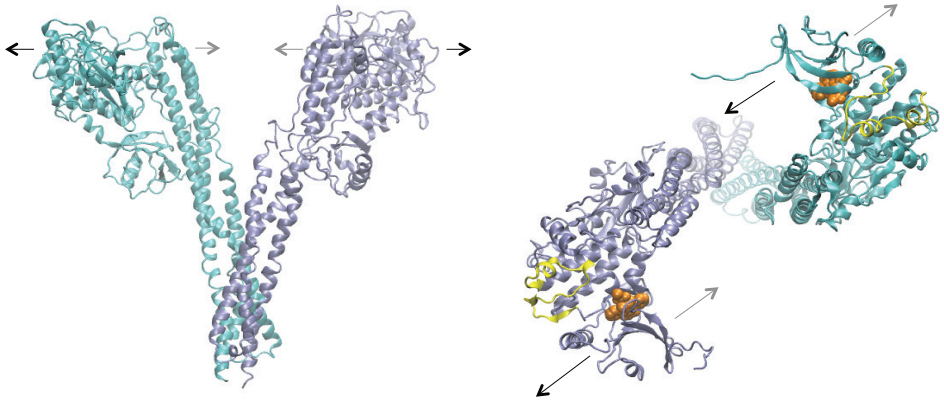


Figure 6.4: The ANM normal modes for the IKK2 dimers are shown with arrow looking down the dimer interface from two different angles. The first image on the left shows the first normal mode motions of flexing in the SDD domains for the AF and DE chains. The image on the right shows the first normal mode motions of rotation about the SDD domains of the KD domains viewed from the KD domain. The black arrays represent one direction of the normal modes and the grey represent the other direction. In orange on the right image is showing the inhibitor crystalized with IKK2 in the PDB 4KIK, a staurosporine analog K252a.[33] In yellow is shown the active site residues of the kinase domain.

The ANM results from the three IKK2 dimer constitutively active structures can be divided into two categories of motion. The first is the movement of the KD and ULD domain together and apart through flexing in the SDD domain (Figure 4A). The second is the rotation of the KD domain about an axis down the center of the SDD dimers with one KD rotating in and the other rotating out (Figure 4B). For the most open dimers, chains AF and DE, the first movement is the first normal mode, and the second movement is the second normal mode. For the more closed dimer, chain BC, the normal modes reverse, suggesting that on slower timescales, once the kinase domains have flexed close, they are more likely to stay closed. Thus we explored the changes in the surface area of the residues predicted to be involved in the protein-protein interface with NEMO to see how this slow motion effect NEMO binding.

The solvent accessible surface area in the anisotropic network modes for the three dimer crystal structures reveal distinct differences in the residues in the plausible protein-protein interfaces as well as the full structural stability. While the SASA of the LVQ pocket found in the docking interface changes increases in the ANMs of chains AF and DE, the SASA stays fairly constant with the ANM of chain BC (Figure 5). However, the SASA of the LVQ pocket is greatest for the BC chains over the AF chains and DE chains. Interestingly, despite the fact that the SASA for the LVQ pocket does not change over the first normal mode of BC chains, the overall SASA does change dramatically over the first normal mode. Additionally, the metaPPISP measurements demonstrated that the likely protein-protein interfaces change as the conformations of the protein changes (Figures S1).

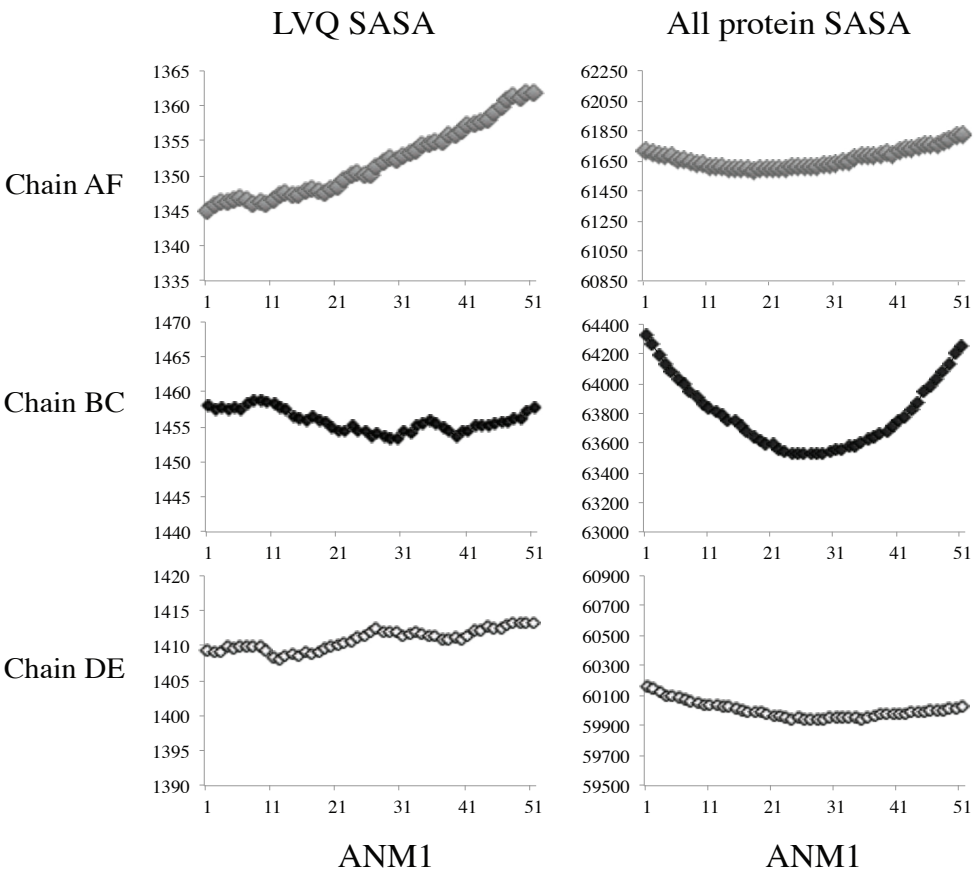


Figure 6.5: Changes in SASA over the first normal modes for each of the three dimer models from the S177E/S181E crystal structures (PDB: 4E3C). The y-axis values on each plot show the SASA for the LVQ pocket (left column) and the full protein (right column). The x-axis values are the frames of each normal mode, with frame 25 being the crystal structures.

Discussion

Here we have built models of the IKK2-NEMO complexes and of the IKK1/2 hexamer. Model of the IKK2-NEMO complexes were built using protein-protein docking and the IKK1/2 hexamers were built using Robetta homology model building. These results demonstrated that NEMO is hypothesized to interact better with IKK2 models with a smaller SDD dimer angle that brings the kinase domains closer together and increases the negative charge at the surface between the two monomers. Additionally, NEMO docks into the IKK2 monomers of the IKK1/IKK2 hexamers. Finally, structure-based dynamics of the IKK2 dimers demonstrate that the SASA of plausible NEMO interface changes with conformational changes. These results all demonstrate that conformational changes may be important in understanding how to best target the IKK signalasome for drug-based therapies for cancer and inflammation diseases. Future directions include mapping the details of conformational space for the IKK2 dimers and IKK1/IKK2 hexamers through implicit solvent molecular dynamics simulations.

References

(1) Hinz, M.; Scheidereit, C. The IκB Kinase Complex in NFκB Regulation and beyond. **2014**, 1–16.

(2) Schmid, J. A.; Birbach, A. IκB Kinase β (IKKβ/IKK2/IKBKB)-A Key Molecule in Signaling to the Transcription Factor NF-κB. *Cytokine Growth Factor Rev.* **2008**, *19* (2), 157–165.

(3) Ferreiro, D. U.; Komives, E. A. Molecular Mechanisms of System Control of NF-kappaB Signaling by IkappaBalpha. *Biochemistry* **2010**, *49*, 1560–1567.

(4) Bergqvist, S.; Croy, C. H.; Kjaergaard, M.; Huxford, T.; Ghosh, G.; Komives, E. A. Thermodynamics Reveal That Helix Four in the NLS of NF-kappaB p65 Anchors IkappaBalpha, Forming a Very Stable Complex. *J Mol Biol*. **2006**, *360*, 421–434.

(5) Kalia, M.; Kukol, A. Structure and Dynamics of the Kinase IKK- B – A Key Regulator of the NF-Kappa B Transcription Factor. *J. Struct. Biol*. **2011**, *176* (2), 133–142.

(6) Xu, G.; Lo, Y.-C.; Li, Q.; Napolitano, G.; Wu, X.; Jiang, X.; Dreano, M.; Karin, M.; Wu, H. Crystal Structure of Inhibitor of κB Kinase B. *Nature* **2011**, *472* (7343), 325–330.

(7) Polley, S.; Huang, D. Bin; Hauenstein, A. V.; Fusco, A. J.; Zhong, X.; Vu, D.; Schröfelbauer, B.; Kim, Y.; Hoffmann, A.; Verma, I. M.; et al. A Structural Basis for IκB Kinase 2 Activation Via Oligomerization-Dependent Trans Auto-Phosphorylation. *PLoS Biol*. **2013**, *11* (6).

(8) Ma, X.; Helgason, E.; Phung, Q. T.; Quan, C. L.; Iyer, R. S.; Lee, M. W.; Bowman, K. K.; Starovasnik, M. A.; Dueber, E. C. Molecular Basis of Tank-Binding Kinase 1 Activation by Transautophosphorylation. **2012**.

(9) Larabi, A.; Devos, J. M.; Ng, S.; Nanao, M. H.; Round, A.; Maniatis, T.; Panne, D. Article Crystal Structure and Mechanism of Activation of TANK-Binding Kinase 1. *CellReports* **2013**, *3* (3), 734–746.

(10) Chen, Z. J.; Parent, L.; Maniatis, T. Site-Specific Phosphorylation of I □ B ⎵ by a Novel Ubiquitination-Dependent Protein Kinase Activity. **1996**, *84*, 853–862.

(11) Rushe, M.; Silvian, L.; Bixler, S.; Chen, L. L.; Cheung, A.; Bowes, S.; Cuervo, H.; Berkowitz, S.; Zheng, T.; Guckian, K.; et al. Structure of a NEMO/IKK-Associating Domain Reveals Architecture of the Interaction Site. *Structure* **2008**, *16* (5), 798–808.

(12) Ageichik, A. V; Cronin, N.; Wallace, B.; Collins, M.; Boshoff, C.; Waksman, G.; Barrett, T. Article Crystal Structure of a vFlip-IKK G Complex : Insights into Viral Activation of the IKK Signalosome. **2008**, 620–631.

(13) Rahighi, S.; Ikeda, F.; Kawasaki, M.; Akutsu, M.; Suzuki, N.; Kato, R.; Kensche, T.; Uejima, T.; Bloor, S.; Komander, D.; et al. Specific Recognition of Linear Ubiquitin Chains by NEMO Is Important for NF- K B Activation. *Cell* **2009**, *136* (6), 1098–1109.

(14) Lo, Y.; Lin, S.; Rospigliosi, C. C.; Conze, D. B.; Wu, C.; Ashwell, J. D.; Eliezer, D.; Wu, H. Article Structural Basis for Recognition of Diubiquitins by NEMO. *Mol. Cell* **2009**, *33* (5), 602–615.

(15) Grubisha, O.; Kaminska, M.; Duquerroy, S.; Fontan, E.; Cordier, F.; Haouz, A.; Raynal, B.; Chiaravalli, J.; Delepierre, M.; Israël, A.; et al. DARPin-Assisted Crystallography of the CC2-LZ Domain of NEMO Reveals a Coupling between Dimerization and Ubiquitin Binding. *J. Mol. Biol.* **2010**, *395* (1), 89–104.

(16) Schrodinger Release 2015-3: Maestro. Schrödinger, LLC: New York, NY 2015.

(17) Rostkowski, M.; Olsson, M. H. M.; Soendergaard, C. R.; Jensen, J. H. Graphical Analysis of pH-Dependent Properties of Proteins Predicted Using PROPKA. *BMC Struct. Biol.* **2011**, *11* (Cc), 6.

(18) Prime. Schrödinger, LLC: New York, NY 2009.

(19) Polley, S.; Passos, D. O.; Huang, D.; Biswas, T.; Inder, M. Structural Basis for the Selective Participation of IKK1 in Non-Canonical NF- ϰ B Signaling.

(20) Arnold, K.; Bordoli, L.; Kopp, J.; Schwede, T. The SWISS-MODEL Workspace: A Web-Based Environment for Protein Structure Homology Modelling. *Bioinformatics* **2006**, *22* (2), 195–201.

(21) Qin, S.; Zhou, H. Structural Bioinformatics Meta-PPISP : A Meta Web Server for Protein-Protein Interaction Site Prediction. **2007**, *23* (24), 3386–3387.

(22) Porollo, A.; Meller, J. Prediction-Based Fingerprints of Protein – Protein Interactions. **2007**, *645* (December 2006), 630–645.

(23) Bakan, A.; Meireles, L. M.; Bahar, I. ProDy: Protein Dynamics Inferred from Theory and Experiments. *Bioinformatics* **2011**, *27* (11), 1575–1577.

(24) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: An Automated Pipeline for the Setup of Poisson-Boltzmann Electrostatics Calculations. *Nucleic Acids Res.* **2004**, *32* (WEB SERVER ISS.), 665–667.

(25) Holst, M. J.; Saied, F. Numerical-Solution of the Nonlinear Poisson-Boltzmann Equation: Developing More Robust and Efficient Methods. *J. Comput. Chem.* **1995**, *16* (3), 337–364.

(26) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (18), 10037–10041.

(27) Comeau, S. R.; Gatchell, D. W.; Vajda, S.; Camacho, C. J. ClusPro: A Fully Automated Algorithm for Protein-Protein Docking. *Nucleic Acids Res.* **2004**, *32* (WEB SERVER ISS.), 96–99.

(28) Nguyen, H.; Roe, D. R.; Simmerling, C. Improved Generalized Born Solvent Model Parameters for Protein Simulations. **2013**.

(29) Nguyen, H.; Maier, J.; Huang, H.; Perrone, V.; Simmerling, C. Folding Simulations for Proteins with Diverse Topologies Are Accessible in Days with a Physics-Based Force Field and Implicit Solvent. **2014**, 1–4.

(30) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23* (3), 327–341.

(31) Lamm, G.; Szabo, A. Langevin Modes of Macromolecules. *J. Chem. Phys.* **1986**, *85* (12), 7334.

(32) Kim, D. E.; Chivian, D.; Baker, D. Protein Structure Prediction and Analysis Using the Robetta Server. *Nucleic Acids Res.* **2004**, *32* (WEB SERVER ISS.), 526–531.

(33) Liu, S.; Misquitta, Y. R.; Olland, A.; Johnson, M. A.; Kelleher, K. S.; Kriz, R.; Lin, L. L.; Stahl, M.; Mosyak, L. Crystal Structure of a Human IκB Kinase β Asymmetric Dimer. *J. Biol. Chem.* **2013**, *288* (31), 22758–22767.

ACKNOWLEDGEMENTS

Chapter 6, in part is currently being prepared for submission for publication of material. Schiffer, Jamie. M, Nguyen, T. T., Chan, G., Demir, O. Ghosh, G., Amaro, R. E., Structure-Based Dynamics of the Inhibitor of κB Kinase β Construe the Complexity at a Plausible NEMO Interface. The dissertation/thesis author in the primary investigator and author of this material.