Computational Modeling of Protein–Biomolecule Interactions with Application to Mechanotransduction and Antibody Maturation

by

Aurore Zyto

Submitted to the Department of Biological Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Biological Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2008

Author .....
Department of Biological Engineering
May 5, 2008

Certified by .....
Bruce Tidor
Professor of Biological Engineering and Computer Science
Thesis Supervisor

Certified by .....
Roger D. Kamm
Germeshausen Professor of Mechanical Engineering and Biological Engineering
Thesis Supervisor

Accepted by .....
Alan J. Grodzinsky
Chairman, Department Committee on Graduate Students

Thesis committee

Accepted by .....

Paul T. Matsudaira
Professor of Biology and Biological Engineering
Chairman of Thesis Committee

Accepted by .....

Bruce Tidor
Professor of Biological Engineering and Computer Science
Thesis Supervisor

Accepted by .....

Roger D. Kamm
Germeshausen Professor of Mechanical Engineering and Biological Engineering
Thesis Supervisor

Accepted by .....

C. Forbes Dewey
Professor of Mechanical Engineering and Biological Engineering
Thesis Committee Member

Computational Modeling of Protein–Biomolecule Interactions with Application to Mechanotransduction and Antibody Maturation

by

Aurore Zyto

Submitted to the Department of Biological Engineering
on May 5, 2008, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Biological Engineering

Abstract

Cell survival, growth, differentiation, migration, and communication all depend on the appropriate combination of specific interactions between proteins and biomolecules. Therefore, understanding the molecular mechanisms influencing protein–biomolecule binding interactions is important both for fundamental knowledge and as a foundation for therapeutic applications and biotechnology. This thesis presents two applications of computational modeling to study protein–biomolecule binding in different contexts. First, we sought to characterize effects of applied mechanical force on protein structural and biochemical properties. Despite growing experimental evidence of force-regulated cell behavior, the molecular mechanisms involved in force sensing and transmission are still largely unknown. We adapted a free energy method to directly compute the change in binding affinity upon force application. Our simulations demonstrated that differential responses in the bound and unbound state of a protein–ligand complex can lead to graded force-modulation of binding affinity. Application to a prototypical protein system — the helical bundle complex of a paxillin fragment bound to the FAT domain of focal adhesion kinase (FAK) — revealed several structural mechanisms responsible. Second, we used computational methods to design individual mutations computed to improve binding affinity of an antibody–small molecule complex with relevance to cancer treatment. Our calculations suggested several beneficial mutations for experimental characterization. The work illustrates the value of computational modeling for understanding protein–biomolecule interactions with application to therapeutic development and advances in biotechnology.

Thesis Supervisor: Bruce Tidor
Title: Professor of Biological Engineering and Computer Science

Thesis Supervisor: Roger D. Kamm
Title: Germeshausen Professor of Mechanical Engineering and Biological Engineering

Acknowledgments

I would like to thank my advisors, Bruce Tidor and Roger Kamm, for their support and guidance throughout my thesis research. Bruce provided fruitful suggestions to tackle new challenges and helped me to develop critical scientific thinking, while trusting me to design and implement my own solutions. Roger's expertise in both experimental and computational techniques was instrumental in laying out the basis of this work. My thesis committee members, Paul Matsudaira and Forbes Dewey, contributed constructive comments and insights on both research content and presentation style.

I would like to thank our experimental collaborators: Matthew Lang and Jorge Ferrer for developing new experimental techniques to probe mechanotransduction at the molecular level, and Dane Wittrup and Kelly Davis for designing the antibody maturation project and for performing experiments to test our predictions.

It was a great pleasure to work with the members of the Tidor and Kamm laboratories throughout the years. Thank you all for creating a productive and enjjoyable environment in which to work and for providing constructive feedback. I would like to thank in particular: Michael Altman for helpful suggestions on any random problem; Woody Sherman for teaching me the basics of molecular dynamics; Shaun Lippow for developing the computational protein design software used in this work; and Caitlin Bever, Mala Radhakrishnan, Kathryn Armstrong, A. Katharina Wilkins, Anya Castillo and Kelly Thayer for helping to create a balanced lab atmosphere.

I am indebted to the MIT Departement of Biological Engineering for allowing me to pursue exciting research opportunities and providing me with excellent training. Thank you to Doug Lauffenburger for fostering such a vibrant scientific community.

I thank my parents, my sister, my entire family, and my friends for their love and support throughout the years. My children, Betsalel and Yehuda, taught me to work more efficiently and to put things in perspective. Finally, there are no words to accurately acknowledge my husband Sacha: thank you for being part of my life.

Contents

List of Figures

Chapter 1

Introduction

Protein–protein and protein–small molecule interactions are a key element of life. Cell survival, growth, differentiation, migration and communication all depend on the appropriate combination of specific interactions between biomolecules. Dynamic complex formation can be integrated into pathways, allowing higher-order processing and specific information flow, from multiple input signals to effectors [1]. At the core of the cell's ability to process information lies tight binding interactions between protein and ligand. Disease can arise from the malfunction of even a single specific protein–ligand interaction. Both activation and inhibition of complex formation can lead to pathological conditions. Mutation of a single amino acid can result in non-physiological oligomerization, leading to protein aggregation and plaque formation, such as in Alzheimer's disease [2], or decrease in red blood cell elasticity, triggering vaso-occlusive crisis for patients with sickle-cell disease [3]. Alternatively, protein mutation can result in loss of binding affinity, interfering with its biological function. For instance, the transcription factor p53 is mutated in 50% of tumors; many oncogenic mutations lie in the DNA-binding domain of p53, preventing it from activating transcription of its target genes [4]. To alleviate a specific disease condition, drugs can be designed to bind specifically to a target protein, either re-activating or inhibiting its function. Therefore, understanding the molecular mechanisms involved in protein–biomolecule complex formation is important both for fundamental reasons and for therapeutic applications.

Protein–biomolecule binding interaction can be tuned by a number of different factors. Binding affinity can be modified by individual point mutations, which can

alter key residues at the binding interface, introduce a conformational rearrangement in the protein structure, or both. Binding of allosteric effectors at a secondary site can trigger a conformational change leading to modifications in binding interactions. In therapeutic applications, binding of the drug to its target in the body can be limited by slow diffusion and rapid clearance of the therapeutic compound [5]. Additionally, increasing experimental evidence has shown that binding affinity can be modulated by the mechanical environment of the proteins interacting with each other [6, 7]. In many cases, subtle changes in protein concentration or binding affinity can have drastic effects in downstream signaling response. For instance, the combination of multiple steps of phosporylation and coupled dephosphorylation in MAPK cascades leads to ultra-sensitivity, where a small change in input concentration can result in full activation of the entire pathway [8]. Alteration of specific molecular interactions can result in enzyme activity modification, thus shifting the dynamic range of the signaling cascade.

Traditionally, molecular biology techniques have been used to identify and characterize specific binding interactions. Experimental assays such as western blots allow detection of individual proteins through staining by a specific antibody [9]. Binding interactions can also be studied by such analytical techniques, through incubation with a secondary probing agent — usually, an antibody to the second binding partner. The resolution is limited, however, and the experimental process requires cell lysis, therefore precluding its use for real-time applications probing the cell dynamic environment. Alternative experimental techniques to probe protein–protein interactions, either at the single molecule level [10], or in living cells, are being developed. Nevertheless, interpretation of experimental data at the molecular level remains a challenge. In the area of protein design, experimental methods such as directed evolution rely on the random generation of protein mutants, followed by selection of clones exhibiting the desired property [11]. The overall sequence space is very large,

therefore many potentially favorable mutations might not be sampled on the experimental time scale. Evolution to sequences requiring cooperative DNA base-pair substitutions is furthermore hindered by the low mutation rate necessary to retain a high fraction of foldable sequences.

Increasing computational power and the development of structure-based computational modeling techniques now allow molecular interactions to be studied in more detail. For example, to study force effects on protein structure and function, single molecule AFM experiments can be used to measure unfolding forces of individual protein domains. Such experiments show that, despite high structure and sequence similarities, fibronectin type III modules exhibit different relative mechanical stability [12]. Computer simulations recreating the experimental setup are able to match the experimental ranking, and more importantly, provide detailed hypotheses for molecular differences giving rise to the observed hierarchy [13]. Computational analysis of a binding interface can separate the contributions of individual atoms. Inspection of the molecular properties of each atom that most strongly influence its contribution can help in the rational design of improved binding interactions.

This thesis presents two applications of computational modeling to study protein–biomolecule interactions. Through the use of molecular models, we investigated the impact of applied forces on protein structure–function relationships. We developed new computational tools of general applicability to understand the molecular mechanisms involved in mechanotransduction, the phenomenon by which cells sense and transduce mechanical cues into biochemical signals that can be interpreted through downstream signaling pathways. We also applied computational protein design to rationally predict individual amino-acid mutations to affinity mature a monoclonal antibody, with applications in cancer therapy. A common underlying theme in this work is that of balancing interactions between different ensembles. Binding occurs when protein–ligand interactions in the bound state are more favorable than protein–

solvent interactions in the unbound state. For instance, polar interactions in the bound state come at the expense of desolvation penalty upon binding [14]. Protein folding arises from beneficial protein–protein contacts in the folded conformation, compared to solvent interaction with individual exposed side-chains in unfolded conformations. Force-induced conformational rearrangement results from balancing protein–protein interactions with the externally applied mechanical force. A detailed analysis of the various ensembles at play, and the difference in molecular interactions between them, is required to understand the trade-off involved. The development and application of such computational tools can lead to further understanding of biological phenomena and better protein engineering for therapeutic applications.

Two theoretical models for biochemical systems make this computational modeling possible. One is a description of molecules as mechanical systems and the other is the use of electrostatics models to speed up the computation of the interactions between charged atoms. Molecular mechanics techniques describe molecules as collections of charged spherical masses (representing atoms), connected by springs (representing covalent bonds). Potential energy for a given molecule configuration is computed by adding contributions from interactions between bonded and non-bonded atoms. Careful parameterization of the interactions between pairs (bonds), triplets (bond angles) and quadruplets (dihedral angles) of atoms, as well as the energetic contribution of interactions between non-connected atoms has produced several useful energetic force fields [15–19]. These can be used to compute the energy of a static atomic structure and to model molecular dynamics by integrating Newton's equations of motion. Biological molecules typically function in an aqueous environment and modeling individual water molecules surrounding a given molecule is computationally very expensive. However, the surrounding water has a significant impact on the interactions between charged atoms, called electrostatics. To model the influence of the water environment more efficiently, a second category of theoretical models, referred to as implicit

solvation models, was developed. These models describe molecules as collections of charges inside a defined molecular volume, wherein the dielectric constant is low and outside of which the dielectric constant is high. The high dielectric region represents water and can also model the presence of mobile ions to more accurately describe a biological environment. Various continuum electrostatic models differ in the methods used to define the molecular shape separating the dielectric regions and to compute the effect of the high dielectric region on interactions inside the low dielectric region. Usually, there is a trade-off between accuracy and computational efficiency. Elaborate models use the exact molecular surface and solve the linearized Poisson–Boltzmann equation to compute the electrostatic potential [20,21]. Generalized Born models use alternative methods to compute the electrostatic solvation energy [22,23], and some achieve good computational efficiency through approximate evaluation of the molecular surface, for instance with a simple smoothing function [24]. Heuristic models make use of several approximations, such as group-based contributions to the solvation free energy and distance-dependent dielectric constant inside the molecular volume, to speed up computation further [25]. In this work, we used models of different complexity as best suited for the particular application.

Search algorithms were also an integral part of the work presented here. In order to computationally make chemical changes and new energy evaluations efficiently, the "inverse design" approach is used to evaluate the sequence and structure search space. Instead of allowing complete conformational freedom, we therefore model individual amino acid mutations in the existing protein backbone structure only. Two search algorithms are used to extract a ranked list of the lowest-energy sequences, in a guaranteed fashion [26,27]. In contrast, molecular dynamics relies on the random exploration of the protein conformational space by integrating Newton's equation of motion based on the potential energy [28]. Protein energy landscapes are multi-dimensional, and have been described as rugged, with many local minima. Rare energy barrier crossing

limits phase space exploration by molecular dynamics at room temperature. Recently, generalized-ensemble algorithms have been developed to enhance sampling [29, 30]. Validation of our free energy computational method on theoretical systems showed that precise sampling of force-induced conformational change had a critical impact on accuracy. Therefore, we incorporated adequate sampling techniques to appropriately simulate force effects on protein structure.

These models for molecular energetics and these search algorithms were used to computationally investigate protein–biomolecule binding properties. Two projects are presented that cover different topics and goals in the area of computational modeling.

In Chapter 2 we developed a new computational method to study force-modulated binding interactions. Despite growing experimental evidence of force-regulated cell behavior, the molecular mechanisms involved in force sensing and transmission are still unknown. We focused on characterizing force-induced changes in protein structure–function relationships. By using an integrated approach, we highlight general principles involved in mechanosensing. Next, we developed, validated, and applied a new free energy method to directly compute the change in binding affinity upon force application. Our simulations suggested force-induced strengthening of binding affinity in a prototypical protein complex. Structural analysis emphasized several general mechanisms involved. Our methodology is directly applicable to other biologically relevant interactions.

In Chapter 3 we used computational protein design to identify mutants with computed improvement of binding affinity over an existing monoclonal antibody to its ligand, Y-DOTA. This system was identified by our experimental collaborators, Kelly Davis and Pr. K. Dane Wittrup, as a potential compound for pretargeted radioimmunotherapy, an antibody-based strategy being developed for cancer treatment. Current application is limited by weak antibody–ligand binding. Our computational design calculations suggested several individual mutations to improve binding affinity.

Selected mutations will be tested experimentally for improved binding, and successful designs could be combined into a higher affinity multiple mutant.

Overall, this work makes use of existing methods and develops new techniques for computational modeling of protein–biomolecule interactions. Detailed understanding of molecular mechanisms involved in tuning binding interactions is important for fundamental reasons and for designing new therapeutic compounds.

Chapter 2

Biomolecular transduction of mechanical signals: Simulation study of graded force-modulated protein affinity

Abstract

The observation that cells respond biochemically to changes in their mechanical environment has prompted investigations into mechanisms for the interconversion of chemical and mechanical signals. Examples include force-induced conformational changes that expose or occlude binding sites in an essentially all-or-nothing manner. Here we investigate a more subtle possibility whereby applied force directed to the vicinity of a protein binding site may result in changes in conformational distributions leading to gradual changes in binding affinity. The investigation presented here includes simple analytical examples and an idealized computational example from which the general principles and properties are developed. The main body of the investigation is a free energy simulation analysis of a protein binding interaction modeled on an atomic resolution structure of the focal adhesion targeting domain (FAT) of focal adhesion kinase (FAK) bound to a portion of paxillin. Together the analysis highlights that binding free energy changes result from differential structural changes between the bound and unbound states, and that often the larger effect is in the unbound state. Moreover, moderate applied force can enhance or diminish binding affinity; one mechanism for enhancement is for applied mechanical force to preconform the binding site in the unbound state closer to the binding conformation than in the absence of applied force. In the protein simulations, such preconformation was observed with concomittent exposure of hydrophobic residues. Increased applied force led to affinity-enhancing distortions in the bound state as well, leading to an increase binding interface area and activation of specific polar contacts, whose energetic role was confirmed through computational mutagenesis. Together these studies probe the properties of graded binding affinity modulation in response to applied force.

2.1 Introduction

Mechanical signals have been shown to regulate various physiological behaviors, including growth, differentiation, apoptosis, motility, and gene expression [31, 32]. Despite experimental characterization of biochemical signaling pathways activated by force, the details of the processes by which mechanical cues result in cellular change, a phenomenon referred to as mechanotransduction, are largely unknown. Abnormal mechanotransduction can arise through changes in cell mechanics, variations in extracellular matrix structure, or deregulation of the molecular mechanisms by which cells sense and respond to mechanical cues. Improper mechanotransduction has been implicated in pathological conditions, such as atherosclerosis, asthma and carcinogenesis [33]. Understanding molecular mechanisms for mechanotransduction is essential fundamental knowledge of cellular behavior and disease and additionally could open new avenues for therapeutic intervention.

Several nano-scale mechanisms for mechanotransduction have been proposed, including changes in membrane fluidity [34], glycocalyx deformation [35, 36], and constrained autocrine signaling [37]. At the molecular level, force-driven conformational change could result in altered protein activity, specific examples of which include force-modulated binding affinity [38–40], enzyme activity [41], or channel conductance [42] (see reviews [6, 7]). In this framework, key protein molecules act as mechanosensors and mechanotransducers; the combined effect of these functions is to convert mechanical signals into chemical ones, thereby connecting to cellular biochemistry and downstream effectors. Despite increasing experimental demonstration of specific binding interactions being activated or inhibited upon mechanical stimulation, the precise details of the relationship between applied force, protein structure, and protein activity are largely unknown and likely to be system-specific.

The adhesion machinery anchors the cell to its surrounding extracellular matrix and adjacent cells, through transmembrane receptors, such as integrins, con-

necting the cytoskeleton with the extracellular matrix, and cadherins, bridging the cytoskeletons of adjacent cells. This connection, established by dynamically interacting proteins, enables the transmission of externally applied mechanical forces and cytoskeleton-generated ones between the cell interior and exterior [43]. Therefore, adhesion complexes have been actively studied for their ability to sense, transduce, and integrate mechanical signals. In particular, focal adhesions serve as mechanosensitive protein complexes, whereby local force leads to greater protein assembly and increased signaling, and relaxation leads to site disassembly and signal termination [44]. Focal adhesion kinase (FAK), an important scaffold and signaling protein in focal adhesions [45,46], has been shown to play a critical role in many aspects of force-regulated signaling, specific examples of which are substrate rigidity sensing [47] and skeletal regeneration [48]. Additionally, FAK is over expressed in a variety of tumors, and its expression correlates with tumor malignancy [49]. FAK contains a FERM domain, a kinase domain, and a focal adhesion targeting (FAT) domain. The FAT domain recruits FAK to focal adhesions by interacting with integrin-associated proteins such as talin and paxillin. The FAT domain structure is a four-helix bundle; helical bundle structures are ubiquitous in proteins involved in adhesion [50], such as vinculin [51], α-catenin [52], and talin [53]. Moreover, there is a commonality of binding mechanism: through bundle association, the paxillin helical LD motif interacts with the FAT domain four-helix bundle, mostly through hydrophobic patches, thereby creating a five-helical bundle structure. A similar binding motif is found in other complexes of the adhesion machinery, such as vinculin head (Vh) binding to vinculin tail domain [54], various Vh–talin complexes [53–55], Vh–α-actinin [56], α-catenin dimerization and α-catenin–β-catenin binding interactions [57]. Despite growing experimental evidence for force-regulated behavior and activated signaling pathways, it is still unclear whether mechanosensitivity is governed by a single protein or by large multi-protein complexes. Additionally, the specific protein interactions

directly involved in mechanosensing and transducing have not yet been identified.

Computational modeling techniques allow specific molecular interactions to be studied in atomistic detail, thus improving our understanding of biological processes at a scale that is not accessible through traditional experimental methods. In particular, steered molecular dynamics simulations, which involve the application of external forces to molecular models, have been widely used to generate hypotheses for how protein structure is affected by applied mechanical forces [58, 59]. Simulations have been used to investigate unfolding pathways for modular proteins resulting from force applied to N- and C-termini of individual domains. Notably, thorough computational investigations of fibronectin modules shed light into phenomena such as force-induced modification of binding site geometry [60], decoupling of synergistic interactions [61], and sequence effects on mechanical stability [62]. Recently, studies have focused on inferring and mimicking biologically relevant environments and force applications in systems where such information is more readily available, such as mechanosensitive ion channels [63, 64] or talin–vinculin binding interactions [65, 66]. Analysis of such simulations has led to descriptions of force-induced conformational changes and estimates of potential of mean force along pre-specified reaction coordinates. Methods have been developed to take into account non-equilibrium sampling through incorporation of Jarzynski's equality [67]. However, when tested on simple protein systems such as a deca-alanine α-helix, the applicability of such techniques to full-size protein systems remains unclear [68]. To the best of our knowledge, no computational study of mechanotransduction has focused on extracting general thermodynamic properties that can be related to biologically meaningful phenomena, such as computing binding free energy changes as a function of applied force.

Here, we present an integrated computational approach to gain insight into force-regulated binding properties. First, we design and characterize simple models representative of classes of mechanosensing scenarios. Highlighting key features involved in

force-dependent protein activity will help to further understanding of detailed mechanisms. Then, we adapt computational free energy methods to study force-modulated binding properties. Using idealized theoretical systems with protein-like features, we validate our methodology and show how both local and global conformational changes can contribute to force-regulated binding. Finally, we apply the free energy method to study the mechanosensing properties of FAT domain binding to a paxillin peptide and observe force-induced enhancement of computed affinity. Currently the field lacks a detailed experimental characterization of which proteins are subject to forces and the properties of those forces (i.e., point of application, direction, magnitude, duration, and time course). Thus, the goal of the current study is not to recapitulate or explain a particular experimental finding, but rather to apply simulation methodology to a common motif, the four-helix bundle, which occurs frequently in proteins associated with mechanotransduction, to explore the relationship between protein structural properties, applied mechanical force characteristics, and binding affinity changes.

2.2 Results and Discussion

2.2.1 Simple mechanical models

A series of simple mechanical block-and-spring models were studied to examine general classes of behavior for force-modulated binding. Figure 2-1 shows these models, and details are given in the Methods section. The general features of the models were selected to be abstractly representative of protein systems, and a particular advantage is that they are analytically solvable. Protein and ligand are each represented by a non-deformable block of material that bind through shape complementarity. Each binding partner has one degree of freedom (the protein translates in the x direction and the ligand in the y direction, each under the influence of a harmonic spring). The

binding energy as a function of force F applied to the protein is given by,

$$\Delta U_{\text{bind}}(F) = U_{\text{bnd}}\left(x_{\text{bnd}}(F)\right) - U_{\text{unb}}\left(x_{\text{unb}}(F)\right) \tag{2.1}$$

where we compute the difference in energies U_i for the bound ($i = \text{bnd}$) and unbound ($i = \text{unb}$) state at their respective global energy minimum, $x_{\text{bnd}}(F)$ and $x_{\text{unb}}(F)$. As force is applied to the protein block, it is opposed by the spring attached to the protein and, additionally, by binding interactions. The balance of forces determines the new equilibrium position in each state. For each applied force, we compute the position of the global minimum in potential energy, and its associated energy $U_i(x_i(F))$. The results of this analysis show how differential effects of force in the bound and unbound state lead to force-modulated binding affinity. The present study includes a small number of model variants that differ in the specific geometry of the protein relative to the ligand and that represent effects such as allostery and exposure of cryptic binding sites.

Scenario 1. The first system illustrates the simplest mode of protein–ligand complex: the protein adopts one conformational state at rest, which is binding-competent. This system is modeled as having a single notch, which is aligned with the ligand block at rest, as shown in Figure 2-1A. Figure 2-1B depicts the corresponding unbound state. We applied a range of forces to the protein block and computed the associated equilibrium conformation for both the bound and unbound states, and the resulting binding energy.

Figure 2-2A shows the global minimum in potential energy $U_{\text{xxx}}\left(x_{\text{xxx}}(F)\right)$ for the bound and unbound states. As the applied force is increased, $x_{\text{unb}}(F)$ increases linearly, resulting in a continuous decrease in $U_{\text{unb}}\left(x_{\text{unb}}(F)\right)$, as described in Methods. In contrast, the position in the bound state $x_{\text{bnd}}(F)$ remains unchanged for sufficiently small forces, as shown in Figure 2-2B (regime I). This behavior stems from

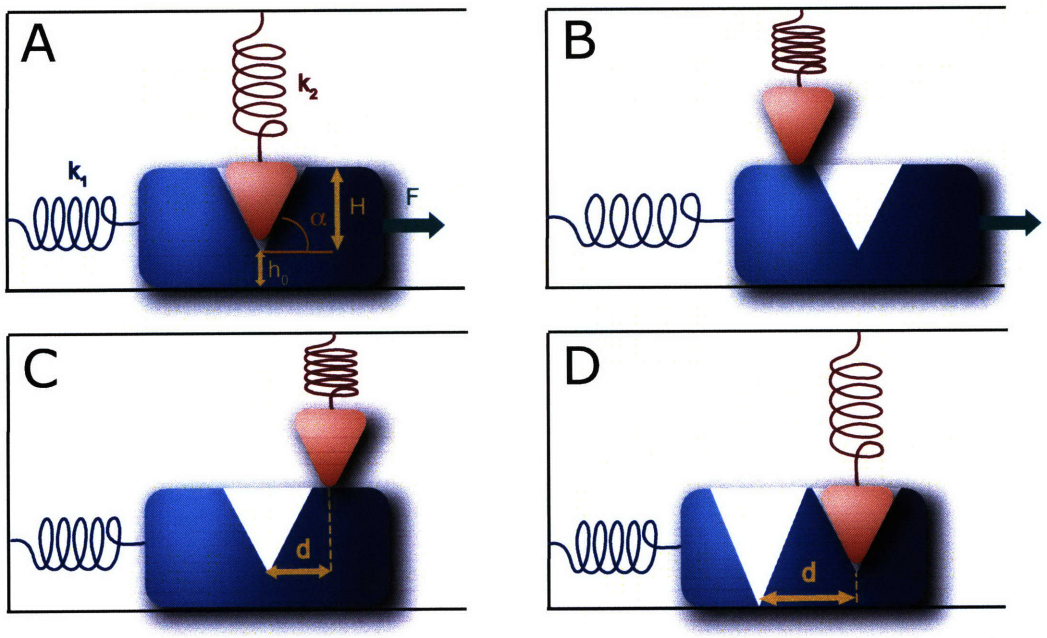


Figure 2-1: Simple mechanical models. (A) Scenario 1 in the bound state: protein (blue) and ligand (pink) bind to each other. The green arrow indicates the force applied. The protein has only the horizontal (x) degree of freedom, and the ligand has the vertical (y) degree of freedom. (B) Scenario 1 in the unbound state: the binding site is displaced. (C) Scenario 2 is representative of a protein that exhibits a cryptic binding site, exposed upon force application. (D) Scenario 3 is representative of an allosteric protein, for which the allosteric transition to the high-affinity state can be triggered by binding to an allosteric effector or by force application.

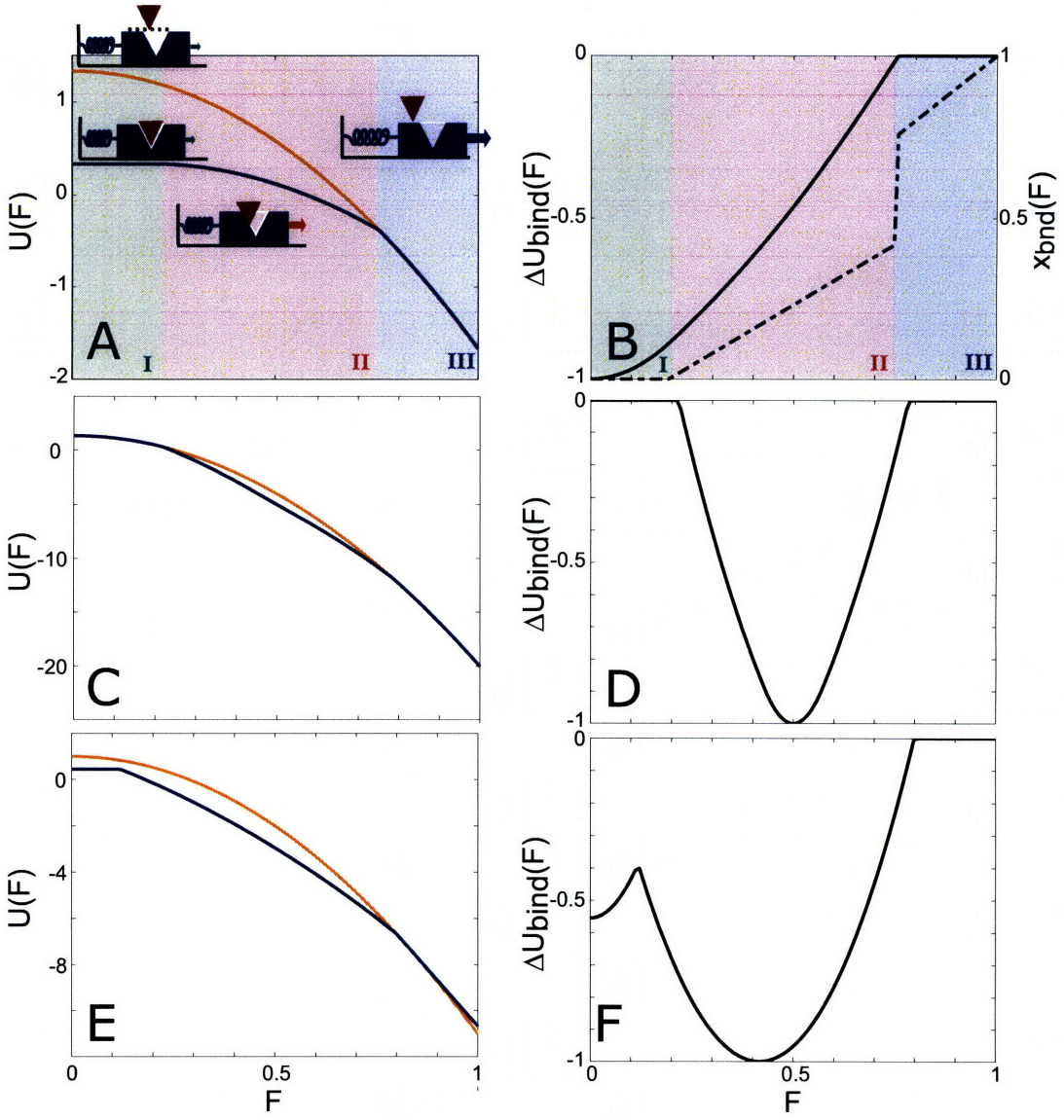


Figure 2-2: Force-modulated energy profiles for the simple mechanical models. (A,C,E) Minimum in unbound state energy $U_{\text{unb}}(x_{\text{unb}}(F))$ is shown in orange, and in bound state energy $U_{\text{bnd}}(x_{\text{bnd}}(F))$ in blue. For each system, the position of the global minimum in energy $x_i(F)$ ($i = \{\text{bnd}, \text{unb}\}$) is computed by setting minus the gradient of Equation 2.8 to zero. We plot the potential energy at this global minimum for each applied force. (B,D,F) Binding energy $\Delta U_{\text{bind}}(F)$, computed with Equation 2.1, is shown in black. (A-B) Scenario 1. The dashed curve in panel B shows the position of the minimum in energy for the bound state $x_{\text{bnd}}(F)$, with corresponding values on the right vertical axis. The 3 regimes of force application are indicated with background colors green (I), pink (II), and blue (III), see text. (C-D) Scenario 2. (E-F) Scenario 3.

favorable interactions with the ligand, which oppose deformation in order to maintain the ground conformation. The binding energy is a balance between bound and unbound state energies, as shown in Equation 2.1. Despite no conformational change in the bound state, the overall binding affinity is decreased, as shown in Figure 2-2B. The increase in binding energy arises from the deformation penalty that needs to be overcome to bring the protein from its distorted unbound configuration ($x_{\text{unb}}(F)$) to its binding-competent original state ($x_{\text{bnd}}(F) = 0$). The overall behavior falls into three regimes, highlighted with background colors in Figures 2-2A and B. For intermediate forces (regime II), a somewhat different behavior results. Both the bound and unbound states are moderately distorted from their resting conformation. Because of the distinct shapes of the potential energy landscapes, this adaptive response is different for both cases and the binding affinity is further decreased. Finally, for high forces (regime III), both states undergo large conformational changes; binding is completely disrupted, and there is no further force-dependent affinity change.

Scenario 2. The second system corresponds to a protein that partially unfolds or otherwise changes conformation upon force application, thereby exposing a cryptic binding site. In such a system, the resting configuration of the protein is not binding competent, which can be achieved, for instance, if key residues involved in binding interactions are buried in the ground conformation. This mode of mechanosensing has been hypothesized for some multi-modular proteins, such as fibronectin. Fluorescence resonance energy transfer (FRET) experiments have demonstrated that this extracellular protein exhibits a wide range of conformations, suggesting force-regulated partial unraveling of FnIII modules [69, 70]. Moreover, several binding sites and catalytic domains have been shown to be active only in extended conformations, indicating that they may be hidden in folded fibronectin [41, 71].

This type of force-sensing system can be recapitulated by a modified model in

which the binding site is blocked at rest, as shown in Figure 2-1C. The force-dependent structural modifications of the bound and unbound state lead to biphasic behavior of the binding affinity, as shown in Figures 2-2C and D. For low forces, both bound and unbound states undergo small conformational rearrangements that do not reveal the binding site. As a result, binding is inhibited. For moderate forces, larger conformational rearrangements lead to the exposure of the cryptic binding site, modeled in this case by the alignment of the protein notch with the ligand block. Favorable interactions are possible in the bound state, leading to a decrease in $U_{\mathrm{bnd}}\left(x_{\mathrm{bnd}}(F)\right)$ compared to $U_{\mathrm{unb}}\left(x_{\mathrm{unb}}(F)\right)$, as shown in Figure 2-2C. As a result, the binding energy decreases, and binding becomes favorable (Figure 2-2D). Finally, high forces induce further conformational change that prevents binding. This situation corresponds to a protein for which exposure of a cryptic binding site is linked to partial unfolding of a domain, but for which complete unfolding disrupts binding. For this scenario, there is an optimal force at which binding is the strongest. We note that, in biological systems where cryptic-site exposure occurs through complete unfolding of a protein domain, there would be a threshold force for activation, but binding might not be disrupted by higher forces, assuming the protein itself could withstand such force magnitudes.

Scenario 3. Finally, we describe a system that illustrates force-modulated binding for an allosteric protein. Allosteric regulation is a common strategy for proteins, where binding of an effector molecule changes protein activity through conformational rearrangement [72]. In a parallel manner, an allosteric protein could respond to applied force rather than effector binding; for instance, the application of an external force could induce the natural allosteric transition between low-affinity and high-affinity states in cases where the high-affinity state corresponds to a more extended conformation. Essentially, applied force takes the place of the allosteric effector. Recently, an allosteric model has been proposed to explain catch-bond behavior of E.

coli adhesin FimH binding to mannose coated surfaces, where increased shear flow leads to bond strengthening [73].

We have modeled allostery in the current framework, using a protein with 2 notches of different depths, corresponding to two main conformations, as shown in Figure 2-1D. The force-modulated protein activity profile is linked to the relative affinities of the two main conformations of the protein, and the force–structure relationship. For illustrative purposes, we describe a system that can adopt either a compact, low-affinity state (state $\mathcal{C}$) or an extended, high-affinity state (state $\mathcal{E}$). The force response of this model is shown in Figures 2-2E and F. In the absence of an applied force, state $\mathcal{C}$ is favored, resulting in $\Delta U_{\text{bind}}(F = 0) = -0.56$ scaled units. For low magnitudes of applied force ($F < 0.1$ scaled units), the protein remains in bound state $\mathcal{C}$. Continuous deformation in the unbound state leads to a slight decrease in binding affinity, as explained previously for Scenario 1. Application of moderate forces induces a conformational transition toward state $\mathcal{E}$. As force is increased, the protein adopts a conformation closer to this higher-affinity state, which is reached for $F_{\text{opt}} = 0.42$ scaled units. This conformational transition results in an increase in binding affinity, leading to tighest binding with $\Delta U_{\text{bind}}(F = F_{\text{opt}}) = -1$ scaled units. Finally, pulling with a force beyond a given threshold leads to conformational changes beyond the allosteric transition, such as partial unfolding, and disruption of binding.

Summary. Analysis of simple mechanical models has elucidated key features of force-modulated binding. The study highlights the point that because binding affinity is a balance between bound and unbound state energies force-modulation of binding strength can arise through modifications of the structures and energetics of the bound state, the unbound state, or both. Large magnitudes of applied force induce significant conformational changes that are generally associated with binding disruption.

This observation does not apply in specific cases where the binding-competent state corresponds to a distorted or an unfolded conformation, which could be the case for certain proteins with cryptic binding sites. Finally, the binding affinity is highest when the unbound state is pre-conformed to the bound state. In cases where the unbound state conformation is distorted away from the binding-competent conformation, a deformation penalty must be paid to bring the unbound state conformation into that configuration. Even when favorable interactions in the bound state are preserved, this energetic penalty leads to an overall binding affinity decrease, as explained for Scenario 1. Here, we have made use of a simple, static, one-dimensional model for which only the global minimum energy conformation was considered for each state (zero-Kelvin limit). In the next subsection we explore an idealized system with protein-like properties of multi-dimensionality and conformational ensembles.

2.2.2 Theoretical systems

To investigate force-regulated binding properties in protein systems, we adapted free energy simulation methodology to compute force-induced changes to the free energy landscape. By independently determining the free energy profile as a function of applied force for the bound and unbound states, we create a thermodynamic cycle that allows the estimation of the change in binding free energy upon mechanical stimulation. In this formulation, applied force is the perturbation to the potential, whereas traditionally chemical changes, such as amino acid mutations, are represented in the perturbation potential [74]. A modified potential energy under force application is constructed,

$$U_\lambda(\vec{x}, \vec{F}_{\text{appl}}) = U_0(\vec{x}) - \lambda \vec{F}_{\text{appl}} \cdot \vec{x} \tag{2.2}$$

such that we can derive the total force as minus the gradient of Equation 2.2, which shows that the perturbed potential results in a modified total force on the system

equal to the unperturbed force $\overrightarrow{F}_0$ plus an applied force $\lambda\overrightarrow{F}_{\text{appl}}$, independent of atomic coordinates:

$$\begin{aligned}
\overrightarrow{F}_\lambda(\overrightarrow{x}, \overrightarrow{F}_{\text{appl}}) &= -\nabla_{\overrightarrow{x}} U_\lambda(\overrightarrow{x}, \overrightarrow{F}_{\text{appl}}) && (2.3)\\
&= -\nabla_{\overrightarrow{x}} U_0(\overrightarrow{x}) + \nabla_{\overrightarrow{x}} \left(\lambda\overrightarrow{F}_{\text{appl}} \cdot \overrightarrow{x}\right) && (2.4)\\
&= \overrightarrow{F}_0(\overrightarrow{x}) + \lambda\overrightarrow{F}_{\text{appl}} && (2.5)
\end{aligned}$$

Here, $\overrightarrow{x}$ is the $3N$–dimensional vector of atomic coordinates, and $\overrightarrow{F}_{\text{appl}}$ is the applied force vector, which we treat as a constant force magnitude and direction that can be scaled by λ. For nomenclature, Equation 2.2 can be recast as,

$$U_\lambda(\overrightarrow{x}, \overrightarrow{F}_{\text{appl}}) = U_0(\overrightarrow{x}) + \lambda U_F \tag{2.6}$$

where we separate the individual contributions of the internal potential energy U_0 and the forcing energy U_F. With this framework, the free energy change as a function of applied force for a given system can be computed using methods in common use, based on the multi-dimensional replica exchange method [75] coupled with the weighted histogram analysis method (WHAM) [76, 77]. Incorporation of high-efficiency sampling algorithms is critical to achieve accuracy, because force application induces substantial conformational rearrangements that may be difficult to sample with traditional simulation protocols. Briefly stated, we perform simulations with M replicas with different values of the parameters $\Lambda_m = (\beta_m, \lambda_m)$, which traverse temperature and applied force space. Here, pairs of replicas adjacent in parameter space are exchanged according to appropriate Boltzmann criteria. We integrate data from different replicas through the WHAM formalism, as presented in Methods, to arrive at $\Delta G(\lambda\overrightarrow{F}_{\text{appl}})$ in the bound and unbound state separately. The difference between the individual state free energy changes provides the change in binding free energy as a function of applied force (see Methods).

To further investigate mechanisms of force-induced binding free energy changes, we designed potential energy landscapes based on experimental and computational characterizations of proteins and studied the force regulation of structural and binding properties. A key advantage of such analytical energy landscapes is that we can independently calculate the free energy change through direct computation of partition functions, thus investigating the overall behavior and convergence properties of our method. We created N-dimensional landscapes by adding individual contributions from each degree of freedom x_i,

$$U_\lambda(\overrightarrow{x}, \overrightarrow{F}_{\text{appl}}) = \left[\sum_{i=1}^{N} U_i(x_i)\right] - \lambda \overrightarrow{F}_{\text{appl}} \cdot \overrightarrow{x} = \sum_{i=1}^{N} [U_i(x_i) - \lambda F_i x_i] \qquad (2.7)$$

As a result, the partition function $Q(\lambda, \overrightarrow{F}_{\text{appl}})$ is decomposable, and thus computable for multi-dimensional landscapes. Each energy function U_i is made up of the combination of a base potential and two sinusoids (see Methods). This structure of the potential energy surface creates many local minima, separated by different energy barrier heights. Due to the rugged nature of the energy landscape and the high number of degrees of freedom simulated, our theoretical systems provide a valuable framework to validate and explore our simulation methodology, and also provide an opportunity to explore force modulated affinity relationships. We created two variants, differing in the specific shape of the energy landscape along the degree of freedom for force application, x_1 (so $\overrightarrow{F}_{\text{appl}} \cdot \overrightarrow{x} = F_{\text{appl}}\, x_1$ here). For both variants, the double-well potential along x_1 creates two main conformational states: state $\mathcal{C}$ ($x_1 < 0$) and state $\mathcal{E}$ ($x_1 \geq 0$). The inset of Figure 2-3A shows the x_1–component of the total potential energy U_λ for $\lambda = 0$ (no force applied) and $\lambda = 1$ (maximum force), in the bound and unbound states. The general shape of the double-well potential, separating the conformational space into two main basins, and the many local minima resulting from the perturbing sinusoids are clearly apparent. As force is applied, Equation 2.2 leads

to a tilting in the energy landscape, inducing a global shift in population distribution between states $\mathcal{C}$ and $\mathcal{E}$. In our design, binding tends to stabilize state $\mathcal{C}$ and applied force tends to stabilize state $\mathcal{E}$. We present results from replica exchange simulations in (β, λ) space for bound and unbound states of a one-thousand dimensional system.

We examined the force effects on the structural and binding properties through analysis of conformation ensembles and application of the free energy method. Figure 2-3A shows the fraction ρ of configurations in compact state $\mathcal{C}$ as a function of λ for the bound and unbound states. ρ is calculated from the theoretical probability and estimated from the simulations at discrete values of λ. In both states, the fraction ρ smoothly decreases as force is applied, from $[\rho(\lambda = 0) = 1]$ to $[\rho(\lambda = 1) = 0]$. Whereas for $\lambda = 0$, the global energy minimum corresponds to state $\mathcal{C}$, state $\mathcal{E}$ becomes the new global energy minimum as force is applied. The fractions computed from the simulations agree very well with the theoretical evaluation of $\rho(\lambda)$ over the whole force range simulated. Fluctuations are larger in the transition region, where states $\mathcal{C}$ and $\mathcal{E}$ have equal probabilities. Operationally, accurate calculation of $\rho(\lambda)$ relies on frequent hopping between the two conformational states, which are separated by a large energy barrier (15 kcal/mol). We define the transition force $\lambda_{\frac{1}{2}}$ as the force at which states $\mathcal{C}$ and $\mathcal{E}$ have equal probability (i.e., $\rho(\lambda_{\frac{1}{2}}) = \frac{1}{2}$). Figure 2-3A shows that $\lambda_{\frac{1}{2}} = 0.5$ for the bound state, whereas $\lambda_{\frac{1}{2}} = 0.25$ for the unbound state. (Forces are reported relative to the maximum force applied.) Hence, the unbound state undergoes its major conformational transition $\mathcal{C} \rightarrow \mathcal{E}$ at a lower force than does the bound state. Figure 2-3B shows the associated change in binding free energy as a function of applied force. The values computed through the free energy methodology agree very well with the theoretical values obtained analytically from the partition functions. The differential response of the bound and unbound states leads to a continuous decrease in binding affinity for this system, which results from its design. At applied forces lower than the transition force for the bound state, this phenomenon is very similar

to that described as regime I for Scenario 1 in the previous section. Indeed, most of the observed loss in binding affinity comes from the existence of distinct global conformations for the bound and unbound states. For binding to occur, the protein needs to first distort its structure into a binding-competent conformation, i.e. from state $\mathcal{E}$ to $\mathcal{C}$, thus overall losing binding free energy through a deformation penalty.

The analysis of simple systems featured unidimensional freedom with harmonic potentials evaluated at zero Kelvin. In contrast, here, the complex structure of the potential energy landscape, as well as the use of physiological temperature (300 K), more realistically probes force–structure relationships in a statistical mechanical framework. Our designed energy surfaces feature many substates, which are local minima separated by small energy barriers and well spacing, co-existing within each global conformational state (i.e., $\mathcal{C}$ or $\mathcal{E}$). Figure 2-4 presents an analysis of the force effects on the distribution of local structural ensembles and the associated change in free energy for the bound and unbound state individually. As shown in Figure 2-4A, the free energy curves exhibit biphasic behavior, increasing until their respective transition forces, and decreasing afterwards. Thus, we observe significant free energy changes independently of the global conformational transition. For instance, for the bound state, for which $\lambda_{\frac{1}{2}} = 0.5$, we computed a free energy change $\Delta G(\lambda = 0.25) = +2.4$ kcal/mol, despite no change in the fraction ρ. Compensating changes occur in the bound and unbound states for forces up to $\lambda = 0.2$, leading to an overall flat dependency of the binding free energy. However, local conformational changes occur in both states, at forces well below their respective transition force $\lambda_{\frac{1}{2}}$. Within state $\mathcal{C}$, the ensemble of accessible configurations is shifted toward more extended states as small levels of force are applied. Figure 2-4B shows histograms of configurations sampled by the simulation along the coordinate x_1 for the bound state with $\lambda = 0.25$ and $\lambda = 0$. Although both ensembles occupy exclusively the compact state (i.e., $x_1 < 0$), the inset showing the difference in histogram counts between $\lambda = 0.25$ and $\lambda = 0$

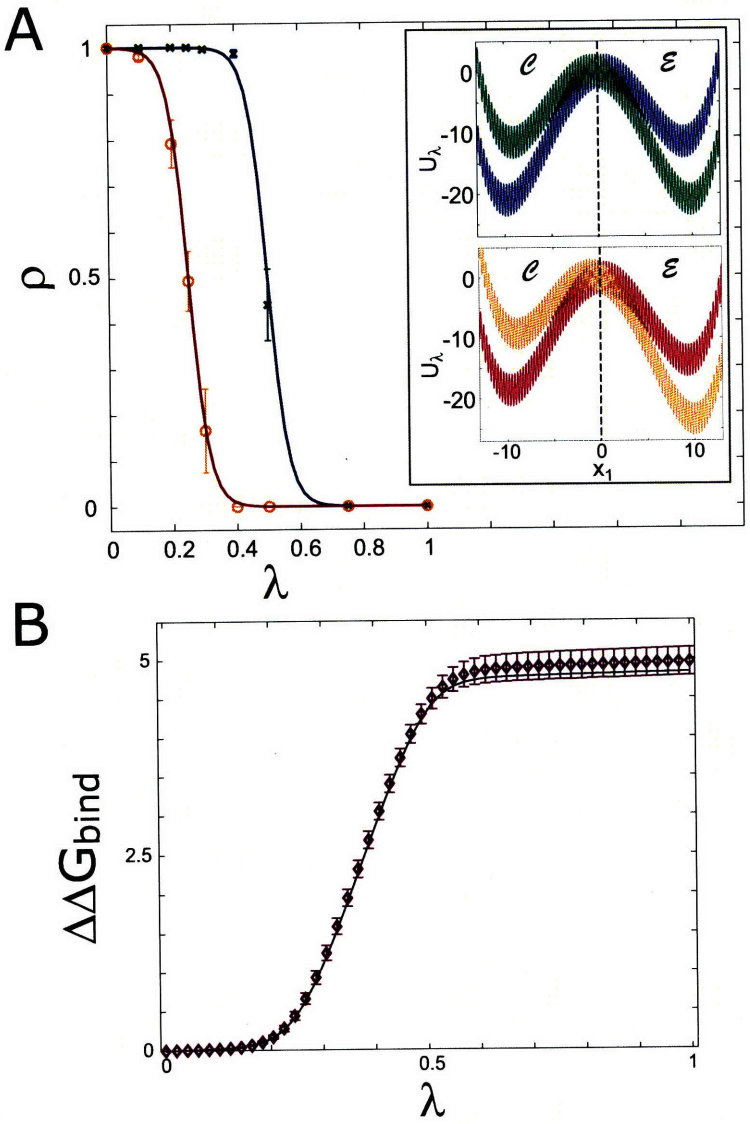


Figure 2-3: Force-induced global conformational change results in binding affinity decrease for a one-thousand dimensional idealized system. (A) Fraction ρ of conformations in compact state $\mathcal{C}$, defined as $x_1 < 0$. Solid curves: fractions computed through numerical integration (bound state: blue, unbound state: red); symbols: simulation estimates, showing average over 3 simulations, with fluctuations (standard deviation), for the bound (green) and unbound (orange) states. Inset: total potential energy along coordinate x_1. Top: bound state, with $\lambda = 0$ (blue) and $\lambda = 1$ (green); bottom: unbound state, with $\lambda = 0$ (red) and $\lambda = 1$ (orange). (B) Change in binding free energy as a function of applied force computed by theory (black curve) and estimated from the free energy simulation method (purple symbols). Error bars represent the standard error estimated with 5 independent sliding windows of data.

reveals systematic local shifts in the conformational ensemble under applied force. Positive values indicate configurations that are sampled more frequently upon force application. Theoretical differences are shown with crosses, and estimates from the simulation with orange error bars representing fluctuations. This Figure illustrates a shift towards more extended configurations, which is demonstrated by negative counts on the left side of the well, and positive ones on the right. (Interestingly, the free energy converges faster than the conformational distribution function, as there is imperfect agreement between simulation and theory.) This subtle rearrangement of the configurational ensemble leads to the significant free energy change observed in Figure 2-4A. Concerted shift in the accessible conformational ensemble upon force application is an important factor for protein applications as well, as will be shown in the next section.

Design and simulations of protein-like energy landscapes serve as a more complex illustrative example to evaluate the free energy methodology and further explore force-modulation of structure–function relationships. Figure 2-3B demonstrates robust, high-accuracy estimates of the binding free energy change upon force application for systems with one thousand degrees of freedom and rugged energy landscapes. Comparable convergence properties were obtained for simulations of ten-thousand-dimensional energy landscapes with similar features (data not shown). For comparison, simulations of a protein domain with implicit solvation typically contain 1000–3000 atoms, each with 3 coordinate degrees of freedom. Accuracy of the free energy profiles simulated here relies on adequate sampling of the phase space accessible to the system, and in particular of the force-induced conformational rearrangement. Our implementation makes use of the replica exchange method, which increases sampling efficiency through coupling with high-temperature simulations. Analysis of our simulation data illustrates that force-modulation of the binding free energy arises from differential responses of bound and unbound states upon mechanical stimulus. More-

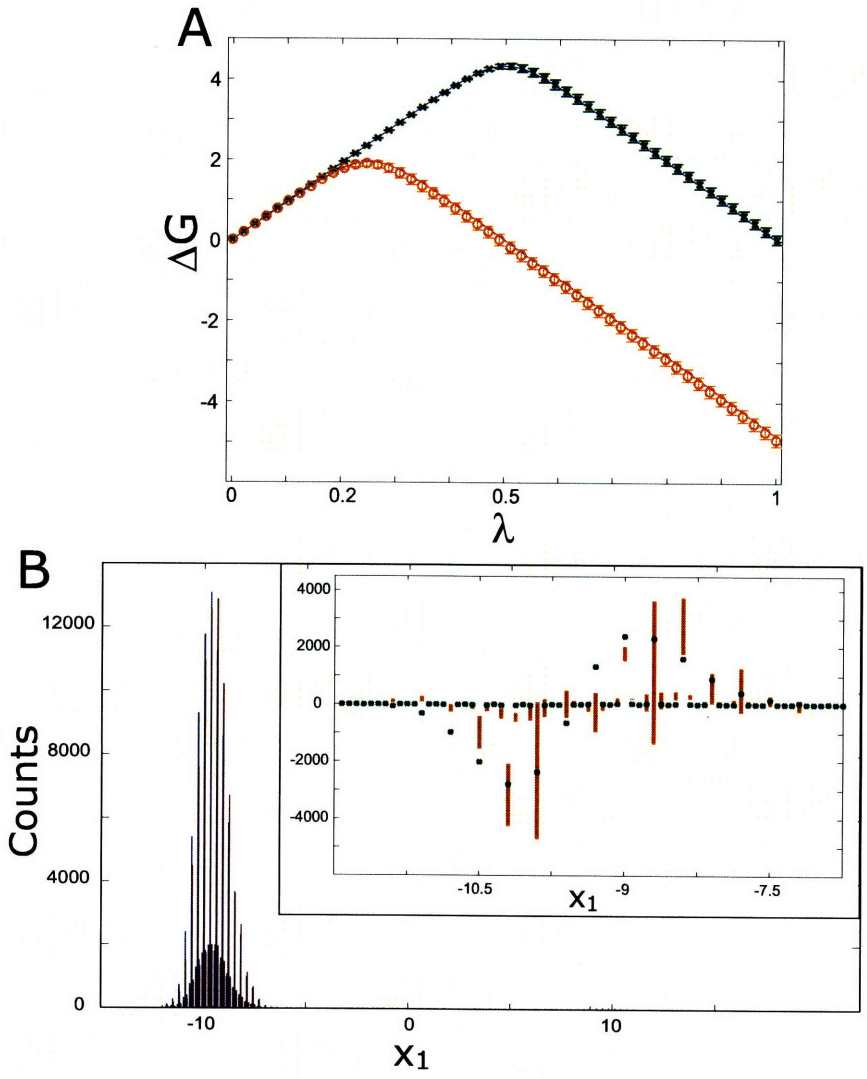


Figure 2-4: Moderate force application results in local structural ensemble shift for theoretical system. (A) Free energy change as a function of applied force for bound and unbound states. Error bars represent the standard error computed over 5 independent sliding windows of data. Color scheme is the same as Figure 2-3A. (B) Histograms of conformations along x_1 coordinate, for the bound state with $\lambda = 0$ (blue) and $\lambda = 0.25$ (red). Inset: difference in histogram counts, for $(\lambda = 0.25) - (\lambda = 0)$, are computed from simulation (orange boxes; length indicates standard deviation), and theory (green crosses). Positive counts indicate conformations sampled more frequently upon force application.

over, significant free energy changes in either state can be brought upon by subtle, yet systematic, shifts in the distribution of conformations at a fine scale, as illustrated in Figure 2-4.

2.2.3 FAT–paxillin binding interactions

As an atomically realistic application, we applied our free energy simulation methodology to explore force-modulated properties of a binding motif commonly found in proteins involved in cellular adhesion machinery. FAK and paxillin both act as scaffold proteins in focal adhesions, and the binding interaction between the C-terminal domain of FAK, referred to as FAT, with paxillin is responsible for FAK localization to focal adhesions. Moreover, its apparent structural features, where a helical peptide inserts itself into a four-helix bundle, are observed in many other protein–protein interactions involved in adhesion machinery. As such, FAT–paxillin binding is a paradigmatic example to study force regulation in a biologically relevant setting.

We performed independent simulations of the FAT–paxillin complex and FAT in its unbound state, with a variety of force magnitudes applied to FAT. Figure 2-5 shows the crystal structure of FAT complexed with a peptide from paxillin; the paxillin LD4 peptide binds to the surface of FAT helices 2 and 3. The applied mechanical force was distributed across the binding interface and the atoms subject to external force are shown as blue spheres in Figure 2-5. Free energy curves as a function of applied force were computed for each state through WHAM analysis, and the change in binding free energy was determined as the difference in the bound and unbound states curves (see Methods). We analyzed force effects on both structural and energetic characteristics.

All the simulations performed maintained the basic fold of the four-helix bundle, as well as binding interactions in the bound state. Figure 2-6A shows the root-mean-square deviation (RMSD) with respect to the starting structure for the bound state, for selected values of applied forces. The RMSD quickly increases as a function of

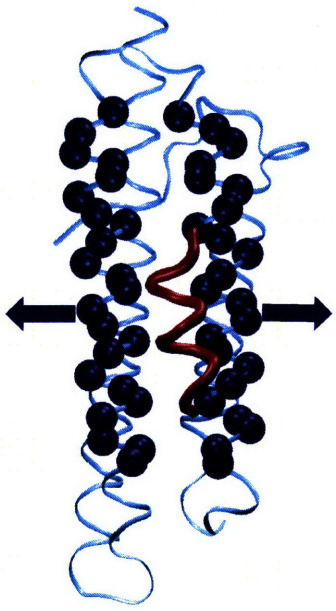


Figure 2-5: Crystal structure of FAT domain (cyan) of FAK binding to the paxillin peptide (purple) (PDB id: 1OW6). The Cα atoms to which mechanical force was applied are shown as blue spheres. Arrows indicate the direction of the applied force.

time, until it reaches a plateau after about 1 ns for all simulations. The value of the equilibrium RMSD depends on the applied force, because larger force magnitudes result in greater conformational change. The trend was similar for all trajectories, both in the bound and unbound state. We note that the RMSD exhibits fluctuations around its mean value, rather than systematically growing with time. To assess the degree of foldedness of the simulations, we examined the proportion of secondary structure elements. Figure 2-6B shows the fraction of residues in α-helices [78] as a function of applied force for the bound and unbound state. It is clear that the secondary structure is conserved upon force application. Figures 2-6C and D show snapshots after 10 ns of simulation for the bound and unbound states at representative force magnitudes. The configurations illustrate the overall stability of our simulations, together with the small force-induced conformational change investigated here.

We then analyzed the force effects on the free energy profiles of the bound and unbound states. Figure 2-7A shows the free energy of the bound state as a function of

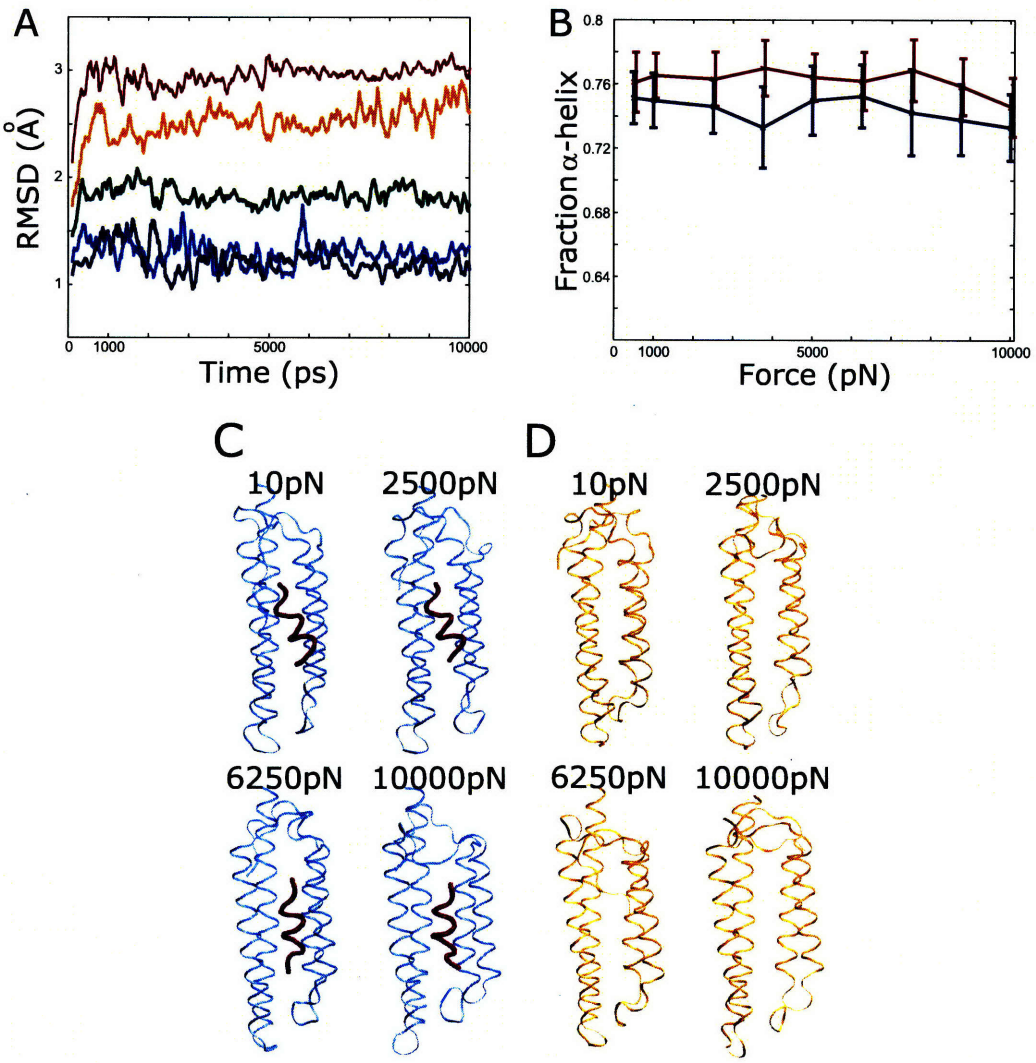


Figure 2-6: Features of FAT–paxillin pulling simulations. (A) RMSD with respect to energy-minimized crystal structure is shown for the FAT–paxillin bound complex as a function of time. The RMSD was computed over backbone atoms of the 4 helices (i.e., excluding loops). Individual curves represent independent constant force simulations performed with $\lambda F_{\text{appl}} = 10$ pN (blue), 2,500 pN (cyan), 5,000 pN (green), 7,500 pN (orange) and 10,000 pN (red). Due to the nature of replica-exchange simulations, curves shown correspond to running averages over 100 frames, for clarity. (B) Fraction of residues in α-helices [78] as a function of applied force. Bound (blue) and unbound (red) states curve are shown with averages and fluctuations (standard deviation). The standard error is less than 0.4%. (C-D) Snapshots after 10 ns of simulation are shown for selected force magnitudes for bound (C) and unbound (D) states.

applied force. Data were computed over 5 ns of production simulation, and error bars (twice the standard error) were estimated from 5 independent sliding windows of 1 ns data collection. The total change, for $F_{\text{appl}} = 10,000$ pN, is computed to be $-2,022 \pm$ 3.4 kcal/mol. Figure 2-7B shows the corresponding binding free energy change. The relative error bars are quite small, given the large magnitude of the actual free energy change for each state individually. Overall, the simulations suggest force-induced strengthening of FAT–paxillin interactions. There is an initial rapid drop in $\Delta\Delta G_{\text{bind}}$ as a function of applied force until $\lambda F_{\text{appl}} = 1,000$ pN, followed by a plateau around $\Delta\Delta G_{\text{bind}} = -9$ kcal/mol until $\lambda F_{\text{appl}} = 3,000$ pN. $\Delta\Delta G_{\text{bind}}$ then decreases smoothly, to reach a broad plateau around $\Delta\Delta G_{\text{bind}} = -16$ kcal/mol, for forces between 5,000 and 10,000 pN. The error bars for $\Delta\Delta G_{\text{bind}}$ were less than 3.5 kcal/mol across the applied force range. We note that the force magnitude applied here is perhaps artificially large, due to a restraining potential added to the simulations, necessary for structural stability in the absence of applied force (see Methods). However, we expect the general trends observed to be robust.

To gain insight into the structural features giving rise to the observed binding free energy change, we performed a detailed analysis of force-induced conformational changes. In particular, we focused on state-specific features and investigated how they contribute to the binding energetics. First, we monitored the average distance between the centers of mass of helices 2 and 3. For both states, we observed a smooth increase in the mean distance with force application. For instance, in the bound state the mean distance is 11.1 ± 0.21 Å with $\lambda F_{\text{appl}} = 500$ pN, and 17.3 ± 0.31 Å with $\lambda F_{\text{appl}} = 10,000$ pN. Figure 2-8A shows an overlay of representative snapshots of the unbound state at selected applied forces. Helices 2 and 3 are pulled apart from each other in a continuous fashion as increasing force is applied. Figure 2-8B shows the distributions of the distance between the centers-of-mass of helices 2 and 3 in the bound state, with warmer colors representing increased force magnitude. Figure 2-

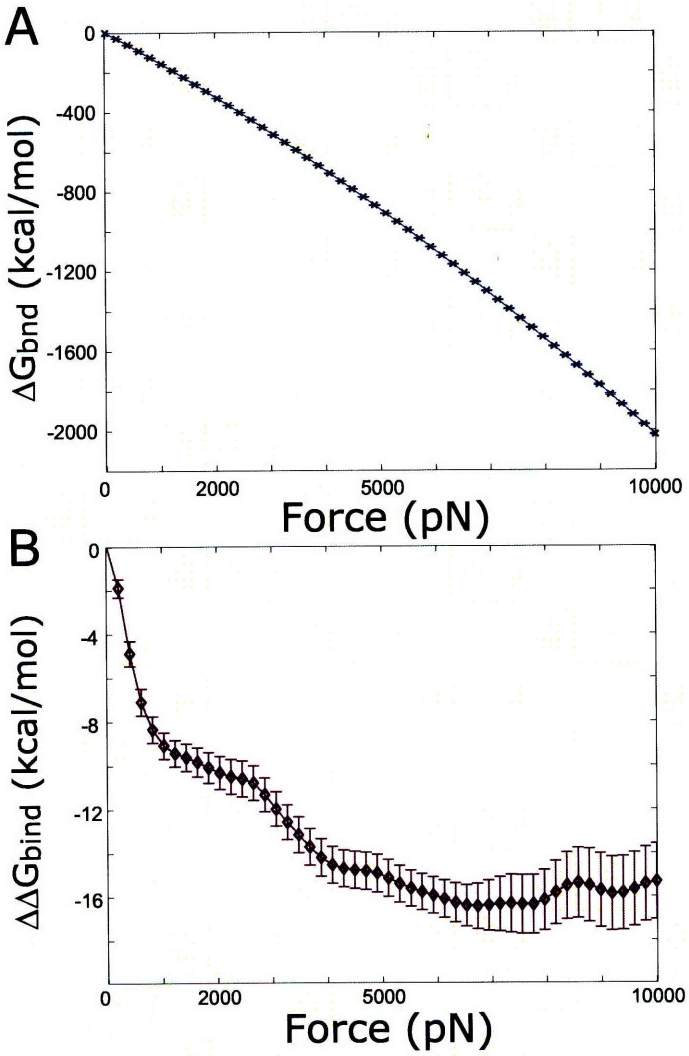


Figure 2-7: Free energy profiles as a function of applied force, for FAT–paxillin simulations. (A) Bound state, (B) binding free energy change. Results are reported in kcal/mol. The average value is computed over 5 ns of data. Error bars represent the standard error estimated with 5 independent sliding windows of 1 ns of data.

8C shows the same analysis for the unbound state simulations. We note that here, unlike in our previous simulations of idealized systems, both states undergo a smooth and comparable force-dependent distortion. Figure 2-8D shows the difference in the histogram counts between the bound and unbound states. For a given applied force, positive counts indicate conformations sampled more frequently in the bound state. Despite a global trend that is similar in the distributions, we observe a systematic shift of the bound state ensemble toward more extended conformations. This trend diminishes as the force magnitude is increased, with the distributions overlapping almost perfectly when $\lambda F_{\text{appl}} = 10,000$ pN. As described in Figure 2-4, even a subtle, yet systematic shift in ensemble distributions can contribute significantly to free energy changes. Hence, the somewhat more extended distribution of the bound state could participate in the observed force-induced strengthening of binding.

As the helices are pried further apart, the hydrophobic core of the four-helix bundle becomes exposed in the unbound state. In the bound state, despite similar increase in helix 2–helix 3 distance, the bound peptide covers those residues, preventing their exposure. Figures 2-9A and B show representative structures of the bound state with $\lambda F_{\text{appl}} = 10$ pN and 10,000 pN, colored by residue type. Figures 2-9C and D show the same analysis for the unbound state. In Figure 2-9D, appearance of hydrophobic (white) residues, which were initially buried at low applied force (Figure 2-9C), indicates the exposure of hydrophobic atoms in the unbound state. In contrast, despite conformational rearrangements, the same residues remain covered by the paxillin peptide in the bound state. To quantify this phenomenon, we computed the solvent-exposed surface area (SASA) over the hydrophobic residues making up the core of the helix-bundle structure (defined as residues with average SASA less than 1.5 Å^2 in the crystal structure). Figure 2-9E shows the SASA over these residues as a function of applied force, for the bound and unbound states. In the unbound state, the SASA shows a statistically significant increase, from 300 ± 111 Å^2

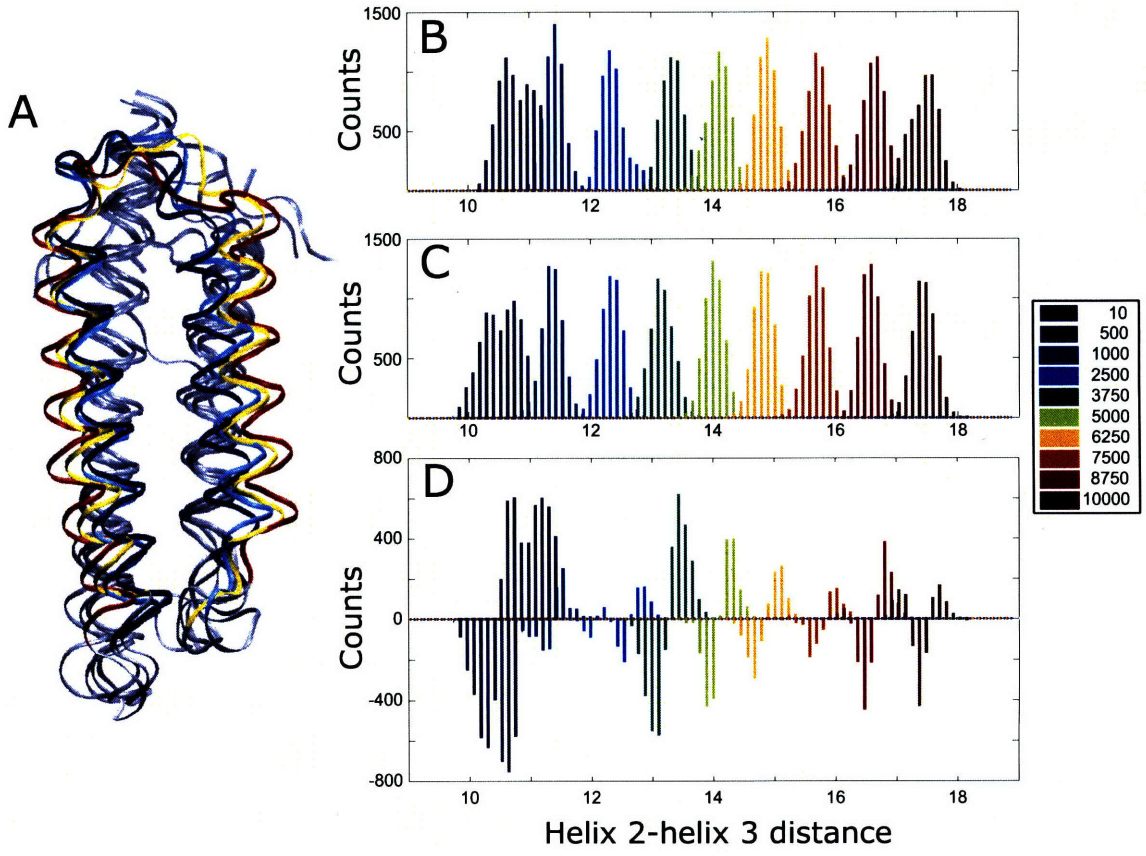


Figure 2-8: Force-induced increase in average helix 2–helix 3 distance. (A) Representative FAT structures from unbound simulations at various forces applied: $\lambda F_{\text{appl}} = 10$ pN (blue), 2,500 pN (cyan), 6,250 pN (yellow) and 10,000 pN (red). Snapshots were selected for which helix 2–helix 3 distance corresponds to the average distance at each applied force. (B-D) Histogram analysis of distance between centers of mass of helices 2 and 3. Histograms (distance in Å) for bound (B) and unbound (C) states. (D) Difference in histogram counts (bnd) – (unb). For each force applied, positive values indicate conformations sampled more frequently in the bound state simulation.

for $\lambda F_{\text{appl}} = 500$ pN to 646 ± 84 Å^2 for $\lambda F_{\text{appl}} = 10,000$ pN. In contrast in the bound state, the SASA remains constant over the force range applied, varying between 254 ± 56 Å^2 for $\lambda F_{\text{appl}} = 500$ pN and 321 ± 64 Å^2 for $\lambda F_{\text{appl}} = 10,000$ pN. The standard error is less than 5.1 Å^2 in all cases.

Increased helix 2–helix 3 distance also allows the paxillin peptide to interact more closely with FAT in the bound state. Figure 2-10 shows characterization of peptide embedding through various metrics. Figures 2-10A and B show representative structures of the bound complex when $\lambda F_{\text{appl}} = 10$ pN and $\lambda F_{\text{appl}} = 10,000$ pN, with a top view to highlight the binding interface. Note an increased distance between helices 2 and 3 making up the binding interface, as discussed previously, together with closer interactions between peptide and protein core. Figure 2-10C shows quantification of peptide embedding through analysis of the distance between the paxillin peptide and the more distant helices 1 and 4. We observe a smooth decrease in the average distance between the peptide and both of these helices. In particular, peptide–helix 1 distance decreases from 19.1 ± 0.33 Å for $\lambda F_{\text{appl}} = 500$ pN to 15.5 ± 0.97 Å for $\lambda F_{\text{appl}} = 10,000$ pN. The standard error was less than 0.45 Å. We define the binding interface area as minus the difference between the complex SASA, and the SASA of each binding partner in the same conformation, but in an unbound state. Figure 2-10D shows an increase in the binding interface area as a function of applied force. Specifically, the binding interface area is $1,278 \pm 44$ Å^2 for simulations run with $\lambda F_{\text{appl}} = 500$ pN, and $1,636 \pm 79$ Å^2 with $\lambda F_{\text{appl}} = 10,000$ pN. The standard error was less than 4 Å^2.

Finally, we analyzed the contribution of polar interactions on force-modulated binding affinity. We monitored the distances of all possible hydrogen-bonding interactions between protein and peptide polar atoms. We selected a distance cutoff of 2.5 Å between polar oxygen and hydrogen atoms to define a contact as present. We then computed, for each polar residue on FAT making interactions with paxillin pep-

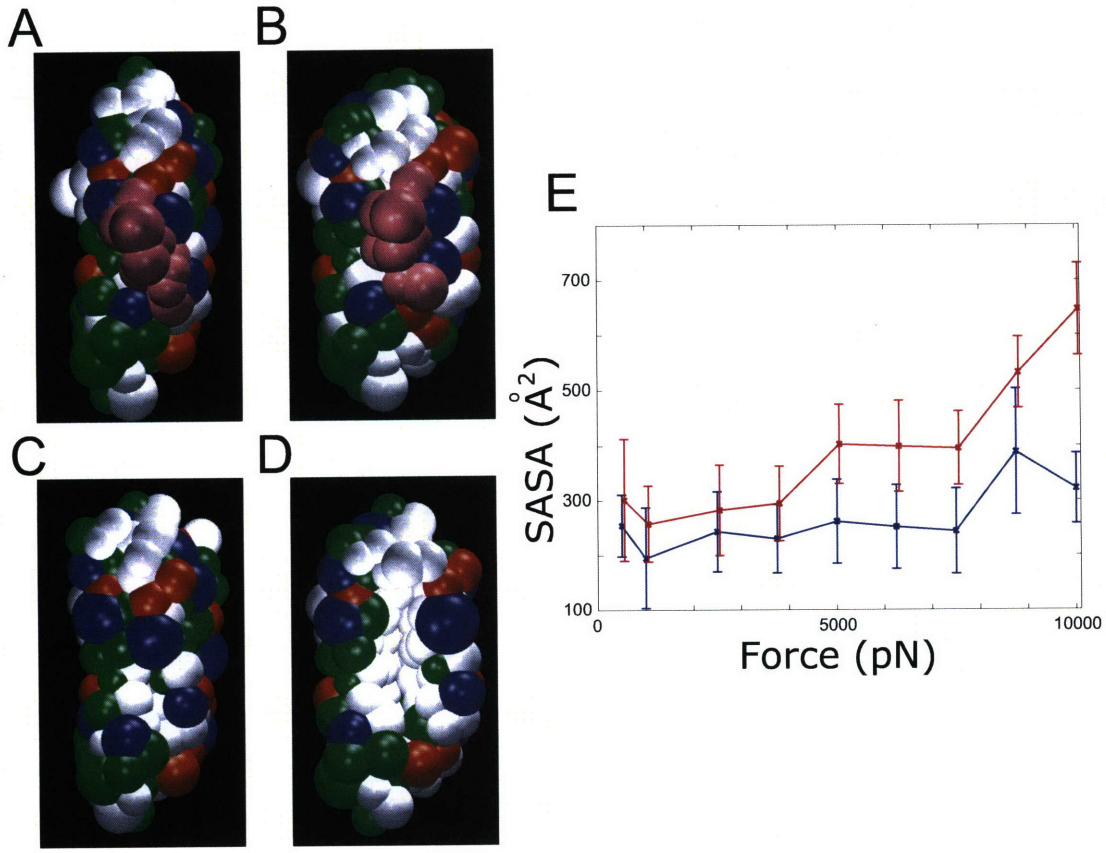


Figure 2-9: Applied force induces exposure of hydrophobic core residues in the unbound state only. (A-D) Characteristic structures, colored by residue types, are shown with one sphere per side chain. Hydrophobic residues are shown in white, polar in green, positively charged in blue and negatively charged in red. The paxillin peptide in shown in pink for the bound state structures. Bound (A-B) and unbound (C-D) states are shown with applied forces $\lambda F_{\text{appl}} = 10$ pN (A and C) and $\lambda F_{\text{appl}} = 10,000$ pN (B and D). Snapshots for which the SASA over hydrophobic core is close to the average at each applied force were selected for illustration. Exposure of the hydrophobic core in the unbound state is apparent through revealing white-colored residues. (E) SASA computed over hydrophobic residues making up the core of the four-helix bundle is plotted as a function of applied force. Bound (blue) and unbound (red) state curves are shown with average and fluctuations (standard deviation) for each applied force. The standard error is less than 5.1 Å^2 in all the simulations.

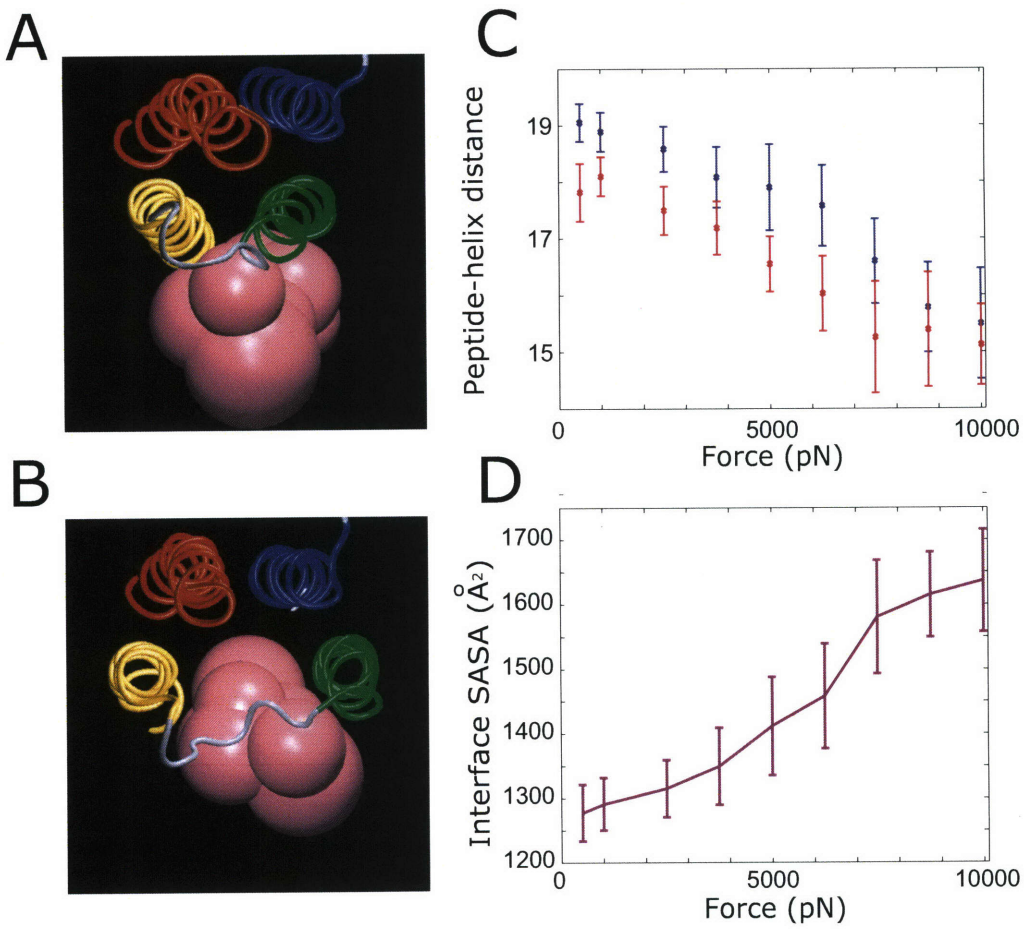


Figure 2-10: Force increases peptide embedding in FAT–paxillin bound state. (A-B) Representative structures of FAT–paxillin binding interface for simulations with $\lambda F_{\text{appl}} = 10$ pN (A) and $\lambda F_{\text{appl}} = 10000$ pN (B). FAT is shown in cartoon representation, with helices 1 (blue), 2 (green), 3 (orange) and 4 (red) individually colored. Paxillin peptide is shown in pink, with one sphere per side chain to highlight the binding surface. (C) Center-of-mass distances as a function of applied force for paxillin–helix 1 (blue) and paxillin–helix4 (red) are shown with average and fluctuations (standard deviation). The standard error is less than 0.45 Å. (D) Binding interface area as a function of applied force. Average is shown with fluctuations; the standard error is less than 4 Å^2.

tide, the fraction of time this residue was involved in polar contacts. Figure 2-11A shows the contact fraction as a function of applied force for the four polar residues at the FAT–paxillin binding interface. Residues Lys 955 and Arg 962 both exhibit a high fraction of polar contact over the whole force range. Interestingly, structural rearrangements lead to a shift in the specific residues that these protein residues interact with, despite roughly constant contact number. In particular for Arg 962, contacts are made predominantly with the paxillin atoms $O_{\delta 1}$ and $O_{\delta 2}$ from residue 6 and the backbone oxygen atom from residue 2 at low forces, with fractions of $90.3 \pm 1.9\%$, $88.4 \pm 0.3\%$, and 100%, respectively, at $\lambda F_{\text{appl}} = 10$ pN. Upon force application of $\lambda F_{\text{appl}} = 10,000$ pN, the contribution of oxygen 6 $O_{\delta 2}$ decreases to $44.6 \pm 6.8\%$, correlated with an increased fraction of contacts with oxygen 2 $O_{\gamma 1}$, from $28.5 \pm 3.9\%$ at $\lambda F_{\text{appl}} = 10$ pN to $68.3 \pm 10.3\%$ at $\lambda F_{\text{appl}} = 10,000$ pN. In contrast, residues Asn 991 and Lys 1002 are initially very far from any polar group on the peptide, with respective minimum distances to peptide polar atoms of 8.3 Å and 5.16 Å, thereby preventing them from making polar contacts with the peptide. Upon force application, increased peptide embedding enables those residues to become closer to the peptide, allowing the formation of binding interactions. As shown in Figure 2-11A, the fraction of contacts for Asn 991 increases from $6.7 \pm 0.8\%$ to $59.5 \pm 12\%$ as force magnitude increases from $\lambda F_{\text{appl}} = 10$ pN to $\lambda F_{\text{appl}} = 10,000$ pN, whereas that for Lys 1002 grows from $17.6 \pm 6.5\%$ to $47.7 \pm 10.6\%$. Figures 2-11B and C illustrate the conformational change in the position of these residues and interacting side chains on the paxillin peptide between the crystal structure and a representative snapshot at $\lambda F_{\text{appl}} = 10,000$ pN. Hence, our simulations suggest that polar interactions involving FAT residues 991 and 1002 are specifically force-activated. To further evaluate the contribution of these residues to the observed force-induced strengthening in binding affinity, we performed additional simulations for a mutant FAT. We computationally mutated residues 991 and 1002 to alanine to abolish their capacity to establish polar

interactions. Simulations were carried out for both bound and unbound states for the mutant FAT, with the same force application procedure as our wild-type calculations. Figure 2-11D shows the resulting binding free energy profile, overlaid with the wild-type curve (identical to Figure 2-7B), for comparison. Overall, the mutant profile shows the same features as wild type, except that the actual magnitude of the binding free energy change is smaller. Our simulations suggest that, for the double-mutant FAT, force still induces binding strengthening; however, this effect is less pronounced than for the wild-type system. Specifically, the total binding free energy change when $\lambda F_{\text{appl}} = 10,000$ pN is -7.93 ± 2.1 kcal/mol for the mutant, whereas it is -15.3 ± 1.7 kcal/mol for the wild type. Uncertainties correspond to the standard error estimated from 5 independent sliding windows of 1 ns of data.

Pulling simulations on a prototypical protein system allowed us to investigate how force application can affect structure–function relationship in a biologically relevant context. The calculations suggest that, upon mechanical distortion of helices 2 and 3, FAT–paxillin binding affinity increases. The protein simulations illustrate several general mechanisms responsible for this force-modulated binding affinity change. In particular, exposure of hydrophobic core in the unbound state destabilized the unbound state with respect to the bound, thus resulting in a binding enhancement. Force-induced activation, or inhibition, of specific polar interactions can also contribute to force-dependent binding free energy profile. In particular, our mutant simulations support the role of two polar FAT residues, Asn 991 and Lys 1002, on force regulation of binding affinity, together with other mutant-independent effects. Such a mutant analysis is particularly valuable, since it allows formulation of specific hypotheses that could be more easily tested experimentally, for instance with single-molecule setups such as optical tweezer experiments [10].

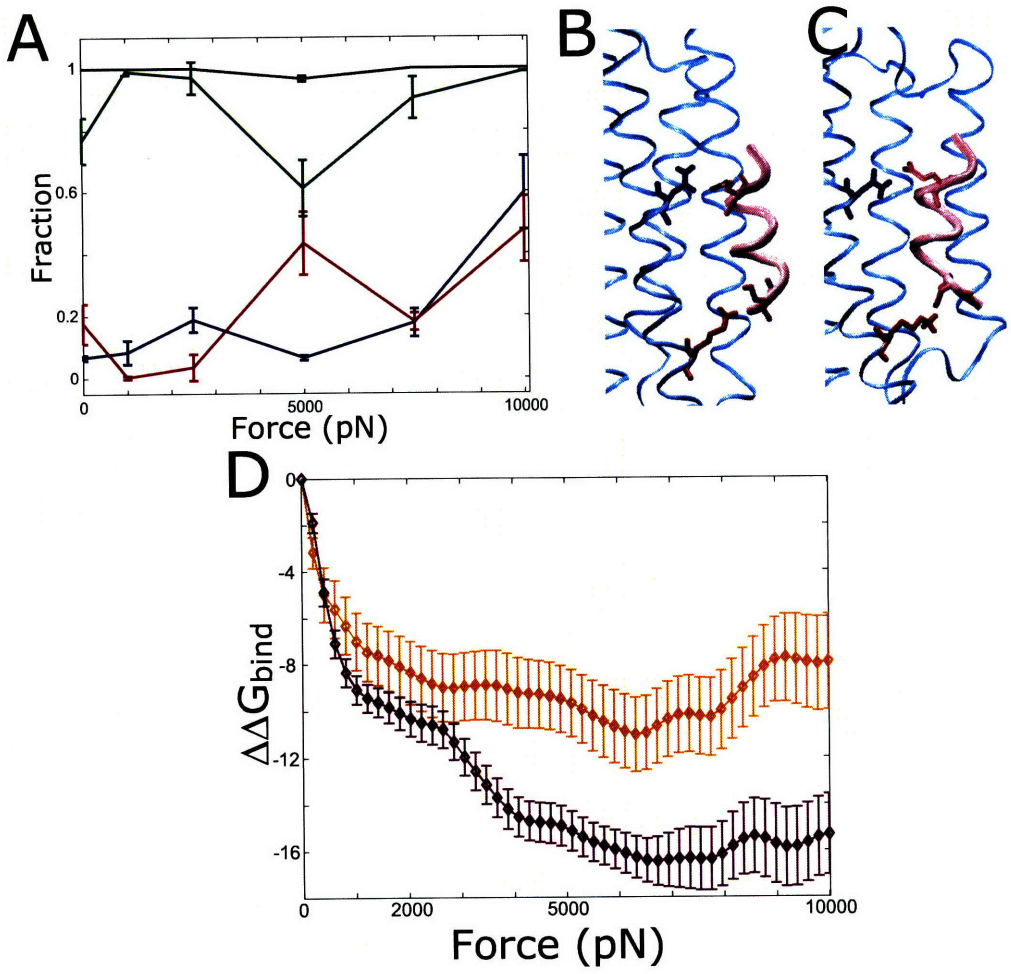


Figure 2-11: Force induces activation of specific polar contacts in the FAT–paxillin complex. (A) Fraction of simulation time polar contacts are made between individual protein residues and the paxillin peptide is shown with fluctuations (standard deviation). FAT residues Lys 955 (green), Arg 962 (black), Asn 991 (blue), and Lys 1002 (red) are shown. (B-C) Binding interface of the FAT–paxillin complex is shown with FAT residues Asn 991 (blue) and Lys 1002 (red) and paxillin polar residues Glu 4 and Ser 11. (B) Crystal structure, (C) representative snapshot with $\lambda F_{\text{appl}} = 10,000$ pN, for which both Asn 991 and Lys 1002 are engaged in polar contacts with the paxillin peptide. (D) Binding free energy change as a function of applied force is computed for wild-type (purple) and double-mutant (orange) FAT. Double-mutant has residues Asn 991 and Lys 1002 both mutated to alanines. Results are reported in kcal/mol. The average value is computed over 5 ns of data. Error bars represent the standard error estimated with 5 independent sliding windows of 1 ns of data.

2.3 Conclusion

We presented an integrated approach to study effects of moderate mechanical force on protein structure and associated binding properties. The levels of applied force were sufficient to induce conformational distortion without complete unfolding, and the statistical mechanical treatment included computational estimation of the thermodynamic stability of the complex as a function of applied force. Our objectives were two-fold: to gain insight into general force-modulation principles, and to illustrate the application of free energy methodology to computationally investigate force-regulated binding affinity. Binding affinity is a balance between bound and unbound state free energy. Here, we showed, in various contexts, that force-modulated changes in binding affinity can arise through limited conformational rearrangements of protein structure with differential energetic consequences in the bound and unbound state. Simple mechanical systems illustrated the fact that binding affinity can be modified by force application in cases when no conformational change occurs in one state. Moreover, force effects on protein energetic properties are highly system dependent, both in terms of the general force-affinity profile and the structural mechanisms involved, whether through a subtle conformational rearrangement or partial unfolding. Through design and simulations of idealized theoretical systems with protein-like features, we demonstrated robust convergence of the free energy protocol in energy landscapes of complexity comparable to that of protein applications. We also showed that subtle redistribution of conformational ensembles within major conformational states can lead to a significant free energy change for individual systems. In FAT–paxillin protein simulations, the bound-state distributions of helix 2–helix 3 distances are skewed to the right of those for the unbound state; we propose that this effect contributes to a certain extent to the observed strengthening in binding affinity.

Several modular proteins, such as the extracellular protein fibronectin, have been hypothesized to partially unfold upon force application, thereby revealing cryptic

binding sites [6,69]. Here, we focused on more limited conformational change, which could be a general mechanosensing scheme [63, 65, 66, 73]. Our protein simulations suggest force-induced strengthening of FAT–paxillin interactions; this could be a contributing factor to the experimentally observed force-stimulated assembly of focal adhesions [44]. Simulations show diminished effect in a double-mutant of FAT that inhibits specific force-activated polar contacts. Finally, we anticipate force direction to be an important factor for the force sensing and integration machinery. As such, different pulling directions could result in different associated conformational changes, triggering alternative responses in affinity and related biochemical signaling. Higher-order signal integration could be achieved if a given protein were subjected to simultaneous mechanical stimuli, leading to either cooperative or anti-cooperative effects in downstream signaling. Molecular mechanisms of mechanotransduction are just beginning to be investigated, and we expect molecular dynamics simulations to make valuable contributions to gaining detailed understanding of the phenomena involved. The free energy methodology is easily applicable to other biologically relevant systems, and expands the modeling toolset for projects focusing on understanding force-regulated structure–function relationships.

2.4 Computational methods

2.4.1 Simple mechanical models

We designed three block-and-spring models of protein-ligand complexes that capture distinct force-transducing properties. Each binding partner has one degree of freedom: the protein translates in the x direction and the ligand in the y direction, each under the influence of a harmonic spring. The potential energy depends on the

conformational variable x and on the force F applied to the protein block as follows:

$$U(x, F) = \frac{k_1}{2}x^2 + \frac{k_2}{2}h_i^2(x) - F \cdot x \tag{2.8}$$

where $h_i(x)$ describes the specific geometry of the system considered. This quantity is model-dependent, and also distinguishes the bound state, for which protein and ligand are allowed to complement each other, from the unbound state.

For all the systems considered, the unbound state is described with: $h_{\text{unb}}(x) = h_o + H$, i.e. the ligand is prevented from binding with the protein block, as shown in Figure 2-1B. We now present the specific shapes of the bound state for three distinct systems.

Scenario 1. The first system represents a protein with only one stable conformational state that is binding-competent at rest. It is made up of a single notch into which the ligand shape is able to fit in the absence of applied force, as shown in Figure 2-1A:

$$h_1(x) = \begin{cases} h_o + |x| \tan\alpha & \text{if } |x| \leq \frac{H}{\tan\alpha} \\ h_o + H & \text{otherwise} \end{cases} \tag{2.9}$$

Using Equation 2.9 in 2.8, we compute the position of the global minimum in potential energy $x_{\text{bnd}}(F)$. The corresponding potential energy $U_{\text{bnd}}(x_{\text{bnd}}(F)) = U(x_{\text{bnd}}(F), F)$ is plotted in Figure 2-2A.

Scenario 2. The second system represents a protein that partially unfolds and reveals a cryptic binding site upon force application. It is also composed of a single notch, albeit displaced from the resting configuration of the protein block, as shown

in Figure 2-1C:

$$h_2(x) = \begin{cases} h_o + |x-d| \tan\alpha & \text{if } |x-d| \leq \frac{H}{\tan\alpha} \\ h_o + H & \text{otherwise} \end{cases} \tag{2.10}$$

Scenario 3. The last system captures the behavior of an allosteric protein, for which the allosteric transition is activated in the presence of a force. It is made up of two separate notches, each one corresponding to one main conformational state, as shown in Figure 2-1D:

$$h_3(x) = \begin{cases} h_o + |x| \tan\alpha & \text{if } |x| \leq \frac{H}{\tan\alpha} \text{ (state } \mathcal{C}) \\ |x-d| \tan\alpha & \text{if } |x-d| \leq \frac{H+h_o}{\tan\alpha} \text{ (state } \mathcal{E}) \\ h_o + H & \text{otherwise} \end{cases} \tag{2.11}$$

Binding energy. For each applied force, we independently compute the position of the global minimum in potential energy for the bound $x_{\text{bnd}}(F)$ and unbound $x_{\text{unb}}(F)$ states. The binding energy $\Delta U_{\text{bind}}(F)$ is computed through Equation 2.1 and plotted in Figures 2-2B, D and F for each scenario.

2.4.2 Free energy calculations

The multidimensional replica-exchange method [75] was adapted for the case where an applied force is the perturbing factor driving the free energy change; we briefly present here the relevant equations. In this framework, the modified potential energy upon force application is given by Equation 2.6. The replica-exchange simulation is performed for M replicas with different values of the parameters $\Lambda_m = (\beta_m, \lambda_m)$, which traverse temperature and force space. Here $\beta_m = \frac{1}{k_B T_m}$, where k_B is the Boltzmann constant, and T_m the absolute temperature of replica m. Discrete values of λ between 0 and 1 effectively correspond to values of applied force between 0 and a target value of magnitude $||\vec{F}_{\text{appl}}||$. We arrange the replicas so that there is a one-

to-one correspondence between replicas and parameter set: the label i ($i = 1...M$) for replicas is a permutation of the label m ($m = 1...M$) for parameter values. We define a conformation $\overrightarrow{X}$ in the generalized ensemble as the collection of configurations $\{\overrightarrow{x}_m^{[i]}\}$ for all the replicas. With this notation, the configuration of replica m at a specified time is $\overrightarrow{x}^{[i]}$. With a user-defined frequency, we attempt an exchange of a pair of replicas in the generalized ensemble. Suppose we exchange replicas i and j which are at parameter values $\Lambda_m = (\beta_m, \lambda_m)$ and $\Lambda_n = (\beta_n, \lambda_n)$, respectively,

$$\overrightarrow{X} = \{..., \overrightarrow{x}_m^{[i]}, ..., \overrightarrow{x}_n^{[j]}, ...\} \rightarrow \overrightarrow{X'} = \{..., \overrightarrow{x}_m^{[j]}, ..., \overrightarrow{x}_n^{[i]}, ...\} \tag{2.12}$$

In order for this exchange process to converge toward the equilibrium distribution, it is sufficient to obey the detailed balance condition on the transition probability $w(\overrightarrow{X} \rightarrow \overrightarrow{X'})$,

$$w(\overrightarrow{X} \rightarrow \overrightarrow{X'}) = \begin{cases} 1, & \text{for} \quad \Delta \leq 0 \\ \exp(-\Delta), & \text{for} \quad \Delta > 0 \end{cases} \tag{2.13}$$

where

$$\Delta = \beta_m \left[U_{\lambda_m}\left(\overrightarrow{x}^{[j]}\right) - U_{\lambda_m}\left(\overrightarrow{x}^{[i]}\right)\right] - \beta_n \left[U_{\lambda_n}\left(\overrightarrow{x}^{[j]}\right) - U_{\lambda_n}\left(\overrightarrow{x}^{[i]}\right)\right] \tag{2.14}$$

In this framework, Equation 2.14 simplifies to

$$\Delta = (\beta_m - \beta_n)\left[U_0\left(\overrightarrow{x}^{[j]}\right) - U_0\left(\overrightarrow{x}^{[i]}\right)\right] - (\lambda_m\beta_m - \lambda_n\beta_n)\left[\overrightarrow{F}_{\text{appl}} \cdot (\overrightarrow{x}^{[j]} - \overrightarrow{x}^{[i]})\right] \tag{2.15}$$

To maximize exchange acceptance ratios, replica swapping is attempted only for adjacent parameter values, either in temperature or in force space. Simulation details are given for each system considered in the following sections.

We analyze data using the general framework of the weighted histogram analysis method (WHAM) [76, 77]. From the simulation, M potential-energy histograms

$N_m(U_0, U_F)$ are obtained, where the data are binned according to their internal U_0 and forcing U_F energies. The free-energy difference as a function of applied force, $\Delta G_l = G_{\lambda=l} - G_{\lambda=0}$, at temperature factor β is calculated from:

$$\exp(-\beta \Delta G_l) = \frac{Z_{\beta,\lambda=l}}{Z_{\beta,\lambda=0}} = \frac{\sum_{U_0,U_F} P_{\beta,\lambda=l}(U_0, U_F)}{\sum_{U_0,U_F} P_{\beta,\lambda=0}(U_0, U_F)} \tag{2.16}$$

where $P_{\beta,\lambda}(U_0, U_F)$ are obtained from the WHAM equations,

$$\begin{cases} P_{\beta,\lambda}(U_0, U_F) = \frac{\sum_{m=1}^{M} g_m^{-1} N_m(U_0,U_F)}{\sum_{m=1}^{M} g_m^{-1} n_m e^{f_m - \beta_m U_{\lambda_m}}} e^{-\beta U_\lambda} \\ e^{-f_m} = \sum_{U_0,U_F} P_{\beta_m,\lambda_m}(U_0, U_F) \end{cases} \tag{2.17}$$

where n_m is the number of analyzed samples obtained for the m-th parameter set Λ_m. Here, we assume that the parameters $g_m = 1 + 2\tau_m$, where τ_m is the integrated auto-correlation time, are equal for all replicas. We solve for the unnormalized probability distribution $P_{\beta_m,\lambda_m}(U_0, U_F)$ and the dimensionless free energies $\{f_m\}$ self-consistently by iteration [76, 77]. Analysis was carried out using a locally implemented C++ package and MATLAB software (The Mathworks, Inc., Natick, MA).

To quantify force-dependent binding properties, independent simulations were carried out for the bound (bnd) and unbound (unb) states. The free energy change as a function of applied force was estimated for each system using Equations 2.16 and 2.17. The binding free energy change as a function of applied force was then computed as

$$\Delta\Delta G_{\text{bind}}(F_{\text{appl}}) = \Delta G_{\text{bnd}}(F_{\text{appl}}) - \Delta G_{\text{unb}}(F_{\text{appl}}) \tag{2.18}$$

Results are reported as binding free energy change with respect to the zero-force value.

2.4.3 Idealized theoretical energy landscapes

We designed idealized potential energy landscapes with protein-like features to validate our free energy method and gain further insight into common force-regulation properties. Previous experimental characterizations and computational studies have led to the observation that protein energy landscapes are complex, with a hierarchical organization of structural states, from taxonomic, to statistical to few-level substates [79]. Protein energy landscapes thus feature a few broad basins in potential energy, separated by large energy barriers (taxonomic states). Each such state is further divided into many statistical substates, separated by smaller energy barriers, and the division continues in a fractal nature with well distance and energy barrier height decreasing as one walks down the tree structure. We incorporated data from experiments [80], providing estimates of energy barrier heights, and simulations [81,82], providing information about well-to-well distances, to create potential energy surfaces of the form:

$$U_0(\vec{x}) = \sum_{i=1}^{N} \left[B_i(x_i) + a_1 \sin\left(\frac{2\pi x_i}{T_1}\right) + a_2 \sin\left(\frac{2\pi x_i}{T_2}\right) \right] \tag{2.19}$$

where $\vec{x} = (x_1, x_2, ..., x_N)$ is the N-dimensional vector of coordinates for the degrees of freedom of the system, and $(a_k,\ T_k)$ $(k = \{1, 2\})$ the amplitudes and periods of sinusoids that describe two levels of conformational substates on the energy landscape. The set of functions B_i captures the taxonomic conformational substates of the system; for each degree of freedom i, B_i takes the form:

$$B_i(x) = \begin{cases} \alpha x^4 + \beta x^2 + \gamma x & \text{if } i \leq n_w \\ kx^2 & \text{otherwise} \end{cases} \tag{2.20}$$

where α, β, γ and k are system-dependent parameters and n_w is the number of dimensions bearing a double-well potential. Thus, the total number of taxonomic states for

a given system is 2^{n_w}. Multi-dimensional replica-exchange simulations were performed using Monte-Carlo moves [83] to update coordinates, with a locally developed C++ software. The simulations presented here correspond to one-thousand-dimensional energy landscapes, with $n_w = 2$; free energy estimates were averaged over 5 independent sliding windows, analyzed after an initial equilibration period. Parameters were as follows: $a_1 = 3$, $T_1 = 0.3$, $a_2 = 0.36$, $T_2 = 0.1$, $k = 2$, $\alpha = 0.002$, $\beta = 0.36$. γ was set to 0.5 for the bound state, and to 0.25 for the unbound state.

2.4.4 Protein simulations

The 2.35 Å crystal structure of the focal adhesion targeting (FAT) domain of focal adhesion kinase bound to paxillin LD motif [84] was obtained from the Protein Data Bank [85] (PDB id 1OW6). The starting structure for the bound state was taken from chains C and F, and for the unbound state from chain B, which is unliganded in the crystal structure. All crystallographic water molecules and ions were removed. Titration states and asparagine, glutamine, and histidine crystallographic uncertainties were resolved based on optimization of hydrogen-bonding in the side-chain local environments. All MD simulations were carried out using the CHARMM program [86] with the CHARMM 19 parameter set [15, 87] with the Effective Energy Function EEF1 solvent model [25], SHAKE algorithm for the bonds involving hydrogen atoms, and a 2-fs time step. Langevin dynamics with a friction coefficient of 50 ps^{-1} was used to control temperature. Trial simulations without restraints were unstable, leading to partial unfolding of the four-helix bundle core. Thus, a harmonic flat-bottom NOE restraint potential [88] was used to penalize stretching beyond the crystal structure value, for two types of contacts. The force constants were set to 5 kcal/(mol $\cdot$ $Å^2$) for helix hydrogen bonds, and 1 kcal/(mol $\cdot$ $Å^2$) for Cα–Cα distances, for which restraints were applied between each atom of the protein core and its nearest neighbor in an adjacent helix. The restraint strength was optimized to minimize disruption to

the force-induced conformational change, while maintaining protein structure in the absence of force application.

The replica exchange simulations were carried out using the Multiscale Modeling Tools in Structural Biology (MMTSB) tool set (available from `http://mmtsb.scripps.edu`) [89] and CHARMM. For each system (bound and unbound states), the crystal structure was subjected to rounds of constrained minimization to alleviate any initial clashes. We used 32 replicas exponentially spaced between 300 and 700 K; a replica exchange was attempted every 1 ps of MD simulation, resulting in an overall exchange ratio between 0.3 and 0.4. Each replica started from a conformation obtained after a 50 ps heating trajectory from 100 K to the target temperature of the replica.

Mechanical force was applied to 23 Cα atoms each of helices 2 (residues 950 to 972) and 3 (residues 980 to 1002), in opposing directions, to produce force-induced opening of the binding site, as shown in Figure 2-5. Independent 10-ns replica-exchange simulations with temperature-exchange only were carried out at 10 fixed values of applied force, ranging from 10 to 10,000 pN. The first 5 ns of simulations were designated as equilibration, and the subsequent 5 ns were kept for data collection. Results were combined using the WHAM methodology described previously. The average free energy change is estimated over the entire 5 ns of data collection. Standard error is computed over 5 independent windows of 1 ns of data each. Additional analysis with incorporation of data gathered from simulations with 2 additional applied force magnitudes showed proper convergence of the computed quantities (data not shown). Figures illustrating protein structures were generated with the programs VMD [90] and Raster3D [91].

Chapter 3

Computational design of affinity maturation for improved pretargeted radioimmunotherapy

Abstract

Pretargeted radioimmunotherapy decouples the pharmacokinetics of targeting and therapy by utilizing a bispecific antibody with molecular recognition sites for both a cancer antigen and a radioactive compound chelated by a small molecule hapten. Despite great potential for tumor treatment, effective application of this strategy has been limited by weak antibody–hapten binding. Y-DOTA radiometal complex exhibits desirable properties for use in pretargeted radioimmunotherapy. Here we carried out computational protein design to identify individual mutations predicted to improve the binding affinity of monoclonal antibody 2D12.5 to its ligand, Y-DOTA. Our calculations suggest several mutation opportunities for affinity maturation, distributed across the entire binding interface. Computational predictions will be tested experimentally to assess binding affinity improvement.

[0]All of the experimental protein engineering and characterization referred to in this collaborative work were performed by Kelly Davis and K. Dane Wittrup.

3.1 Introduction

Cancer is the leading cause of death in the United States for people younger than 85, with over 500,000 deaths per year. However, developing targeted, effective cancer therapies remains a challenge due to side effects and immune reactions. One possible therapeutic strategy uses monoclonal antibodies. Antibodies are proteins made by the immune system to specifically bind to individual target molecules, called antigens, from foreign pathogens. Monoclonal antibodies are multiple copies of identical proteins, as opposed to polyclonal antibodies, which consist of multiple proteins binding to a given target through different mechanisms. Antibodies can be engineered for specificity to surface receptors on cancer cells; they can then be introduced as a therapeutic agent either "naked" (i.e., unbound) or conjugated to drugs, toxins or radioactive metal atoms. In 1997, the first monoclonal antibody was approved by the FDA for use in cancer therapy, and thereafter eight more antibodies have been approved for the treatment of this disease [92]. Naked antibodies can target cancer cells for recognition and destruction by the immune system. As such, their tumor toxicity lies primarily in the intrisic ability of antibodies to trigger an immune response. By specifically binding to cancer surface receptors, they can additionally block their interactions with growth factors and signaling molecules, thereby reducing proliferation and metastasis.

Conjugated antibodies deliver toxic substances specifically to tumor cells, while minimizing toxicity to normal cells. Antibodies conjugated to radioactive metal atoms, or radionuclides, are used clinically for the treatment of non-Hodgkin's lymphoma [93]. Current radioimmunotherapy has shown great potential, but has been limited by toxicity. Because of low diffusivity into the tissues, antibodies have long half-lives in the blood, resulting in non-specific radiation of normal tissues. An alternative strategy, referred to as pretargeted radioimmunotherapy (PRIT), consists of separately administering the antibody and a chelated radionuclide. Hence, this

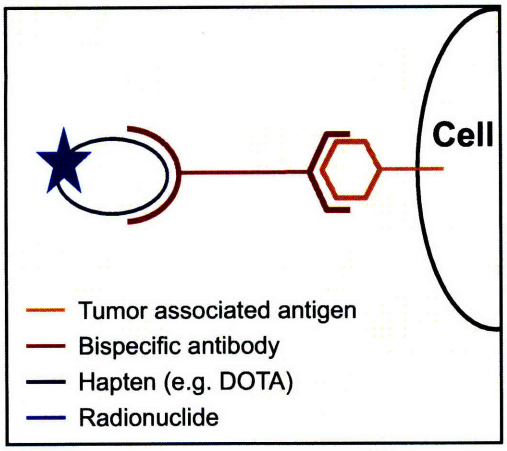


Figure 3-1: Illustrative schematic of pretargeted radioimmunotherapy. In a first stage, the bispecific antibody is administered and allowed to bind specifically to the tumor surface receptors. Subsequently, the radiometal–hapten complex is administered; tight binding allows hapten capture by the antibody, thereby targetting radiation to the tumor site.

approach decouples the pharmacokinetics of antibody targeting and radionuclide delivery, and has been shown to increase efficacy and decrease toxicity in both preclinical [94–97] and clinical models [98–100]. A key aspect of this technology is the use of a bispecific antibody with molecular recognition sites for both the cancer antigen and a chelated radionuclide, as shown in Figure 3-1. The antibody is administered first, and sufficient time is allowed for the antibody to specifically accumulate at the cancer site and clear from the blood stream and normal tissues. In the second step, the chelated radionuclide is administered and is captured only by the tumor-localized antibody. One of the main challenges in PRIT technology is efficient radionuclide capture, which requires tight binding between the antibody and the small molecule carrying the radionuclide. A hapten is a small molecule that is not recognized by the immune system, unless attached to a large carrier protein. The second pretargeting molecule that carries the radionuclide will be called a hapten.

For human clinical application, radiometal haptens require a number of characteristics. Diffusion must be rapid so that the radiometal can reach the targeted cancerous cells before excretion from the body. The plasma half-life must be short and

specific tissue accumulation minimal to reduce nonspecific radiation of healthy tissues. The radiometal–hapten complex must not be metabolized or transformed such that its affinity for the bispecific antibody is altered. Finally, there must be minimal inherent toxicity of the compound. Therefore, focusing on a molecule for which prior clinical investigation demonstrated those characteristics is valuable. DTPA (diethylenetriaminepentaacetate) and DOTA (tetraazacyclododecane-tetraacetic acid) are the two metal chelating agents most commonly used in radioimmunotherapy applications. Gadolinium DOTA (Gd-DOTA) has extensive clinical history as an NMR contrast agent and exhibits the desired pharmacokinetic properties in both rats and humans [101–103]. DOTA can chelate a variety of lanthanide metals with comparable and very tight affinity. An antibody binding to a DOTA analog specific for yttrium (Y) complexes has been made and characterized [104], and the crystal structure of the bound complex is available [105]. Among the metallic radionuclides for radioimmunotherapy, ^{90}Y is especially attractive due to its availability, pure high energy β-emission and the high dose yield per pmol [106]. Moreover, a model developed in the Wittrup lab was recently used to determine the therapeutic window for pretargeting with ^{90}Y [107]. Therefore, Y-DOTA exhibits promising properties for use as a radiometal hapten in pretargeted radioimmunotherapy.

One of the remaining challenges, however, is to re-engineer the existing monoclonal antibody for increased binding affinity to Y-DOTA. A single-chain variable fragment (scFv) molecule with picomolar affinity for the carcinoembryonic antigen (CEA), a protein used as a marker for certain types of cancer, has already been engineered in the Wittrup lab [108]. The two antibodies could then be synthesized into a bispecific molecule for use in PRIT. As mentioned above, radionuclide capture through high-affinity interaction is one of the key steps of PRIT still hindering application of this technology to cancer treatment. Recently, experimental efforts have been carried out in the Wittrup lab at MIT to improve antibody affinity. Starting with a library

composed of variations of an initial protein, directed evolution methods can select the proteins containing a desired function, for instance binding to a specific compound. Subsequently, mutagenesis is performed to mutate the selected proteins, and a second selection is conducted. This cyclic process is iterated, with increasingly stringent selection constraints until a protein with the desired properties is isolated. There are many different display technologies that can be used with directed evolution [11]. In particular, yeast surface display, which was developed in the Wittrup lab, enables linking genotype with phenotype during directed evolution [109]. It also provides efficient folding of proteins and can be used with fluorescence-activated cell sorting (FACS), a quantitative method of screening and sorting cells based on their fluorescence levels. When screening for improved ligand binding, cells displaying individual variants of the protein of interest on their surface are incubated with the ligand; conjugation of the ligand with a fluorescent dye enables fluorescence level to be correlated with the binding affinity of the protein expressed by that clone [110]. Detailed protocols have been developed for protein engineering utilizing yeast display [111], which was used in this project. A library of scFvs was created to increase the affinity of the existing monoclonal antibody, referred to as 2D12.5 antibody, to Y-DOTA. After several rounds of directed evolution, random mutagenesis through error-prone PCR [112], and selection, the observed improvement in binding affinity was still modest.

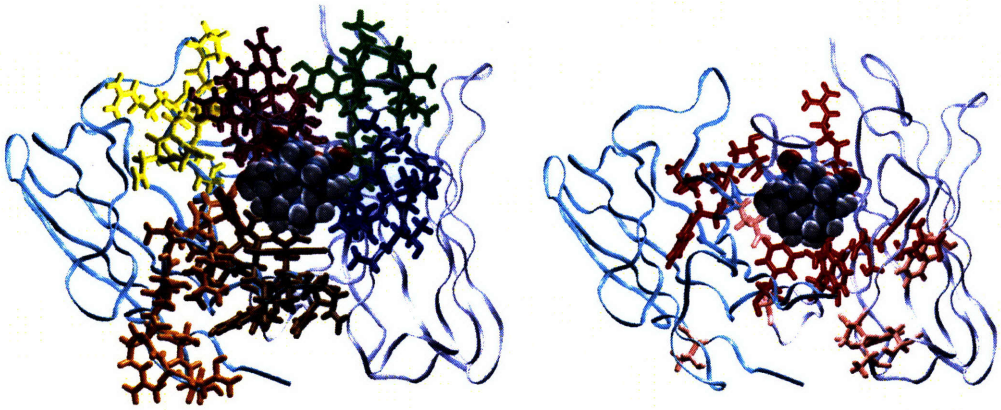
In contrast with random mutagenesis, computational protein design methods allow for the rational selection of mutations to improve binding affinity. Computational design relies critically on accurate energetic evaluation and thorough conformational search. In many cases, the design of improved antigen-binding affinity has shown limited success [113–116]. Challenges for protein–protein affinity design include conformational rearrangement upon binding, interfacial trapped water molecules, and accurate modeling of the trade-off of protein–solvent and protein–protein interactions

between the unbound and the bound states. Recently however, Lippow *et al.* [117] developed a new approach to computationally design protein–protein interactions for improved binding affinity. The method was applied to several antibody systems and tested experimentally. Overall, the authors achieved significant improvements for both model and therapeutic antibodies, demonstrating the robustness and general applicability of this computational design protocol.

The computational design procedure developed in-house by Lippow *et al.* was used to identify mutations predicted to improve 2D12.5 antibody binding affinity for Y-DOTA. Our design positions span all the complementary determining region residues in 2D12.5, together with additional positions that were either within a distance cutoff of the ligand, or found to be experimentally mutated during the directed evolution experimental maturation. Selected single mutations and cooperative double mutations predicted to improve binding affinity will be experimentally characterized in the Wittrup lab using site directed mutagenesis.

3.2 Results and Discussion

We attempted to redesign the antibody 2D12.5 for improved binding to its antigen, Y-DOTA. The effect of single mutation at 77 design positions was calculated using a two-stage hierarchical procedure, as described in Methods. As shown in Figure 3-2, the design positions included all of 57 non-proline complementary determining region (CDR) positions, and 20 additional positions belonging to the framework regions. These positions correspond to residues close to the Y-DOTA ligand, or residues found to be mutated in the directed evolution experiments. In the context of the complete design procedure and full energy function for side-chain conformational search, we used only the electrostatic term of the computed binding free energy to predict improvements in affinity, because it was shown to provide better correlation with



(a)				(b)	
CDR-H1	CDR-H2	CDR-H3	CDR-L1	CDR-L2	CDR-L3
H26-35	H50-58	H98-108	L23-36	L52-58	L91-99
green	blue	purple	orange	yellow	ochre

Figure 3-2: 2D12.5 design positions. Protein backbone is depicted by ribbons (ice blue: antibody heavy chain; cyan: antibody light chain). Y-DOTA is shown with van der Waals representation, in atom colors. (A) The 2D12.5 wild-type side chains at the CDR positions are color coded. (B) Additional design positions from the framework regions: positions nearby Y-DOTA ligand (at least one side-chain atom within 11 Å) are shown in red, and experimental design positions are shown in pink. All figures illustrating protein structures were generated with the programs Visual Molecular Dynamics (VMD) [90] and Raster3D [91].

experimentally measured binding affinity [117].

Overall, our predictions for binding affinity improvement include 11 single mutations across 9 design positions. Individual mutations predicted to enhance binding affinity fell into three main categories. Improvements can result from modification of the net balance of binding interactions of residues forming intermolecular contacts at the binding interface in two distinct ways. (1) The removal of unsatisfied existing interactions, for instance by eliminating an unsatisfied buried polar group, and (2) the introduction of a residue forming new favorable interactions constitute two means of improving affinity. A third mode of action, as described by Joughin *et al.* [118], is through noncontacting residues that can make substantial contributions to binding affinity, usually through changes in net charge. Table 3.1 displays the predicted ener-

Table 3.1: Predicted 2D12.5 single mutation binding affinities. We classified predictions based on their mode of action, i.e. contacting residues removing unsatisfied interactions (mode 1), contacting residues introducing new favorable interactions (mode 2), or noncontacting residues ("action-at-a-distance", mode 3).

Position		Mut'n	Mode of Action	$\Delta\Delta G^a_{\text{binding}}$ Elec	$\Delta\Delta G^a_{\text{binding}}$ Total	$\Delta\Delta G^a_{\text{folding}}$ Elec	$\Delta\Delta G^a_{\text{folding}}$ Total
L53	Asn	Ala	1	−0.28	+0.02	+0.38	−3.55
L93	Trp	Phe	1	−0.29	+0.66	+0.15	+4.68
H53	Ser	His	2	−0.59	−1.85	+3.66	+2.95
H100	Ser	Arg	2	−1.78	−4.89	+3.39	+0.79
L34	Tyr	Gln	2	−0.97	+0.15	+1.66	+3.54
H106	Phe	Tyr[b]	3	−0.26	−0.41	+1.60	+20.96
L34	Tyr	Lys	3	−0.42	+0.37	+1.90	+2.80
L53	Asn	His[b]	3	−0.14	−1.49	+1.14	−0.02
H31	Asp	Arg	3	−0.44	−0.43	+4.56	+0.70
L32	Ser	Arg	3	−0.25	−0.52	+1.96	−0.46
H56	Gly	Lys	3	−0.26	0.16	+1.38	+10.74

[a]All values in kcal/mol.
[b]Mutation observed in the directed evolution experiments.

getics for the selected single mutations, together with their associated mode of action. Positions are referred to with their chain (L, light or H, heavy) and residue number. At two positions (L34 and L53), we selected mutants to two different amino acids, to probe different modes of action. Figure 3-3 shows the positions of single mutations predicted to improve binding, with colors corresponding to the indicated modes of action. Interestingly, favorable predictions were spread across the entire binding interface. Note that all the favorable design predictions are overlaid in this figure, but our predictions correspond to single amino-acid substitutions from the wild-type sequence. Therefore, the apparent clashes of side-chain atoms between neighboring design positions are not part of the designed structures.

Figure 3-4 displays the structures of selected mutation examples for contacting residues. In the first example, shown in Figures 3-4A and B, mutation of residue L93 from tryptophane to phelylalanine removes unsatisfied polar contacts, while preserving van der Waals interactions with the ligand. The effect of mutating residue

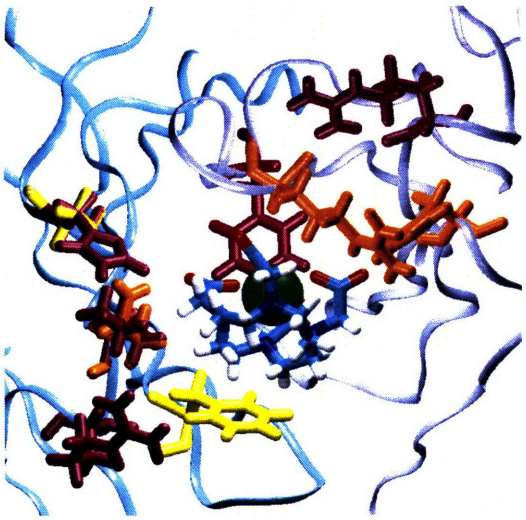


Figure 3-3: 2D12.5 design predictions. Protein backbone is depicted by ribbons (ice blue: antibody heavy chain; cyan: antibody light chain). Y-DOTA is shown in atom colors. Predictions for binding improvement are shown with the designed mutant side chain. Colors correspond to modes of action: contacting residues removing unsatisfied interactions (mode 1, yellow), contacting residues introducing new favorable interactions (mode 2, orange), or noncontacting residues (mode 3, purple). Calculations were done for single mutations, and therefore any apparent clash of side chains does not correspond to actual designed structures.

H53 from serine to histidine is shown in Figures 3-4C and D. Whereas the wild-type serine is solvent-exposed, the designed histidine makes a new hydrogen bond with a carboxylate oxygen from Y-DOTA.

Figure 3-5 displays the structures of selected mutation examples for non-contacting residues. For 2D12.5 antibody, most mutations of this type were to positively charged amino acids, i.e. lysine or arginine. In our prepared structure, the total charge on the wild-type antibody heavy chain is $+2$, and that on the light chain is -3, resulting in 2D12.5 carrying a total charge of -1; Y-DOTA was modeled in its net charge state of -1. As such, increasing the total net charge of the antibody at positions moderately far from the binding interface was often computed to improve binding energetics.

During the experimental affinity maturation procedure carried out in the Wittrup lab, mutations were accumulated across 22 positions of 2D12.5 antibody. The average number of mutations increases as more rounds of mutagenesis and selection

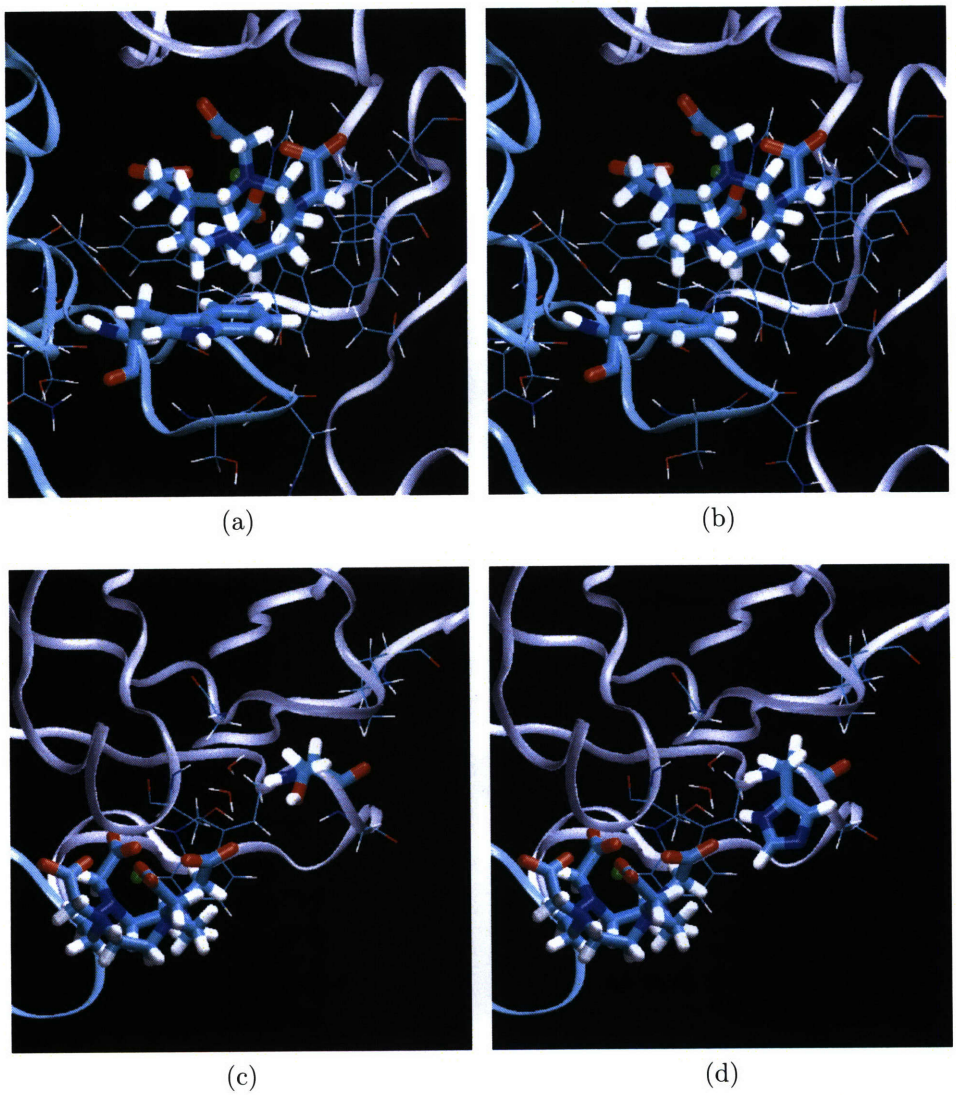


Figure 3-4: Examples of 2D12.5 single mutation designs: contacting residues. Each pair of figures displays the wild type and a design. Y-DOTA and mutated position are shown in licorice; positions given conformational flexibility are shown with lines. (A-B) Removal of unsatisfied interactions: mutation of L93 from tryptophane to phenylalanine. (C-D) Introduction of new favorable interactions: mutation of H53 from serine to histidine, creating a hydrogen bond with Y-DOTA.

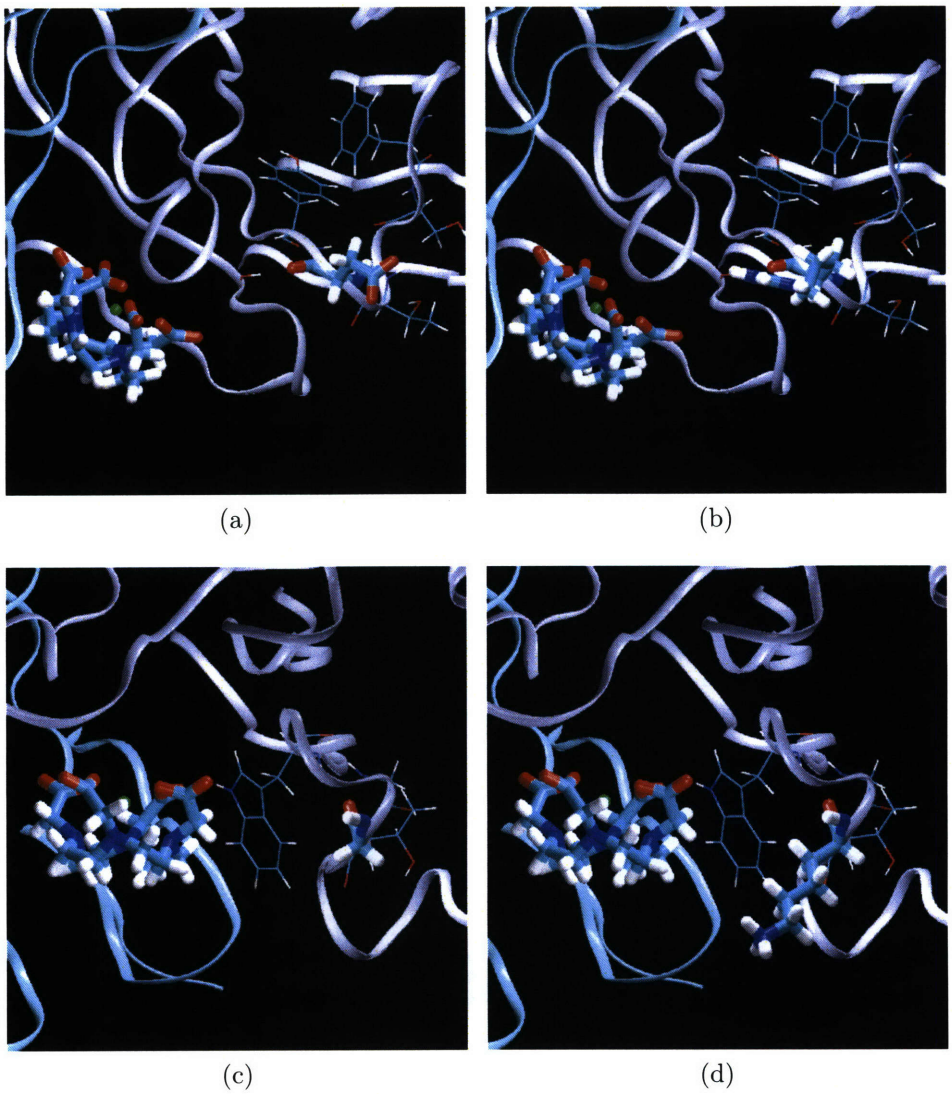


Figure 3-5: Examples of 2D12.5 single mutation designs: non-contacting residues. Each pair of figures displays the wild type and a design. Y-DOTA and mutated position are shown in licorice; positions given conformational flexibility are shown with lines. (A-B) Mutation of H31 from aspartate to arginine introduces a net charge increase of +2. (C-D) Mutation of H56 from glycine to lysine.

are performed. After 6 rounds, the sequenced clones had accumulated, on average, 12 mutations. Due to the randomness of the mutagenesis protocol used, mutations are generally either beneficial, i.e. contributing to binding improvement and allowing the clone to be selected for the next round, or neutral. For therapeutic applications, introducing the minimum number of mutations necessary to achieve the targeted binding affinity is desirable to reduce potential problems with immunogenicity. This is especially important for mutations outside of the complementary determining regions, which composed half of the observed mutations here. Therefore, computationally identifying the individual mutations contributing significantly to the affinity improvement is valuable as a first step to eliminate unnecessary mutations.

Single mutation designs, in the context of the wild-type 2D12.5, were carried out to computationally recreate experimentally observed mutations. Using our electrostatic-based metric, only two individual mutations were computed to contribute to binding affinity improvement, as shown in Table 3.1. Interestingly, mutation of residue H106 from phenylalanine to tyrosine was predicted to improve binding affinity, at the detriment of folding stability. This mutation was observed in early rounds of selection, but disappeared from the experimental library at round 6. Experimentally, scFv 2D12.5 was observed to be only midly stable (melting temperature $T_m = 43.5°C$). Our calculations suggest that clones containing the F(H106)Y mutation, despite predicted favorable contribution to the binding energetics, were under selective pressure for better folding stability. The other mutation computed to enhance binding affinity was N(L53)H. Our calculations suggest many favorable mutations at this position, including mutating to alanine. From the wild-type DNA sequence of 2D12.5, mutating to histidine can be achieved with one DNA base-pair mutation, whereas mutating to alanine requires two. One of the parameters of random mutagenesis is the average mutation rate. A low mutation rate leads to poor exploration of sequence space, while a high mutation rate can result in many non-foldable sequences. Standard error-prone

PCR protocols introduce a relatively low mutation rate, to maintain structural integrity. Therefore, it is very unlikely to achieve two DNA base-pair substitutions within the same codon in the same mutagenesis round. Here, we propose to compare the binding affinity of two mutants of L53: one experimentally observed (histidine), and one computationally designed (alanine). Our calculations suggest that, while mutating to histidine is beneficial, alanine substitution, less likely to be observed in directed evolution experiment, is more favorable.

Double mutations were computationally designed in 2D12.5. Instead of selecting all 2,926 pairs of positions (77 choose 2), where most pairs would be out of contact with each other, we selected only pairs of nearby positions. In the interest of focusing our computational efforts to he most promising regions of search space, we additionally required that at least one of the design positions was contacting Y-DOTA ligand, as degined by one side-chain atom within 4.75 Å. With this selection criterion, we carried out a total of 91 double mutation designs. For each pair of positions, $18^2 = 324$ sequences were considered. Predictions of improved binding affinity were filtered to select cooperative double mutations with energy improvement exceeding either single mutation or their additive sum.

The difficulty of obtaining lowest-energy conformations of each sequence is related to the number of protein positions and crystallographic water molecules allowed conformational flexibility. A comparison of the number and general difficulty of designs between single and double mutations is shown in Figure 3-6. For any given design, the number of rotamers allowed at each design position is a function of the amino-acid identity, larger residues having more fixed conformations to sample. With the computational library used here, each mutatable residue is allowed to adopt any of 4,025 side-chain rotamers. Therefore, double mutation designs require much more computational time than single mutation designs.

We identified two double mutations predicted to cooperatively improve binding

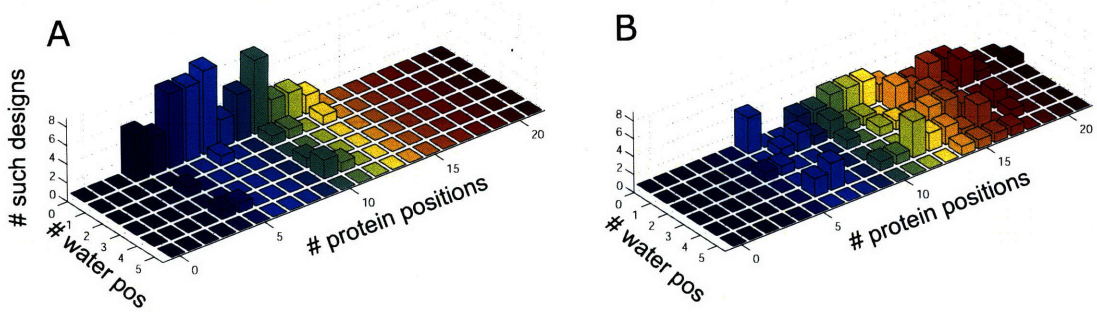
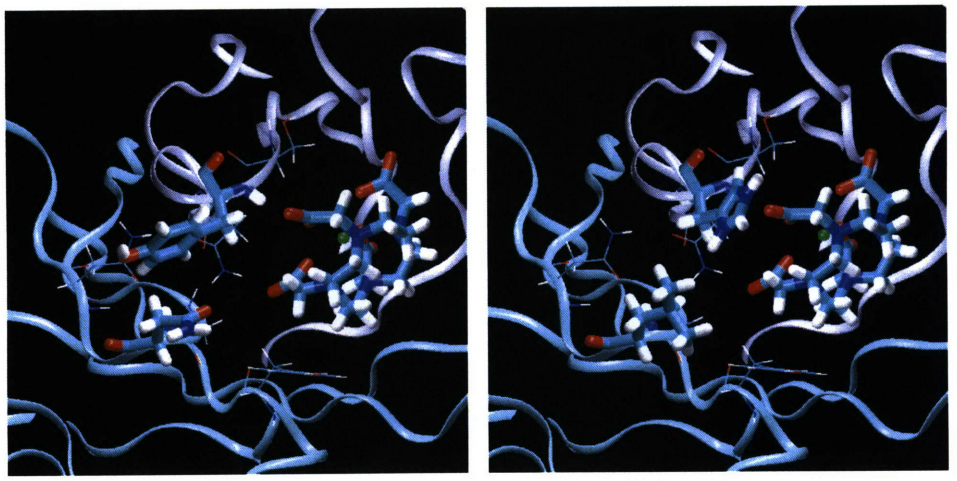
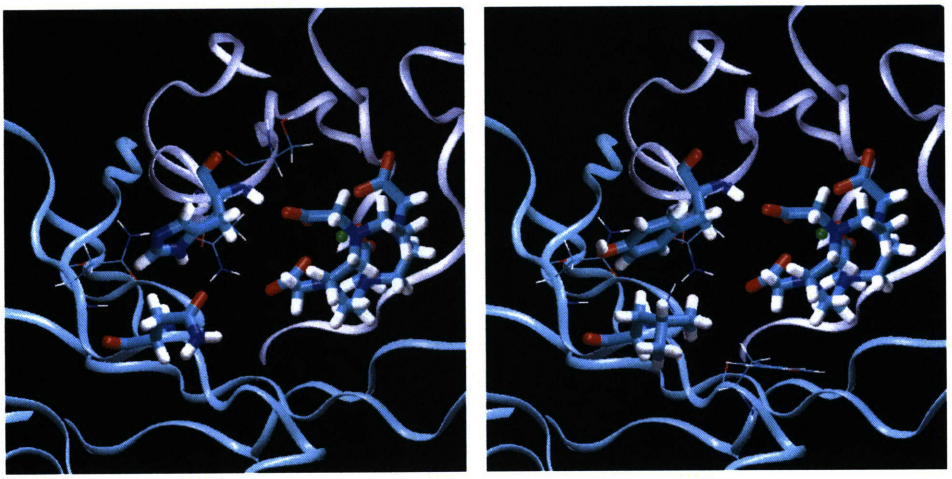


Figure 3-6: Number and difficulty of single (A) and double mutation (B) designs. Each design is characterized by the number of positions, water and protein, with conformational flexibility.

affinity of 2D12.5 to Y-DOTA. The first design consisted of mutating (H50 Val, H58 Ala) to (H50 Gly, H58 Ser). In the designed structure, a large cavity was created inside the protein core underneath the binding interface. Therefore, we eliminated this mutation (see Methods). Figure 3-7 displays the structure of the second designed mutation (H101 Tyr, L53 Asn) to (H101 His, L53 Leu). The protonated form of histidine was computed to provide for the optimal binding energy in the double mutation design. Neutral forms of histidine were also computed to increase the binding affinity, both in the single (H101) and double (H101–L53) mutation designs. The wild-type L53 asparagine pays a desolvation penalty upon binding that is not compensated by electrostatic interactions with the ligand. The proposed mutation to leucine eliminates this desolvation penalty, while maintaining van der Waals interactions. The designed histidine points toward the ligand, introducing favorable long-range electrostatic interactions. Interestingly, both single mutations were predicted to be favorable, but the side-chains adopt different rotamers in the designed double mutation than they did in the individual single mutation designs, resulting in cooperative improvement of the binding affinity.



(a) (b)



(c) (d)

H101	L53	$\Delta\Delta G^a_{\text{binding}}$ Elec	$\Delta\Delta G^a_{\text{binding}}$ Total	$\Delta\Delta G^a_{\text{folding}}$ Elec	$\Delta\Delta G^a_{\text{folding}}$ Total
Tyr	Leu	−0.24	−0.60	+1.81	+4.08
His	Asn	−0.15	−0.12	+1.51	+1.08
His	Leu	−0.88	−1.01	+2.82	+5.20

[a]All values in kcal/mol.

Figure 3-7: Predicted structure for 2D15.5 cooperative double mutation. (A) Wild-type (H101 Tyr, L53 Asn), (B) designed cooperative double mutation (H101 His, L53 Leu). The individual single mutation designs are also shown: (H101 His, L53 Asn) in (C) and (H101 Tyr, L53 Leu) in (D). Y-DOTA and mutated positions are shown in licorice; positions given conformational flexibility are shown with lines.

3.3 Conclusions

In this work, we computationally designed single and double mutations to improve the binding affinity of existing monoclonal antibody 2D12.5 to Y-DOTA. Several opportunities for binding improvement were predicted, despite the small size of the ligand. As shown in Figure 3-3, the single mutations were distributed surrounding the entire binding interface. Designed mutations are predicted to enhance electrostatic binding free energy either through removal of poorly-satisfied polar group, or through introduction of new interactions. These mutations can act either through direct hydrogen-bonding with the ligand, or by addition of a net charge at the periphery of the binding interface.

The overall magnitude of the predicted improvement in electrostatic binding energies are usually small, however. Only three designed single mutations were calculated to improve the binding affinity by more than 0.5 kcal/mol, as shown in Table 3.1. All of these mutations are predicted to introduce new direct interactions with the ligand, a design scenario that requires precise estimation of specific energetic contributions. Even though previous work using this methodology demonstrated a greater than 60% success rate across different systems, predictions with higher computed improvement in binding affinity are more robust [117, 119].

Binding-induced conformational change can be important in biological interactions. In this work, we assumed a fixed backbone conformation of the antibody, both in the bound and unbound states. Moreover, we fixed the coordinates of Y-DOTA ligand to those determined in the crystallographic structure. Previous work using the methodology developed by Lippow *et al.* focused mostly on design of protein–protein interfaces. During the design process, specific positions of the protein ligand were given rotameric conformational freedom, if they were neighbors of design positions. This computational design procedure was applied to a small molecule ligand only in the context of antibody 4-4-20 binding to fluorescein. The goal of the design

was to validate the computational methodology by recapitulating experimentally observed mutations obtained through directed evolution techniques [120]. Independent computational analysis showed that four substitutions contributed favorably to the electrostatic component of the binding affinity. Computational design identified only two of those four positions, probably because of the limitations of the fixed backbone and ligand conformations [119].

Our designed mutations will be experimentally tested for binding improvement in the Wittrup lab. Single mutations measured to enhance binding affinity could then be combined into a multiple mutant. This strategy was shown to be very effective for other antibody systems, where for instance, combining six mutations led to a 140-fold improvement in binding affinity of the anti-hen egg-white lysozyme antibody D44.1 [117]. One of the main advantages of computational design over directed evolution experimental techniques is the systematic evaluation of mutations to any amino acid, across a large number of positions. In the computer, one can model mutations requiring the simultaneous substitution of two or three DNA base pairs, an event very unlikely to be achieved in random mutagenesis setup. In this system, achieving 100-fold improvement in binding affinity from the wild-type affinity ($K_d = 0.16$ nM), as measured in the Wittrup lab, would be a significant contribution to the developement of efficient pretargeted radioimmunotherapy, with potential to greatly improve cancer treatment.

In the event that predicted binding affinity improvements are not achieved experimentally, an alternative computational design procedure could be developed. As mentioned above, the current procedure used fixed protein backbone and ligand coordinates. Parallel designs could be carried out with fixed poses of Y-DOTA, using various rotation/translation transformations from the crystallographic structure. The drug design methodology developed in-house by Michael Altman [121] could be used to generate individual poses in the context of the wild-type antibody, and independent

protein designs could be carried out for each ligand conformation. Due to the special coordination state of the yttrium metal atom by four carboxylate oxygens in DOTA, rotameric flexibility of the ligand is most likely not a limiting factor here. Further conformational relaxation could include incorporating different backbone conformations for the antibody, with CDR loops of various lengths modeled from existing antibody structures [122].

3.4 Methods

3.4.1 Structure preparation

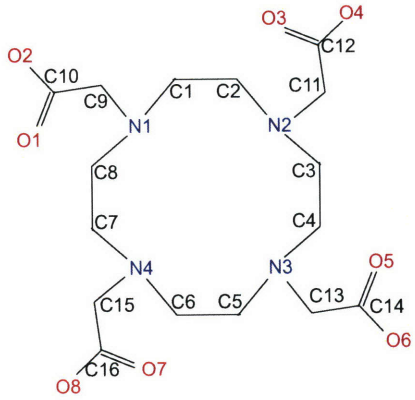
The 2.1 Å crystal structure of antibody 2D12.5 Fab bound to an yttrium-DOTA analog [105] was obtained from the Protein Data Bank [85] (PDB ID 1NC2). Only the variable region (Fv) of the antibody (heavy chain: residues 1 through 119 and light chain: residues 1 through 110) was kept for calculations. Heavy chain C-terminus was patched as an N-methylamide C-terminus, and light chain N-terminus as an acetylated N-terminus, for consistency with planned experimental testing using single chain variable fragment (scFv) expression system. Pyroglutamic acid originally present at the N-terminus of the light chain was removed and replaced by wild-type glutamine. Five water molecules bridging the binding interface and four water molecules buried away from bulk solvent (water numbers 11, 25, 32, 350) were retained. For calculations of two-state rigid binding, water molecules were either assigned to the antibody (water numbers 5, 11, 25, 27, 32, 138, 350) or to Y-DOTA (water numbers 8, 19). All other water molecules and ions were removed from the structure file and were modeled implicitly. Titration states, and asparagine, glutamine, and histidine carbon/nitrogen/oxygen crystallographic uncertainties were resolved based on optimization of hydrogen-bonding in the side-chain local environments. Coordinates for the heavy atoms of Y-DOTA were obtained from the crystal structure, bound to yttrium-

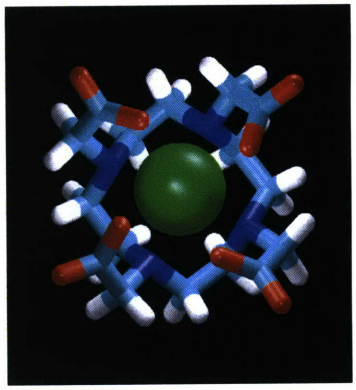
(*S*)HETD, by removing (*S*)HETD side chain atoms (carbon atoms C17 through C27, oxygen atoms O9 and O10, and sulfur atom S1).

Molecular mechanics was used to prepare the protein structure. The HBUILD facility [123] in the program CHARMM [86] was used to build all hydrogen atoms onto each structure. The CHARMM22 all-atom parameter set [16] with the CHARMM-adapted TIP3P water model was used, and appropriate general atom types for Y-DOTA were assigned (Figure 3-8). We incorporated two atom types from the CHARMm22 parameter set [124] to better describe tetrahedral aliphatic carbon (CT) and tetrahedral nitrogen (NT) atoms. Y-DOTA was modeled in its net charge –1 state. Van der Waals parameters describing yttrium interactions were borrowed from the calcium ion. Y-DOTA partial atomic charges (Figure 3-8) were obtained as described by Green and Tidor [125] by first using the program GAUSSIAN98 [126] with restricted Hartree–Fock and the LanL2DZ basis set [127] to optimize the geometry of Y-DOTA starting from that in the 2D12.5 structure. Subsequently, the electrostatic potential was fitted using the restrained-fitting methods (RESP) of Bayly *et al.* [128], creating two sets of partial atomic charges. One was an all-atom representation, and the other had all non-polar hydrogen atoms assigned a charge of zero.

3.4.2 Search space

The Kabat definition for antibody complementarity determining region (CDR) positions was used to select antibody sites for potential computational mutation [129]. This presented 60 positions spanning the six contiguous stretches in primary sequence, as shown in Figure 3-2. Additionally, we included 13 framework positions for which at least one side-chain atom (non-hydrogen atoms beyond Cβ, or the Hα1 atom for glycine) was within 11 Å of Y-DOTA in the crystal structure. These positions consist of residues H36, H37, H47, H49, H69, H71, H96, H97, L37, L47, L50, L51 and L100. We also included positions found to mutate during the directed evolution experiments





Atom Name	Atom Type	Partial Charge[a]	Partial Charge[b]	Atomic Radius	Atom Name	Atom Type	Partial Charge[a]	Partial Charge[b]	Atomic Radius
Y	Y	2.36205	2.106	1.367	O7	OC	-0.98174	-0.916	1.0
C1	CT[c]	-0.06305	0.231	2.0	O8	OC	-0.81375	-0.801	1.0
C2	CT[c]	-0.17141	0.091	2.0	H1	HA	0.16014	0.0	0.0
C3	CT[c]	-0.06306	0.242	2.0	H2	HA	0.09606	0.0	0.0
C4	CT[c]	-0.14414	0.067	2.0	H3	HA	0.11488	0.0	0.0
C5	CT[c]	-0.07290	0.221	2.0	H4	HA	0.12795	0.0	0.0
C6	CT[c]	-0.17118	0.070	2.0	H5	HA	0.14948	0.0	0.0
C7	CT[c]	-0.06589	0.226	2.0	H6	HA	0.08262	0.0	0.0
C8	CT[c]	-0.18225	0.083	2.0	H7	HA	0.09520	0.0	0.0
C9	CT[c]	-0.19027	0.164	2.0	H8	HA	0.10309	0.0	0.0
C10	C	1.04273	0.938	1.7	H9	HA	0.14472	0.0	0.0
C11	CT[c]	-0.24120	0.160	2.0	H10	HA	0.07416	0.0	0.0
C12	C	1.05192	0.949	1.7	H11	HA	0.13455	0.0	0.0
C13	CT[c]	-0.17716	0.151	2.0	H12	HA	0.11584	0.0	0.0
C14	C	1.02579	0.923	1.7	H13	HA	0.16909	0.0	0.0
C15	CT[c]	-0.27170	0.135	2.0	H14	HA	0.09317	0.0	0.0
C16	C	1.07484	0.958	1.7	H15	HA	0.13610	0.0	0.0
N1	NT[c]	-0.47477	-0.503	1.5	H16	HA	0.12907	0.0	0.0
N2	NT[c]	-0.37419	-0.510	1.5	H17	HA	0.09595	0.0	0.0
N3	NT[c]	-0.26001	-0.449	1.5	H18	HA	0.15368	0.0	0.0
N4	NT[c]	-0.42741	-0.444	1.5	H19	HA	0.16417	0.0	0.0
O1	OC	-0.95693	-0.901	1.4	H20	HA	0.09212	0.0	0.0
O2	OC	-0.81396	-0.797	1.4	H21	HA	0.13547	0.0	0.0
O3	OC	-0.96548	-0.913	1.4	H22	HA	0.06368	0.0	0.0
O4	OC	-0.80930	-0.799	1.4	H23	HA	0.17630	0.0	0.0
O5	OC	-0.96799	-0.871	1.4	H24	HA	0.10601	0.0	0.0
O6	OC	-0.81110	-0.794	1.0					

[a] All atoms are assigned partial atomic charges.
[b] All non-polar hydrogen atoms are assigned zero charges.
[c] CHARMm22 general atom type [124].

Figure 3-8: Computational parameterization of Y-DOTA. The appropriate general atom types from the CHARMM22 all-atom parameter set, together with two atom types from CHARMm22 parameter set, were used for molecular mechanics calculations. The partial atomic charge distribution was obtained by quantum-mechanical geometry minimization followed by restrained fitting of the electrostatic potential. Radii for continuum electrostatic calculations were assigned following PARSE convention, assigning yttrium radius to that parameterized for calcium.

carried out in the Wittrup lab. Those positions are referred to as experimental design positions. The experimental design positions that were specifically added, independently of our other selection criteria, consist of the 7 residues H44, H60, H61, H63, H80, L5 and L40. Due to the difficulty of generating an appropriate rotamer library for proline, the prolines at positions H102, L57 and L58 were left as wild type in all designs. Hence, we designed a total of 77 single mutations. For the design of single mutants, each of these residues was individually mutated to 17 other amino acids (proline and cysteine excluded). For the design of double mutations, the only pairs of positions considered were those with at least one pair of side-chain atoms within 4.75 Å and at least one position with one side-chain atom within 4.75 Å of Y-DOTA . For each independent design, the one or two mutated positions were given rotameric degrees of freedom, as were side chains at nearby positions (same definition of nearby as used above), including crystallographic water molecules. The two neutral tautomers and the protonated form of histidine were allowed.

3.4.3 Design of mutations

Each protein sequence was explicitly considered, for example, all 1,242 single mutations from 69 design positions with 18 amino acids each. Coordinates for Y-DOTA atoms were kept fixed for all the designs. We used the computational design protocol as described in Lippow *et al.* [117].

Conformational search is initially simplified by assuming a rigid protein backbone and allowing only discrete side chain rotamers. The rotamer library was based on the backbone-independent May 2002 library from Dunbrack [130, 131], expanded by $\pm$ 10° in both χ_1 and χ_2. The library contained 4,025 side-chain rotamers. The novel water library developed by Lippow *et al.* provided for conformational freedom of crystallographic water molecules, one rotamer allowing for the water to no longer exist in the structure. In addition, each crystallographic wild-type rotamer was added in

a position-specific manner to the library, using the complete cartesian representation of the side chain.

We used a two-step strategy, where we first generate a list of lowest-energy sequences based on a fast energy function, and then reevaluate the structures of each sequence using a more accurate, yet more computationally intensive, energy function. For the energy function for initial search ("low-resolution"), we used the CHARMM PARAM22 all-atom parameter set [16], augmented with Y-DOTA-specific atom types as presented in Figure 3-8, with no cut-offs for non-bonded interactions and a distance-dependent dielectric constant of 4. The objective function was the difference between the bound state energy and a sum of isolated, model compounds. Each model compound conformation was the lowest energy of all side-chain rotamers with the local single amino acid backbone with an acetylated N-terminus and an N-methylamide C-terminus. The model compounds are important for canceling intrinsic side-chain energies when comparing the state energy of different protein sequences. Side chain rotamers that clashed, either with backbone atoms or with the side chains of positions not given conformational flexibility, were eliminated from the search space. This low-resolution energy and objective function is pairwise-decomposable, permitting application of dead-end elimination and A* search algorithms [26, 27]. An in-house implementation of dead-end elimination and A* (DEE/A*) was at the core of the conformational search [121]. For each protein sequence, we found its global minimum energy conformation (GMEC), and if this energy was within a specified energy cut-off of the wild-type GMEC energy, then a continued list of lowest-energy structures was found for that sequence. We used the low-resolution energy function to rank-order structures that only differ by 10° dihedral angle rotations. We then used a high-resolution energy function to reevaluate different low-energy side chain conformations that may, for example, exhibit different burial and trade-off between electrostatic interaction and desolvation. The expansion of χ_1 and χ_2 by $\pm$ 10° gen-

erally created families of nine rotamers (a "fleximer" [132]); we reevaluated the 50 lowest-energy structures within 12 kcal/mol that contain at least one rotamer from a new rotamer family (contain at least one new fleximer).

Second, we reevaluated those lowest-energy structures using more accurate, yet more computationally-demanding models. We used Poisson–Boltzmann continuum electrostatics as a substitute for the distance-dependent electrostatic energy from the molecular mechanics force field. The PARSE parameters were used for partial atomic charges and radii [133], with Y-DOTA specific parameters as described in Figure 3-8. A locally modified version of the DELPHI computer program was used to solve the linearized Poisson–Boltzmann equation [20, 21, 134, 135]. A dielectric constant of 4 was used for protein and explicit water, and 80 for implicit solvent regions. Ionic strength was set to match experiments at 0.145 M, modeled with a 2.0-Å Stern layer and a molecular surface generated with a $1.4 - $ Å probe sphere. Calculations were performed with two-step focusing from 23% to 92% molecular fill of the grid, and one translation of a $129 \times 129 \times 129$ grid. The nonpolar component of the solvation free energy was independently computed using two different models. In method 1, the "SASA" model, it was added as a solvent-accessible surface area (SASA) term of 5 cal/mol/Å^2 [133], calculated using the analytical surface area routine in the program CHARMM. In method 2, the "CvdW" model, cavity formation and solute–solvent van der Waals interaction were modeled separately, as the attractive interaction is insufficiently modeled by a surface area term [136–140]. The cavitation term was modeled as linearly proportional to the molecular (solvent-excluding) surface area [141], calculated using the program MSMS [142]. We used the continuum van der Waals (CvdW) energy framework from Levy and co-workers [139], modified for the three-atom CHARMM-adapted TIP3P water model and the PARAM22 parameter set [16]. There were three free parameters in total, consisting of cavitation slope, cavitation intercept, and CvdW solvent probe radius, which were set as determined

in Lippow *et al.* [117] through parametrization on solvation free energies of alkanes (continuum van der Waals probe radius = 0.75 Å, cavitation slope = 52.5 cal/mol/Å^2, cavitation intercept = −1.6 kcal/mol). In total, the typical "high-resolution" energy function for rigid binding was comprised of van der Waals, electrostatic, and nonpolar terms (computed with SASA or CvdW model, as described above). Folding energies include additional molecular mechanics covalent bonding terms for conformational change. Structures were reranked based on these calculations.

A standard calculation yielded prediction of binding affinity relative to wild type, using the predicted bound-state structure and assuming rigid binding, as well as prediction of folding stability relative to wild type, using the same structure and subtracting off energies for isolated model compounds. In general, mutations predicted to destabilize the antibody by more than 3 kcal/mol were disregarded. In addition, conformational change upon binding was modeled by approximating the lower-energy unbound state with a second side-chain conformational search. The same positions from the bound-state design were allowed conformational relaxation, again with a fixed backbone but without the binding partner present. Binding free energy was the difference between the bound and unbound states. Lippow *et al.* [117] showed that, with this design procedure, the calculated electrostatic term for binding was a better predictor for improvement than the total calculated binding free energy. As such, we used this criterion to predict improvements in binding affinity. We eliminated by hand mutations that opened significant cavities at the interface or its neighborhood, as a continuum model with a dielectric constant of 80 may be inadequate in a confined space. The removal of mutations predicted to lose significant folding or binding van der Waals energy is usually sufficient to identify these positions.

Double mutations predicted to be favorably cooperative were distinguished by ensuring that predicted favorability was greater than the predicted energetics for any mutation subset. Double mutation energetics were required to exceed each single

mutation as well as the computed sum of each single mutation. Within the design of $18^2 = 324$ sequences, the $18-1 = 17$ single mutations were used for computing cooperativity of the $324-18 = 306$ double mutations. Cooperativity was required to exceed 0.2 kcal/mol to avoid marginal positives, e.g. a double mutation at -0.7 kcal/mol was not considered cooperative if the two single mutations were -0.3 kcal/mol each, but would be designated as cooperative if the double mutation exceeded -0.8 kcal/mol. Overall, our criterion for selecting favorably cooperative double mutations can be summarized as follows:

$$\Delta G_{\mathrm{mut}_1,\mathrm{mut}_2} \leq -0.2 + min(\Delta G_{\mathrm{mut}_1} + \Delta G_{\mathrm{mut}_2}, \Delta G_{\mathrm{mut}_1}, \Delta G_{\mathrm{mut}_2}) \tag{3.1}$$

where $\Delta G_{\mathrm{mut}_1,\mathrm{mut}_2}$ is the binding free energy of the double mutant, and $\Delta G_{\mathrm{mut}_k}$ is the binding free energy of each single mutant ($k = \{1, 2\}$).

Chapter 4

General conclusions

In conclusion, computational modeling was applied to study various aspects of protein–biomolecule binding interactions. Detailed understanding of the molecular mechanisms leading to improved or modified binding affinity is important for fundamental reasons and for therapeutic applications. Existing computational techniques and a new method were used to tackle different themes in protein–biomolecule binding.

In Chapter 2, we presented the development of a new free energy method to study force-regulated binding in a quantitative fashion. Our work illustrates the usefulness of integrated computational approaches and assessment of convergence with idealized systems, whenever possible. Simple mechanical systems highlighted key features that are expected to be biologically relevant. In particular, we showed that force-modulation of binding affinity arises from the differential response of bound and unbound states to force application. Moreover, initial technology development was coupled with careful validation on theoretical systems of increasing difficulty. Failure of traditional simulation techniques to provide robust convergence led to the refinement of our protocol through incorporation of more advanced sampling algorithms. Detailed structural analysis of simulations on a prototypical protein complex highlighted several molecular mechanisms resulting in force-modulated binding affinity. The new methodology presented in this work is easily applicable to study other mechanically regulated protein interactions.

In Chapter 3, we applied existing computational design techniques to improve the binding affinity of an antibody–hapten complex. This interaction was selected for its potential application in pretargeted radioimmunotherapy, a promising strategy for

cancer treatment. Through exhaustive evaluation of individual mutations at a large number of positions, our calculations identified several beneficial mutations, to be tested experimentally.

Hypothesis generation and validation were at the core of our work. The sequence of a protein specifies its three-dimensional structure, and the intricate interactions between neighboring side chains and their ability to interact with other biomolecules encode its function, whether enzyme catalysis or ligand binding. As such, a useful strategy to dissect biologically relevant interactions is by introducing point mutations associated with specific, computationally predicted, modifications in protein activity. In protein engineering, the goal by itself is to design mutants with altered activity, for instance improved binding affinity. In many other applications, experimental verification of the effects of sequence modifications is instrumental in confirming computationally-derived molecular hypotheses. In this thesis, we make use of this principle in two instances. In the first project, application of our free energy method to FAT–paxillin complex suggested that mechanical stretching of FAT could strengthen its binding affinity. Analysis of structural data identified potential mechanisms, including activation of specific polar contacts. Additional simulations on a mutant FAT protein confirmed our hypothesis computationally. Single-molecule pulling assays are being developed to study force-regulated binding, and could provide a platform to test such hypotheses, by comparing wild-type to mutant protein activity under controlled force application [143]. In the second project, computational design identified several opportunities for binding improvement by introduction of individual single mutations. Experimental directed evolution methods are sometimes limited by the need for cooperative mutations at the DNA level to enhance the targeted function. Experimental testing of our mutants could make substantial contribution to this project. Rationally combining independently beneficial mutations was shown to lead to substantial affinity improvement in other systems [117].

Computational modeling requires both accurate energetic evaluation and efficient search algorithms. In this work, we used a variety of molecular mechanics techniques, as best suited for each application. The computational design work presented in Chapter 3 used a discrete search space where side chains of mutatable and flexible design positions are given rotameric degrees of freedom, in the context of a fixed backbone. We used a hierarchical strategy, where lowest-energy structures obtained from fast and guaranteed search algorithms, based on a pairwise-decomposable energy function, are re-evaluated using detailed electrostatic calculations. While this method has met with considerable success in several applications, incorporation of backbone flexibility is expected to be important in adressing more challenging design problems. In contrast, advanced simulation techniques were used to investigate collective conformational rearrangements upon force application, as presented in Chapter 2. Starting with random velocity assignments at a given temperature, the conformations are updated based on integration of the equations of motion. Information transfer between parallel, independent simulations at increasing temperatures greatly increased the overall robustness of the sampling procedure. In this context, existing energetic models, or force fields, were found to disrupt the overall protein fold, necessitating the introduction of restraints to maintain the general structural features. Development of more accurate, yet computationally efficient, energy models is an active area of research that should greatly benefit the field of protein computational modeling [144].

As shown in this work, computational modeling can provide molecular level understanding of specific factors regulating protein–biomolecule interactions, whether point mutation or force-induced conformational rearrangements. A prerequisite, however, is the availability of appropriate data, namely structural information and minimum biological insight. Three-dimensional coordinates for protein–biomolecule complexes are usually determined either by X-ray crystallography [145] or NMR spectroscopy [146]. To apply the free energy methodology developed in Chapter 2 to other protein pairs

biologically involved in mechanotransduction requires availability of structural data and hypotheses for pulling mechanisms (contacts for application of mechanical force). Currently, there is a lack of validated, detailed, structural mechanistic data to support such a study. As such data become available, we anticipate that these methods will be of general applicability.

This thesis has presented applications of novel and existing computational methods to study protein–biomolecule interactions. The computational design of antibody–Y-DOTA affinity maturation, with potential application in cancer treatment, is one example of how molecular computational modeling can help advance therapeutic development. We also developed and applied a new method to computationally estimate force-modulated binding free energy. Increasing experimental evidence suggests that, in addition to biochemical signals, mechanical signals play a significant role in regulating both physiological and pathological cell behavior. Molecular analysis of force effects on protein structure and associated biochemical function is thus important for fundamental reasons. Additionally, detailed understanding of force-modulation, using computational methods such as those presented in this thesis, could be applied to identify new protein targets with novel mechanisms of therapeutic intervention.

Appendix A

Analysis of protein global motions in bound and unbound states

Abstract

Mechanical forces regulate many cellular processes, such as growth, differentiation and motility. The cell response arises from sensing mechanical cues and transforming them into biochemical signals, but little is known about how this is achieved. A proposed mechanism for mechanotransduction is force-induced conformational changes, resulting in modifications of protein activity. In this context, one hypothesis, consistent with robust force sensing and response, is that zones that are easy to deform in the absence of mechanical strain are likely to be involved in response to mechanical stimulus. Hence, studying the natural global motions of a protein is a useful first step towards understanding its force response. In this work, we focused on a likely mechanosensor protein, the FAT domain of Focal Adhesion Kinase (FAK), binding to a paxillin peptide. Normal mode analysis carried out in the bound and unbound states highlighted subtle differences in the protein dynamics. Binding was found to affect the magnitude of residue fluctuations, as well as the correlated motions between different regions of the protein. Small alterations in the frequencies of many normal modes below 500 cm^{-1} contributed to a vibrational entropy increase in the bound state that could partly compensate for lost rigid body degrees of freedom.

A.1 Introduction

Mechanical forces have been shown to be of critical importance in biological processes, such as endothelial cells orientation following shear stress and anchorage of the cell to its surrounding extracellular matrix through focal adhesion complexes. Cells can sense their mechanical environment and respond to externally applied forces, as well as cell-generated forces; force-regulated behaviors include growth, differentiation, apoptosis, motility and gene expression [31]. One hypothesis for the mechanistic basis of force-induced biological responses is that they are carried out by key proteins acting as "mechanosensors" and "mechanotransducers"; the combined effect of these functions is to convert a mechanical cue into a chemical signal that the cell can interpret through propagation with biochemical signaling pathways. Modification of biochemical properties of specific proteins upon force application has been reported experimentally; these include a force-dependence lifetime profile for P-selectin complexes [147], stretch-dependent binding of cytoplasmic proteins to cytoskeleton, including focal adhesion kinase (FAK) and paxillin [148], and stretch-activated gating of ion channels [149]. One mechanism for achieving mechanotransduction is through force-driven conformational changes, and a more generalized mechanism is force-modulated binding affinity.

Zones that are easy to deform in the absence of mechanical strain are likely to be involved in response to mechanical stimulus. This would ensure that a protein response to an applied force is somewhat robust to the details of the forcing direction. This idea was formulated in a study on the mechanosensitive channel (MscL) of *E. coli*, where normal mode analysis was performed in an attempt to elucidate the major conformational change occuring during the gating mechanism, a process known to be regulated by membrane strain [150]. Hence, studying the natural global motions of a protein constitutes a useful step towards understanding its force response.

Focal adhesions are known to act as mechanosensory devices, where internal con-

tractile forces and externally applied force can regulate the assembly of the site and trigger adhesion-dependent signaling [151]. These dynamic complexes link the cell to its surrounding extracellular matrix, through integrin binding; they are made up of more than 50 different molecules, many interacting with each other in an intricate and dynamic fashion [43]. Focal adhesion kinase (FAK) plays a central role in integrin-mediated signaling through its many protein–protein interaction motifs, which serve as a scaffold for the formation of focal adhesion complexes [152]. The C-terminal focal adhesion targeting (FAT) region of FAK causes its localization to sites of integrin clustering, by interacting with talin and paxillin through binding to the so-called LD motifs. The FAT domain structure is a four-helix bundle; helical bundle structures are ubiquitous in proteins involved in adhesion [50], such as vinculin [51], α-catenin [52], and talin [53]. Moreover, there is a commonality of binding mechanism: through bundle association, the paxillin helical LD motif interacts with the FAT domain four-helix bundle, mostly through hydrophobic patches, thereby creating a five-helical bundle structure. A similar binding motif is found in other complexes of the adhesion machinery, such as vinculin head (Vh) binding to vinculin tail domain [54], various Vh–talin complexes [53–55], Vh–α-actinin [56], α-catenin dimerization and α-catenin–β-catenin binding interactions [57].

The precise identity of mechanosensory proteins in the focal adhesion system remains unknown. Interactions between FAT and paxillin play an important role in the assembly of focal adhesions, and could be part of the mechanosensing machinery [153]. In this work, we studied the global motions of the FAT domain in the presence and absence of a bound paxillin peptide, by carrying out normal mode analysis. We sought to assess the impact of binding on protein dynamics in this model system, as a way to gain general insight on force-regulated binding properties.

Table A.1: Normal mode calculations results. Protein refers to the unbound four-helix bundle domain from FAT, while peptide refers to the unbound paxillin peptide.

	Bound	Protein	Peptide
Average energy gradient (kcal/(mol Å))	9×10^{-5}	6×10^{-5}	6×10^{-5}
Atom RMS difference (Å)			
Backbone	1.20	1.08	0.75
All atoms	1.70	1.56	1.83

A.2 Results

Bound and unbound state structures were prepared from available crystallographic data. For each system, local energy minimization was carried out to bring the structure into a local minimum, where the gradient in potential energy is zero. Normal mode analysis was performed on the energy-minimized structures. The results of the energy-minimization and normal mode calculations are sumarized in Table A.1. For both the bound and the unbound states, the average gradient in potential energy is close to zero (less than 10^{-4} kcal/(molÅ)). The normal mode calculations resulted in the expected number of zero-frequency modes for each system (six for the bound state, twelve for the unbound state, with calculated magnitudes of 2.9×10^{-2} cm^{-1} or less). These modes correspond to overall translation and rotation of the system. Because no negative eigenvalues were obtained, each structure is at a local minimum, rather than a saddlepoint. The root-mean-square difference in atom coordinates with respect to the crystallographic structure is around 1 Å over backbone atoms, indicating that each structure is essentially at a stationary point on its potential energy surface close to the crystal structure.

RMS fluctuations of the $C\alpha$ atoms were computed over all the normal modes, according to Equation A.1 (see Methods). Figure A-1 shows the RMS fluctuations in the bound and unbound states. The N-terminal loop, which is fairly flexible, appears to have even greater motions in the bound state. However, the fluctuations in the

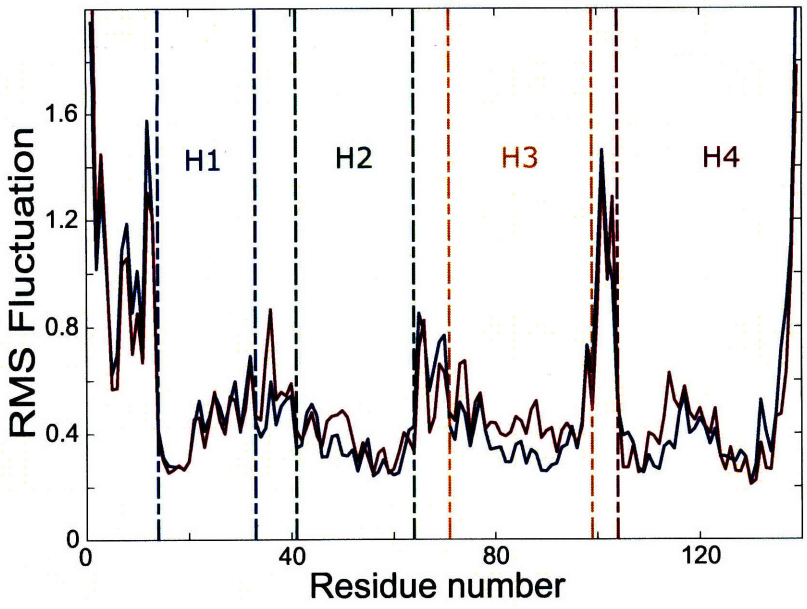


Figure A-1: RMS fluctuations in Å of the $C\alpha$ atoms, for the bound (blue) and unbound (red) states, computed over all the normal modes. The labels H1, H2, H3 and H4 indicate the span of the α-helices of the four-helix bundle domain.

bound state overall tend to be reduced compared to those in the unbound state. Specific regions of the four-helix bundle domain exhibit binding-induced decrease in their dynamic motions. These include the loop connecting helices 1 and 2, the middle regions of helices 2 and 3, and part of helix 4. These regions are enriched in residues that are either directly binding to the paxillin peptide, such as Lys 955, Gly 958, Leu 959, Gln 987, Asn 991, Leu 994, Gly 995, Ile 998 and Asn 999, or contacting such residues. We also observed more distant effects, for instance residues 1019–1023 in helix 4, which appear to have hindered motions in the bound state even though they are on the opposite side of FAT.

The extent of correlation in the motions of pairs of atoms is an alternative metric to study the dynamical properties of each system. Cross-correlation coefficients for the $C\alpha$ atoms of the four-helix bundle domain over all its normal modes were computed according to Equation A.2, in the bound and unbound states. Figure A-2A shows the cross-correlation matrices for each state independently. Since the FAT do-

main consists of four anti-parallel α-helices, two general features arise. First, strong correlations are observed along the main diagonal, i.e. for residues in the same helix. Second, residues in adjacent helices contact each other in the three-dimensional structure, resulting in correlated motions. This phenomenon is visualized as "stripes" of positively correlated residues, perpendicular to the main diagonal. Correlation between residues in helix pairs 1-2, 2-3, and 1-4 is apparent in Figure A-2A for both states. Binding affects the correlation coefficients throughout the structure, as shown with the cross-correlation difference matrix in Figure A-2B. The extent of correlation in the motions of residues at the binding interface (helices 2 and 3) are not altered by binding, but rather stay well correlated. Coupled motions of those helices with their neighbors, however, are modified. In particular, we observed decreased correlations in the relative motions of helices 1 and 2, and strengthening of the correlated motions between helices 3 and 4. The correlated motions of the loops relative to the core of the domain are also greatly altered by binding, in particular displacements of certain residues of helices 3 and 4 relative to their connecting loop, as well as the neighboring loop connecting helices 1 and 2.

Next, we examined the distribution of normal modes frequencies. The cumulative vibrational density of states for the bound and unbound states is shown in Figure A-3. We observed a slight shift towards lower frequency normal modes in the bound state compared to the unbound state, even though the overall distributions of frequencies are very similar. Based on the normal modes frequency distribution, we computed the vibrational entropy according to Equation A.3. The subtle difference in frequency distribution was calculated to give rise to a very large vibrational entropic contribution. The bound state exhibits a decrease of about 150 kcal/mol in $-T\Delta S$ at 300 K, compared to the unbound state. This change arises from small alterations in the frequencies of many modes, mostly below 500 cm^{-1}, as shown in Figure A-4B. The fact that many modes contribute to the entropy change, instead of a small number

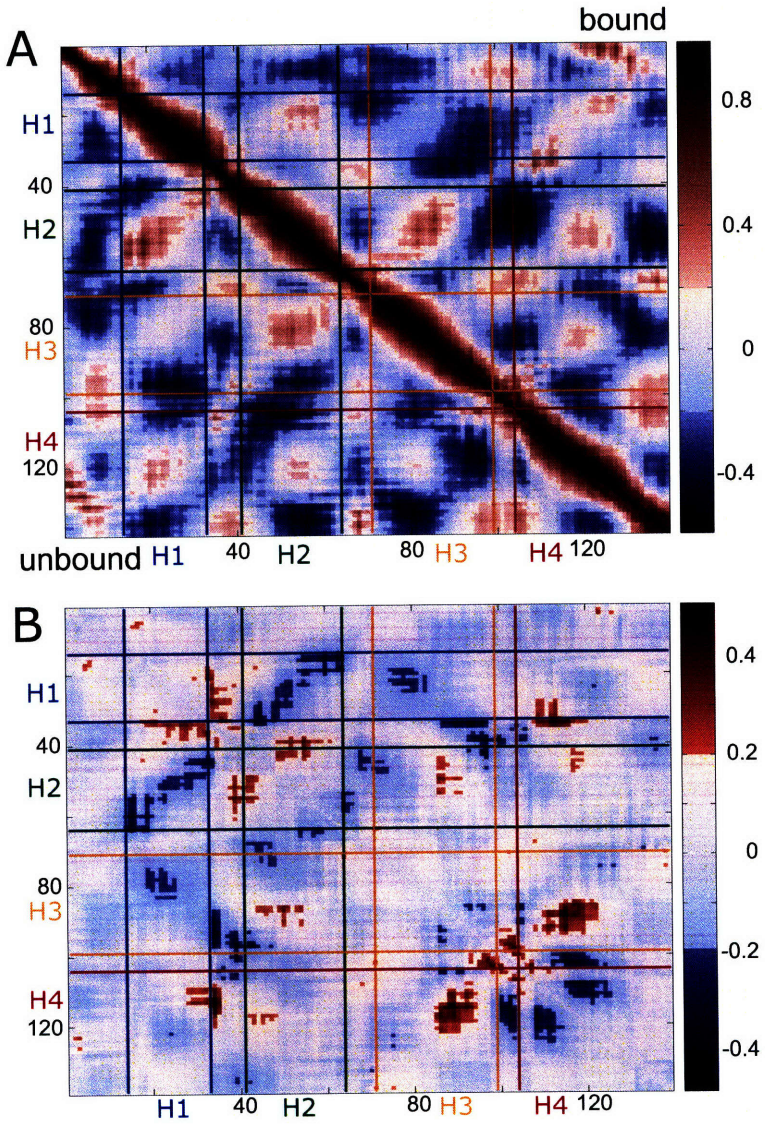


Figure A-2: Cross-correlation coefficients of the $C\alpha$ atoms, computed over all the normal modes. (A) Cross-correlation matrices. The matrices being symmetric, we plot the coefficients for the bound state in the upper triangular part, and those for the unbound state in the lower triangular part for comparison. (B) Difference between the bound and unbound correlation coefficients (full symmetric matrix).

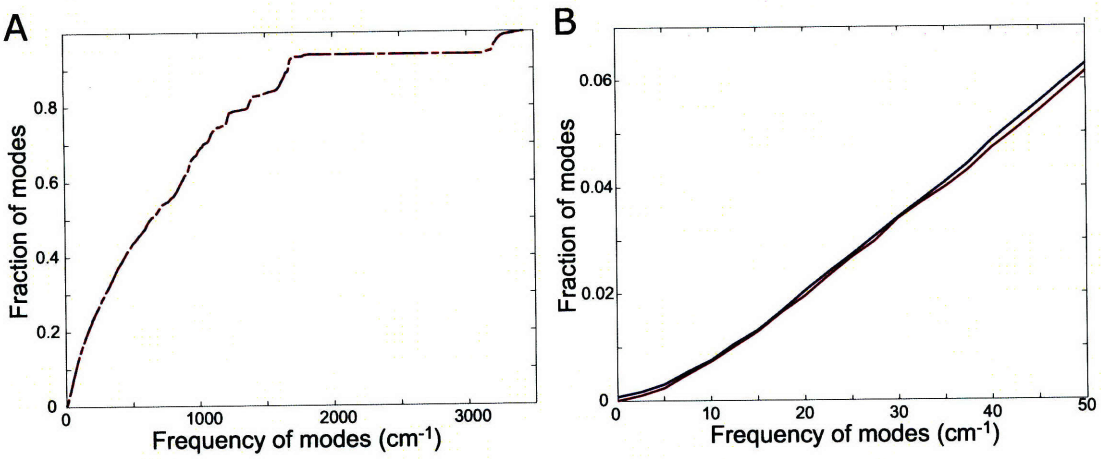


Figure A-3: Cumulative vibrational density of states for the bound (blue) and unbound (red) states. The curves overlap almost perfectly in panel (A). Panel (B) shows an enlargement of the low-frequency region along the abcissa.

of specific modes, was observed in other systems [154].

A.3 Discussion

Natural global motions of a protein are expected to be preferentially activated upon mechanical force application. Therefore, exploring a protein low-energy motions is a useful tool to understand force-induced conformational changes. Analyzing how binding alters those natural motions can provide with hypotheses for force-regulated binding mechanisms. In this work, we studied the global motions of the FAT region of FAK, in the presence and absence of a bound paxillin peptide.

Binding was found to alter the protein dynamics in several different fashions. The global fluctuations of individual residues throughout the protein structure were modified. This effect included, but was not restricted to, residues directly involved in the binding process. The binding interface is more constrained in its motions, whereas some loop and terminal regions are more flexible. The extent of correlation in relative motions of parts of the FAT domain were also altered. Effects were mostly observed for the correlations between residues at the binding interface and their neighbors.

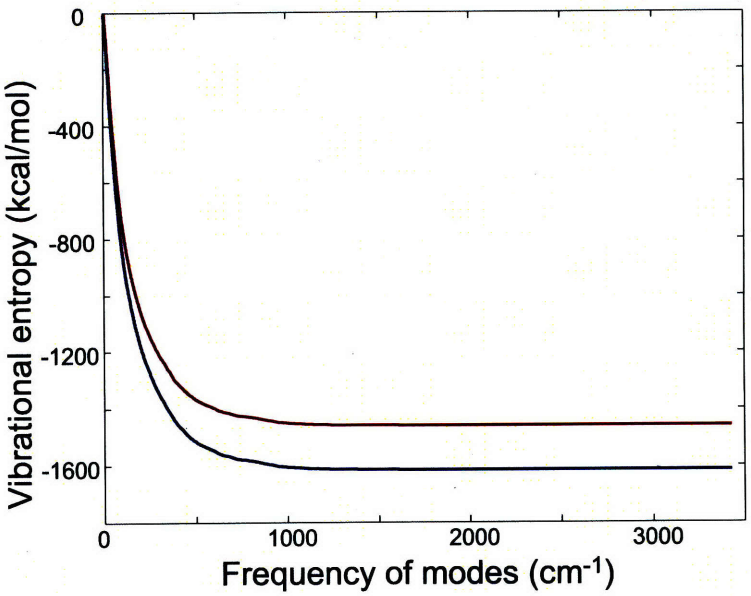


Figure A-4: Cumulative vibrational entropy for the bound (blue) and unbound (red) states. Vibrational entropy is plotted as $-T\Delta S$ at 300 K.

Interestingly, the extent of correlation among pairs of residues inside the binding site was not visibly affected, but stayed high.

We computed the vibrational entropy for the bound and unbound states based on the distribution of normal modes frequencies. Our calculations suggest a difference of 150 kcal/mol upon binding, the bound state having higher entropy due to a global shift towards lower normal modes frequencies. The binding reaction of two molecules to form a single complex must overcome an entropic barrier due to the loss of translational and rotational degrees of freedom. Estimating the vibrational entropy in the bound and unbound states can provide insight into ways to compensate for this entropy loss [155]. The entropy loss due to rigid-body motions scales as the logarithm of molecular size. For insulin dimerization, the entropy loss was estimated to be around 30 kcal/mol, and the compensating difference in vibrational entropy was 7 kcal/mol [154]. In our study, we modeled the paxillin LD4 peptide as helical, based on NMR data for the related LD2 motif [156]. If the unbound peptide was unfolded in solution, it would have a much higher vibrational entropy than the

one computed here, assuming a folded state. Therefore, the unlikely high value for the entropy difference between the bound and unbound states computed in this work (150 kcal/mol) might come from an underestimation of the vibrational entropy of the unbound peptide.

To gain further insight into binding-induced changes in protein dynamics, a detailed comparison of the normal mode eigenvector directions would be valuable. In the case of HIV-1 reverse transcriptase, normal mode analysis with a corse-grained model suggested a reorganization of the directions of the eigenvectors upon binding [157]. Ultimately, global protein motions and their rearrangements upon binding could be used to guide the study of protein response to applied forces. Specifically, analysis of force-induced conformational changes could be carried out either by projecting forcing direction onto normal mode eigenvectors, or by using molecular dynamics simulations.

A.4 Conclusion

Normal mode analysis was carried out in the bound and unbound states of FAT–paxillin complex. Our calculations suggested several differences in the natural motions of both states, including the overall flexibility of individual residues and the network of correlated motions between pairs of residues. The vibrational entropy of the bound state was found to be significantly higher than our modeled unbound state, providing a mechanism for compensating entropy loss due to rigid motions freezing upon binding. Ultimately, analysis of the directions of protein natural motions and their rearrangement upon binding could help further our understanding of force-modulated protein activity.

A.5 Methods

A.5.1 Structure preparation

The 2.35-Å crystal structure of the focal adhesion targeting (FAT) domain of focal adhesion kinase bound to paxillin LD motif [84] was obtained from the protein data bank [85] (PDB 1OW6). The starting structure for the bound state was taken from chains C and F, and for the unbound state from chain B, which is unliganded in the crystal structure. In the text, we refer to residues according to their numbering in this structure; there is an offset of 908 compared to the relative residue numbering that appears in the supporting figures. Based on NMR data which showed that paxillin LD2 motif forms an α-helix in solution [156], we modeled the unbound peptide with the helical configuration it adopts in the bound complex. All crystallographic water molecules and ions were removed. Titration states and asparagine, glutamine, and histidine carbon/nitrogen/oxygen crystallographic uncertainties were resolved based on optimization of hydrogen-bonding in the side-chain local environments. All molecular mechanics calculations were carried out using the CHARMM program [86] with the CHARMM 19 parameter set [15, 87] and the Effective Energy Function EEF1 solvent model [25]. Positions for polar-hydrogen atoms were generated with HBUILD [123].

A.5.2 Normal mode calculations

The normal mode theory is based on the harmonic approximation of the potential energy function, around a minimum energy conformation. This approximation allows an analytical solution to the equations of motion by diagonalizing the Hessian matrix (mass-weighted second derivatives of the potential energy). The eigenvectors are the normal modes, and the eigenvalues are the squares of the associated frequencies. The protein motion can be represented as a superposition of normal modes, fluctuating

around a minimum energy conformation. The lowest-frequency normal modes are collective motions of large groups of atoms, and provide insights about the flexible and rigid parts of the protein, as well as their correlated motions [158].

The bound and unbound structures were energy-minimized using 1000 steps of the steepest-descent (SD) algorithm followed by 10000 steps of the adopted-basis Newton-Raphson (ABNR) algorithm. This extensive minimization is required to achieve a local minimum so that there are no negative modes. The mass-weighted second-derivative matrix was calculated and diagonalized to give the normal modes [159]. Using the normal modes representation, thermodynamic averages can be calculated efficiently using analytical formulae. The RMS atom fluctuations were computed by evaluation of the expression [160]:

$$\left\langle \Delta r_i^2 \right\rangle = \frac{k_B T}{m_i} \sum_k \frac{|\overrightarrow{a}_{ik}|^2}{\nu_k^2} \tag{A.1}$$

where $\overrightarrow{a}_{ik}$ is the vector of the projections of the kth normal mode with frequency ν_k on the Cartesian components of the displacement vector for the ith atom, k_B is the Boltzmann constant, and T is the absolute temperature. Note that the sum is over the vibrational modes only. The cross-correlations between pairs of atoms were computed as follows:

$$\frac{\langle \Delta r_i \cdot \Delta r_j \rangle}{\sqrt{\langle \Delta r_i^2 \rangle \left\langle \Delta r_j^2 \right\rangle}} = \frac{\sum_k \frac{\overrightarrow{a}_{ik} \cdot \overrightarrow{a}_{jk}}{\nu_k^2}}{\sqrt{\left(\sum_k \frac{|\overrightarrow{a}_{ik}|^2}{\nu_k^2}\right)\left(\sum_k \frac{|\overrightarrow{a}_{jk}|^2}{\nu_k^2}\right)}} \tag{A.2}$$

The vibrational entropic contribution to the free energy at a given temperature is obtained through the expression [161]:

$$TS = \sum_i \frac{h\nu_i}{\exp\left(\frac{h\nu_i}{k_B T}\right) - 1} - k_B T \ln\left(1 - \exp\left(\frac{-h\nu_i}{k_B T}\right)\right) \tag{A.3}$$

where h is the Planck constant.

Analysis of atomic motions was performed for the bound state and the unbound state consisting of FAT by itself. For the frequency distribution of the normal modes and associated properties, such as the vibrational entropy, the unbound state consisted of both the FAT domain and the paxillin peptide, treated separately.

Bibliography

[1] P Vivekanand and I Rebay. Intersection of signal transduction pathways and development. *Ann. Rev. Genet.*, 40:139–157, 2006.

[2] N Ertekin-Taner. Genetics of Alzheimer's disease: a centennial review. *Neurol. Clin.*, 25:611–667, 2007.

[3] A Ashley-Koch, Q Yang, and RS Olney. Sickle hemoglobin (Hb S) allele and sickle cell disease: a HuGE review. *Am. J. Epidemiol.*, 151:839–845, 2000.

[4] E Kim and W Deppert. Transcriptional activities of mutant p53: When mutations are more than a loss. *J. Cell Biochem.*, 93:878–886, 2004.

[5] O Cars. Pharmacokinetics of antibiotics in tissues and tissue fluids: a review. *Scand. J. Infect. Dis. Suppl.*, 74:23–33, 1990.

[6] V Vogel. Mechanotransduction involving multimodular proteins: Converting force into biochemical signals. *Annu. Rev. Biophys. Struct.*, 35:459–488, 2006.

[7] V Vogel and M Sheetz. Local force and geometry sensing regulate cell functions. *Nat. Rev. Mol. Cell Biol.*, 7:265–275, 2006.

[8] C-YF Huang and JE Ferrell. Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc. Natl. Acad. Sci. U. S. A.*, 93:10078–10083, 1996.

[9] WN Burnette. Western Blotting: Electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein. *Anal. Biochem.*, 112:195–203, 1981.

[10] KC Neuman and SM Block. Optical trapping. *Rev. Sci. Instrum.*, 75:2787–2809, 2004.

[11] EV Shusta, J VanAntwerp, and KD Wittrup. Biosynthetic polypeptide libraries. *Curr. Opin. Biotechnol.*, 10:117–122, 1999.

[12] AF Oberhauser, C Badilla-Fernandez, M Carrion-Vazquez, and JM Fernandez. The mechanical hierarchies of fibronectin observed with single-molecule AFM. *J. Mol. Biol.*, 319:433–447, 2002.

[13] D Craig, M Gao, K Schulten, and V Vogel. Tuning the mechanical stability of fibronectin type III modules through sequence variations. *Structure*, 12:21–30, 2004.

[14] LP Lee and B Tidor. Optimization of electrostatic binding free energy. *J. Chem. Phys.*, 106:8681–8690, 1997.

[15] E Neria, S Fischer, and M Karplus. Simulation of activation free energies in molecular systems. *J. Chem. Phys.*, 105:1902–1921, 1996.

[16] AD MacKerell, D Bashford, M Bellott, RL Dunbrack, JD Evanseck, MJ Field, S Fischer, J Gao, H Guo, S Ha, D Joseph-McCarthy, L Kuchnir, K Kuczera, FTK Lau, C Mattos, S Michnick, T Ngo, DT Nguyen, B Prodhom, WE Reiher, B Roux, M Schlenkrich, JC Smith, R Stote, J Straub, M Watanabe, J Wiorkiewicz-Kuczera, D Yin, and M Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102:3586–3616, 1998.

[17] WL Jorgensen and J Tirado-Rives. The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.*, 110:1657–1666, 1988.

[18] J Hermans, HJC Berendsen, WF van Gunsteren, and JPM Postma. A consistent empirical potential for water-protein interactions. *Biopolymers*, 23:1513–1518, 1984.

[19] JW Ponder and DA Case. Force fields for protein simulations. *Adv. Prot. Chem.*, 66:27–85, 2003.

[20] MK Gilson, KA Sharp, and BH Honig. Calculating the electrostatic potential of molecules in solution - method and error assessment. *J. Comput. Chem.*, 9:327–335, 1988.

[21] KA Sharp and B Honig. Calculating total electrostatic energies with the nonlinear Poisson–Boltzmann equation. *J. Phys. Chem.*, 94:7684–7692, 1990.

[22] WC Still, A Tempczyk, RC Hawley, and T Hendrickson. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.*, 112:6127–6129, 1990.

[23] W Im, D Beglov, and B Roux. Continuum solvation model: Computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Comput. Phys. Commun.*, 111:59–75, 1998.

[24] W Im, MS Lee, and CL Brooks III. Generalized Born model with a simple smoothing function. *J. Comput. Chem.*, 24:1691–1702, 2003.

[25] T Lazaridis and M Karplus. Effective energy function for proteins in solution. *Proteins*, 35:133–152, 1999.

[26] J Desmet, M Demaeyer., B Hazes, and I Lasters. The dead-end elimination theorem and its use in protein side-chain positioning. *Nature*, 356:539–542, 1992.

[27] AR Leach and AP Lemon. Exploring the conformational space of protein side chains using dead-end elimination and the A* algorithm. *Proteins*, 33:227–239, 1998.

[28] M Karplus and JA McCammon. Molecular dynamics simulations of biomolecules. *Nature Struct. Biol.*, 9:646–652, 2002.

[29] A Mitsutake, Y Sugita, and Y Okamoto. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers*, 123:96–123, 2001.

[30] Y Sugita and Y Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, 314:141–151, 1999.

[31] ME Chicurel, CS Chen, and DE Ingber. Cellular control lies in the balance of forces. *Cur. Opin. Cell Biol.*, 10:232–239, 1998.

[32] AJ Engler, S Sen, HL Sweener, and DE Discher. Matrix elasticity directs stem cell lineage specification. *Cell*, 126:677–689, 2006.

[33] DE Ingber. Mechanobiology and diseases of mechanotransduction. *Ann. Med.*, 35:564–577, 2003.

[34] MA Haidekker, N L'Heureux, and JA Frangos. Fluid shear stress increases membrane fluidity in endothelial cells: A study with DCVJ fluorescence. *Am. J. Physiol. Heart. Circ. Physiol.*, 278:H1401–H1406, 2000.

[35] S Weinbaum, X Zhang, Y Han, H Vink, and SC Cowin. Mechanotransduction and flow across the endothelial glycocalyx. *Proc. Natl. Acad. Sci. U. S. A.*, 100:7988–7995, 2003.

[36] Y Yao, A Rabodzey, and CF Dewey Jr. Glycocalyx modulates the motility and proliferative response of vasular endothelium to fluid shear stress. *Am J Physiol Heart Circ Physiol*, 293:H1023–H1030, 2007.

[37] DJ Tschumperlin, G Dai, IV Maly, T Kikuchi, LH Laiho, AK McVittie, KJ Haley, CM Lilly, PTC So, DA Lauffenburger, RD Kamm, and JM Drazen. Mechanotransduction through growth-factor shedding into the extracellular space. *Nature*, 429:83–86, 2004.

[38] CS Chen, J Tan, and J Tien. Mechanotransduction at the cell-matrix and cell-cell contacts. *Annu. Rev. Biomed. Eng.*, 6:275–302, 2004.

[39] PA Janmey and DA Weitz. Dealing with mechanics: Mechanisms of force transduction in cells. *Trends Biochem. Sci.*, 29:364–370, 2004.

[40] C Zhu, G Bao, and N Wang. Cell mechanics: Mechanical response, cell adhesion and molecular deformation. *Annu. Rev. Biomed. Eng.*, 2:189–226, 2000.

[41] KJ Langenbach and J Sottile. Identification of protein-disulfide isomerase activity in fibronectin. *J. Biol. Chem.*, 274:7032–7038, 1999.

[42] OP Hamill. Twenty odd years of stretch-sensitive channels. *Pflugers Arch.*, 453:333–351, 2006.

[43] B Geiger, A Bershadsky, R Pankov, and KM Yamada. Transmembrane extracellular matrix-cytoskeleton crosstalk. *Nat. Rev. Mol. Cell Biol.*, 2:793–805, 2001.

[44] A Bershadsky, M Kozlov, and B Geiger. Adhesion-mediated mechanosensitivity: A time to experiment, and a time to theorize. *Cur. Opin. Cell Biol.*, 18:472–481, 2006.

[45] SK Mitra, DA Hanson, and DD Schlaepfer. Focal adhesion kinase: In command and control of cell motility. *Nat. Rev. Mol. Cell Biol.*, 6:56–68, 2005.

[46] RW Tilghman and JT Parsons. Focal adhesion kinase as a regulator of cell tension in the progression of cancer. *Sem. Cancer Biol.*, 18:45–52, 2008.

[47] HB Wang, M Dembo, SK Hanks, and Y Wang. Focal adhesion kinase is involved in mechanosensing during fibroblast migration. *Proc. Natl. Acad. Sci. U. S. A.*, 98:11295–11300, 2001.

[48] P Leucht, J-B Kim, JA Currey, J Brunski, and JA Helms. FAK-mediated mechanotransduction in skeletal regeneration. *PLoS ONE*, 2:e390, 2007.

[49] V Gabarra-Niecko, MD Schaller, and JM Dunty. FAK regulates biological processes important for the pathogenesis of cancer. *Cancer Metastasis Rev.*, 4:359–374, 2003.

[50] I Hayashi, K Vuori, and RC Liddington. The focal adhesion targeting (FAT) region of focal adhesion kinase is a four-helix bundle that binds paxillin. *Nat. Struct. Biol.*, 9:101–106, 2002.

[51] RA Borgon, C Vonrhein, G Bricogne, PRJ Bois, and T Izard. Crystal structure of human vinculin. *Structure*, 12:1189–1197, 2004.

[52] J Yang, P Dokurno, NK Tonks, and D Barford. Crystal structure of the M-fragment of alpha-catenin: Implications for modulation of cell adhesion. *EMBO J.*, 20:3645–3656, 2001.

[53] E Papagrigoriou, AR Gingras, IL Barsukov, N Bate, IJ Fillingham, B Patel, R Frank, WH Ziegler, GC Roberts, DR Critchley, and J Emsley. Activation of a vinculin-binding site in the talin rod involves rearrangement of a five-helix bundle. *EMBO J.*, 23:2942–2951, 2004.

[54] T Izard, G Evans, RA Borgon, CL Rush, G Bricogne, and PR Bois. Vinculin activation by talin through helical bundle conversion. *Nature*, 427:171–175, 2004.

[55] I Fillingham, AR Gingras, E Papagrigoriou, B Patel, J Emsley, DR Critchley, GCK Roberts, and IL Barsukov. A vinculin binding domain from the talin rod unfolds to form a complex with the vinculin head. *Structure*, 13:65–74, 2005.

[56] PRJ Bois, RA Borgon, C Vonrhein, and T Izard. Structural dynamics of α-actinin–vinculin interactions. *Mol. Cell. Biol.*, 25:6112–6122, 2005.

[57] S Pokutta and WI Weis. Structure of the dimerization and β-catenin binding region of α-catenin. *Mol. Cell*, 5:533–543, 2000.

[58] B Isralewitz, M Gao, and K Schulten. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.*, 11:224–230, 2001.

[59] M Sotomayor and K Schulten. Single-molecule experiments in vitro and in silico. *Science*, 316:1144–1148, 2007.

[60] M Gao, D Craig, V Vogel, and K Schulten. Identifying unfolding intermediates of Fn-III10 by steered molecular dynamics. *J. Mol. Biol.*, 323:939–950, 2002.

[61] A Krammer, D Craig, WE Thomas, K Schulten, and V Vogel. A structural model for force regulated integrin binding to fibronectins RGD-synergy site. *Matrix Biol.*, 21:139–147, 2002.

[62] D Craig, M Gao, K Schulten, and V Vogel. Tuning the mechanical stability of fibronectin type III modules through sequence variations. *Structure*, 12:21–30, 2004.

[63] J Gullingsrud and K Schulten. Gating of MscL studied by steered molecular dynamics. *Biophys. J.*, 85:2087–2099, 2003.

[64] GR Meyer, J Gullingsrud, K Schulten, and B Martinac. Molecular dynamics study of MscL interactions with a curved lipid bilayer. *Biophys. J.*, 91:1630–1637, 2006.

[65] SE Lee, RD Kamm, and MRK Mofrad. Force-induced activation of talin and its possible role in focal adhesion mechanotransduction. *J. Biomech.*, 40:2096–2106, 2007.

[66] V Hytönen and V Vogel. How force might activate talins vinculin binding sites: SMD reveals a structural mechanism. *PLoS Comput. Biol.*, 4:e24, 2008.

[67] C Jarzynski. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78:2690–2693, 1997.

[68] S Park, F Khalili-Araghi, E Tajkhorshid, and K Schulten. Free energy calculation from steered molecular dynamcs simulations using Jarzynski's equality. *J. Chem. Phys.*, 119:3559–3566, 2003.

[69] G Baneyx, L Baugh, and V Vogel. Coexisting conformations of fibronectin in cell culture imaged using fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci. U. S. A.*, 98:14464–14468, 2001.

[70] G Baneyx, L Baugh, and V Vogel. Fibronectin extension and unfolding within cell matrix fibrils controlled by cytoskeletal tension. *Proc. Natl. Acad. Sci. U. S. A.*, 99:5139–5143, 2002.

[71] JL Sechler, H Rao, AM Cumiskey, I Vega-Colón, MS Smith, T Murate, and JE Schwarzbauer. A novel fibronectin binding site required for fibronectin fibril growth during matrix assembly. *J. Cell Biol.*, 154:1081–1088, 2001.

[72] HF Lodish, A Berk, CA Kaiser, M Krieger, MP Scott, A Bretscher, H Ploegh, and P Matsudaira. *Molecular Cell Biology*, chapter 3. WH Freeman and Company, 6th edition, 2007.

[73] W Thomas, M Forero, O Yakovenko, L Nilsson, P Vicini, E Sokurenko, and V Vogel. Catch-bond model derived from allostery explains force-activated bacterial adhesion. *Biophys. J.*, 90:753–764, 2006.

[74] DL Beveridge and FM DiCapua. Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.*, 18:431–492, 1989.

[75] Y Sugita, A Kitao, and Y Okamoto. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.*, 113:6042–6051, 2000.

[76] AM Ferrenberg and RH Swendsen. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.*, 63:1195–1198, 1989.

[77] S Kumar, D Bouzida, RH Swendsen, PA Kollman, and JM Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.*, 13:1011–1021, 1992.

[78] W Kabsch and C Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22:2577–2637, 1983.

[79] F Frauenfelder and BH McMahon. Energy landscape and fluctuations in proteins. *Ann. Phys.*, 9:655–667, 2000.

[80] A Ansari, J Berendzen, SF Bowne, H Frauenfelder, IET Iben, TB Sauke, E Shyamsunder, and RD Young. Protein states and proteinquakes. *Proc. Natl. Acad. Sci. U. S. A.*, 82:5000–5004, 1985.

[81] E Elber and M Karplus. Multiple conformational states of proteins: A molecular dynamics analysis of myoglobin. *Science*, 235:318–321, 1987.

[82] JM Troyer and FE Cohen. Protein conformational landscapes: Energy minimization of a long molecular dynamics trajectory. *Proteins*, 23:97–110, 1995.

[83] N Metropolis, AW Rosenbluth, MN Rosenbluth, AH Teller, and E Teller. Equation state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1092, 1953.

[84] MK Hoellerer, MEM Noble, G Labesse, ID Campbell, JM Werner, and ST Arold. Molecular recognition of paxillin LD motifs by the focal adhesion targeting domain. *Structure*, 11:1207–1217, 2003.

[85] HM Berman, J Westbrook, Z Feng, G Gilliland, T Bhat, H Weissig, I Shindyalov, and PE Bourne. The Protein Data Bank. *Nucl. Acids Res.*, 300:197–212, 2000.

[86] BR Brooks, RE Bruccoleri, BD Olafson, DJ States, S Swaminathan, and M Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, 4:187–217, 1983.

[87] WE Reiher, III. *Theoretical Studies of Hydrogen Bonding.* PhD thesis, Harvard University, Cambridge, MA, USA, 1985.

[88] M Nilges, AM Gronenborn, AT Brünger, and GM Clore. Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints. Application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. *Protein Eng.*, 2:27–38, 1988.

[89] M Feig, J Karanicolas, and CL Brooks, III. MMTSB tool set: Enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graph. Model.*, 22:3777–3795, 2004.

[90] W Humphrey, A Dalke, and K Schulten. VMD – Visual Molecular Dynamics. *J. Mol. Graph.*, 14:33–38, 1996.

[91] EA Merritt and DJ Bacon. Raster3D: Photorealistic molecular graphics. *Methods Enzymol.*, 277:505–524, 1997.

[92] TA Waldmann. Immunotherapy: past, present and future. *Nature Med.*, 9:269–277, 2003.

[93] RM Sharkey and Goldenberg DM. Perspectives on cancer therapy with radiolabeled monoclonal antibodies. *J. Nucl. Med.*, 46:115S–127S, 2005.

[94] DB Axworthy, JM Reno, MD Hylarides, RW Mallett, LJ Theodore, LM Gustavson, F-M Su, LJ Hobson, PL Beaumier, and AR Fritzberg. Cure of human carcinoma xenografts by a single dose of pretargeted yttrium-90 with negligible toxicity. *Proc. Natl. Acad. Sci. U. S. A.*, 97:1802–1807, 2000.

[95] JM Pagel, N Hedin, K Subbiah, D Meyer, R Mallet, D Axworthy, LJ Theodore, DS Wilbur, DC Matthews, and OW Press. Comparison of anti-CD20 and anti-CD45 antibodies for conventional and pretargeted radioimmunotherapy of B-cell lymphomas. *Blood*, 101:2340–2348, 2003.

[96] OW Press, M Corcoran, K Subbiah, DK Hamlin, DS Wilbur, T Johnson, L Theodore, E Yau, R Mallett, DL Meyer, and D Axworthy. A comparative evaluation of conventional and pretargeted radioimmunotherapy of CD20-expressing lymphoma xenografts. *Blood*, 98:2535–2543, 2001.

[97] K Subbiah, DK Hamlin DK, JM Pagel, DS Wilbur, DL Meyer, DB Axworthy, RW Mallett, LJ Theodore, PS Stayton, and OW Press. Comparison of immunoscintigraphy, efficacy, and toxicity of conventional and pretargeted radioimmunotherapy in CD20-expressing human lymphoma xenografts. *J. Nucl. Med.*, 44:437–445, 2003.

[98] A Forero, PL Weiden, JM Vose, SJ Knox, AF LoBuglio, J Hankins, ML Goris, VJ Picozzi, DB Axworthy, HB Breitz, RB Sims, RG Ghalie, S Shen, and RF Meredith. Phase 1 trial of a novel anti-CD20 fusion protein in pretargeted radioimmunotherapy for B-cell non-Hodgkin lymphoma. *Blood*, 104:227–236, 2004.

[99] HB Breitz, PL Weiden, PL Beaumier, DB Axworthy, C Seiler, F-M Su, S Graves, K Bryan, and JM Reno. Clinical optimization of pretargeted radioimmunotherapy with antibody-streptavidin conjugate and ^{90}Y-DOTA-biotin. *J. Nucl. Med.*, 41:131–140, 2000.

[100] DM Goldenberg, RM Sharkey, G Paganelli, J Barbet, and J-F Chatal. Antibody pretargeting advances cancer radioimmunodetection and radioimmunotherapy. *J. Clin. Oncol.*, 24:823–834, 2006.

[101] MM Le Mignon, C Chambon, S Warrington, R Davies, and B Bonnemain. Gd-DOTA. pharmacokinetics and tolerability after intravenous injection into healthy volunteers. *Invest. Radiol.*, 25:933–937, 1990.

[102] J-C Bousquet, S Saini, DD Stark, PF Hahn, M Nigam, J Wittenberg, and JT Ferrucci Jr. Gd-DOTA: Characterization of a new paramagnetic complex. *Radiology*, 166:693–698, 1988.

[103] M Allard, D Doucet, P Kien, B Bonnemain, and JM Caille. Experimental study of DOTA- gadolinium. pharmacokinetics and pharmacologic properties. *Invest. Radiol.*, 23:S271–S274, 1988.

[104] TM Corneillie, PA Whetstone, AJ Fisher, and CF Meares. A rare earth-DOTA-binding antibody: probe properties and binding affinity across the lanthanide series. *J. Am. Chem. Soc.*, 125:3436–3437, 2003.

[105] TM Corneillie, AJ Fisher, and CF Meares. Crystal structures of two complexes of the rare-earth-DOTA-binding antibody 2D12.5: ligand generality from a chiral system. *J. Am. Chem. Soc.*, 125:15039–15048, 2003.

[106] B Wessels and RD Rogus. Radionuclide selection and model absorbed dose calculations for radiolabeled tumor associated antibodies. *Med. Phys.*, 11:639–645, 1984.

[107] SC Zajic. *Improved methods and reagents for pretargeted radioimmunotherapy of cancer.* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, U.S.A., 2006.

[108] CP Graff, K Chester, R Begent, and KD Wittrup. Directed evolution of an anti-carcinoembryonic antigen scfv with a 4-day monovalent dissociation half-time at 37 degrees c. *Protein Eng. Des. Sel.*, 17:293–304, 2004.

[109] ET Boder and KD Wittrup. Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol.*, 15:553–557, 1997.

[110] JJ VanAntwerp and KD Wittrup. Fine affinity discrimination by yeast surface display and flow cytometry. *Biotechnol. Prog.*, 16:31–37, 2000.

[111] G Chao, WL Lau, BJ Hackel, SL Sazinsky, SM Lippow, and KD Wittrup. Isolating and engineering human antibodies using yeast surface display. *Nature Protocols*, 1:755–768, 2006.

[112] DW Leung, E Chen, and DW Goeddel. A method for random mutagenesis of a defined dna segment using a modified polymerase chain reaction. *Techniques*, 1:11–15, 1989.

[113] RW Dixon, RJ Radmer, B Kuhn, PA Kollman, J Yang, C Raposo, CS Wilcox, LA Klumb, PS Stayton, C Behnke, I Le Trong, and R Stenkamp. Theoretical and experimental studies of biotin analogues that bind almost as tightly to streptavidin as biotin. *J. Org. Chem.*, 67:1827–1837, 2002. Number = 6, Dk49655/dk/niddk Gm-29072/gm/nigms P41-rr01081/rr/ncrr Journal Article Research Support, U.S. Gov't, P.H.S. United States.

[114] G Song, GA Lazar, T Kortemme, M Shimaoka, JR Desjarlais, D Baker, and TA Springer. Rational design of intercellular adhesion molecule-1 (ICAM-1) variants for antagonizing integrin lymphocyte function-associated antigen-1-dependent adhesion. *J. Biol. Chem.*, 281:5042–5049, 2006.

[115] VD Sood and D Baker. Recapitulation and design of protein binding peptide structures and sequences. *J. Mol. Biol.*, 357:917–927, 2006.

[116] LA Clark, PA Boriack-Sjodin, J Eldredge, C Fitch, B Friedman, KJ Hanf, M Jarpe, SF Liparoto, Y Li, A Lugovskoy, S Miller, M Rushe, W Sherman, K Simon, and H Van Vlijmen. Affinity enhancement of an in vivo matured

therapeutic antibody using structure-based computational design. *Protein Sci.*, 15:949–960, 2006. Number = 5, Evaluation Studies Journal Article Research Support, Non-U.S. Gov't United States a publication of the Protein Society.

[117] SM Lippow, KD Wittrup, and B Tidor. Computational design of antibody-affinity improvement beyond *in vivo* maturation. *Nat. Biotechnol.*, 25:1171–1176, 2007.

[118] BA Joughin, DF Green, and B Tidor. Action-at-a-distance interactions enhance protein binding affinity. *Protein Sci.*, 14:1363–1369, 2005.

[119] SM Lippow. *Computational analysis, design, and experimental validation of antibody binding affinity improvements beyond in vivo maturation.* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, U.S.A., 2007.

[120] ET Boder, KS Midelfort, and KD Wittrup. Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proc. Natl. Acad. Sci. U. S. A.*, 97:10701–10705, 2000.

[121] M. D. Altman. *Computational ligand design and analysis in protein complexes using inverse methods, combinatorial search, and accurate solvation modeling.* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, U.S.A., 2006.

[122] LC Allcorn and ACR Martin. SACS - a Self-maintaining database of Antibody Crystal Structures. *Bioinformatics*, 18:175–181, 2002.

[123] AT Brünger and M Karplus. Polar hydrogen positions in proteins: empirical energy placement and neutron diraction comparison. *Proteins*, 4:148–156, 1988.

[124] FA Momany and R Rone. Validation of the general-purpose QUANTA(R)3.2/CHARMM(R) force-field. *J. Comput. Chem.*, 13(7):888–900, 1992.

[125] DF Green and B Tidor. Evaluation of ab initio charge determination methods for use in continuum solvation calculations. *J. Phys. Chem. B*, 107:10261–10273, 2003.

[126] MJ Frisch, GW Trucks, HB Schlegel, GE Scuseria, MA Robb, JR Cheeseman, VG Zakrzewski, JA Montgomery Jr., RE Stratmann, JC Burant, S Dapprich, JM Millam, AD Daniels, KN Kudin, MC Strain, Ö Farkas, J Tomasi, V Barone, M Cossi, R Cammi, B Mennucci, C Pomelli, C Adamo, S Clifford, J Ochterski, GA Petersson, PY Ayala, Q Cui, K Morokuma, P Salvador, JJ Dannenberg, DK Malick, AD Rabuck, K Raghavachari, JB Foresman, J Cioslowski, JV Ortiz, AG Baboul, BB Stefanov, G Liu, A Liashenko, P Piskorz, I Komáromi, R Gomperts, RL Martin, DJ Fox, T Keith, MA Al-Laham, CY Peng, A Nanayakkara, M Challacombe, PMW Gill, B Johnson, W Chen, MW Wong, JL Andres, C Gonzalez, M Head-Gordon, ES Replogle, and JA Pople. *Gaussian 98* (Gaussian, Inc., Pittsburgh, PA, 1998).

[127] TH Dunning Jr. and PJ Hay. *Modern Theoretical Chemistry*, volume 3, pages 1–28. Plenum, New York, 1976.

[128] CI Bayly, P Cieplak, WD Cornell, and PA Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges — the RESP model. *J. Phys. Chem.*, 97:10269–10280, 1993.

[129] EA Kabat, TT Wu, HM Perry, KS Gottesmann, and C Foeller. *Sequences of Proteins of Immunological Interest.* Public Health Service, National Institutes of Health, Bethesda, MD, fifth edition, 1991.

[130] RL Dunbrack Jr. and FE Cohen. Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein Sci.*, 6:1661–1681, 1997. Journal Article Research Support, U.S. Gov't, P.H.S. United states a publication of the Protein Society.

[131] RL Dunbrack. Rotamer libraries in the 21(st) century. *Curr. Opin. Struct. Biol.*, 12:431–440, 2002.

[132] J Mendes, AM Baptista, MA Carrondo, and CM Soares. Improved modeling of side-chains in proteins with rotamer-based methods: A flexible rotamer model. *Proteins*, 37:530–543, 1999.

[133] D Sitkoff, KA Sharp, and B Honig. Accurate calculation of hydration free-energies using macroscopic solvent models. *J. Phys. Chem.*, 98:1978–1988, 1994.

[134] MK Gilson and B Honig. Calculation of the total electrostatic energy of a macromolecular system — solvation energies, binding-energies, and conformational-analysis. *Proteins*, 4:7–18, 1988.

[135] KA Sharp and B Honig. Electrostatic interactions in macromolecules — theory and applications. *Annu. Rev. Biophys. Bio.*, 19:301–332, 1990.

[136] E Gallicchio, MM Kubo, and RM Levy. Enthalpy–entropy and cavity decomposition of alkane hydration free energies: Numerical results and implications for theories of hydrophobic solvation. *J. Phys. Chem. B*, 104:6271–6285, 2000.

[137] JW Pitera and WF van Gunsteren. The importance of solute–solvent van der Waals interactions with interior atoms of biopolymers. *J. Am. Chem. Soc.*, 123:3163–3164, 2001.

[138] E Gallicchio, LY Zhang, and RM Levy. The SGB/NP hydration free energy model based on the surface generalized born solvent reaction field and novel nonpolar hydration free energy estimators. *J. Comput. Chem.*, 23:517–529, 2002. Gm 30580/gm/nigms Rr 06892/rr/ncrr Journal Article Research Support, U.S. Gov't, P.H.S. United States.

[139] RM Levy, LY Zhang, E Gallicchio, and AK Felts. On the nonpolar hydration free energy of proteins: surface area and continuum solvent models for the solute-solvent interaction energy. *J. Am. Chem. Soc.*, 125:9523–9530, 2003. Gm30580/gm/nigms Journal Article Research Support, U.S. Gov't, P.H.S. United States.

[140] J Wagoner and NA Baker. Solvation forces on biomolecular structures: A comparison of explicit solvent and Poisson–Boltzmann models. *J. Comput. Chem.*, 25:1623–1629, 2004.

[141] J Pitarch, V Moliner, J-L Pascual-Ahuir, E Silla, and I Tunon. Can hydrophobic interactions be correctly reproduced by the continuum models? *J. Phys. Chem.*, 100:9955–9959, 1996.

[142] MF Sanner, AJ Olson, and JC Spehner. Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers*, 38:305–320, 1996.

[143] MJ Lang an PM Fordyce and SM Block. Combined optical trapping and single-molecule fluorescence. *J. Biol.*, 2:6, 2003.

[144] J Chen, CL Brooks III, and J Khandogin. Recent advances in implicit solvent-based methods for biomolecular simulations. *Curr. Opin. Struct. Biol.*, 18:140–148, 2008.

[145] J Drenth. *Principles of protein X-ray crystallography.* Springer-Verlag, New York, 1999.

[146] K Wuthrich. Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science*, 243:45–50, 1989.

[147] BT Marshall, M Long, JW Piper, T Yago, RP McEver, and C Zhu. Direct observation of catch bonds involving cell-adhesion molecules. *Nature*, 423:190–193, 2003.

[148] Y Sawada and MP Sheetz. Force transduction by Triton cytoskeletons. *J. Cell Biol.*, 156:609–615, 2002.

[149] F Sachs. Stretch-sensitive ion channels: an update. *Soc. Gen. Physiol. Ser.*, 47:241–260, 1992.

[150] H Valadié, JJ Lacapčre, Y-H Sanejouand, and C Etchebest. Dynamical properties of the MscL of *escherichia coli*: a normal mode analysis. *J. Mol. Biol.*, 332:657–674, 2003.

[151] B Geiger and A Bershadsky. Assembly and mechanosensory function of focal contacts. *Cur. Opin. Cell Biol.*, 13:584–592, 2001.

[152] DD Schlaepfer, CR Hauck, and DJ Sieg. Signaling through focal adhesion kinase. *Prog. Biophys. Mol. Biol.*, 71:435–478, 1999.

[153] CE Turner. Paxillin and focal adhesion signaling. *Nat. Cell Biol.*, 2:E231–E236, 2000.

[154] B Tidor and M Karplus. The contribution of vibrational entropy to molecular association. *J. Mol. Biol.*, 238:405–414, 1994.

[155] S Fischer, JC Smith, and CS Verma. Dissecting the vibrational entropy change on protein/ligand binding: Burial of a water molecule in bovine pancreatic trypsin inhibitor. *J. Phys. Chem.*, 105:8050–8055, 2001.

[156] G Liu, CD Guibao, and J Zheng. Structural insight into the mechanisms of targeting and signaling of focal adhesion kinase. *Mol. Cell. Biol.*, 22:2751–2760, 2002.

[157] NA Temiz and I Bahar. Inihibitor binding alters the directions of domain motions in HIV-1 reverse transcriptase. *Proteins*, 49:61–70, 2002.

[158] B Brooks and M Karplus. Harmonic dynamics of proteins: Normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. U. S. A.*, 80:6571–6575, 1983.

[159] BR Brooks, D Janezic, and M Karplus. Harmonic analysis of large systems. I. Methodology. *J. Comput. Chem.*, 16:1522–1542, 1995.

[160] M Levitt, S Sander, and PS Stern. Protein normal-mode dynamics: Trypsin inhibitor, crambin, ribonuclease an lysozyme. *J. Mol. Biol.*, 181:423–447, 1985.

[161] DA McQuarrie. *Statistical Mechanics.* Harper and Row, New York, 1976.