UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
The Role of Conformational Energetics in Ligand Binding and Thermal Sensation

Permalink
https://escholarship.org/uc/item/2z65j8wf

Author
Koulechova, Diana

Publication Date
2014

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

The Role of Conformational Energetics in Ligand Binding and Thermal Sensation

by

Diana Koulechova

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

Doctor of Philosophy
in
Molecular and Cell Biology
in the
GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, BERKELEY

Committee in charge:
Professor Susan Marqusee, Chair
Professor John Kuriyan
Professor David Wemmer
Professor Andreas Martin

Spring 2014

Abstract

Funnel and Function: The Physiological Relevance of the Energy Landscape

by

Diana Koulechova

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professor Susan Marqusee, Chair

A major outstanding question in protein science is how different regions of a protein communicate with one another. Although we can identify action-at-a-distance phenomena when they occur, a generalized mechanism and an understanding sufficiently thorough as to allow for *de novo* design remain works in progress. This is particularly true for natively disordered proteins and channel proteins, both of which have traditionally been more technically difficult to characterize biophysically. In this work, I explore the functional relevance of their unique energetics and dynamics.

The first project investigates the functional relevance of the hydrophobic core of disordered transcription factor MarA. We randomized the MarA hydrophobic core and selected for variants able to bind the consensus sequence. We find that MarA is highly intolerant of core mutation; this is in contrast to what is seen for the well-folded transcription factor λ-repressor. Furthermore, core variants that do retain the ability to bind consensus sequence have differentially altered affinities for different binding partners. We propose that this can be explained by taking into account the varying energetic impact of these mutations on different MarA conformations and posit that residues distant from the active site can alter both binding affinity and specificity in natively disordered proteins.

The second project aims to shed light on the mechanisms of thermosensation. Protein channel TRPA is a heat sensor in rattlesnakes but exhibits no heat-dependent activation in mammals. This discrepancy was mapped to select ankyrin repeats within the protein's cytoplasmic N-terminal domain. We hypothesized that temperature-dependent conformational changes within these repeats propagated to the pore region may be the thermosensing mechanism employed by TRPA1. The isolated repeats were purified and analyzed biophysically. Rattlesnake ankyrin repeats 3-8 are unique in that they do not appear to unfold with temperature on the timescale tested. The mechanism of this remarkable thermotolerance and its physiological relevance are currently under investigation.

Table of Contents

List of Figures & Tables

Acknowledgements

It goes without saying that the work in this thesis has been a collaborative effort. The study described in Chapter 2 was done in collaboration with Geoffrey Horner and Katherine Tripp; I continue to be thankful that it did not have to be a solitary effort. Katie, in particular, has continued to be both mentor and friend long after leaving the Marqusee lab. None of this work would exist without my advisor Susan Marqusee. Susan, thank you for making me into a scientist. I'm not sure it could have happened without your help. The entire Marqusee lab has been a source of both joy and wisdom, but none more so than Laura Rosen and Sabriya Rosemond. Learning how to go to work without you is likely to be the biggest adjustment I have to make after leaving graduate school.

Outside the lab, my family has been a constant source of love and support in everything except my predilection for voting Democrat. Whatever sanity I have left exists thanks in large part to my friends, especially Beatrice Rabkin and Sarah Papp. It seems somewhat strange to put you girls in this section, because I long ago began thinking of you as family. The remainder of my sanity was preserved by Mal, who is undoubtedly the cutest and weirdest dog anyone has ever had. Finally, thank you to my soon-to-be husband, Brooks Bond-Watts. For your love alone, the past six years would have been well worth it.

Chapter 1: INTRODUCTION

1.1 Introduction

Proteins are the major actors in the cell. They carry out a vast array of functions, including accelerating chemical reaction rates, recognizing ligands, driving the transport of molecules or entire cells by forming complex molecular motors, and sensing the environment. Despite their diverse functions, many proteins make use of similar motifs, modules, or domains. Understanding how these different modules or parts respond to different environmental conditions and combine to create different functions is an important goal in biochemistry and synthetic biology. Currently, we are limited both in our understanding of and our ability to predict how a particular protein or protein part will behave in a novel context. This becomes especially true as protein engineers and synthetic biologists combine proteins and protein domains to create novel machines and systems.

Particularly relevant are systems that hold great promise as engineering targets in terms of usefulness but which have proven challenging to work with because they do not behave like the traditional model systems that have been the focus of most biophysical studies. Among these are the two protein types that have been the focus of my thesis work: intrinsically disordered or marginally stable proteins and the cytoplasmic region of ion channels. The folded structures of these proteins resemble the typical globular proteins, thus the differences must lie in their energetics and dynamics (their energy landscapes, see below for a more detailed exploration). The specifics of their energy landscapes and how they impact protein function is a rapidly expanding area of research with many outstanding questions. This introduction will summarize our current understanding of protein energy landscapes, their significance for protein function, and how our understanding of energy landscapes is expanding to accommodate novel model systems.

1.2 Beyond a Single Structure

Proteins are often referred to as molecular machines, playing a variety of functions in the cell that range from structural to enzymatic. Just as machines made from metal or plastic have distinct three-dimensional structures that enable them to perform their functions, so do proteins. For example, cytoskeletal proteins like keratin and tubulin are monomers that fit together into long, twisted fibers that are able to withstand stress and provide the cell with shape and structural integrity[1,2]. Many DNA-binding proteins employ a helix-turn-helix motif, enabling the protein to insert a helix into the DNA major groove and make extensive contacts that enable sequence-specific recognition[3].

A single static structure, however, is insufficient for function. Just like cranes and motors move as they operate, proteins frozen in one conformation are unable to catalyze reactions[4] or interact with their binding partners[5]. Even structural proteins like microtubules are highly dynamic, continuously assembling and disassembling[6]. A single protein can adopt multiple conformations with their relative probabilities determined by their Boltzmann energies. Their energetics relative to the native conformation determine their relative populations under any given set of equilibrium conditions, dictating the

effective concentrations of binding-competent or reaction-competent proteins in a system. Thus, a better way to look at proteins' conformations is to look their energy landscapes, which can be represented as a funnel that depicts internal energy on the z-axis and entropy in the xy-plane[7] (Fig. 1.1). One begins at the top of the funnel as a high-entropy, high-energy random coil, and then continues down the funnel, losing conformational freedom and decreasing in internal energy. The native conformation – what is frequently referred to as the protein's structure – is located at the global minimum of this funnel.

The energy landscape details another important aspect of a protein – dynamics. As discussed above, protein binding or enzymatic catalysis require protein movement – this is another way of saying that they require access to high energy conformations on the landscape The height of the barriers between states dictates the rate of interconversion and thus determines the accessibility of each state. By altering either the populations of different conformations or the dynamics between them, one can alter protein function without fundamentally changing the native structure. For example, biasing ubiquitin towards one of two major conformations shifts its binding specificity and dramatically affects the viability of *S. cerevisiae*[8]. Similarly, Levin *et al.* engineered an IL-2 "superkine" by stabilizing IL-2 in its optimal receptor-binding conformation, thereby eliminating the requirement for cofactor CD25, which increased IL-2 activity by stabilizing the high energy binding-competent conformation[9].

Of-course, conformations on the energy landscape are important for more than just the protein's specific function. After being synthesized, a protein must fold correctly or risk degradation or aggregation. Misfolded and aggregated proteins decrease the amount of the protein available for proper function, tax the cell's proteostasis machinery, and can even lead to disease[10]. Additionally, proteolysis rates can vary dramatically depending on conformation, so the rates of interconversion and, by extension, the amount of time the protein spends in each conformation are a large factor in determining the lifetime of a protein with a cell. While proteins are generally biased towards the native conformation, some fraction will always exist in the unfolded conformation; thus, traversing the entire height of the funnel is a path many protein molecules will need to travel more than once in their lifetimes. Although it is a rapidly growing area of research, a lot of work remains to be done characterizing the correlation of different aspects of energy landscapes with phenotype.

1.3 Probing the Energy Landscape

Until the mid 1990's, protein folding was most commonly conceptualized as a pathway[11]. The energy landscape model was an expansion of this way of thinking, with parallel pathways that vary in their flux, like water flowing down a mountain[7]. To characterize the different wells, gullies, and hills of the mountain, there are a number of experimental approaches one can take: hydrogen exchange[12-15], NMR[16-18], fluorescence[19-21], and molecular dynamics simulations[22-24] to name a few. Those that are used in the work in this thesis are detailed below.

Circular Dichroism

Circular dichroism (CD) spectroscopy reports on protein secondary structure and, as such, can be used as a measure of "foldedness." This technique relies on the differential absorption of right and left circularly polarized light by chiral molecules. By their intrinsic nature, polypeptides are chiral. Regular secondary structures like α-helices and β-sheets have distinctive CD spectra, as do unfolded polypeptide chains. Unordered peptides are characterized by a minimum near 200 nm[25]. β-sheets typically have minima near 180 and 216 nm and a maximum near 195 nm[26], although these spectra are less constant than those of α-helices due to varying extent of twisting of β-sheets in different contexts[27]. α-helices have the strongest and most consistent spectral signal: minima at 208 and 222 nm and a maximum at 190 nm[25].

Since protein conformation is sensitive to environment, we can alter the environment and monitor conformational changes by CD. Therefore, CD is a great probe of the energy landscape. To determine the free energy difference (ΔG) between two conformations, one can monitor how the CD signal changes with perturbations of the energy landscape. The landscape can be perturbed with denaturants such as urea, guanidine, or temperature. The change in the relative populations of the two conformations as a function of denaturant can then be related to their ΔG under native conditions[28].

The height of the barriers between the different conformations determines the rates of interconversion and can therefore be probed by kinetic experiments. In these experiments, protein under conditions favoring one conformation is rapidly placed under conditions favoring a different conformation, and the change in CD signal is monitored over time. In the case of protein folding, while diluting concentrated unfolded protein into folding buffer would report on the folding reaction, whereas diluting folded protein into unfolding buffer would report on unfolding.

Mutational Analysis

As I alluded to in the previous section, mutations can be a powerful tool for probing the energy landscape. Classically, people have looked at structure and used mutational analysis to predict which residues are important for chemical interactions. For example, one approach for the identification of functionally critical residues is alanine scanning mutagenesis[29]. With this technique, the energetic or functional consequences of alanine substitution are assessed sequentially for an amino acid sequence to identify functional hot spots (see below). Similarly, mutations are often used to probe the nature of the transition state of a folding reaction, termed ϕ-value analysis[30]. Here, the energetic impact of a mutation on the native conformation relative to the transition state is taken to indicate whether that residue is structured in the transition state. By looking at a number of mutations distributed throughout the protein, one can determine the subset of native structure that is thought to exist in the transition state.

Of-course, not all mutations need to be experimentally designed; studying homologous proteins provides an excellent approach to investigate which biophysical characteristics are preserved for a particular structure and function and how those might be encoded within the amino acid sequence[31]. This has been further extended to look

beyond extant proteins to their ancestors, observing how changes in the energy landscape have been brought about by changes in the sequence and which of the landscape's many parameters are most important for function[32,33]. Another, more highly targeted approach is to employ directed evolution. Here, a gene is randomized and the resultant variants undergo selection for a function of choice, with the mutations that allow evolution of a novel function or optimization of an existing function providing insight into the conformations and motions necessary[34,35].

In sum, regardless of the particular technique used, altering the amino acid sequence enhances our understanding of the relationship between sequence, energy landscape, and function.

1.4 Modules of Protein Structure

Are there units of structure smaller than a single domain? If so, how do they communicate to create the overall protein fold? Protein domains are known to fold and unfold cooperatively; typically, at equilibrium under standard physiological conditions, no more than 5% of the population comprises partially folded intermediates[36]. Even in larger, multi-domain proteins, domains or regions of the protein are typically coupled in some manner[37,38]. How does the communication between these different regions of a protein allow information sensed on one side of a protein to be transmitted to the other? This is a central question in the work described in this thesis. For instance, proteins of therapeutic interest include a slew of receptors where a conformational change in one region of the protein is triggered by a binding or phosphorylation event at another region[39,40]. In addition, one promising approach for the development of drugs with high specificity and low side effects is to design drugs that allosterically regulate their targets[41]. Finally, understanding the components that make up the domain of a protein allows the recombination of regions and domains, useful as a way of redirecting or recombining functions[42-44]. Recent computational studies have attempted to compute the energetic and structural coupling between regions of a protein in order to predict function[45,46], but further study will be required if we are to arrive at a thorough and predictive understanding.

1.5 The Hot Spot Hypothesis

Do specific regions of proteins determine the energetics and specificity of ligand binding? Studies directed at this question let to the hot spot hypothesis, which arose out of the development of alanine scanning mutagenesis by Clackson and Wells[29] (mentioned above). The hypothesis states that protein-ligand binding is driven by a select number of residues at the binding interface. In essence, this presents the protein as a collection of one or more functional epitopes grafted onto an overdetermined structural scaffold. This view is entirely consistent a number of studies that have looked at the functional import of the hydrophobic core. The most significant aspect for residues within the core seems to be hydrophobicity, with specific packing between residues appearing to be a form of structural fine tuning rather than structure determination[47,48]. One striking illustration of this principle is a designed variant of the protein barnase with an entirely random hydrophobic core that nevertheless maintains the wild-type level of function[49].

These studies have been hugely influential, becoming the dominant school of thought when discussing protein-ligand binding. As protein engineering has continued to increase in popularity for industrial and therapeutic purposes, the potential of treating a protein as a collection of separable functional epitopes that can be mixed and matched to create novel functions and phenotypes has continued to become increasingly more attractive. This is the basis of techniques like protein grafting, wherein the binding epitope of one protein is grafted onto another scaffold that may be more pharmaceutically tractable or otherwise more attractive. For instance, Sia and Kim[50] grafted the non-continuous binding epitope of the HIV C-peptide, a potent inhibitor of HIV entry into cells, onto a leucine zipper scaffold. The result was a peptide with antiviral activity comparable to the C-peptide but with improved stability and protease resistance. Similarly, Kritzer *et al.*[51] developed miniature proteins that bind the oncoprotein hDM2 with nanomolar affinity, inhibiting hDM2 binding to p53. The design of enzymes with novel functions has also benefited from an analogous approach, where one computationally optimizes the residue identity and geometry of the catalytic site in isolation and then identifies a protein scaffold that can accommodate such an active site[52].

While the ability to dissect proteins into separable regions of structure and function has been thought of as a general principle, it has heretofore been applied to only a small subset of proteins, and the degree to which it can be generalized remains to be tested. Even in systems like antibodies, once thought to be a textbook example of proteins with distinct functional and structural regions, recent work has found a previously underappreciated role for protein motions and interdomain coupling[53]. Of particular interest for the work in this thesis, there have been few descriptions of modularity in intrinsically disordered or marginally stable proteins.

1.6 The Role of Disorder

Intrinsically disordered proteins (IDPs) are proteins that are natively unstructured over large portions – perhaps even the entirety – of their sequence[54]. They are responsible for a diverse host of functions, either in their unstructured state or by via a disorder-to-order transition. Their relevance has become increasingly more apparent as the number of IDPs identified has increased in recent years, due in large part to technical and theoretical advances in bioinformatic disorder prediction and NMR sensitivity[54]. Current estimates predict that, in any given eukaryote, an average 30% of proteins contain disordered regions longer than 30 amino acids[55].

IDPs exist in a state of dynamic conformational ensembles. For many, this translates to marginal stability of a structured conformation relative to an unstructured ensemble. Computational folding studies have shown decreasing protein stability increases the size of the cooperative core, where the cooperative core comprises residues that are all either folded or unfolded in states with the highest probabilities[45]. This would imply that those disordered proteins required to temporarily assume a structured conformation for function have a greater portion of their sequence acting as determinants of binding. The universality of the hot spot hypothesis is thus called into question: as the cooperative core increases, does the structural scaffold begin to overlap with the binding interface?

In the work outlined in this thesis, this question is explored using transcription factors MarA and the N-terminal domain of λ-repressor (herein called λ-repressor) as model systems. Both are small, helix-turn-helix proteins. λ-repressor is a canonical globular protein that was used in work that first posited the redundancy of hydrophobic core residues[47]. It has a well defined native structure[56] and binds to three highly similar DNA sequences[57]. MarA, on the other hand, is thought to be largely unfolded in the absence of its ligand[58], suggesting that its native conformation is at best marginally stable. It binds to at least 40 different binding sites *in vivo*[59], suggesting that under native conditions MarA populates, instead of a single native structure, an ensemble comprising multiple rapidly interconverting conformations. As summarized below, we find that these proteins' divergent landscapes impact the predictive power of the hot spot hypothesis and suggest conditions under which it should be applied.

1.7 Protein Folding of Protein Channels

Like IDPs, protein channels also fall outside of the predictive power of theories based on work done with archetypal globular proteins.

Phospholipid membranes serve to separate: they create the concept of self at a cellular level, delineating where the interior of an organelle or cell ends and where the outside world begins. This cellular self, however, continues to exist within this outside world and thus must be able to communicate with it. This communication can occur via protein channels embedded in the membrane that are able to detect environmental conditions and changes. Different channels are activated by different types of stimuli; these can be chemical, mechanical, or thermal[60]. Channels can serve as the first wave of a cellular response to the stimulus detected or as conduits for molecular transfer into or out of the cell. Many channel proteins have intra- or extracellular globular domains responsible for ligand binding or sensing the environment. How these domains transmit information to effect channel opening is poorly understood.

Protein channels have a unique set of factors governing their folding and dynamics. Their native conformation exists in a environment of striated hydrophobicity; channels in an animal cell's plasma membrane are surrounded by solvent in the cytoplasm and extracellularly while the regions in between are embedded within phospholipids. Furthermore, the channel opening, or pore, is formed by the transmembrane regions of the protein, which means the motions that open and close the pore may be heavily influenced by the composition of the surrounding lipids[61,62].

The distinctive factors that play an important role in membrane protein folding create technical hurdles – both for biology and for researchers – but conform to the same theory that motivates soluble protein folding work: that the native conformation is an equilibrium structure entirely encoded within the amino acid sequence[63]. *In vivo*, an extensive array of machinery able to get newly synthesized membrane proteins to their target membranes has been identified[64], but a detailed understanding of folding determinants and how these forces layer within the translocation pathways continue to be questions under active investigation[65]. Our knowledge of the conformational dynamics that are important for channel proteins and the impact had on these dynamics by various membrane components continue to increase as researchers continue to develop and

refine techniques for studying membrane protein folding *in vitro*[63] and integrate these studies with work mutational and phenotypic analyses[66].

The recognition of an environmental change and the opening of a channel protein present a highly refined example of conformational changes at one end of a protein driving conformational changes at the other. The covalent modification of cysteines within the cytoplasmic N-terminal domain of ion channel TRPA1 (see next section for further background) somehow induces pore opening to allow the influx of cations[67]. Ion channels that are involved in mechanotransduction, like those involved in hearing or touch, are activated by deflection of an external structure – i.e. the extracellular matrix – relative to an internal structure – i.e. the cytoskeleton[68]. Thus in addition to any therapeutic application of channel research (many are potential drug targets[69,70]), they have the potential to clarify the mechanism underlying information transfer between regions of a protein.

1.8 Thermosensor Meets Ankyrin Repeats

The TRP channel family of proteins is found in both vertebrates and invertebrates across a wide variety of tissues. They have been shown to be involved in nearly all aspects of somatosensation; including thermo- and nociception and mechanotransduction. Furthermore, these channels have been clinically linked to diabetes, chronic cough, Familial Alzheimer Disease, cancer, and Familial Episodic Pain Syndrome, just to name a few[69]. Their ubiquitous presence, impressively diverse range of agonists, and key positioning in cellular and organismal environmental sensing and response makes them attractive subjects of study.

The TRP superfamily has over 85 members among seven subfamilies: TRPC, TRPA, TRPV, TRPM, TRPN, TRPP, and TRPML[71]. All are cation channels thought to form tetramers, and all contain six putative transmembrane segments with intracellular N- and C-termini[60]. Residues critical for chemical agonist activation have been functionally mapped for some of the more highly studied channels[67,72,73], but a sensing and gating mechanism remains unknown.

Similarly, several regions have been identified as important for thermal response, but whether these are functionally conserved among homologues and how they behave in concert with the rest of the protein to detect and respond to temperature remains to be determined. In the *Drosophilia* TRPA1, residues in the pore region have been found to be important for thermal activation[74], while the rattlesnake TRPA1 sensitivity to thermal stimuli has been mapped to a subset of the ankyrin repeats within the cytoplasmic N-terminal domain[75] (Fig. 1.2A).

Ankyrin repeats are motifs of roughly 33 amino acids in a helix-turn-helix joined to adjacent repeats by a longer loop (Fig. 1.2B). While they get their name from the cytoskeletal protein ankyrin, which contains an impressive 24 of these repeats[76], they have been found within the proteins of organisms from viruses to humans[77]. Ankyrin repeat-containing proteins carry out an incredible variety of biological functions and are the ninth most common protein family in the human genome[78].

Ankyrin repeats are found in the cytoplasmic N-terminal domains of TRPC, TRPV, TRPA, and TRPN subfamily channels, with the number of repeats ranging from 2 – 29[78-80]. To our knowledge, no enzymatic activity has ever been attributed to ankyrin

repeats, and they are thought to function as structural, protein-binding, or ligand-binding domains. In TRPV1, the ankyrin repeats form a competitive binding site for ATP and calmodulin that regulates channel desensitization[81]. In TRPA1, a subset of cysteines within and adjacent to the ankyrin repeats is required for channel activation by its chemical agonists[67]. How information about the presence of ligand is propagated remains unknown in all cases. Regions within ankyrin repeats have similarly been shown through functional assays to be critical for thermal sensitivity[75] and hypothesized to be involved in mechanotransduction by acting as a mechanical spring[82]. To date, however, no sensing or gating mechanisms have been identified and the precise function of ankyrin repeats within TRP channels continues to be an active area of research.

1.9 Summary of Work in this Thesis

The work summarized in this thesis comprises two projects that aim to increase our understanding of the energetics of proteins outside the globular canon. In the first, we investigate whether the hydrophobic cores of natively disordered proteins play a significant role in protein-ligand binding (Chapter 2). In the second, we ask whether ankyrin repeat conformational dynamics in the cytoplasmic N-terminal domain of TRPA1 are involved in temperature sensing (Chapter 3).

Chapter 2: The Role of the Hydrophobic Core in Ligand Binding

Eight residues within the hydrophobic core of MarA were randomized in clusters of three. When the resulting libraries were selected for binding to the MarA consensus binding site, we found that MarA is highly intolerant to core mutation. This is in direct contrast to what we would expect based on the λ-repressor results described previously[47], which suggested that a protein's core should be highly tolerant to mutations as long as they are hydrophobic. We further show that the mutations identified in the selection differentially destabilize different conformations within the MarA native state ensemble, such that a single core mutation can actually alter MarA binding specificity.

Chapter 3: Ankyrin Repeats and Thermosensation

As mentioned previously, ankyrin repeats within the N-terminal domain of rattlesnake TRPA1 are crucial for its ability to respond to heat[75]. In order to determine whether ankyrin repeats sense and respond to heat via conformational changes, we expressed and purified the previously identified thermal response modules, repeats 3-8 and 10-15, from both human and rattlesnake. Remarkably, the fragment comprising rattlesnake repeats 3-8, but not the other three fragments, remains folded up to at least 95˚C. The mechanism for this thermotolerance and the potential physiological implications are still under investigation.

1.10 References

1. Nogales, E., Wolf, S. G. & Downing, K. H. Structure of the alpha beta tubulin dimer by electron crystallography. *Nature* **391,** 199–203 (1998).
2. Lee, C.-H., Kim, M.-S., Chung, B. M., Leahy, D. J. & Coulombe, P. A. Structural basis for heteromeric assembly and perinuclear organization of keratin filaments. *Nat. Struct. Mol. Biol.* **19,** 707–715 (2012).
3. Brennan, R. G. & Matthews, B. W. Structural basis of DNA-protein recognition. *Trends Biochem. Sci.* **14,** 286–290 (1989).
4. Rasmussen, B. F., Stock, A. M., Ringe, D. & Petsko, G. A. Crystalline ribonuclease A loses function below the dynamical transition at 220 K. *Nature* **357,** 423–424 (1992).
5. Goh, C.-S., Milburn, D. & Gerstein, M. Conformational changes associated with protein-protein interactions. *Curr. Opin. Struct. Biol.* **14,** 104–109 (2004).
6. Mitchison, T. & Kirschner, M. Dynamic instability of microtubule growth. *Nature* **312,** 237–242 (1984).
7. Dill, K. A. & Chan, H. S. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4,** 10–19 (1997).
8. Phillips, A. H. *et al.* Conformational dynamics control ubiquitin-deubiquitinase interactions and influence in vivo signaling. *Proc. Natl. Acad. Sci. U.S.A.* **110,** 11379–11384 (2013).
9. Levin, A. M. *et al.* Exploiting a natural conformational switch to engineer an interleukin-2 'superkine'. *Nature* **484,** 529–533 (2012).
10. Bucciantini, M. *et al.* Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* **416,** 507–511 (2002).
11. Sali, A., Shakhnovich, E. & Karplus, M. How does a protein fold? *Nature* **369,** 248–251 (1994).
12. Fazelinia, H., Xu, M., Cheng, H. & Roder, H. Ultrafast hydrogen exchange reveals specific structural events during the initial stages of folding of cytochrome c. *J. Am. Chem. Soc.* **136,** 733–740 (2014).
13. Hu, W. *et al.* Stepwise protein folding at near amino acid resolution by hydrogen exchange and mass spectrometry. *Proceedings of the National Academy of Sciences* **110,** 7684–7689 (2013).
14. Wang, G., Abzalimov, R. R., Bobst, C. E. & Kaltashov, I. A. Conformer-specific characterization of nonnative protein states using hydrogen exchange and top-down mass spectrometry. *Proceedings of the National Academy of Sciences* **110,** 20087–20092 (2013).
15. Wildes, D. & Marqusee, S. Hydrogen-exchange strategies applied to energetics of intermediate processes in protein folding. *Meth. Enzymol.* **380,** 328–349 (2004).
16. Cembran, A., Kim, J., Gao, J. & Veglia, G. NMR mapping of protein conformational landscapes using coordinated behavior of chemical shifts upon ligand binding. *Phys Chem Chem Phys* **16,** 6508–6518 (2014).
17. Moschen, T. & Tollinger, M. A kinetic study of domain swapping of Protein L. *Phys Chem Chem Phys* **16,** 6383–6390 (2014).
18. Gupta, S. & Bhattacharjya, S. NMR Characterization of the Near Native and

Unfolded States of the PTB Domain of Dok1: Alternate Conformations and Residual Clusters. *PLoS ONE* **9,** e90557 (2014).
19. Sarkar, S. & Chattopadhyay, K. Studies of Early Events of Folding of a Predominately β-Sheet Protein Using Fluorescence Correlation Spectroscopy and Other Biophysical Methods. *Biochemistry* **53,** 1393–1402 (2014).
20. Lamprou, P., Kempe, D., Katranidis, A., Büldt, G. & Fitter, J. Nanosecond Dynamics of Calmodulin and Ribosome-Bound Nascent Chains Studied by Time-Resolved Fluorescence Anisotropy. *Chembiochem* (2014). doi:10.1002/cbic.201400014
21. Banerjee, P. R. & Deniz, A. A. Shedding light on protein folding landscapes by single-molecule fluorescence. *Chem Soc Rev* **43,** 1172–1188 (2014).
22. Chong, L. T., Snow, C. D., Rhee, Y. M. & Pande, V. S. Dimerization of the p53 oligomerization domain: identification of a folding nucleus by molecular dynamics simulations. *Journal of Molecular Biology* **345,** 869–878 (2005).
23. Hinczewski, M., Gebhardt, J. C. M., Rief, M. & Thirumalai, D. From mechanical folding trajectories to intrinsic energy landscapes of biopolymers. *Proceedings of the National Academy of Sciences* **110,** 4500–4505 (2013).
24. Lindorff-Larsen, K., Piana, S., Dror, R. O. & Shaw, D. E. How fast-folding proteins fold. *Science* **334,** 517–520 (2011).
25. Woody, R. W. Circular dichroism. *Meth. Enzymol.* **246,** 34–71 (1995).
26. Brahms, S., Brahms, J., Spach, G. & Brack, A. Identification of beta,beta-turns and unordered conformations in polypeptide chains by vacuum ultraviolet circular dichroism. *Proc. Natl. Acad. Sci. U.S.A.* **74,** 3208–3212 (1977).
27. Manning, M. C., Illangasekare, M. & Woody, R. W. Circular dichroism studies of distorted alpha-helices, twisted beta-sheets, and beta turns. *Biophys. Chem.* **31,** 77–86 (1988).
28. Greene, R. F. & Pace, C. N. Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin, and beta-lactoglobulin. *J. Biol. Chem.* **249,** 5388–5393 (1974).
29. Clackson, T. & Wells, J. A. A hot spot of binding energy in a hormone-receptor interface. *Science* **267,** 383–386 (1995).
30. Matouschek, A., Kellis, J. T., Serrano, L. & Fersht, A. R. Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **340,** 122–126 (1989).
31. Trudeau, D. L., Smith, M. A. & Arnold, F. H. Innovation by homologous recombination. *Curr Opin Chem Biol* **17,** 902–909 (2013).
32. Thornton, J. W. Resurrecting ancient genes: experimental analysis of extinct molecules. *Nat. Rev. Genet.* **5,** 366–375 (2004).
33. Carroll, S. M., Ortlund, E. A. & Thornton, J. W. Mechanisms for the evolution of a derived function in the ancestral glucocorticoid receptor. *PLoS Genet.* **7,** e1002117 (2011).
34. Brinkmann-Chen, S. *et al.* General approach to reversing ketol-acid reductoisomerase cofactor dependence from NADPH to NADH. *Proceedings of the National Academy of Sciences* **110,** 10946–10951 (2013).
35. Wu, I. & Arnold, F. H. Engineered thermostable fungal Cel6A and Cel7A

cellobiohydrolases hydrolyze cellulose efficiently at elevated temperatures. *Biotechnol. Bioeng.* **110,** 1874–1883 (2013).
36. Freire, E., Murphy, K. P., Sanchez-Ruiz, J. M., Galisteo, M. L. & Privalov, P. L. The molecular basis of cooperativity in protein folding. Thermodynamic dissection of interdomain interactions in phosphoglycerate kinase. *Biochemistry* **31,** 250–256 (1992).
37. Shepherd, C. M. & Vogel, H. J. A molecular dynamics study of Ca(2+)-calmodulin: evidence of interdomain coupling and structural collapse on the nanosecond timescale. *Biophys. J.* **87,** 780–791 (2004).
38. Chong, P. A., Lin, H., Wrana, J. L. & Forman-Kay, J. D. Coupling of tandem Smad ubiquitination regulatory factor (Smurf) WW domains modulates target specificity. *Proceedings of the National Academy of Sciences* **107,** 18404–18409 (2010).
39. Kuriyan, J. Allostery and coupled sequence variation in nuclear hormone receptors. *Cell* **116,** 354–356 (2004).
40. Christopoulos, A. & Kenakin, T. G protein-coupled receptor allosterism and complexing. *Pharmacol. Rev.* **54,** 323–374 (2002).
41. Christopoulos, A. Allosteric binding sites on cell-surface receptors: novel targets for drug discovery. *Nat Rev Drug Discov* **1,** 198–210 (2002).
42. Paduch, M. *et al.* Generating conformation-specific synthetic antibodies to trap proteins in selected functional states. *Methods* **60,** 3–14 (2013).
43. Dueber, J. E., Mirsky, E. A. & Lim, W. A. Engineering synthetic signaling proteins with ultrasensitive input/output control. *Nat. Biotechnol.* **25,** 660–662 (2007).
44. Sallee, N. A., Yeh, B. J. & Lim, W. A. Engineering modular protein interaction switches by sequence overlap. *J. Am. Chem. Soc.* **129,** 4606–4611 (2007).
45. Hilser, V. J., Dowdy, D., Oas, T. G. & Freire, E. The structural distribution of cooperative interactions in proteins: analysis of the native state ensemble. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 9903–9908 (1998).
46. Smock, R. G. *et al.* An interdomain sector mediating allostery in Hsp70 molecular chaperones. *Mol. Syst. Biol.* **6,** 414 (2010).
47. Lim, W. A. & Sauer, R. T. Alternative packing arrangements in the hydrophobic core of lambda repressor. *Nature* **339,** 31–36 (1989).
48. Lim, W. A. & Sauer, R. T. The role of internal packing interactions in determining the structure and stability of a protein. *Journal of Molecular Biology* **219,** 359–376 (1991).
49. Axe, D. D., Foster, N. W. & Fersht, A. R. Active barnase variants with completely random hydrophobic cores. *Proc. Natl. Acad. Sci. U.S.A.* **93,** 5590–5594 (1996).
50. Sia, S. K. & Kim, P. S. Protein grafting of an HIV-1-inhibiting epitope. *Proc. Natl. Acad. Sci. U.S.A.* **100,** 9756–9761 (2003).
51. Kritzer, J. A. *et al.* Miniature protein inhibitors of the p53-hDM2 interaction. *Chembiochem* **7,** 29–31 (2006).
52. Jiang, L. *et al.* De novo computational design of retro-aldol enzymes. *Science* **319,** 1387–1391 (2008).
53. Sela-Culang, I., Kunik, V. & Ofran, Y. The Structural Basis of Antibody-Antigen Recognition. *Front Immunol* **4,** 302 (2013).
54. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions.

Nat. Rev. Mol. Cell Biol. **6,** 197–208 (2005).
55. Peng, Z., Mizianty, M. J. & Kurgan, L. Genome-scale prediction of proteins with long intrinsically disordered regions. *Proteins* **82,** 145–158 (2014).
56. Jordan, S. R. & Pabo, C. O. Structure of the lambda complex at 2.5 A resolution: details of the repressor-operator interactions. *Science* **242,** 893–899 (1988).
57. Ptashne, M. *et al.* How the lambda repressor and cro work. *Cell* **19,** 1–11 (1980).
58. Dangi, B. *et al.* Structure and dynamics of MarA-DNA complexes: an NMR investigation. *Journal of Molecular Biology* **314,** 113–127 (2001).
59. Barbosa, T. M. & Levy, S. B. Differential expression of over 60 chromosomal genes in Escherichia coli by constitutive expression of MarA. *J. Bacteriol.* **182,** 3467–3474 (2000).
60. Latorre, R., Zaelzer, C. & Brauchi, S. Structure-functional intimacies of transient receptor potential channels. *Q. Rev. Biophys.* **42,** 201–246 (2009).
61. Andersen, O. S. & Koeppe, R. E. Bilayer thickness and membrane protein function: an energetic perspective. *Annu Rev Biophys Biomol Struct* **36,** 107–130 (2007).
62. Phillips, R., Ursell, T., Wiggins, P. & Sens, P. Emerging roles for lipids in shaping membrane-protein function. *Nature* **459,** 379–385 (2009).
63. Stanley, A. M. & Fleming, K. G. The process of folding proteins into membranes: challenges and progress. *Arch. Biochem. Biophys.* **469,** 46–66 (2008).
64. Wickner, W. & Schekman, R. Protein translocation across biological membranes. *Science* **310,** 1452–1456 (2005).
65. Delaney, E., Khanna, P., Tu, L., Robinson, J. M. & Deutsch, C. Determinants of pore folding in potassium channel biogenesis. *Proceedings of the National Academy of Sciences* (2014). doi:10.1073/pnas.1324274111
66. Vásquez, V., Krieg, M., Lockhead, D. & Goodman, M. B. Phospholipids that Contain Polyunsaturated Fatty Acids Enhance Neuronal Cell Mechanics and Touch Sensation. *Cell Rep* (2013). doi:10.1016/j.celrep.2013.12.012
67. Macpherson, L. J. *et al.* Noxious compounds activate TRPA1 ion channels through covalent modification of cysteines. *Nature* **445,** 541–545 (2007).
68. Gillespie, P. G. & Walker, R. G. Molecular basis of mechanosensory transduction. *Nature* **413,** 194–202 (2001).
69. Kaneko, Y. & Szallasi, A. Transient Receptor Potential (TRP) channels: a clinical perspective. *Br. J. Pharmacol.* (2013). doi:10.1111/bph.12414
70. Hutt, D. M. *et al.* FK506 binding protein 8 peptidylprolyl isomerase activity manages a late stage of cystic fibrosis transmembrane conductance regulator (CFTR) folding and stability. *J. Biol. Chem.* **287,** 21914–21925 (2012).
71. Montell, C. The TRP superfamily of cation channels. *Sci. STKE* **2005,** re3 (2005).
72. Johnson, D. M. *et al.* Functional mapping of the transient receptor potential vanilloid 1 intracellular binding site. *Mol. Pharmacol.* **70,** 1005–1012 (2006).
73. Wang, L., Cvetkov, T. L., Chance, M. R. & Moiseenkova-Bell, V. Y. Identification of in vivo disulfide conformation of TRPA1 ion channel. *J. Biol. Chem.* **287,** 6169–6176 (2012).
74. Wang, H., Schupp, M., Zurborg, S. & Heppenstall, P. A. Residues in the pore region of Drosophila transient receptor potential A1 dictate sensitivity to thermal stimuli. *J. Physiol. (Lond.)* **591,** 185–201 (2013).

75. Cordero-Morales, J. F., Gracheva, E. O. & Julius, D. Cytoplasmic ankyrin repeats of transient receptor potential A1 (TRPA1) dictate sensitivity to thermal and chemical stimuli. *Proc. Natl. Acad. Sci. U.S.A.* **108,** E1184–91 (2011).
76. Lux, S. E., John, K. M. & Bennett, V. Analysis of cDNA for human erythrocyte ankyrin indicates a repeated structure with homology to tissue-differentiation and cell-cycle control proteins. *Nature* **344,** 36–42 (1990).
77. Sedgwick, S. G. & Smerdon, S. J. The ankyrin repeat: a diversity of interactions on a common structural framework. *Trends Biochem. Sci.* **24,** 311–316 (1999).
78. Gaudet, R. A primer on ankyrin repeat function in TRP channels and beyond. *Mol Biosyst* **4,** 372–379 (2008).
79. Walker, R. G., Willingham, A. T. & Zuker, C. S. A Drosophila mechanosensory transduction channel. *Science* **287,** 2229–2234 (2000).
80. Story, G. M. *et al.* ANKTM1, a TRP-like channel expressed in nociceptive neurons, is activated by cold temperatures. *Cell* **112,** 819–829 (2003).
81. Lishko, P. V., Procko, E., Jin, X., Phelps, C. B. & Gaudet, R. The ankyrin repeats of TRPV1 bind multiple ligands and modulate channel sensitivity. *Neuron* **54,** 905–918 (2007).
82. Lee, G. *et al.* Nanospring behaviour of ankyrin repeats. *Nature* **440,** 246–249 (2006).
83. McCleverty, C. J., Koesema, E., Patapoutian, A., Lesley, S. A. & Kreusch, A. Crystal structure of the human TRPV2 channel ankyrin repeat domain. *Protein Sci.* **15,** 2201–2206 (2006).

FIGURE LEGENDS

Figure 1.1. Energy landscape for a hypothetical protein, with conformational freedom on the horizontal axes and internal free energy on the vertical axis. Generated by the Dill lab[7].

Figure 1.2. A) Cartoon schematic of TRPA1, with cylinders representing the transmembrane domains and circles representing ankyrin repeats. B) Crystal structure of the TRPV2 channel ankyrin repeat domain, with each ankyrin repeat individually colored. PDB ID: 2F37[83]

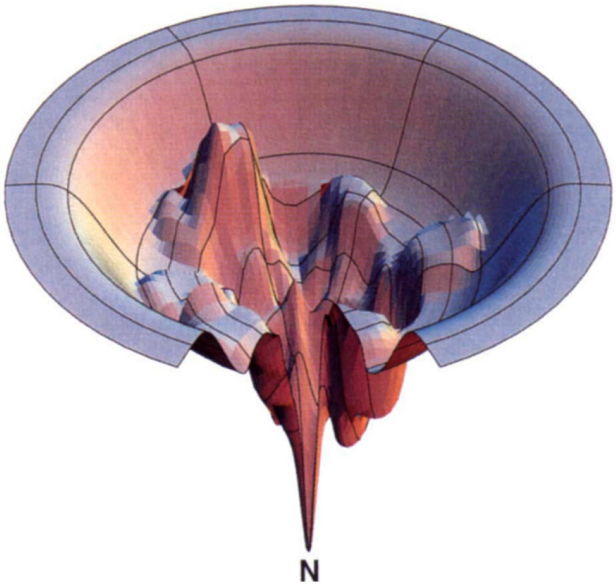


Figure 1.1. Energy landscape for a hypothetical protein, with conformational freedom on the horizontal axes and internal free energy on the vertical axis. Generated by the Dill lab[7].

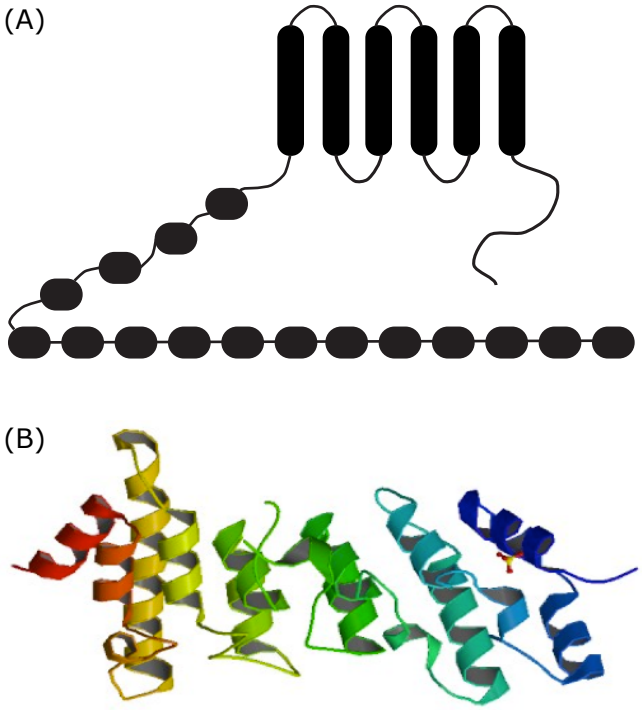


Figure 1.2. A) Cartoon schematic of TRPA1, with cylinders representing the transmembrane domains and circles representing ankyrin repeats. B) Crystal structure of the TRPV2 channel ankyrin repeat domain, with each ankyrin repeat individually colored. PDB ID: 2F37[83]

Chapter 2: When the scaffold can't be ignored: the role of the hydrophobic core in ligand binding and specificity

2.1 Abstract

The traditional view of protein-ligand binding treats a protein as comprising regionally distinct binding epitopes on the surface of a degenerate structural scaffold, largely ignoring the impact of a protein's energy landscape. To determine the robustness of this simplification, we compared two small helix-turn-helix transcription factors with dissimilar energy landscapes. λ-repressor is stable and well folded, while MarA appears marginally stable with multiple native conformations (molten). We find MarA drastically less tolerant to core mutation than λ-repressor. Moreover, core mutations change the relative affinities of its binding partners, altering ligand specificity. These results can be explained by taking into account the mutational effects on protein stability together with their differential destabilization of MarA's native conformations. Thus, for proteins with multiple conformations that are close in energy, such as many intrinsically disordered proteins, residues distant from the active site can alter both binding affinity and specificity.

2.2 Introduction

Informational content per residue is a way of describing the degree to which a residue contributes to a protein's function. For instance, binding is typically thought about locally, as regions of high informational content that are superimposed upon a structural scaffold that is informationally poor. The majority of binding energy is proposed to come from a minority of residues that function as binding hot spots[1]; these are treated as functionally separable from an over-determined protein core with minimal sequence requirements[2,3].

This assumption of separability between binding and fold has been hugely important for protein design and related research. Techniques like protein grafting[4-6] explicitly treat proteins as collections of distinct modules that can be recombined to create new functions. Work investigating the evolution of protein folds or structure has also frequently been based on the assumption that a protein's general fold can be treated as a scaffold for the active site, largely independent of the identity of active site residues[7,8]. Even the use of alanine scanning mutagenesis to identify sites for biochemical study or drug design[9] works best for proteins that can be effectively modeled by separable functional epitopes. Many of these techniques rely on the ability to screen a large number of candidates, both *in silico* and via experimental means. The potential for engineering enzymes with new functions or new inhibitors for medically relevant targets would be greatly enhanced if we could better predict which residues are information rich and which are not.

One obvious limitation to the assumption of modularity is that the native structure does not provide the entire picture. Proteins are not static; under native conditions they populate an ensemble of conformations, including unfolded conformations, according to their relative stability. Any experiment to measure binding will also be affected by this pre-existing equilibrium between binding competent and binding incompetent conformations. Thus, it is more accurate to talk about an apparent dissociation constant, $K_{d,app} = K_d (1 + K_f)$, that incorporates both folding and binding energetics (see Discussion). This scenario is further complicated for proteins with

multiple binding targets, as these proteins are likely to have an ensemble of native conformations that are close in energy. The proteins' interactions with each of their substrates may be differentially affected by mutation. Although this limitation has been acknowledged, its exploitation or circumvention has, with few exceptions[10], remained largely elusive.

The degree to which aspects of a protein's energy landscape correlate with the distribution of informational content (how much a residue contributes to a protein's function) across its amino acid sequence remains unclear. Here, we begin to address this question by characterizing how proteins with highly divergent energy landscapes differ in their abilities to modulate binding via regions distant from the binding site.

MarA and λ-repressor's DNA-binding domain are both small helix-turn-helix transcription factors with very different energy landscapes (Fig. 2.1A). For both, the binding interface geometry and binding sites have been well documented. λ-repressor binds to only three different sites, all quite similar to one another[11], and has a well defined native conformation of archetypal stability[12]. Structural information about MarA is available from both X-ray crystallography[13] and NMR[14,15], and its sequence logo has been experimentally determined[16]. *In vivo*, it regulates expression of over 60 genes[17], which have different binding-site (marbox) permutations at their promoters. Marboxes are fairly degenerate, but MarA is quite sensitive to their orientation and location relative to the transcriptional start site and thus binds different promoters with varying affinity [18]. It requires 50% glycerol for MarA to remain soluble in the absence of its ligand, suggesting that, in isolation, it is largely unfolded. Taken together, these data indicate that MarA may be molten, i.e. marginally stable with multiple native conformations.

We show that if a protein is on the brink of stability, then mutations *throughout* the entire protein will have dramatic effects on function, e.g. ligand binding; here, sites distant from the active site play a role that goes beyond defining the fold. Moreover, we show that for promiscuous proteins (proteins with many high affinity binding partners), a single site substitution within the hydrophobic core, distant from the binding site, is sufficient to alter binding specificity.

2.3 Results

Library construction

To probe the functional impact of subtle changes to the energy landscape without directly affecting the binding surface, we created three libraries of MarA variants with diverse stabilities by changing residues within the hydrophobic core. Each library was randomized at three clustered positions within the core, defined as comprising residues that are 95% or more solvent inaccessible (Fig. 2.1B): I13, F48, and L56 comprise cluster 1; I68, L94, and F98, cluster 2; and L72, F98, and Y109, cluster 3. The core positions chosen are diverse both in their location within the protein and in their conservation among homologues (20 – 99%); none contact DNA.

To confirm unbiased library compositions, the individual libraries were evaluated for their sequence diversity. Each library was transformed into cells and grown in the absence of selection; 86 – 94 transformants from each library were then sequenced. The results were compared to a simulation of the above protocol; the simulation was

performed with the assumption that there is no bias for any codon over another and repeated 1000 times (Fig. 2.2A). Our experimental data correspond to the results of the simulation, consistent with the presence of a diverse and unbiased library.

GENETIC SELECTION TO IDENTIFY FUNCTIONAL CORE VARIANTS OF MARA

To identify functional variants within each library, we employed a genetic selection system wherein growth on tetracycline reports on the presence of functional MarA protein (Fig. 2.2B). Surprisingly, of the roughly 24,000 MarA library variants evaluated, only 17 retain wild-type function sufficient to pass the selection; these 17 feature almost exclusively conservative substitutions (Table I). The number of molecules subjected to selection is more than 3.4 times the size of our library, which corresponds to a greater than 95% chance of having evaluated any random mutant at least once. Additionally, most positive hits have been picked up repeatedly, indicating we have effectively sampled the available sequence space.

The positive hits identified in our selection can be divided into functional classes based on growth on solid media (Fig. 2.3A). Class 1 variants (10 members) are defined as those that demonstrate wild type-like growth on plates following a 24-hour incubation when transformed into *E. coli*. The four class 2 variants similarly result in numerous colonies but require a 48-hour incubation. Finally, the three class 3 variants exhibit the least robust growth, resulting in very few colonies after a 24-hour incubation and showing no inclination to be rescued by extended incubation.

These functional differences, however, are not reflected in growth rates obtained from growth-curve analyses. Measured in this way, even the most dramatic differences in growth rates among variants are less than two fold, and the standard error of all measured growth rates is smaller than that of any one variant's replicates (Fig. 2.3B). *In vitro*, class 3 variants have a dramatically increased propensity for aggregation compared to class 1 or even class 2 (Fig. 2.3C). The qualitative correlation between the penetrance of tetracycline resistance and protein aggregation suggests the latter as a possible explanation of the phenotypic differences between functional variants.

DESTABILIZED Λ-REPRESSOR RECAPITULATES MARA'S INTOLERANCE TO CORE SUBSTITUTION

Our results with MarA seem at odds with previous published studies using the protein λ-repressor. If hydrophobicity were the sole or nearly sole requirement for core residues, as has been reported for proteins like λ-repressor[19], we would expect 1,022 variants to pass the selection. However, it appears that MarA is strikingly sensitive to core mutation. To determine whether marginal stability is a cause of this intolerance, we tested a subset of previously identified functional λ-repressor variants[19] in a destabilized background. If our hypothesis is correct, a destabilized λ-repressor should be tolerant of only the least disruptive of substitutions.

To create a destabilized background, we selected the λ-repressor variant R17A/S77A[12]. This variant destabilizes the protein by 2.8 $\pm$ 0.3 kcal/mol with a resulting ΔG_u = 2.5 $\pm$ 0.1 kcal/mol (Fig. 2.4). The mutations disrupt a salt bridge interaction distant from the active site; the resulting protein is folded and active. We evaluated the effects of three previously studied core variants V36I, V36I/M40F, and V36I/M40I/V47I in both the wild-type and destabilized λ-repressor backgrounds. While V36I was active in both,

V36I/M40F and V36I/M40I/V47I yielded active proteins only in the wild-type background (Table II). Taken together, these data indicate that decreasing the global stability of λ-repressor is sufficient to alter its tolerance to core mutations – that is to say, the distribution of informational content per residue – to more closely resemble that of MarA.

CORE MUTATION DIFFERENTIALLY AFFECTS MARA BINDING TO DIFFERENT PROMOTERS, ALTERING BINDING SPECIFICITY

The specificity of an interaction reflects the preference for a protein to bind one particular ligand over another; specificity is a measure of a protein's relative affinity for different ligands[20]. Because mutations within MarA's hydrophobic core can abolish binding by destabilizing the native conformation relative to the unfolded conformation, we hypothesized that core mutations could actually alter binding specificity by destabilizing certain native conformations relative to others.

To test this, we purified wild type MarA and one of the single core variant isolated in the functional selection, and we determined their relative affinities to three different known binding sites: consensus, which contains the most frequently seen base at each position, and naturally occuring *micF* and *zwf*[18]. Binding was detected by an electrophoretic mobility shift assay (EMSA) (Fig. 2.5A). Both the wild type and the core variant bind to the consensus and *micF* oligos with micromolar affinity (Fig. 2.5B). Wild-type MarA binds to the *zwf* at an affinity of 16 μM; I68V-*zwf* binding is undetectable in our assay, corresponding to a K_d > 95mM. Thus, these single-site core variants appear to selectively affect binding to the *zwf* sequence; the minimum concentration required for either single variant is at least 54 times that required for wild type-consensus binding.

Both variants bind to the consensus sequence most strongly, as expected since both were isolated in a selection based on consensus binding. While specificity changes are frequently illustrated as changes in binding *preference*, altering the relative affinity (K_{d1}/K_{d2}) to go from a number greater than 1 to a number less than 1, we believe this threshold is arbitrary and is actually asking much less than what we have shown. For instance, a change from 1.2 to 0.8 actually reflects a very small change in the relative energetics of binding (ΔΔG ~.2 kcal/mol). On the other hand, our variants show a larger change in the relative energetics of >1.4 kcal/mol.

To observe this change in specificity more directly, we monitored MarA's relative binding to each of the two naturally occurring promoters in a competition format. In this competition assay, one oligo is labeled with a green 6-FAM dye, while the second oligo is labeled with a red TYE 665 dye. As shown in figure 2.6, MarA has a slight preference for the red-labeled *zwf* oligo (~3-fold decrease in K_d), and the preference goes up an additional five-fold when compared to red-labeled *micF* oligo (for a K_d(*zwf*-FAM)/K_d(*micF*-TYE)~15). The variant I68V, however, only shows a slight (2-fold) preference for the green 6-FAM-labeled *micF* oligo and loses all detectable binding to the *zwf* oligo. Hence, while the wild-type protein has an approximately five-fold preference for *micF* over *zwf*, consistent with the affinities determined for each binding site individually, I68V retains binding to *micF* with no detectable binding to *zwf*, even at the micromolar concentrations of protein and DNA used for the competition. Taken

together, these data confirm that mutations within the hydrophobic core can alter MarA specificity by decreasing affinity for one binding site while maintaining affinity for another. Thus, remarkably, core mutations can alter both specificity and affinity.

2.4 Discussion

We have demonstrated that for proteins with multiple conformations close in energy, substitutions at sites distant from the active site can have dramatic ramifications for protein function. Whereas canonical well-folded globular proteins like λ-repressor[3,19] and barnase[2] have been shown to have hydrophobic cores with little informational content, MarA possesses a hydrophobic core that is intensely intolerant to substitution. Moreover, conservative substitutions within MarA's hydrophobic core can alter its relative ligand specificity.

MarA expresses insolubly, requires 50% glycerol to remain in solution, and is incredibly prone to aggregation. This suggests that it does not respond to core substitutions like a canonical well-folded globular protein because it is, in fact, better modeled as being marginally stable and molten.

A difference in global stability is alone sufficient to explain MarA's intolerance to core mutation. For a stable globular protein, small changes in stability do not change the concentration of the binding-competent state to a meaningful degree. For a marginally stable protein, the same small changes can appreciably shift the equilibrium towards binding-incompetent states and, in so doing, have profound functional consequences. Let us outline these comments in a quantitative manner for clarity. For a simple two-state system:

$$U + L \xleftrightarrow{K_f} N + L \xleftrightarrow{K_d} NL,$$

where L is the ligand, U is the unfolded polypeptide, and N is the native protein,

$$K_f = \frac{[U]}{[N]} \qquad K_d = \frac{[N][L]}{[NL]}.$$

The apparent dissociation constant of the system, $K_{d,app}$, can thus be described as:

$$K_{d,app} = \frac{([U]+[N])[L]}{[NL]} = K_f K_d + K_d.$$

For ordered proteins, K_f is small and the above effect becomes largely invisible experimentally. For example, for a hypothetical well-ordered globular protein with $K_d = 500nM$ and $\Delta G_f = 8.2$ kcal/mol (and thus $K_f = 10^{-7}$),

$$\begin{aligned} K_{d,app} &= 5\times10^{-7}(1+5\times10^{-7}) \\ &= 5\times10^{-7} + 5\times10^{-14} \approx 5\times10^{-7} \end{aligned}.$$

If this protein is destabilized by 3 kcal/mol, the resultant variant has a $K_f = 1.6\times10^{-5}$ and still,

$$\begin{aligned} K_{d,app} &= 5\times10^{-7} + 8\times10^{-12} \\ &\approx 5\times10^{-7} \end{aligned}.$$

Conversely, proteins that are marginally stable will have K_f's that are not insignificant. For a protein with $K_d = 500nM$ but $\Delta G_f = 0$ kcal/mol (and thus $K_f = 1$),

$$K_{d,app} = 5 \times 10^{-7}(1+1)$$
$$= 1 \times 10^{-6}$$

Destabilizing this protein by 3 kcal/mol leads to a $K_f = 160$ and now,

$$K_{d,app} = 5 \times 10^{-7}(1+160)$$
$$= 5 \times 10^{-7} + 8 \times 10^{-5} \approx 8.05 \times 10^{-5}$$

This simple two-state thermodynamic perspective outlines how the ability to retain function changes drastically across distinct regimes of global stability. However, many marginally stable proteins bind to multiple targets and thus could be considered molten with the potential to populate multiple native conformations. For these proteins, the above analysis can be extended to changes in stability between any functionally distinct conformations; for conformations that are close in energy, mutations at a distance can alter the population of one native conformation relative to others[21,22] and, expanding the discussion above from merely maintaining binding to actually altering specificity.

We directly show this to be true for MarA. As previously mentioned, MarA binds to a number of different targets and is thus likely to have a number of distinct native conformations that are likely to be differentially affected by core mutations. We have identified two MarA core variants whose relative affinities for two naturally occurring promoter sequences differ from that of the wild type, providing a direct demonstration that specificity can be altered by mutations located solely within the hydrophobic core. This is consistent with previous work showing that alanine substitutions at multiple positions within MarA differentially impact binding to different promoters[23]. The disparate impact of core or other mutations on different native states is, of-course, not limited to marginally stable proteins[24]. However, the high informational content per residue in these proteins means that even subtle changes that would be functionally invisible in stable globular proteins can alter specificity in a protein whose energy landscape is akin to that of MarA.

A major protein class likely to be enriched in MarA-like energy landscape is intrinsically disordered proteins (IDPs), a significant percentage of the eukaryotic proteome[25]. Although IDPs have been proposed to be more tolerant to mutation than well-folded proteins[26], their energy landscapes more closely resemble that of MarA than of λ-repressor, indicating that the converse may actually be true for those whose function requires a disorder-to-order transition. Thus, this work provides a potential functional role for a subset of the high percent of the human genome that encodes disorder. Additionally, it offers some tantalizing clues as to what factors may have favored the evolution of such proteins, as several studies have implicated functional plasticity as a critical factor in the molecular evolution of new function[27-29]. In this work, we've shown that marginally stable proteins have a broad distribution of informational content over their amino acid sequence, increasing the probability that a random mutation will have an impact on binding. The increased ease of effecting appreciable change may explain the favorability of marginal stability to the evolution of novel binding targets.

It has been proposed that the maintenance of organismal evolvability may depend on the ability to have large-scale functional impact with a small number of mutations[30].

Multi-gene regulators like MarA present the possibility to accomplish exactly this, as even small changes in specificity for promiscuous proteins could affect multiple pathways and result in global changes within a cell. Biophysical and biological plasticity thus represent two related but distinct nodes upon which evolution can act to rapidly achieve phenotypic diversity. This may be part of the reason why transcription factors, which frequently bind multiple targets[31], are enriched for intrinsically disordered regions[32,33].

Our experimental results demonstrate that the shape of the energy landscape and, in particular, global stability are indicative of the distribution of informational content over a protein's sequence. In essence, stable proteins are overdetermined and can accommodate mutations more easily, whereas molten proteins have less informational overlap between residues. This relationship between energy landscape and informational content can guide protein design in two ways. Firstly, it can increase protein design efficiency by identifying proteins with distinct regions high in informational content that can be treated as functionally modular. Secondly, it can suggest a path for engineering proteins with dispersed informational content that have previously been problematic for protein engineering, highlighting the importance of incorporating non-interface residues into design strategies.

2.5 Materials & Methods

Gene synthesis

Libraries of MarA genes with selected codons randomized were purchased from GeneArt®, PCR-amplified, and inserted into the selection plasmid [16]. These were transformed into XL10-Gold cells for library propagation, plated on 30 μg/ml ampicillin LB plates. ~200,000 colonies per library were resuspended in liquid media, and the plasmids were extracted using the Qiagen Midi-prep Kit. All genes were subcloned into pET28 plasmid for purification.

We used the λ-repressor N-terminal domain from plasmid pWL104[19]. Specific point mutations were introduced using QuikChange mutagenesis.

Functional selections

MarA libraries were transformed into *E. coli* strain N8453 (Δmar, Δsox-8::cat, Δrob::kan variant of GC4468) made by J.L. Rosner and R.G. Martin and obtained from M.B. Eisen. MarA expression was induced with 0.1% L-arabinose after one hour of recovery; cells were allowed to grow for an additional four hours before plating on 30 μg/ml tetracycline plates + 0.1% L-arabinose and incubated at 37°C for 48 hours.

λ-repressor functional assays were performed in X90 cells. Single colonies were streaked sequentially across lambdaphage strains KH54, KH54h80, and imm[21]c on 200 μg/ml ampicillin plates + 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG).

Protein purification

Proteins were overexpressed in Rosetta2 DE3 pLysS cells induced with 1 mM IPTG and purified using nickel column chromatography. For MarA purification, cells were lysed by sonication in 50 mM HEPES, pH 8, 500 mM NaCl, and 1 mM TCEP. MarA was expressed in inclusion bodies, which were isolated by centrifugation and resuspended in lysis buffer supplemented with 8 M urea and 10 mM imidazole. Imidazole

concentration was increased to 350 mM to elute, and protein was dialyzed into storage buffer (50% glycerol, 50 mM HEPES, pH 8, 500 mM NaCl, 1 mM TCEP). For λ-repressor purification, cells were lysed by sonication in 50 mM sodium phosphate, pH 8, 500 mM NaCl, 20 mM imidazole. λ-repressor was found entirely in supernatant; to elute, imidazole concentration was increased to 250 mM.

Binding assays

Serial dilutions of MarA variants were incubated for 30 minutes at room temperature with 0.5 nM oligo in 50 mM Tris, pH 8.5, 50 mM NaCl, and 10% glycerol. Experiments using dot-blot apparatus were performed as described in Chakravarthy *et al.*[34] with 5' ^{32}P-radiolabeled double-stranded consensus oligo (5' – ATT CGA TTT AGC AAA ACG TGC CAT CGG T – 3')[16]. Electrophoretic mobility shift assays (EMSA) used 3' FAM-labeled consensus, micF (5' – ATT CGA CAG CAC TGA ATG TCA AAA CCG GT – 3'), or zwf (5' – ATT CGA TCG CAC GGG TGG ATA AGC GCG GT – 3') MarA-binding sites[23]; after equilibration, samples were loaded onto 5% PAGE gel (29:1 acrylamide:bis-acrylamide and 10% glycerol in TAE) and run at approximately 80V at 4°C. Dissociation constants were determined from fractions folded at 0.2, 0.5, 2, and 5 μM, averaged over 2-3 experiments at each concentration. The lowest consistently detected fraction folded was used to calculate the threshold of detectable K_ds.

Procedure for competition assays was identical to EMSA protocol for 5 μM MarA protein equilibrated with 2.5 μM 6-FAM-labeled oligo and 2.5 μM 3' TYE 665-labeled oligo. Quantitation of each gel allowed the calculation of relative affinities, since

$$\frac{K_{d(FAM)}}{K_{d(TYE)}} = \frac{[DNA - FAM_{free}] \times [DNA - TYE_{bound}]}{[DNA - FAM_{bound}] \times [DNA - TYE_{free}]}.$$

All binding assays were done in triplicate and then averaged.

Circular dichroism

N-terminal domain of λ-repressor was dialyzed into 50 mM sodium phosphate, pH 7, 50 mM NaCl, and 0.2 mM TCEP. Circular dichroism studies were carried out at 25 °C on an Aviv 410 spectrapolarimeter with Peltier temperature control using a Microlab titrator. ~30 μg/ml protein in ~10 M urea was titrated into 0 M urea sample of identical protein concentration, maintaining 2.5 ml total volume. After each titration step, sample was equilibrated for 5 minutes and CD signal was monitored at 222, averaged over a 60-second time period. The change in Gibbs free energy upon unfolding (ΔG) was obtained by plotting the CD signal as a function of urea concentration and fitting the data using a two-state linear extrapolation model to obtain ΔG in the absence of urea[35].

Growth curves

MarA mutants were transformed into N8453 cells as described for the library selections. 5 ml 30 μg/ml tetracycline LB + 0.1% L-arabinose were inoculated with resulting colonies and allowed to grow at 37°C for approximately 20 hours. For each growth curve, 200 ml 30 μg/ml tetracycline LB + 0.1% L-arabinose was inoculated with 2 ml of overnight cultures and incubated at 37°C with agitation. Cell growth was monitored by measuring optical density (OD) at 600 nm, with samples taken every half hour during

log phase. For OD_{600} between 0.12 and 2, the $\log_{10}$ of the OD was plotted against the time and fit to a line; the slope of the line corresponds to the growth rate in tennings/hour.

2.6 Acknowledgements

We thank Ryan Shultzaberger for the MarA plasmids and strains, Jim Hu and Adrienne Zweifel for lambda phage strains and protocols, Sam Sternberg and David Steakley for technical assistance, Mike Botchan for comments on the manuscript, and the entire Marqusee lab for helpful discussions. Molecular graphics images were produced using the UCSF Chimera package from the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIH P41 RR001081). This work was supported by a grant from the NIH and NSF.

No competing financial interests.

2.7 References

1. Clackson, T. & Wells, J. A. A hot spot of binding energy in a hormone-receptor interface. *Science* **267,** 383–386 (1995).
2. Axe, D. D., Foster, N. W. & Fersht, A. R. Active barnase variants with completely random hydrophobic cores. *Proc. Natl. Acad. Sci. U.S.A.* **93,** 5590–5594 (1996).
3. Lim, W. A. & Sauer, R. T. Alternative packing arrangements in the hydrophobic core of lambda repressor. *Nature* **339,** 31–36 (1989).
4. Sia, S. K. & Kim, P. S. Protein grafting of an HIV-1-inhibiting epitope. *Proc. Natl. Acad. Sci. U.S.A.* **100,** 9756–9761 (2003).
5. Jiang, L. *et al.* De novo computational design of retro-aldol enzymes. *Science* **319,** 1387–1391 (2008).
6. Kritzer, J. A. *et al.* Miniature protein inhibitors of the p53-hDM2 interaction. *Chembiochem* **7,** 29–31 (2006).
7. Anantharaman, V., Aravind, L. & Koonin, E. V. Emergence of diverse biochemical activities in evolutionarily conserved structural scaffolds of proteins. *Curr Opin Chem Biol* **7,** 12–20 (2003).
8. Iyer, L. M., Zhang, D., Rogozin, I. B. & Aravind, L. Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. *Nucleic Acids Res.* **39,** 9473–9497 (2011).
9. Thorn, K. S. & Bogan, A. A. ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. *Bioinformatics* **17,** 284–285 (2001).
10. Levin, A. M. *et al.* Exploiting a natural conformational switch to engineer an interleukin-2 'superkine'. *Nature* **484,** 529–533 (2012).
11. Ptashne, M. *et al.* How the lambda repressor and cro work. *Cell* **19,** 1–11 (1980).
12. Marqusee, S. & Sauer, R. T. Contributions of a hydrogen bond/salt bridge network to the stability of secondary and tertiary structure in lambda repressor. *Protein Sci.* **3,** 2217–2225 (1994).
13. Rhee, S., Martin, R. G., Rosner, J. L. & Davies, D. R. A novel DNA-binding motif in MarA: the first structure for an AraC family transcriptional activator. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 10413–10418 (1998).
14. Dangi, B. *et al.* Structure and dynamics of MarA-DNA complexes: an NMR investigation. *Journal of Molecular Biology* **314,** 113–127 (2001).
15. Dangi, B., Gronenborn, A. M., Rosner, J. L. & Martin, R. G. Versatility of the carboxy-terminal domain of the α subunit of RNA polymerase in transcriptional activation: use of the DNA contact site as a protein contact site for MarA. *Molecular Microbiology* **54,** 45–59 (2004).
16. Shultzaberger, R. K., Maerkl, S. J., Kirsch, J. F. & Eisen, M. B. Probing the informational and regulatory plasticity of a transcription factor DNA-binding domain. *PLoS Genet.* **8,** e1002614 (2012).
17. Barbosa, T. M. & Levy, S. B. Differential expression of over 60 chromosomal genes in Escherichia coli by constitutive expression of MarA. *J. Bacteriol.* **182,** 3467–3474 (2000).

18. Martin, R. G., Gillette, W. K., Rhee, S. & Rosner, J. L. Structural requirements for marbox function in transcriptional activation of mar/sox/rob regulon promoters in Escherichia coli: sequence, orientation and spatial relationship to the core promoter. *Molecular Microbiology* **34,** 431–441 (1999).
19. Lim, W. A. & Sauer, R. T. The role of internal packing interactions in determining the structure and stability of a protein. *Journal of Molecular Biology* **219,** 359–376 (1991).
20. Kuriyan, J., Konforti, B. & Wemmer, D. The molecules of life: KURIYAN John, KONFORTI Boyana, WEMMER David: Librairie Lavoisier. (2013).
21. Hilser, V. J., Dowdy, D., Oas, T. G. & Freire, E. The structural distribution of cooperative interactions in proteins: analysis of the native state ensemble. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 9903–9908 (1998).
22. Liu, T., Whitten, S. T. & Hilser, V. J. Functional residues serve a dominant role in mediating the cooperativity of the protein ensemble. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 4347–4352 (2007).
23. Gillette, W. K., Martin, R. G. & Rosner, J. L. Probing the Escherichia coli transcriptional activator MarA using alanine-scanning mutagenesis: residues important for DNA binding and activation. *Journal of Molecular Biology* **299,** 1245–1255 (2000).
24. Haririnia, A. *et al.* Mutations in the hydrophobic core of ubiquitin differentially affect its recognition by receptor proteins. *Journal of Molecular Biology* **375,** 979–996 (2008).
25. Uversky, V. N. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* **11,** 739–756 (2002).
26. Babu, M. M., Kriwacki, R. W. & Pappu, R. V. Structural biology. Versatility from protein disorder. *Science* **337,** 1460–1461 (2012).
27. Tokuriki, N. & Tawfik, D. S. Protein dynamism and evolvability. *Science* **324,** 203–207 (2009).
28. Meier, S. & Ozbek, S. A biological cosmos of parallel universes: does protein structural plasticity facilitate evolution? *Bioessays* **29,** 1095–1104 (2007).
29. Ancel, L. W. & Fontana, W. Plasticity, evolvability, and modularity in RNA. *J. Exp. Zool.* **288,** 242–283 (2000).
30. Jarosz, D. F. & Lindquist, S. Hsp90 and environmental stress transform the adaptive value of natural genetic variation. *Science* **330,** 1820–1824 (2010).
31. MacQuarrie, K. L., Fong, A. P., Morse, R. H. & Tapscott, S. J. Genome-wide transcription factor binding: beyond direct target regulation. *Trends Genet.* **27,** 141–148 (2011).
32. Liu, J. *et al.* Intrinsic disorder in transcription factors. *Biochemistry* **45,** 6873–6888 (2006).
33. Minezaki, Y., Homma, K., Kinjo, A. R. & Nishikawa, K. Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation. *Journal of Molecular Biology* **359,** 1137–1149 (2006).
34. Chakravarthy, S., Sternberg, S. H., Kellenberger, C. A. & Doudna, J. A. Substrate-specific kinetics of Dicer-catalyzed RNA processing. *Journal of Molecular Biology* **404,** 392–402 (2010).

35. Greene, R. F. & Pace, C. N. Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin, and beta-lactoglobulin. *J. Biol. Chem.* **249,** 5388–5393 (1974).
36. Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25,** 1605–1612 (2004).

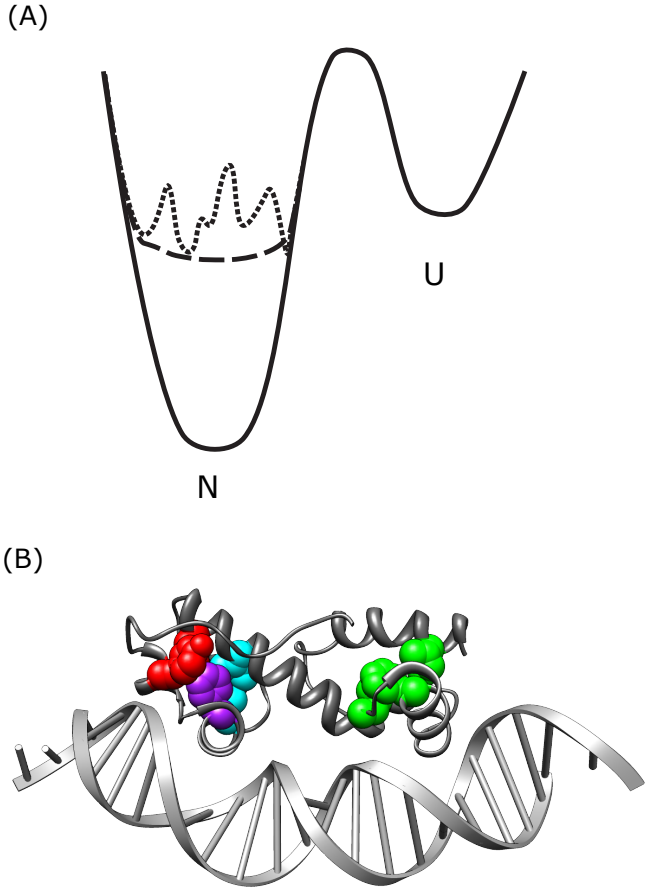
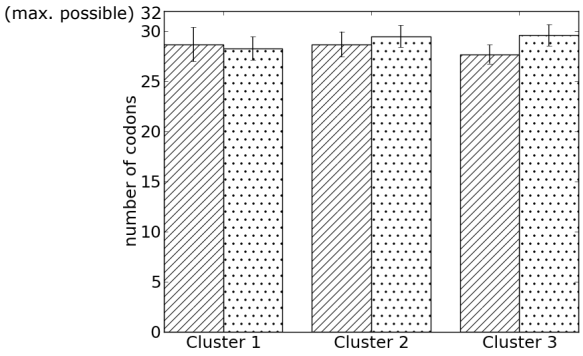


Figure 2.1. A) Two-dimensional representations of idealized energy landscapes for a well-folded protein with a single native ground state (solid line), a marginally stable protein with a defined native state (dashed line), and a marginally stable protein with multiple minima (dotted line). B) Representation of the crystal structure of MarA binding to DNA. Cluster 1 residues I13 (76%), F48 (96%), and L56 (59%) are indicated in green; cluster 2 residues I68 (20%) and L94 (20%), in cyan; cluster 3 residues L72 (99%) and Y109 (75%), in red. F98 (98%), which is included in both clusters 2 and 3, is indicated in purple[13,35]. Numbers in parentheses after residue number indicate degree of conservation among homologues.

(A)



(B)

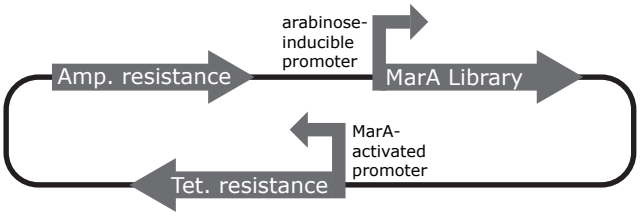


Figure 2.2. A) Libraries do not appear biased. Dotted bars indicate results of simulated sequencing of 86 (Cluster 1), 93 (Cluster 2), or 94 (Cluster 3) individual clones, repeated 1000 times, with error bars indicating one standard deviation. Striated bars show results of actual sequencing, with the results of three randomized positions averaged for each cluster and error bars indicating one standard deviation. B) The MarA selection system contains MarA mutants under control of an arabinose-inducible promoter and the MarA consensus binding sequence as the promoter for tetracycline resistance.

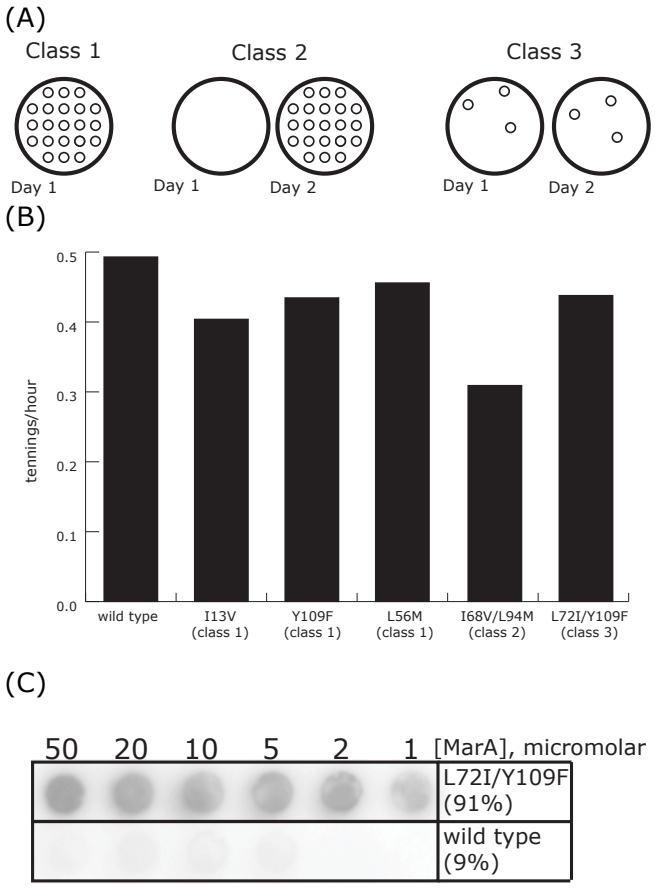


Figure 2.3. Characterization of functional MarA variants. A) Class 1 variants exhibit wild type-like growth on plates following a 24-hour incubation when transformed into E. coli. Class 2 variants also result in numerous colonies but require a 48-hour incubation. Class 3 variants exhibit the least robust growth, resulting in very few colonies regardless of incubation time. B) Mean log-phase growth rates of E. coli transformed with the variant indicated. Standard errors for wild type, Y109F, and I68V/L94M replicates were 0.033, 0.051, and 0.011, respectively. Standard error among all growth rates was 0.025. C) Increasing concentrations of wild type or L72I/Y109F (class 3) MarA were mixed with radioactively labeled target DNA, equilibrated for 30 minutes, and passed through Tuffryn size exclusion membrane. Percentages indicate amount of total signal present on the Tuffryn membrane.

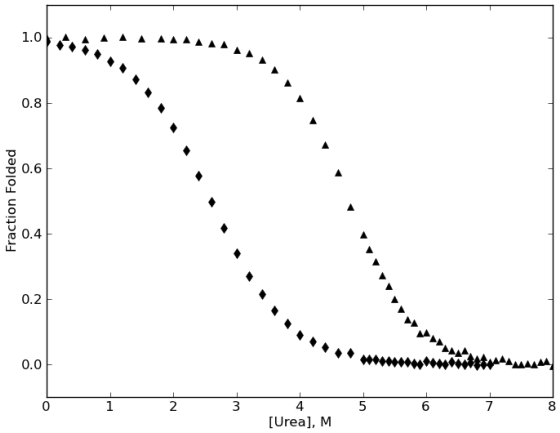
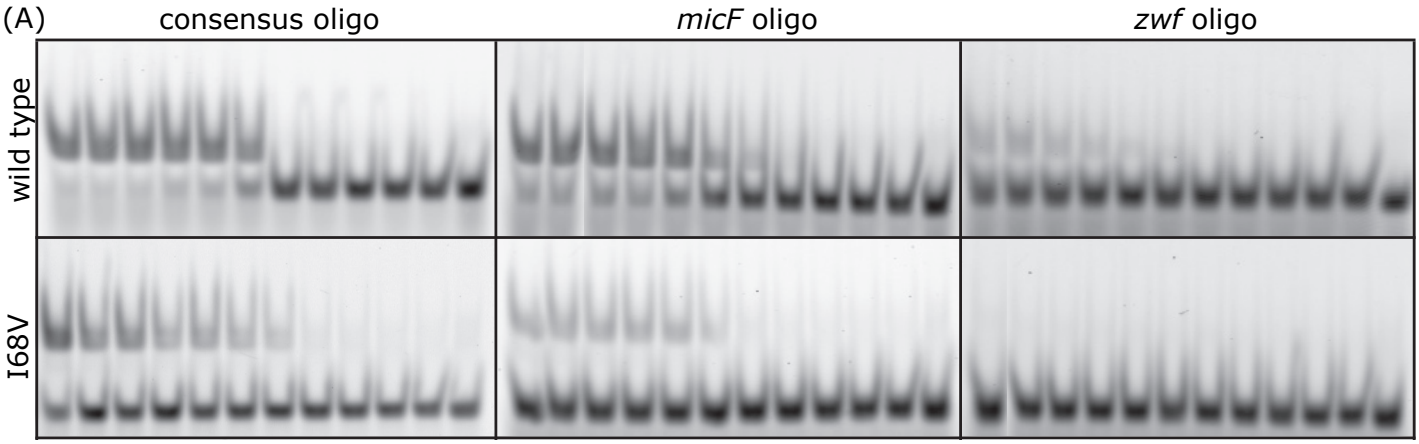


Figure 2.4. Urea denaturation of wild type (triangles) and R17A/S77A (diamonds) λ repressor, as measured by CD [25°C, 0.2 μM tris(2-carboxyethyl)phosphine (TCEP), 50 mM sodium chloride, 50 mM sodium phosphate, pH 7].

(A)



(B)

	consensus	*micF*	*zwf*
wild type	0.91 ± .2 μM	1.5 ± .3 μM	16.1 ± 5 μM
I68V	2.9 ± .7 μM	3.1 ± 1 μM	>95 mM

Figure 2.5. MarA variants binding to different promoters. A) For each gel, the concentration of protein in each lane, from left to right: 20 μM, 10 μM, 5 μM, 2 μM, 1 μM, 500 nM, 200 nM, 100 nM, 50 nM, 20 nM, 10 nM, none. Top row shows wild type MarA binding to consensus (left), micF (middle), and zwf (right). Bottom row shows I68V binding to consensus (left), micF (middle), and zwf (right). B) Dissociation constants for binding of the MarA variant indicated in the leftmost column to the oligos indicated in the top row.

(A)

FAM-labeled oligo	TYE-labeled oligo	$K_{d(FAM)}/K_{d(TYE)}$	
		wild type	I68V
zwf	*zwf*	3.3 ± 0.7	*
micF	*micF*	1.1 ± 0.4	0.5 ± 0.1
zwf	*micF*	15.5 ± 5.9	*
* no detectable *zwf* binding			

(B)

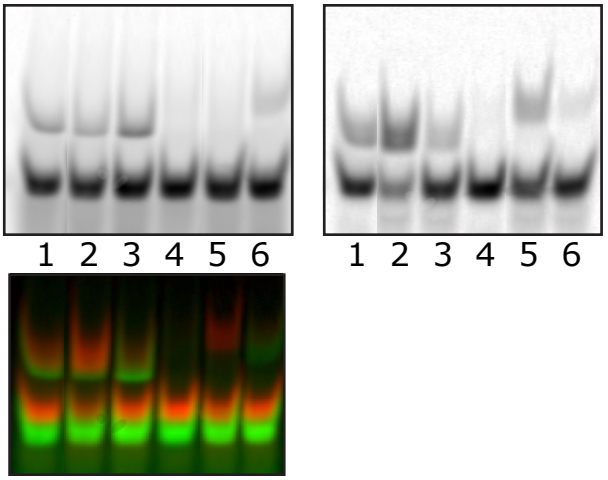


Figure 2.6. MarA variants binding to different promoters under competition conditions, with 2.5 μM protein and 2.5 μM each of the specified oligos. A) Relative binding affinities of the differently labeled oligos to either wild type or I68V MarA protein; the error reported is one standard deviation. B) 6-FAM channel, TYE 665 channel, and overlay of a representative gel. Lanes 1–3 are with wild type; lanes 4–6, with I68V. Lanes 1 and 4 contain FAM- and TYE-labeled zwf, lanes 2 and 5 contain FAM-labeled micF and TYE-labeled zwf, and lanes 3 and 6 contain FAM- and TYE-labeled micF.

Table 2.1. MarA core variants surviving functional selection.

13	48	56	class
I	F	L	wt
V	F	L	1
A	F	V	2
V	F	E	3

72	98	109	class
L	F	Y	wt
L	F	F	1
L	L	F	2
I	F	F	3
P	I	I	3

68	94	98	class
I	L	F	wt
I	F	F	1
L	L	F	1
V	F	F	1
A	F	F	1
C	F	F	1
F	L	F	1
M	F	F	1
V	L	F	1
L	F	F	2
V	M	F	2

Classes are described in Figure 3A. 'wt' denotes wild type sequence.

Table 2.2. λ repressor variant activity monitored *in vivo* by immunity to lambdaphage infection.

mutation(s):	**background:**	
	wild type	**R17A/S77A**
V36I	+	+
V36I/M40F	+	-
V36I/M40I/V47I	+	-
+ indicates resistance to KH54 and KH54h80 lambdaphage strains; - indicates sensitivity to both. All were sensitive to $imm^{21}c$; none were resistant to KH54 and not KH54h80.		

Chapter 3: WHAT MAKES A THERMOSENSOR?: THE ROLE OF ANKYRIN REPEATS IN TRPA1 HEAT RESPONSE

3.1 Abstract

The ability of rattlesnakes to detect infrared radiation (IR) is dependent on the heat-activated transient receptor potential ankyrin 1 (TRPA1) channel. In contrast, the human homologue shows no activation by heat; this difference has been linked to a subset of ankyrin repeats within the protein's cytoplasmic N-terminal domain. To investigate this, we have studied the stability and folding of isolated ankyrin repeats 3-8 from the human and rattlesnake TRPA1 proteins. Both fragments fold in isolation, and we find that the fragment isolated from the rattlesnake but not human protein gains stability with increasing temperature up to at least 70˚C. This appears to arise from an unusual dependence of stability on temperature; when fit to the Gibbs-Helmholtz equations, the rattlesnake fragment appears to have a maximal stability near 75˚C and a low ΔC_p. We propose that this remarkable response to temperature may be important for temperature detection when the fragment is in the context of the TRPA1 protein.

3.2 Introduction

How do organisms sense and respond to temperature? Although many details of the molecular mechanism still remain to be discovered, we know that in many cases the first step is the direct activation of protein channels within the Transient Receptor Potential (TRP) family by temperature[1,2]. This family has over 80 members in both vertebrate and invertebrate animals, with distant relatives in multiple fungi[3]. The channels respond not only to thermal but also mechanical and chemical signals within the environment. They assemble as tetramers; each monomer has six transmembrane regions, the pore located between the fifth and sixth[4] (Fig. 3.1A).

In this work, I will focus on TRPA1, the only mammalian member of the TRP Ankyrin (TRPA) protein channel subfamily. In humans, this protein has been shown to be involved in a wide variety of signaling pathways, including itch[5], pain[6], and inflammation[7]. It has also been linked to cold sensitivity, both as a channel directly activated by noxious cold[8] and as a key player in cold hyperalgesia[9]. Its role in cold sensitivity, however, is unique to mammalian TRPA1. Drosophila, for example, has three TRPA channels, all activated by temperatures between 25-45˚C[10-12]. Snakes similarly possess heat-sensitive TRPA1 homologues, which ancient snakes have exploited to detect infrared radiation[13,14]. How TRPA1 homologues are able to specifically respond to such a wide range of agonists, including a variety of distinct temperature ranges, remains an active area of investigation.

One insight into this question came from a 2011 study that pointed to regions of ankyrin repeats within the rattlesnake TRPA1 cytoplasmic N-terminal domain as sufficient to confer heat sensitivity to the human homologue[15]. Ankyrin repeats (ARs) are amino acid motifs approximately 33 residues in length, formed by a helix-turn-helix with a longer loop linking to adjacent repeats[16]. TRPA1 has a large N-terminal cytoplasmic domain comprising 16 ankyrin repeats that, based on electron microscopy, form a basket-like structure below the pore[17]. Cordero-Morales *et al.* found that chimeras of human TRPA1 with the rattlesnake sequence for ankyrin repeats 3-8 or 10-15 largely recapitulated the heat response of rattlesnake TRPA1[15].

We hypothesized that TRPA1 may detect heat via temperature-dependent conformational changes within the ankyrin repeats. To investigate this, we expressed and

purified the human and rattlesnake variants of one of the thermal response modules identified in the Cordero-Morales study[15]. The fragments comprising rattlesnake repeats 3-8 exhibits remarkable thermotolerance, remaining folded up to temperatures of at least 95˚C. Preliminary results suggest that this arises from a right-shifted and broadened stability curve and may have important consequences for rattlesnake TRPA1 IR detection.

3.3 Results

Expression and Purification

In order to investigate temperature-dependent conformational changes in TRPA1 ankyrin repeats using CD spectroscopy, we analyzed the isolated thermal response modules comprising rattlesnake or human repeats 3-8[15]. For solubility, AR fragments were capped with N- and C-terminal ankyrin repeats with an increased number of polar and charged residues[18]; the sequence of these capping repeats was identical for all fragments assayed (Fig. 3.1B). In addition, a 6xHis tag was added for purification. Each construct was expressed in *E. coli* cells; the resulting recombinant proteins were affinity purified from inclusion bodies using a nickel column and refolded. From hereon out, human ankyrin repeats 3-8 with solubility caps will be referred to as h3c; the rattlesnake, as s3c. Sequences of constructs are provided in the Materials & Methods.

Fragment Structure and Stability

Based on their CD spectra, both h3c and s3c are folded (Fig. 3.2). We expected the spectra to reflect primarily α-helix, with two minima of comparable magnitudes at 208 nm and 222 nm[19], as has been seen for other ankyrin repeat proteins[20,21]. Neither has this classic spectrum; rather than having their major minimum at 222 nm, both fragments have their major minima near 218 nm and another at 208 nm. In addition to differences in spectral features, the mean residue ellipticity (MRE) for h3c is less negative than anticipated and than has been seen for other ankyrin repeats[20,21]. While we assume that these fragments fold into ankyrin-like repeats, these spectra raise the possibility that they may be folding into alternative conformations. Fragments will undergo higher resolution structural studies to definitely confirm whether they are adopting the expected fold.

To investigate the stability of these fragments, we carried out chemical induced denaturation experiments. Both fragments were studied by urea denaturation and the denaturation curves were fit with the standard two-state linear extrapolation method[22]. h3c had $\Delta G_{unfolding}$ = 2.7 $\pm$ 0.3 kcal/mol and an m-value of 0.7 $\pm$ 0.1 kcal mol^{-1} M^{-1} (Fig. 3.3A). For h3c, we observed an unusual increase in signal at low denaturant concentrations, prior to the observed denaturation transition, points with less than 0.5 M urea were not considered in the fit. We hypothesize that this irregularity is due to the presence of soluble aggregates in h3c solution. s3c showed $\Delta G_{unfolding}$= 6.0 $\pm$ 0.3 kcal/mol and m-value of 1.31 $\pm$ 0.1 kcal/mol (Fig. 3.3B).

For s3c, we also monitored stability by guanidinium chloride in order to determine stability curves (see below). The rattlesnake construct had $\Delta G_{unfolding}$ = 5.6 $\pm$ 0.2 kcal/mol and m-value of 2.4 $\pm$ 0.1 kcal mol^{-1} M^{-1} (Fig. 3.3C). In addition, to determine whether s3c structure formation is monomeric, CD spectra were measured at 50, 250, and 500 μg/ml (Fig. 3.3D). The spectra change slightly with concentration but do not seem to indicate large-scale multimerization or aggregation.

Temperature Dependence of Thermal Response Module Unfolding

To determine if and what conformational changes might occur at the temperatures that the ankyrin repeats are charged with detecting, we followed thermal denaturation by monitoring the CD signal of the capped fragments at 222 nm (Fig. 3.4A,B). h3c showed the expected cooperative temperature-dependent unfolding. After incubation at high temperature, however, we observed visible precipitate, with a largely irreversible loss of CD signal after cooling the sample back to 25˚C. s3c, however, showed no temperature-dependent unfolding. No change in signal was observed over the entire range (4˚C to 95˚C). Spectra were taken at different temperatures to make sure that this unusual behavior is not a reflection of an unusual isosbestic point at 222nm. The spectra after a five-minute equilibration at 25˚C, 45˚C, 65˚C, and 85˚C are coincident; moreover, no change in this spectrum was seen after a twelve-hour incubation at 85˚C (Fig. 3.4C). Thus, s3c appears to adopt a folded conformation that is resistant to prolonged exposure to high temperatures, implying a melting temperature above the measured range.

In order to investigate the temperature-dependence of s3c stability, we carried out GdnCl-denaturation studies of s3c at three temperatures: 15˚C, 25˚C, and 45˚C. At all three temperatures, the m-value remains constant at 2.4 $\pm$ 0.1 kcal/mol. The $\Delta G_{unfolding}$ = 5.0 kcal/mol at 15˚C, 5.6 $\pm$ 0.2 kcal/mol at 25˚C, and 7.3 $\pm$ 0.5 kcal/mol at 45˚C (Fig. 3.3C, 3.5A). Protein stability should depend on temperature according to the Gibbs-Helmholtz equation (see Materials & Methods). With only three data points, we modeled our data with a range of ΔC_p values in order to extrapolate a potential T_m for the protein (Fig. 3.5B). ΔC_p values were estimated from expectations based on size and m-value. While typically, both ΔC_p and m-values correlate with size[23], our m-value is lower than expected for a 300-residue protein; thus, we estimated a value for ΔC_p based on size and another based solely on the m-value. 300-residue s3c would be predicted to have $\Delta C_{p,unfolding}$ ~ 4 kcal mol^{-1} K^{-1}, a value consistent with others that have been seen for ankyrin repeat proteins of comparable size[24] but which gives a poor fit of the Gibbs-Helmholtz to the data. Based on its m-value, s3c would be predicted to have $\Delta C_{p,unfolding}$ ~ 1.1 4 kcal mol^{-1} K^{-1} [23]. Even this appears too high; in order to fit these data, $\Delta C_{p,unfolding}$ must be close to 0.5 4 kcal mol^{-1} K^{-1}. The resulting fit suggests that s3c is maximally stable near 75˚C, with an extrapolated melting temperature (T_m) of 173˚C.

The modeling suggests that the s3c fragment has a very unusual protein stability curve. To further test this fit, we tested a hypothesis suggested by this modeling: the s3c fragment should increase in stability from 25˚C to 85˚C. Consistent with this, the s3c CD spectrum in 3.6 M guanidinium chloride appears more folded as the temperature is increased from 25˚C to 85˚C (Fig. 3.5C).

Kinetics of thermal response module unfolding

This thermodynamic analysis assumed that samples were completely equilibrated to reflect true equilibrium conditions. To test that the folding and unfolding of our fragments were sufficiently fast to justify this assumption, we examined h3c and s3c (un)folding kinetics. These were determined by monitoring CD signal at 222 nm after rapid dilution (using either manual or stop-flow mixing) of a concentrated unfolded or folded protein stock into buffer favoring folding or unfolding, respectively. h3c unfolding

is completed within the 20 second dead time of manual mixing (Fig. 3.6A). s3c folding is completed within the 15 millisecond dead time of stop-flow mixing (Fig. 3.6B). s3c unfolding fits to two kinetic phases, one with a relaxation time of 200 seconds and the second with a relaxation time of 1800 s (Fig. 3.6C-E).

CHIMERAS

To determine how this unusual thermal resistance is encoded in the s3c sequence, we are currently making chimeras of s- and h3c and evaluating their temperature-dependent unfolding. This work is ongoing, with constructs and progress towards their analysis detailed in Table 3.1. None of the s3c-based chimeras assayed to date unfold with typical temperature dependence. Additionally, we are evaluating a range of N-terminal domain fragments to determine how the temperature response of rattlesnake repeats 3-8 is altered by the presence of additional regions of the TRPA1 protein.

3.4 Discussion

The ability of rattlesnake TRPA1 to respond to heat, a function absent in mammalian homologues, has been linked to ankyrin-repeat modules within its cytoplasmic N-terminal domain human[15]. In an attempt to determine a molecular mechanism for this response, we have studied the human and rattlesnake variants of one of these modules, repeats 3-8, in isolation. While the conformation of human repeats 3-8 shows clear temperature dependence, the analogous rattlesnake protein fragment appears resistant to thermal denaturation and shows an unusual gain of stability with increasing temperature. The data presented here suggest that its maximum stability is near 75˚C and its melting temperature near 170˚C, unexpected for a fragment of a protein from a cold-blooded mesophile. It's interesting to note that the equilibrium constant moves in the direction favoring folding when going from ambient temperatures to activating temperatures for rattlesnake TRPA1.

The Gibbs-Helmholtz equation suggests that protein stability has a parabolic relationship to temperature, with a stability maximum and two T_ms. To gain thermostability (i.e. increase their T_ms), proteins can up-shift, right-shift, or broaden their stability curves[25]. The s3c stability curve appears both right shifted and broadened relative to a typical mesophilic protein. Of-course, these data will need to be confirmed by measuring a full protein stability curve. To that end, more data reporting $\Delta G_{unfolding}$ over a range of temperatures will be collected. These data can then be reliably fit by the Gibbs-Helmholtz equation to determine ΔC_p, T_m, and the temperature of maximum s3c stability.

The breadth of the curve is a direct consequence of the ΔC_p. Our model indicates that a lower-than-expected ΔC_p is associated with s3c unfolding. Previous studies have identified a similar phenomenon for some homologues of RNase H; these have correlated the low ΔC_p value to residual structure in the unfolded state[26]. This may be what is occurring with the ankyrin repeats, although mutational analysis or other strategies to confirm this interpretation remain to be done. While much of this analysis assumes the fragment adopts a native-like, ankyrin fold, the observed ΔC_p is lower than that expected based on size alone and is therefore not dependent on this assumption. However, x-ray crystallography or NMR is required to confirm the fold of this fragment.

There are also many remaining questions about the mechanism by which s3c is achieving thermostability. To further understand the molecular mechanism of this high T_m, we are making chimeras of s3c with regions encoded by the human sequence. These will shed light as to whether there is a particular region of s3c responsible for its unique response to temperature or it is an emergent property diffused over the sequence as a whole.

Our hypothesis is that this unique temperature response plays a role in TRPA1 IR detection. We went into this looking for an unusual temperature-dependent conformational change in the s3c fragment, expecting an order to disorder transition. In this scenario, induced disorder would correlate with function. What we found, however, was an unusual change in the equilibrium constant from less to more folded over the relevant temperature range. This suggests that, contrary to our expectation, order would correlate with function. Moreover, in the context of the whole protein, it may well go from a predominantly unfolded to predominantly folded conformation.

The kinetics observed for s3c support this model. The unusually long unfolding kinetics after rapid unfolding would allow the ankyrin repeats to remain in the folded conformation long enough for large-scale conformational changes that lead to pore opening. In this scenario, once folding of the ankyrin repeats is induced, the channel would be activated even if the activating signal were discontinued.

To begin to address how robust our results are within the context of the entire protein, we have begun making constructs that include additional ankyrin repeats from the rattlesnake N-terminal domain. If s3c is destabilized or made more sensitive to temperature when in the presence of additional regions of the N-terminal domain, that would support the model outlined above. Conversely, if s3c thermotolerance is propagated to adjacent repeats, a different hypothesis would be required to explain rattlesnake TRPA1 IR detection. While the original study[15] identified two thermal response modules - repeats 3-8 and 10-15 – repeats 3-8 were chosen for detailed analysis because they are the predicted to be more ordered[17] and, empirically, appear more amenable to conditions required for biophysical analysis. Evaluating constructs that comprise both modules will help us dissect whether both use the same mechanism to induce temperature-dependent TRPA1 activation.

3.5 Materials & Methods

Cloning

DNA encoding the capping ankyrin repeats, separated by NheI and SacI sites, was synthesized by IDT and subcloned into the pET28 vector. Desired ankyrin repeats were PCR-amplified from full length TRPA1 and inserted between the capping repeats using either standard sub- or Gibson[27] cloning techniques. The XL10-gold cell line was used for all cloning.

s3c DNA sequence:

```
ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGT
GCCGCGCGGCAGCCATATGAGCAAAGATGGTAATACCCCGCTGC
ATAATGCAGCAAAAAATGGTCATGCAGAAGAAGTTAAAAAACTG
CTGAGCAAAGGTGCAGATGTTAATGCACGTGAGCTCAATATGAT
```

TGCTCCACTTCACTGGGCTCTGCATTATCTTCTTGACGATTTAG
TTACGATTTTCCTTGAATGCAGCAATACGAATATAAATCTGGAA
GGAGAAGGTGGGAATACACCCATACTATTAGCATGTTATAAAGA
TAATCCTACCGCACTGAAAATTCTGATTGAAAAAGGCGGTGACA
TTTGTAAAGTGAACAATATGGGCTGTATGCCAGTCCATGCAGCT
GCCTTTTCTGGTTCGAAATTATGTTTGGAGATGATTATAAAGCG
AGGTGAACAACTTGGGTATTCACCTAAAAATCATATTAATTTTA
TTAATAATGAGAAAAGTAGTCCTCTTCATCTAGCTGTCCAAAGC
AGAGATGTGGAAATGATTAAAATGTGCATTGAATATGGAGCACA
GATTGACCTGAAACAGAGTGACAACTGCACAGCTCTTCATATTG
CTGCTATCCAAGGAGCTACTGAGATTATCGAGTTACTGATGTCA
GCTTATTCAGGGGAGGAGTGTCTCATTAATGCATCAGATGAAAA
TAAGGAAACATTGCTCCATAGGGCTGCATTGTTTGATCACGATG
AAATGACAGATTATCTCATTTCAAAGGGAGCAAATATTGATAGT
GTTGACATTGCTAGCAAAGATGGTAATACCCCGGAACATCTGGC
AAAAAAAAATGGTCATCATGAAATTGTTAAACTGCTGGATGCAA
AAGGTGCAGATGTTAATGCACGTTAATAA

h3c DNA sequence:

ATGGGCAGCAGCCATCATCATCATCACAGCAGCGGCCTGGT
GCCGCGCGGCAGCCATATGAGCAAAGATGGTAATACCCCGCTGC
ATAATGCAGCAAAAAATGGTCATGCAGAAGAAGTTAAAAAACTG
CTGAGCAAAGGTGCAGATGTTAATGCACGTGCTAGCAACATGAT
GGCTCCTCTCCACATAGCTGTGCAGGGCATGAATAATGAGGTGA
TGAAGGTCTTGCTTGAGCATAGAACTATTGATGTTAATTTGGAA
GGAGAAAATGGAAACACAGCTGTGATCATTGCGTGCACCACAAA
TAATAGCGAAGCATTGCAGATTTTGCTTAAAAAAGGAGCTAAGC
CATGTAAATCAAATAAATGGGGATGTTTCCCTATTCACCAAGCT
GCATTTTCAGGTTCCAAAGAATGCATGGAAATAATACTAAGGTT
TGGTGAAGAGCATGGGTACAGTAGACAGTTGCACATTAACTTTA
TGAATAATGGGAAAGCCACCCCTCTCCACCTGGCTGTGCAAAAT
GGTGACTTGGAAATGATCAAAATGTGCCTGGACAATGGTGCACA
AATAGACCCAGTGGAGAAGGGAAGGTGCACAGCCATTCATTTTG
CTGCCACCCAGGGAGCCACTGAGATTGTTAAACTGATGATATCG
TCCTATTCTGGTAGCGTGGATATTGTTAACACAACCGATGGATG
TCATGAGACCATGCTTCACAGAGCTTCATTGTTTGATCACCATG
AGCTAGCAGACTATTTAATTTCAGTGGGAGCAGATATTAATAAG
ATCGATTCTGCTAGCAAAGATGGTAATACCCCGGAACATCTGGC
AAAAAAAAATGGTCATCATGAAATTGTTAAACTGCTGGATGCAA
AAGGTGCAGATGTTAATGCACGTTAATAA

PROTEIN PURIFICATION

Proteins were overexpressed in BL21 (DE3) pLysS cells induced with 1 mM IPTG. Cells were lysed by sonication in 50 mM potassium phosphate buffer pH 7, 50 mM NaCl, 0.5 mM TCEP. Fragments were expressed in inclusion bodies; these were isolated by centrifugation and resuspended in lysis buffer supplemented with 6 M urea and 20 mM imidazole. The protein was purified from the solubilized pellet using nickel column

chromatography and eluted from the column by increasing the imidazole concentration to 350 mM. The proteins were then refolded by dialysis into the lysis buffer specified above.

CIRCULAR DICHROISM

Stop-flow circular dichroism studies were carried out on an Aviv 222 spectrapolarimeter. All other circular dichroism studies were carried out on an Aviv 410 spectrapolarimeter with Peltier temperature control. Unless otherwise specified, denaturant and temperature melts were performed with 50 μg/ml protein in a 1 cm quartz cuvette, while CD spectra were collected for 250 μg/ml protein in a 0.2 cm quartz cuvette. Protein stability curves are described by the Gibbs-Helmholtz equation[28]:

$$\Delta G_{unf} = \Delta H^\circ - T\Delta S^\circ + \Delta C_p[T - T^\circ - T \ln(T/T^\circ)],$$

where T° is some reference temperature and ΔH° and ΔS° are, respectively, the enthalpy and entropy at that temperature. ΔC_p is the difference in heat capacity of unfolded and folded protein and is assumed to be constant over the relevant temperature range. By using the melting temperature T_m as the reference temperature, the above can be modified to

$$\Delta G_{unf} = \Delta H^\circ - T\Delta H^\circ/T_m + \Delta C_p[T - T_m - T \ln(T/T_m)]^{29}.$$

Data were qualitatively evaluated for fit to this modified Gibbs-Helmholtz with one of four fixed ΔC_p values.

3.6 Acknowledgements

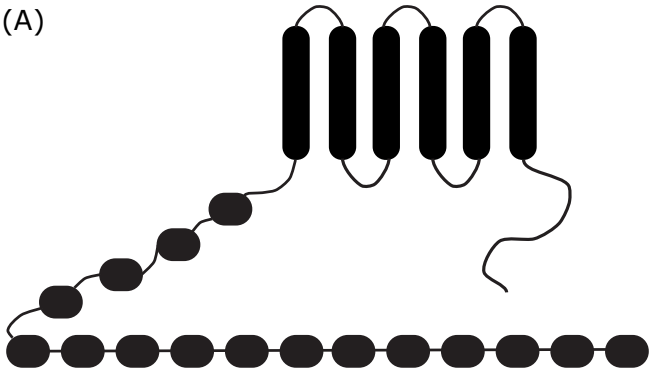
We thank Diana Bautista and David Julius for TRPA1-containing plasmids.

3.7 References

1. Montell, C. The TRP superfamily of cation channels. *Sci. STKE* **2005,** re3 (2005).
2. Caterina, M. J. *et al.* The capsaicin receptor: a heat-activated ion channel in the pain pathway. *Nature* **389,** 816–824 (1997).
3. Montell, C. Drosophila TRP channels. *Pflugers Arch.* **451,** 19–28 (2005).
4. Latorre, R., Zaelzer, C. & Brauchi, S. Structure-functional intimacies of transient receptor potential channels. *Q. Rev. Biophys.* **42,** 201–246 (2009).
5. Wilson, S. R. *et al.* The ion channel TRPA1 is required for chronic itch. *J. Neurosci.* **33,** 9283–9294 (2013).
6. Andersson, D. A. *et al.* Methylglyoxal evokes pain by stimulating TRPA1. *PLoS ONE* **8,** e77986 (2013).
7. Bautista, D. M., Pellegrino, M. & Tsunozaki, M. TRPA1: A gatekeeper for inflammation. *Annu. Rev. Physiol.* **75,** 181–200 (2013).
8. Story, G. M. *et al.* ANKTM1, a TRP-like channel expressed in nociceptive neurons, is activated by cold temperatures. *Cell* **112,** 819–829 (2003).
9. Obata, K. *et al.* TRPA1 induced in sensory neurons contributes to cold hyperalgesia after inflammation and nerve injury. *J. Clin. Invest.* **115,** 2393–2401 (2005).
10. Viswanath, V. *et al.* Opposite thermosensor in fruitfly and mouse. *Nature* **423,** 822–823 (2003).
11. Lee, Y. *et al.* Pyrexia is a new thermal transient receptor potential channel endowing tolerance to high temperatures in Drosophila melanogaster. *Nat. Genet.* **37,** 305–310 (2005).
12. Tracey, W. D., Wilson, R. I., Laurent, G. & Benzer, S. painless, a Drosophila gene essential for nociception. *Cell* **113,** 261–273 (2003).
13. Ebert, J. Infrared sense in snakes – behavioural and anatomical examinations (. 1–108 (2008).
14. Gracheva, E. O. *et al.* Molecular basis of infrared detection by snakes. *Nature* **464,** 1006–1011 (2010).
15. Cordero-Morales, J. F., Gracheva, E. O. & Julius, D. Cytoplasmic ankyrin repeats of transient receptor potential A1 (TRPA1) dictate sensitivity to thermal and chemical stimuli. *Proc. Natl. Acad. Sci. U.S.A.* **108,** E1184–91 (2011).
16. Gaudet, R. A primer on ankyrin repeat function in TRP channels and beyond. *Mol Biosyst* **4,** 372–379 (2008).
17. Cvetkov, T. L., Huynh, K. W., Cohen, M. R. & Moiseenkova-Bell, V. Y. Molecular architecture and subunit organization of TRPA1 ion channel revealed by electron microscopy. *J. Biol. Chem.* **286,** 38168–38176 (2011).
18. Aksel, T., Majumdar, A. & Barrick, D. The contribution of entropy, enthalpy, and hydrophobic desolvation to cooperativity in repeat-protein folding. *Structure* **19,** 349–360 (2011).
19. HOLZWARTH, G. & Doty, P. THE ULTRAVIOLET CIRCULAR DICHROISM OF POLYPEPTIDES. *J. Am. Chem. Soc.* **87,** 218–228 (1965).
20. Tripp, K. W. & Barrick, D. The tolerance of a modular protein to duplication and deletion of internal repeats. *Journal of Molecular Biology* **344,** 169–178 (2004).

21. Michaely, P. & Bennett, V. The membrane-binding domain of ankyrin contains four independently folded subdomains, each comprised of six ankyrin repeats. *J. Biol. Chem.* **268,** 22703–22709 (1993).
22. Greene, R. F. & Pace, C. N. Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, alpha-chymotrypsin, and beta-lactoglobulin. *J. Biol. Chem.* **249,** 5388–5393 (1974).
23. Myers, J. K., Pace, C. N. & Scholtz, J. M. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4,** 2138–2148 (1995).
24. Zweifel, M. E. & Barrick, D. Studies of the ankyrin repeats of the Drosophila melanogaster Notch receptor. 2. Solution stability and cooperativity of unfolding. *Biochemistry* **40,** 14357–14367 (2001).
25. Sterner, R. & Liebl, W. Thermophilic adaptation of proteins. *Crit. Rev. Biochem. Mol. Biol.* **36,** 39–106 (2001).
26. Robic, S., Guzman-Casado, M., Sanchez-Ruiz, J. M. & Marqusee, S. Role of residual structure in the unfolded state of a thermophilic protein. *Proc. Natl. Acad. Sci. U.S.A.* **100,** 11345–11349 (2003).
27. Gibson, D. G., Smith, H. O., Hutchison, C. A., Venter, J. C. & Merryman, C. Chemical synthesis of the mouse mitochondrial genome. *Nat. Methods* **7,** 901–903 (2010).
28. Becktel, W. J. & Schellman, J. A. Protein stability curves. *Biopolymers* **26,** 1859–1877 (1987).
29. Hollien, J. & Marqusee, S. A thermodynamic comparison of mesophilic and thermophilic ribonucleases H. *Biochemistry* **38,** 3831–3836 (1999).

(A)



(B)

N–TERMINAL CAP SEQUENCE:
SKDGNTPLHNAAKNGHAEEVKKLLSKGADVNAR
C–TERMINAL CAP SEQUENCE:
SKDGNTPEHLAKKNGHHEIVKLLDAKGADVNAR

Figure 3.1. A) Schematic of TRPA1 channel. Cylinders represent transmembrane domains; ovals, ankyrin repeats. B) Amino acid sequences of the terminal capping ankyrin repeats.

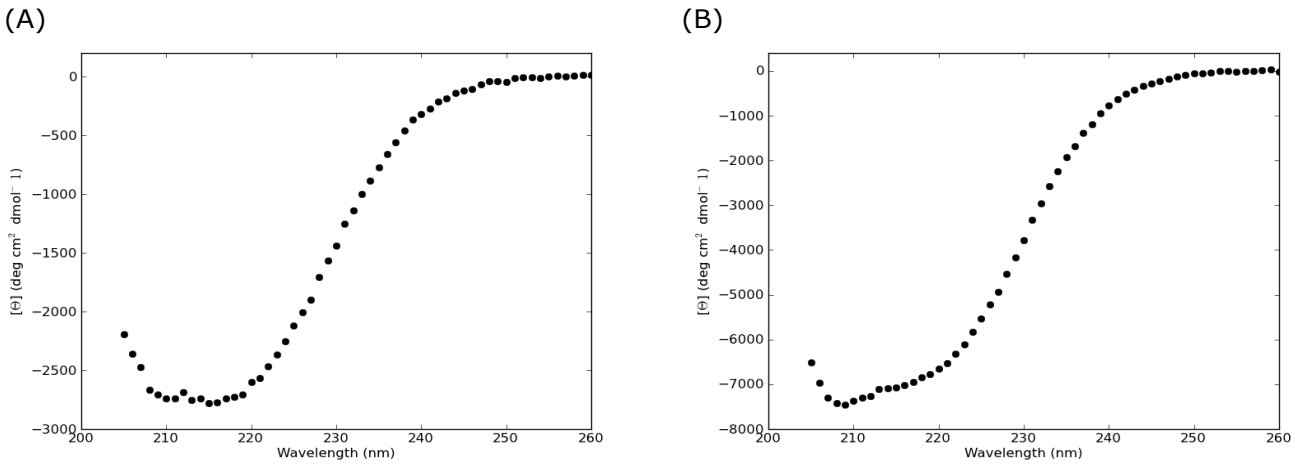


Figure 3.2. CD spectra of capped human (A) and rattlesnake (B) ankryin repeats 3-8, referred to in the text as h3c and s3c, respectively. 50 mM potassium phosphate buffer pH 7, 50 mM NaCl, 0.2 mM TCEP, 25˚C.

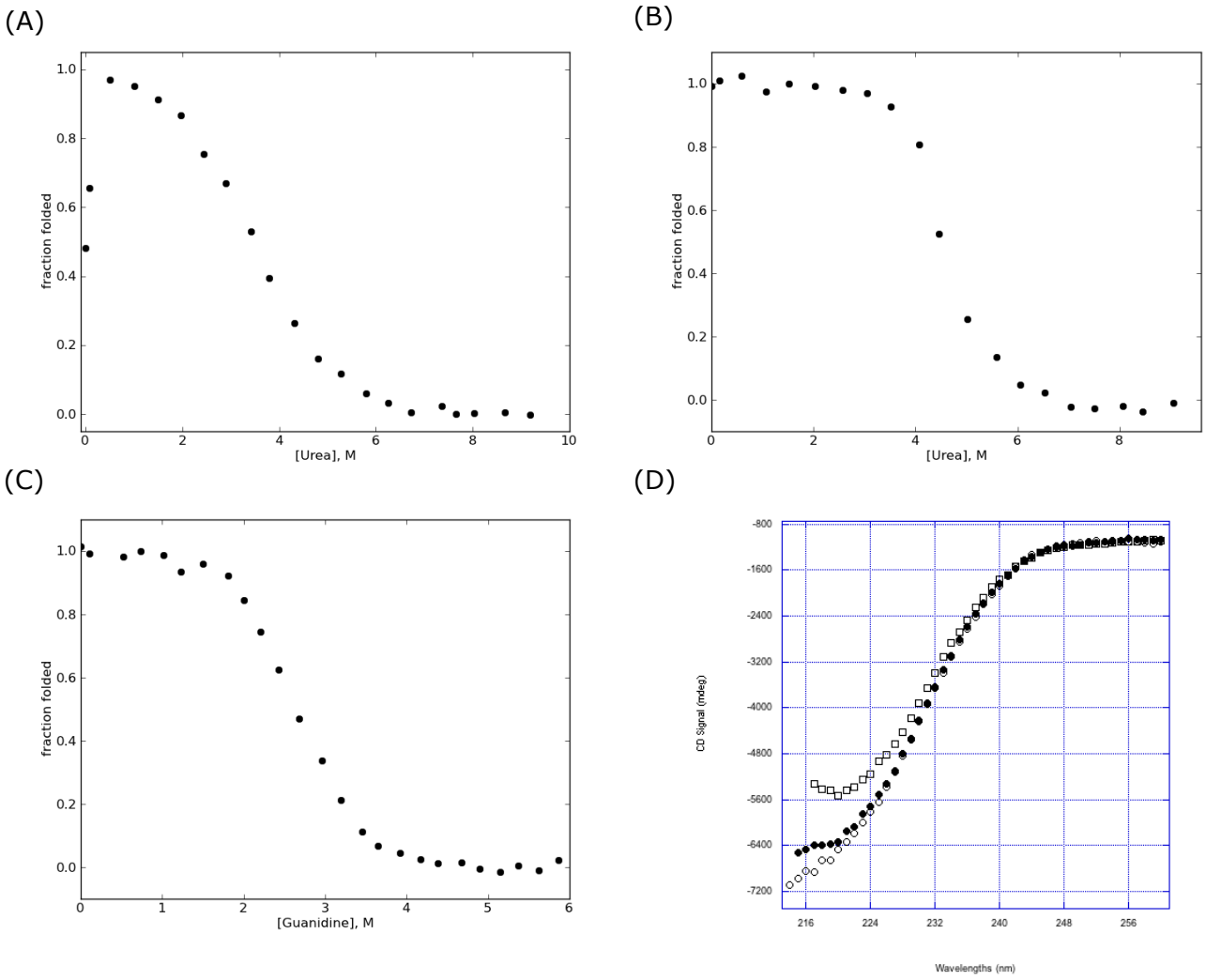


Figure 3.3. A) Fraction folded of h3c determined from urea melts fit to a two-state approximation. B) Fraction folded of s3c determined from urea melts fit to a two-state approximation. C) Fraction folded of s3c determined from guanidine melts fit to a two-state approximation. D) s3c CD spectra at 50 μg/ml (open circles), 250 μg/ml (filled circles), and 500 μg/ml (filled triangles). 50 mM potassium phosphate buffer pH 7, 50 mM NaCl, 0.2 mM TCEP. All experiments performed at 25˚C.

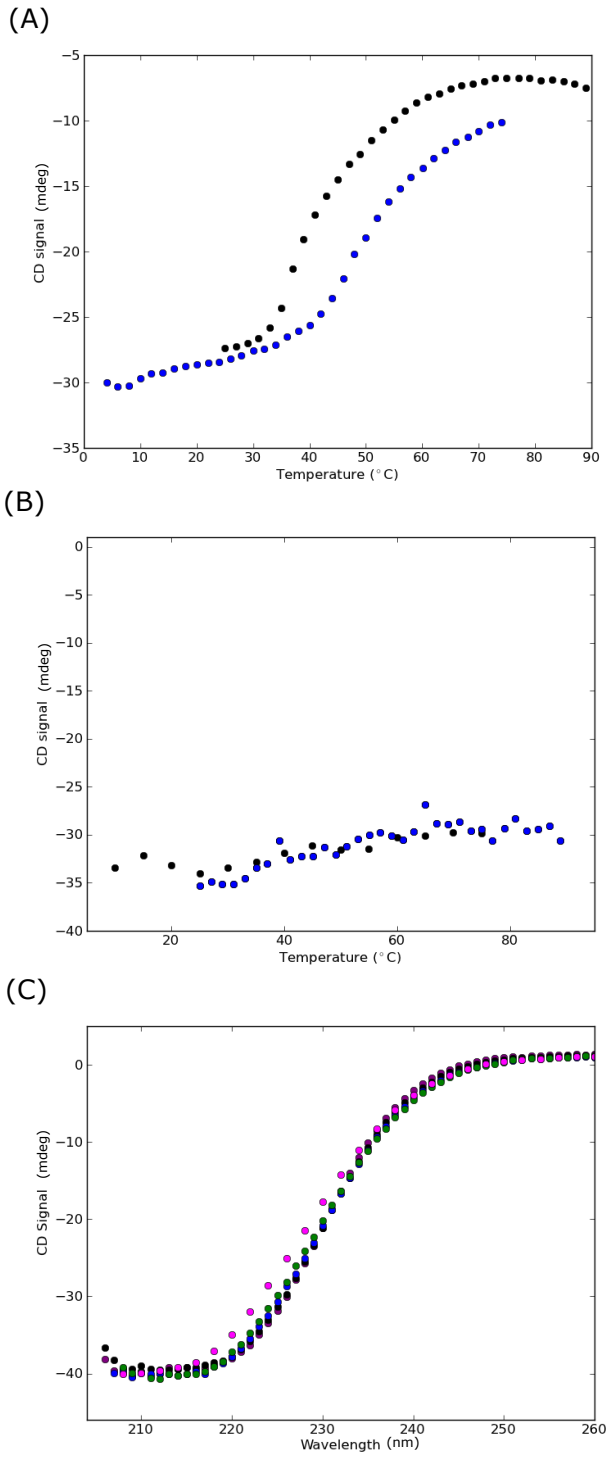


Figure 3.4. Two temperature melts (blue and black) of h3c (A) and s3c (B) CD signal at 222 nm in response to temperature. C) CD spectra of s3c following a 5-minute incubation at 25˚C (purple), 45˚C (black), 65˚C (blue), or 85˚C (green), or a twelve-hour incubation at 85˚C (magenta). 50 mM potassium phosphate buffer pH 7, 50 mM NaCl, 0.2 mM TCEP. Spectra were taken at 25˚C.

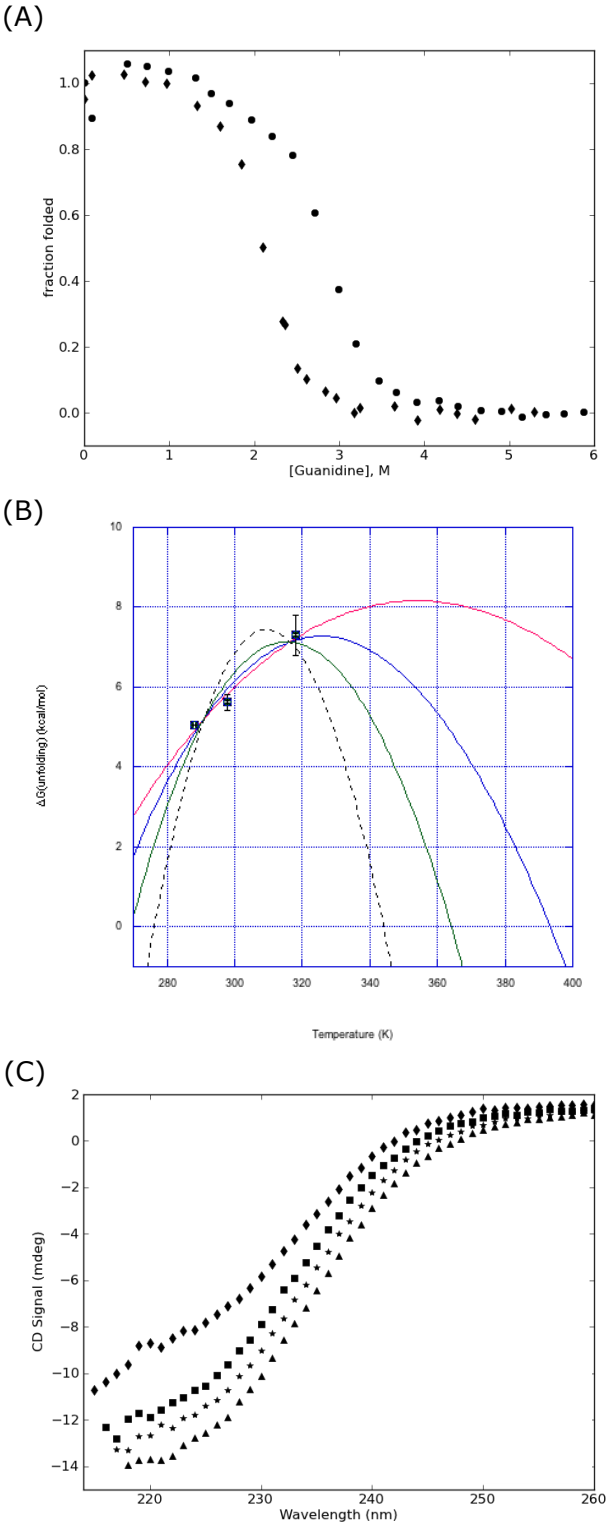


Figure 3.5. A) Fraction folded of s3c at 15˚C (diamonds) and 45˚C (circles) determined from guanidine melts fit to a two-state approximation. B) s3c $\Delta G_{unfolding}$ as a function of temperature, fit to Gibbs-Helmholtz equation by assuming ΔC_p = 4 kcal/mol (dashed black line), 2 kcal/mol (green line), 1 kcal/mol (blue line), or 0.5 kcal/mol (red line). C) CD spectra of s3c in 3.6M guanidine at 25˚C (diamonds), 45˚C (squares), 65˚C (stars), and 85˚C (triangles). 50 mM potassium phosphate buffer pH 7, 50 mM NaCl, 0.2 mM TCEP.

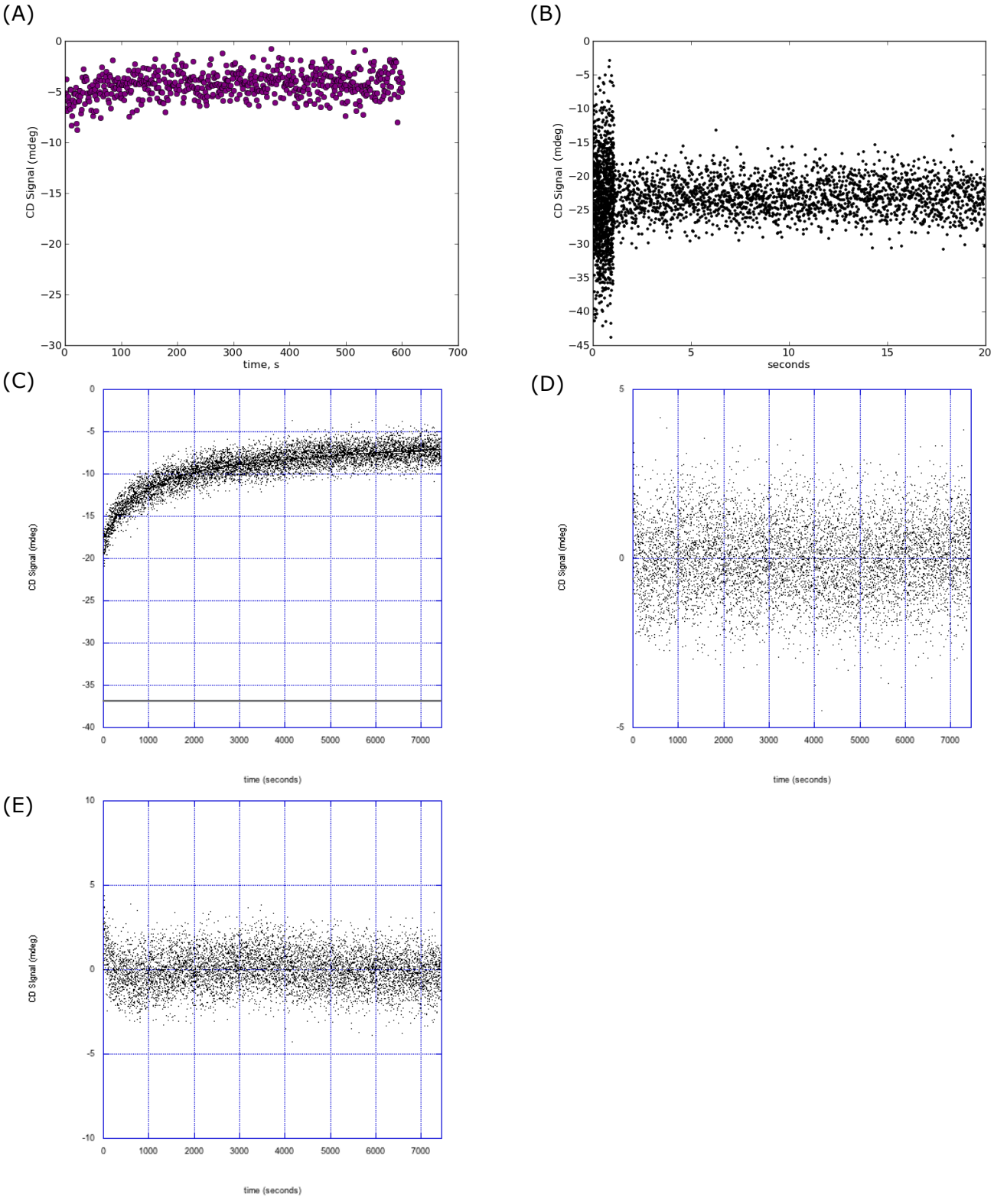


Figure 3.6. Kinetic traces with mdeg CD signal at 222nm plotted as a function of time after A) h3c in 0M urea was diluted into 9.5M urea buffer, B) s3c in 7M urea was diluted into 0M urea, and C) s3c in 0M urea was diluted into 9.5M urea. Line through data indicates fit; horizontal line indicates signal of the fully folded protein. Plots A and C were done with manual mixing (~20 second dead time); plot B was done with stop-flow instrument (~15 millisecond dead time). 50 mM potassium phosphate buffer pH 7, 50 mM NaCl, 0.2 mM TCEP, 25˚C. Residuals for biphasic (D) and monophasic (E) fits.

Table 3.1. Progress summary of s3c chimera evaluation		
Construct	Cloning	Temperature-dependent unfolding
s3c3*	completed	Not observed
s3c4*	completed	Not observed
s3c5*	in progress	TBD
s3c6*	in progress	TBD
s3c7*	in progress	TBD
s3c8*	in progress	TBD
s3-15c	in progress	TBD
s2-9c	in progress	TBD
s3cN* connotes s3c construct with human sequence of Nth ankyrin repeat. sY-Zc connotes capped rattlesnake ankyrin repeats Y-Z.		