# CHAPTER 1

# Protein folding: how, why, and beyond

**Athi N. Naganathan**
Department of Biotechnology, Bhupat & Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras (IITM), Chennai, India

## Outline



## Abbreviations

**H/D** hydrogen/deuterium
**IDP** intrinsically disordered protein
**IDR** intrinsically disordered region
**MD** molecular dynamics
**NMR** nuclear magnetic resonance
**PolyQ** polyglutamine
**WSME** Wako-Saitô-Muñoz-Eaton

## Introduction

A universal feature of biological polymers is their conformational complexity. This is more apparent in proteins, the molecular machines that perform the majority of functions within the cell. With an arsenal of 20 amino acids with very different chemical and structural properties, the unstructured heteropolymers synthesized by the ribosomes pack themselves into unique structures. The time taken for this self-organization and the shape eventually adopted depend, in most cases, almost entirely on the precise patterning of amino acids in the polymer sequence and the length of the sequence. In other words, if the patterning is different even in one or a few positions, it is highly likely that the polymer does not fold or that it populates several competing structures. Thus, the complexity of protein folding can be broadly classified as originating from two different phenomena: the large conformational entropy of the polymeric chain and the multitude of noncovalent interactions that hold the structure together.

*Protein Homeostasis Diseases.* http://dx.doi.org/10.1016/B978-0-12-819132-3.00001-4

The phase space (i.e., the array of conformations) accessible to a protein chain is astronomical, and this arises primarily from the backbone Ramachandran angles[1] that impart tremendous flexibility limited only by steric effects of the main-chain and side-chain atoms. For example, it is known that the entropic penalty for fixing an alanine in a folded-like conformation is $\sim 17$ J mol$^{-1}$ K$^{-1}$;[2] this translates to eight times more conformations (or main-chain dihedral possibilities) in the unfolded state compared to the folded state. If one were to consider a sequence of five alanines this would translate to a collection of $>28{,}000$ possible conformations apart from the fully folded conformation. This entropic stabilization of the polymeric chain needs to be substantially outweighed by favorable energetics to promote folding.[3-7] Such elementary considerations are also evidence that simple chemical models of folding would not do justice to the underlying conformational heterogeneity. Therefore, the probability of randomly arriving at a heteropolymer sequence that folds to a specific conformation is next to impossible.

The difference between a random heteropolymer and the evolutionarily selected extant protein sequence thus lies in the ability of the latter to fold in a biologically relevant timescale to one or a specific collection of conformations. How is this possible? Energy landscape theory, the currently accepted paradigm for protein folding, states that natural selection has fine-tuned the sequence patterning such that proteins fold by avoiding traps (misfolded nonnative states) that could arise due to conflicting energetics (or "frustration").[3,8] This helps them to more likely form favorable interactions, thus simultaneously decreasing energy and entropy and speeding up folding. The resulting conformational landscape in three dimensions is conventionally represented as a funnel with a deep minimum or a collection of a few minima. Protein folding landscapes are considered to be "minimally frustrated,"[3,9] as natural selection would have eliminated most of the nonnative interactions through sequence changes to enable rapid folding and optimal functionality.

Simple lattice models of protein folding that use polar (P) and hydrophobic (HP) partitioning of amino acid residue types are able to reproduce many of the basic thermodynamic and kinetic features observed experimentally.[10-14] In such simulations, it is clear that random placements of hydrophobic and polar amino acids result in multitudes of traps, and no specific organized structure is populated. In fact, one has to predesign such lattice sequences to ensure that they fold to specific conformations following which other folding-related features are studied. Folding mechanisms in lattices can also be distinctly tuned through design of sequences with different local and nonlocal content.[13] These simulations suggest that even if we have the computational power to sample all possible conformations for a real protein, the problem of protein structure-stability prediction boils down to the precise estimation of noncovalent energetics that hold the protein together. In other words, which one or a selected few has the lowest (free) energy while also accounting for solvent-associated energies and entropies?

The noncovalent interactions that stabilize proteins are varied and include van der Waals, charge-charge and charge-dipole, hydrogen bonds, and cation-$\pi$ interactions apart from hydrophobic free-energetic factors arising from polymer and solvent entropies (bulk and first shell of solvent).[15] While a majority of these terms are routinely included in simulations and structure-prediction algorithms, the precise magnitude or even the relative balance between the different terms is challenging to estimate. This is a critical component of any predictive modeling, as folded states of most proteins exhibit stabilities of ~10-30 kJ mol$^{-1}$ at, say, 298 K. In this regard, polymer chain entropic free-energy components are of the order of ~500 kJ mol$^{-1}$ for a 100-residue polypeptide. Many of the noncovalent interactions are of the order of just thermal energy or *RT* (~2.5 kJ mol$^{-1}$) or weaker but are very many in number and eventually dominate over the destabilizing entropic terms, thus stabilizing the folded state (or rather, just about stabilizing it).

These conflicting free-energetic terms also complicate the prediction of protein structures from sequence information, as not only should the conformational space be sampled extensively but the (free-) energy functions should be reasonably detailed and accurate to account for the diversity in noncovalent energetics. In this regard, it is pertinent to note that the success of the fragment-assembly method pioneered by the group of Baker and coworkers highlights the importance of local energetics and backbone geometry.[16] However, large-scale mini-protein design from the same group highlights the importance of buried non-polar surface area as a prominent feature to enhance protein stability potentially through higher-order effects.[17] Lessons from ancestral protein design and studies on mesophilic-thermophilic protein pairs reinforce the role of hydrophobicity and a highly interconnected network of charge-charge interactions in determining stability.[18-20] These studies underpin the role of multiple energetic terms in determining stability, designability, and foldability of proteins. However, designed proteins generally exhibit more complex folding thermodynamics and kinetics[21–23] compared to the natural proteins, suggesting that there are still rules to be learned that are hidden in the layers of compensating energetic-entropic effects.

How does one go about probing the folding mechanism of a protein? Experimentally, this involves constructing unfolding curves from varied experimental probes (circular dichroism, fluorescence, infrared) with temperature or denaturant as a first step followed by a simple chemical two- or three-state analysis to estimate stabilities and populations. This is followed by kinetic studies (stopped-flow kinetics or ultrafast spectroscopy)[24,25] to probe for kinetic evidence to the proposed mechanism from the number of relaxation rates, amplitudes, and their dependence on the perturbation employed. This is generally complemented by various studies, including differential scanning calorimetry[26,27] (information on stabilities, thermodynamic barriers, order of conformational transitions), nuclear magnetic resonance[28-31] (NMR; structure, dynamics, population of folded and partially structured states), hydrogen-deuterium (H/D) exchange coupled

with mass spectrometry[32,33] (rates, local and global stability estimates for most residues in the protein), and a host of single-molecule methods[34-38] (kinetics, distances between specific positions, populations, pathway heterogeneity, transition path times). Experimental constraints (e.g., solubility) preclude the possibility of doing all possible experiments on the same protein, but the overall theme is to identify a self-consistent mechanism of folding that explains the observed experimental signals. Site-specific mutations are then introduced to test the hypothesis on folding mechanisms and functional roles of specific regions (*vide infra*).

## Protein conformational landscapes

While it is common to employ two- or three-state models to fit the equilibrium unfolding curves, it is important to realize that they are, at best, approximations with little predictive power.[39,40] As discussed earlier, a statistical description that accounts for the diversity of conformations and noncovalent interaction types is best suited for explaining the experimental data (e.g., see the original classical works on statistical treatments of folding).[41-43] In fact, if the precise magnitude of each of the noncovalent interactions from the perspective of the local dielectric environment is known, it should be possible to predict protein stabilities and mechanisms, explore the role of mutations, and identify competing conformations. Some of these conformations could exhibit different catalytic rates or activities, substrate promiscuity, and misfolding tendencies, thus enabling a detailed view of protein sequence-structure-function-misfolding connection. However, it is impossible to sample all possible conformations of a protein chain, and this is where computational modeling comes to the fore.

All-atom molecular dynamics (MD) simulations in which all atoms are explicitly represented and a majority of the energetics are accounted for are arguably the ideal way to sample protein conformational space starting from the folded state.[44,45] Recent developments in sampling, Markov model-based approaches, hardware and better force fields, and water models have made all-atom MD an attractive avenue to explore conformational transitions with great temporal and structural detail.[46-49] Coarse graining, wherein residues are represented as single or multiple beads, can enhance conformational sampling and can transcend time-scale issues that typically plague all-atom MD simulations.[50-53] An alternative is to employ statistical mechanical models that treat residues as units and, based on specific assumptions, construct a predefined physical collection of possible conformational states, the relative probabilities of which are determined by the energetics.[54-57] While researchers can choose from any of the earlier-mentioned methods to sample conformational states, it is important to underscore the fact that coarse-grained and statistical models are arguably more predictive and quantitative but make more assumptions with fewer parameters.[52,58] Particularly, the latter two methods generally assume that the interactions observed in the native structure determine the folding mechanism (Gō-model)[59] and hence the relative population of partially structured states.

Long-time-scale MD simulations have provided strong evidence that native contacts determine folding mechanisms,[60] giving a recent fillip to Gō-model based approaches.

Once the states are sampled adequately or predicted unfolding curves calibrated against experiments, the conformations are generally projected onto one- or two-order parameters (that can also serve as reaction coordinates), and this constitutes the conformational landscape of a protein (Fig. 1.1). They provide a simple view of the complex hyperdimensional space associated with polymers, information on the relative population

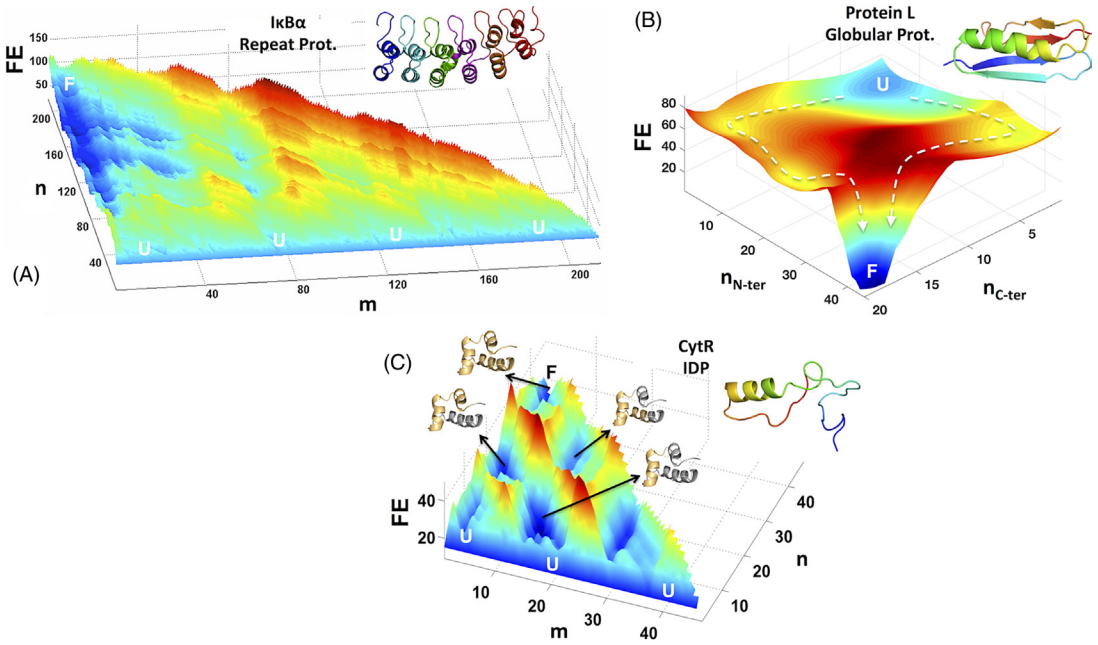


**Figure 1.1** ***Protein conformational landscapes (color online).*** All representations have been generated from the Wako-Saitô-Muñoz-Eaton (WSME) statistical mechanical model. A spectral color-coding is employed going from low free energy (FE in kJ $mol^{-1}$; *blue*) to high *(red)*. Unfolded and folded-like conformations are indicated as *U* and *F*, respectively. The cartoon on the top right of each panel depicts the native conformation. (A) Experimentally consistent conformational landscape of the repeat protein IκBα.[75] The coordinates *m* and *n* represent the starting structured residue and the number of structured residue in a single stretch, respectively. Numerous partially structured states are observed apart from the folded state that represent different numbers of structured repeats arising from weak packing of interrepeat interfaces. (B) The folding conformational landscape of the globular protein L, highlighting its funneled nature. Two macroscopic folding paths to the folded state starting from the unfolded state can also be seen *(dashed curves)*, in agreement with single-molecule experiments that point to intermediates.[76] The coordinates are the number of residues structured in the N and C termini, respectively. (C) Same as in panel A, but for the disordered protein CytR highlighting a flat landscape with multiple competing minima or conformations.[77] The *golden-colored* regions in the cartoons represent residues assuming a folded-like conformational status. The model predicts even a fully folded conformation to coexist in the disordered ensemble that was recently validated experimentally.[105] *Source: Adapted with permission from Ref. [77]. Copyright (2013) American Chemical Society.*

of varied states (that can be intermediates, excited, or partially structured states) and the barriers between them (Fig. 1.1). It should be possible to perform diffusive calculations of even simple Monte Carlo simulations on these landscapes to probe for time scales of interconversion. If landscapes are generated using all-atom MD, then it should be possible to identify even misfolded states that can serve as starting points for aggregation. Thus, conformational landscapes, if properly constructed, carry critical information and not just folding pathways but also the functionalities and regions of protein that are more prone to early or late unfolding. It is important to note that a careful choice of order parameters is needed, but in many cases, simple metrics like *RMSD* (root mean squared deviation from the starting structure), $R_g$ (radius of gyration), $Q$ (fraction of native contacts), or number of structured residues, are sufficient to provide a detailed picture of conformational landscapes (Fig. 1.1).

Conformational landscapes have been constructed for a large number of proteins and have shown to be extremely useful in understanding protein folding mechanisms (number of macrostates, possible routes, and fluxes), function, allostery, fold switching, posttranslational modification, misfolding, and environment-dependent modulation of populations (the reference list is not exhaustive!).[61-67] One-dimensional free-energy profiles have also been extensively employed to predict folding rates, identify intermediates, and probe for excited states that could be functionally relevant.[55,68-70] Successes of these methods are, in turn, a validation of the "minimum frustration principle" described in the energy landscape theory of protein folding. One aspect learned from many of these simulations with a strong theoretical underpinning is that frustration has not been or cannot be completely eliminated, as there are constraints on a protein to not only fold reasonably quickly but to also function and exhibit a certain kinetic and thermodynamic stability for efficient regulation. For example, DNA-binding domains need to necessarily display a collection of positive-charged residues close in space to bind the anionic DNA. Proteins that need to bind aromatic ligands would have a patch of hydrophobic residues on their surface, and this would evolve at the expense of a higher aggregation tendency. Enzyme active sites have also been predicted to be frustrated from structural analysis.[71] Such functionally driven frustration, in turn, makes certain regions of the protein more prone to local unfolding that can appear as intermediates or excited states in conformational landscapes (Fig. 1.1B). Importantly, they could determine the flexibility and hence the dynamics that is required for optimal catalytic activity. On the other hand, such partially structured states could determine the aggregation propensity of a protein, as has been shown on several systems.[72-74]

## Mutational perturbations to probe folding mechanisms and function

To understand many of the features discussed earlier, it is necessary to perturb proteins and probe for how they react (thermodynamic and kinetic stability, function, organismal fitness) to such perturbations. This is generally carried out through mutational studies

that have revealed a gold mine of information on protein folding, function, evolution, cooperativity, and allostery.[78-84] Mutations in proteins drive the adaptability of organisms to different environmental stresses and to generate new functionality or antibiotic resistance. Importantly, a majority of the inherited diseases have their origins in missense mutations that not only modulate stability but also have a silent (not yet quantified) impact on associated cellular processes including abundance, protein-protein/DNA interactions, sensitivity to changes in cytoplasmic composition, efficiency of degradation, and posttranslational modifications.

Numerous studies have established that proteins are marginally stable, and random mutational events constantly change the stability of proteins in cellular populations.[85,86] However, in many cases, it is seen that there is a stability threshold beyond which the protein cannot tolerate mutations and they unfold. The extent to which they tolerate mutations depends on the degree or order of a disorder in proteins and the interplay between them.[87] For example, while most random mutations in protein tend to destabilize proteins, certain mutations can enhance protein stability or activity in the presence of other mutations. Such high-order nonadditive mutational effects, termed epistasis, can be either beneficial (positive epistasis) or detrimental (negative epistasis) to the organismal fitness that can be quantified based on their functionality.[88] Ancestral protein reconstruction methods, deep mutational scanning, and directed evolution point to tremendous variation in the extent to which catalytic activity can be modulated upon mutations, highlighting their epistatic effects.[82,84,89] Remarkably, epistatic interactions alone can be employed to calculate three-dimensional structures of proteins through deep mutational scans, thus tremendously reshaping conventional applications of mutational studies.[90,91]

Importantly, it is now established that protein mutations in interior (i.e., buried residues) not only alter properties of the residues in their immediate neighborhood (first shell of residues) but also distant residues (second shell and beyond).[78] Such a propagative effect (in an equilibrium sense) is a hallmark of the network of noncovalent interactions within proteins that can readjust their packing and dynamics depending on the nature of the perturbation. Studies indicate dramatic changes in the dynamics, hydrogen-exchange protection factors, and chemical shifts for a large number of residues in proteins (at distances even >15-20 Å; Fig. 1.2A,C,E) on mutational perturbations.[92-94] Repacking of the protein interior on mutations affects the specificity-affinity ratio for ligands on the surface,[95] highlighting the "connectedness" of the protein interior with the surface. It is possible to accurately capture a majority of these equilibrium propagatory features via a variety of models, including network analysis of protein structures, MD simulations, and even detailed quantitative statistical mechanical modeling.[96]

This tightrope walk of protein sequences in the stability-activity phase space provides a peek into how nature could have engineered proteins with different functionalities. At a fundamental level, such tunability and context dependence of missense variants has its origins in the pliable nature of the intraprotein interaction network and the degree

of partitioning of stabilizing versus destabilizing interactions in specific regions within proteins. When interactions within a protein are perturbed, different regions react differently, depending on the strength and malleability of the local interaction network. This in turn (de)stabilizes certain interactions and modulates the population of partially structured states apart from the unfolded state, thus affecting the conformational landscape (Fig. 1.2B,D,F). Single-point mutations thus reshape the landscape irrespective

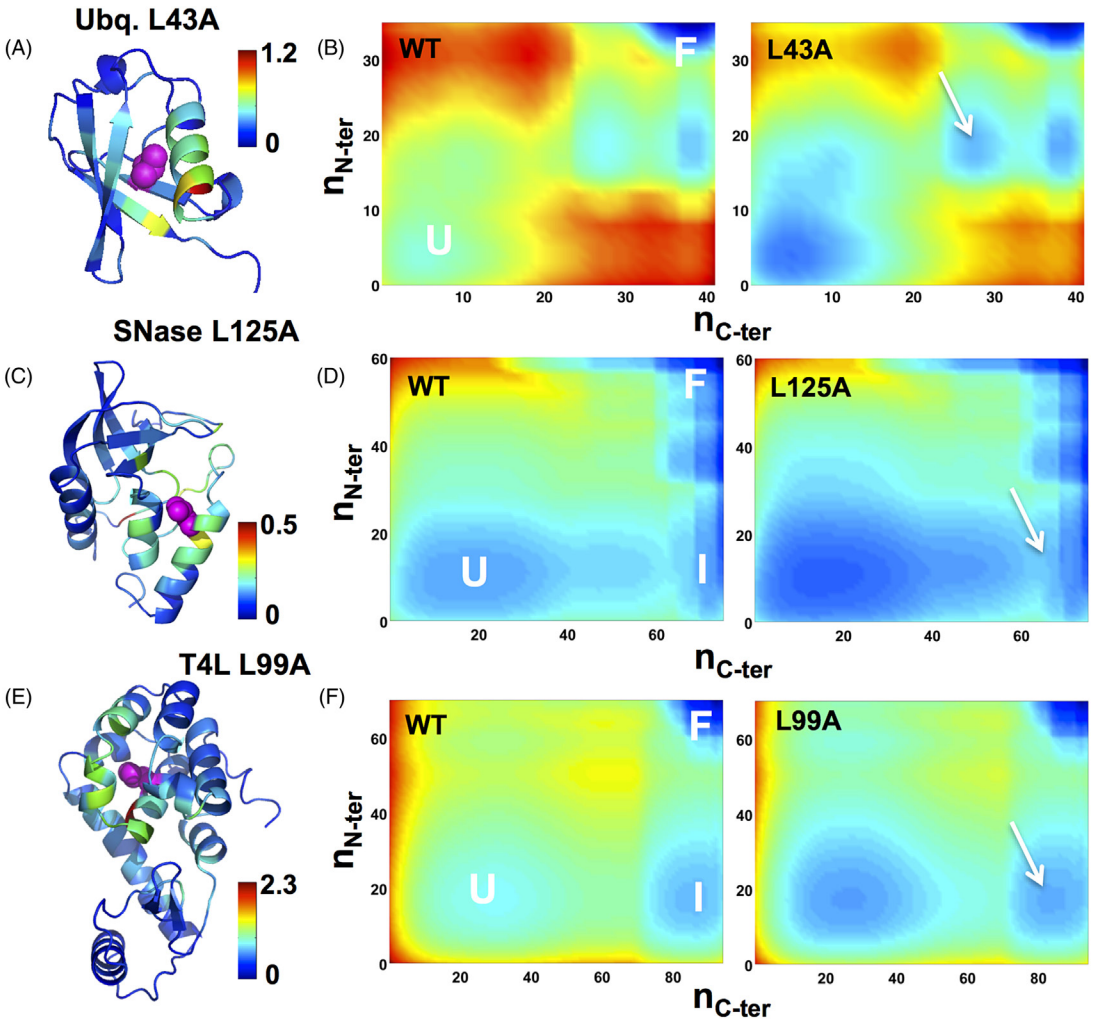


**Figure 1.2** ***Mutations in protein interior modulate the population of partially structured states.*** The left-most column (A, C, and E) depicts cartoons of proteins with the effect of specific mutations (i.e., chemical shifts) mapped onto the structure.[93,97,98] The middle and right-most columns (B, D, and F) are the folding landscapes of the wild type (WT) and the representative mutants from the perspective of the Wako-Saitô-Muñoz-Eaton (WSME) model.[96] *Arrows* point to the appearance or a relatively higher population of partially structured states compared to the WT. *Source: Adapted with permission from Ref. [96]. Copyright (2017) American Chemical Society.*

of whether they are buried or exposed to the solvent (Figs. 1.2, 1.3A) and potentially contribute to epistatic effects in the presence of a secondary mutation, aspects that can be extracted only from an intimate and quantitative modeling of protein conformational behavior. The modulation of the population need not be restricted to mutations but can also be achieved through ion-binding events (Fig. 1.3B), temperature changes (Fig. 1.3C,D), pH variations (Fig. 1.3E,F), generic charge screening from ionic-strength changes, and posttranslational modifications (e.g., phosphorylation, Fig. 1.3G,H), and therefore even upon ligand binding. It is also likely that certain mutations at specific positions can predispose proteins to undergo misfolding and aggregation by populating alternate conformations or even through rerouting of folding flux.

## Disordered proteins-regions and unfolded states

If the driving force for compaction is small due to weak sequence hydrophobicity, spatially proximity of similarly charged residues, or excess of glycine/proline or polar residues, a protein or regions of proteins can be completely disordered, even in native conditions.[106,107] Disordered proteins can therefore be seen as an extreme manifestation of frustration at not only the level of sequence but through degeneracy in interactions; in fact, this is expected of polymeric systems that can be stabilized by either entropy and specific interactions alone or through a combination of both. Such intrinsically disordered proteins (IDPs) or regions (IDRs) have redefined the structure-function paradigm due to their ubiquity in organisms at different levels. Many IDPs act as hubs in protein-protein interaction networks and acquire structure only in the presence of their partners, thus regulating function in a conditional manner. Being disordered can help in easier regulation (through degradation) and can potentially promote promiscuity in binding, aspects that are currently of great interest to protein scientists.[108-111]

Much of the focus over the last decade has now shifted from working on ordered domains to IDPs or proteins with a large chunk of IDRs. Many IDRs and IDPs have larger mutational rates compared to their ordered counterparts, suggesting that they are the hotbeds of evolution.[112] Experiments and simulations are now increasingly identifying that IDPs populate pockets of local structures (and even fully folded conformations) while at the same time exhibiting random coil dimensions.[113-115] For example, α-synuclein, which is largely expressed in the nervous system and implicated in Parkinson's disease, is an IDP that can adopt helical conformation near membranes and beta-sheet-like structures in aggregates. It can also undergo extensive posttranslational modifications ranging from phosphorylation to glycosylation, despite its size of just 140 amino acids.[116-120] Similarly, polyglutamine (polyQ) diseases can have complex origins depending on the length of the polyQ tract and the associated regions.[121,122] Such a feature is not intrinsic to just α-synuclein or proteins rich in polyQ tracts but to a large number of disordered proteins arising from the diverse nature of conformations present in equilibrium. Thus, even simple sequence-based approaches can provide a detailed

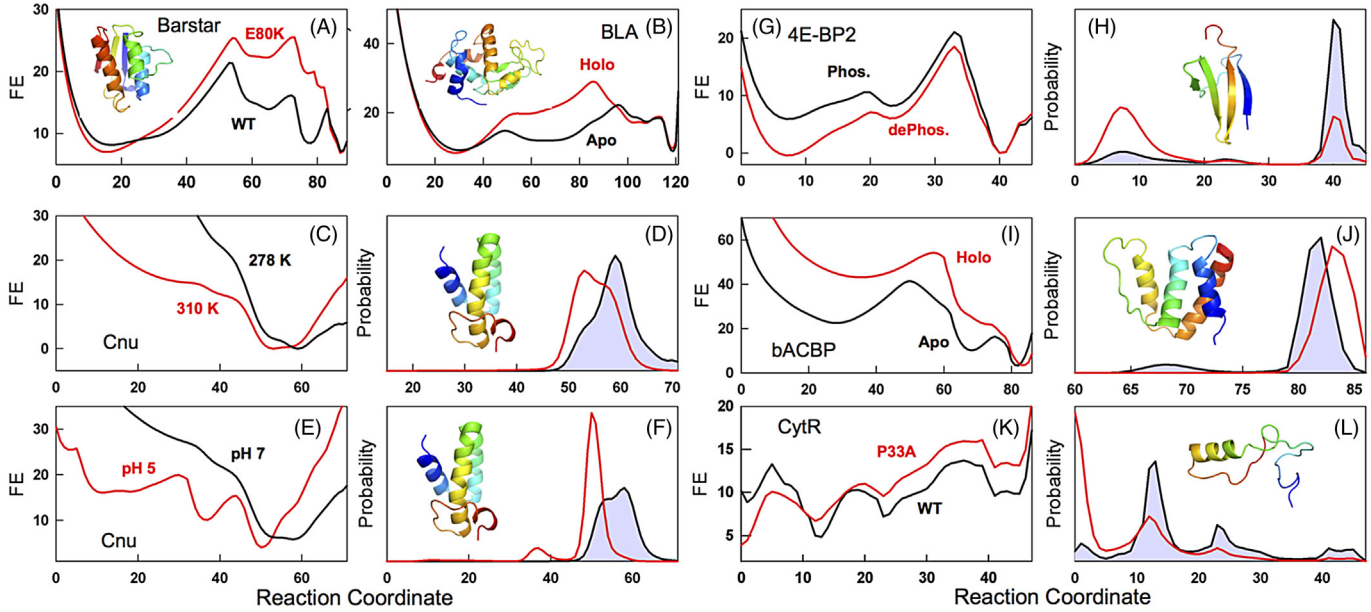


**Figure 1.3** ***Role of mutations-solvent-environment in determining folding thermodynamic behaviors.*** All simulations are done at a (semi)quantitative level with the Wako-Saitô-Muñoz-Eaton (WSME) model[56] that employs the number of structured residues as the reaction coordinate (RC). (A) Wild-type (WT) Barstar exhibits multistate unfolding behavior[99] that can be eliminated by a specific mutation that relieves electrostatic frustration.[69] (B) Free-energy profiles, at the respective midpoint conditions for the apo- and holo-forms of bovine lactalbumin,[100] with the ion binding mimicked by a D82N mutation that again relieves electrostatic frustration.[56] (C, D) Free-energy profiles (panel C) and probability densities (panel D) of the thermosensory protein Cnu that displays population redistribution within the native well upon temperature rise from 278 to 310 K.[101] (E, F) Cnu is also sensitive to pH changes in the range between 5 and 7 due to protonation of a partially buried histidine that promotes electrostatic frustration in its vicinity.[102] (G, H) Experimentally observed phosphorylation-driven conformational switch between folded- and unfolded-like conformations in the disordered protein 4E-BP2[103] can be quantitatively captured by the WSME model.[104] (I, J) Apo- and holo-forms of the bovine Acyl-CoA-binding protein (ACBP) exhibit different packing and electrostatic interaction energy distributions contributing to differences in the free-energy profiles (panel I) that are suggestive of a conformational selection mechanism of binding.[70] Note the depopulation of a partially structured state on binding acyl-coA ligand (at RC value ~67, panel J). (K, L) Disordered CytR populates an excited, folded-like conformation in its native ensemble whose population can be reduced by a specific mutation of P33A.[105] *Source: Part A, Adapted with permission from Ref. [69]. Copyright (2015), American Chemical Society; Part B, Adapted with permission from Ref. [56]. Copyright (2012), American Chemical Society; Part C, D, Adapted with permission from Ref. [101]. Copyright (2017), American Chemical Society; Part E, F, Adapted with permission from Ref. [102]. Copyright (2018), American Chemical Society; Part G, H, Adapted from Ref. [104] with permission from the PCCP Owner Societies; Part I, J, Adapted from Ref. [70] with permission from the PCCP Owner Societies.*

view of the phase space accessible to IDPs based purely on the partitioning of charged residues in sequence.[123-125] On the other hand, recent experiments point to how disorder in some proteins is agnostic to sequence composition or the resulting microstructure, pointing to an entropic mechanism for driving allostery,[126] as expected from purely thermodynamic considerations.[127-129]

The heterogeneous landscapes accessible to IDPs in turn translates to an array of binding mechanisms. IDPs can therefore bind their partners on folding or fold upon binding, but simulations hint that a combination of the two mechanisms is also possible.[77,130,131] Therefore, reshaping of conformational landscapes on mutations need not be restricted to folded proteins but can have a distinct effect on disordered proteins and regions, affecting the population of partially structured states and dimensions in a nonintuitive manner. For example, a single P33A mutation in the disordered CytR DNA-binding domain reduces the population of the excited folded conformation while weakening its binding affinity to the cognate site (Fig. 1.3K,L).[105] Similarly, disordered tails in DNA-binding domains determine their specificity-affinity ratio to varied DNA sequences[132,133] and help in one-dimensional sliding-transfer between different DNA strands,[134,135] while tuning their charge patterning can effectively determine transcriptional activity.[136]

Similar to IDPs, the unfolded states of folded proteins have received equivalent attention, as the presence of specific structural elements or interactions can predispose folding reactions and determine stabilities. Unfolded states are more compact under folding conditions compared to unfolding conditions, can have very different structural preferences depending on the nature of the denaturant, and thus can speed up or slow down folding through various nonintuitive effects, aspects that are still being explored in detail.[114,115,137-142] Studies on unfolded states have thus challenged the limits of experiments, theoretical arguments, and simulations, promoting a fresh look into numerous aspects, ranging from the assumptions implicit in experimental analyses to the nature of force fields in simulations.[141-144] They have had positive knock-on effects in understanding the phase space of IDPs and the approaches employed to characterize them.

## Folding, stability, and binding in vivo

The protein folding-stability-function measurements discussed earlier are usually carried out under controlled conditions in vitro, thus simplifying the ease of experimentation and interpretation. However, in reality, proteins fold and function in highly crowded nonequilibrium environments within a cell (see Chapter 2). The cytoplasmic composition of prokaryotic or eukaryotic cells are highly varied with numerous ions, solutes, cofactors, and sugars. It is important to realize that the basic interactions that define folding also define the degree of association with different solutes and ligands. Accordingly, the protein conformation is affected not only by steric (excluded-volume) effects but also confinement, specific interactions, and nonspecific binding (mostly transient), all

of which fall under the category of "quinary" interactions.[145-153] Moreover, the cell frequently undergoes shape changes that modulate the cytoplasmic composition, concentration of solutes, and hence crowding driven by a variety of factors including osmotic environment, temperature, cell division status, or even when being motile. These features are further compounded by complex structures such as membranes, DNA, and membrane-less organelles that contribute to a radically different microenvironment around them arising from a combination of electrostatic effects, solvation, and enrichment of specific solutes/ions.

In-cell effects on folding, stability, and binding are now increasingly being studied through a combination of live-cell NMR ($^{1}H$-$^{15}N$ or $^{19}F$-labeled proteins) and fluorescence microscopy.[145,154] As expected, the extent to which a protein is stabilized and destabilized within a cell compared to in vitro conditions is highly protein dependent and has its origins in the nature, composition, and partitioning of charged residues on the protein surface.[155] Remarkably, protein stability and kinetics are also affected by phases of the cell cycle and the compartment in which a protein is localized.[156,157] Proteins exhibit a larger spread in their folding relaxation times in vivo[158] when compared to in vitro conditions, potentially arising from the highly variable environment that (de)stabilizes specific conformations along the folding pathway.

These studies suggest that there is an additional selection pressure in the evolution of proteins governed by the quinary environment. This is more evident in recent experiments in which eukaryotic proteins were reported to feel a stickier environment (i.e., slower diffusion) in *E. coli* while an endogenous counterpart is able to move relatively faster.[159] The situation is more complicated when extrapolating IDPs' behavior from in vitro experiments, as they can exhibit very different structures or dynamics in the variable cellular environment. In fact, recent experiments-simulations on disordered DNA-binding proteins propose that the protein is fully folded in the vicinity of DNA[160] owing to the large negative electrostatic potential of nucleic acids[161] that compensates for unfavorable electrostatic interactions within the protein. Such conditional order has potentially been selected for efficient regulation but highlight the problems one encounters when extrapolating function from in vitro experiments.

## "Real proteins" and beyond

A large chunk of protein folding studies have been on small single-domain proteins that are either single-gene products or independently folding domains (~30-100 residues) from larger proteins. The functional units of most proteins are significantly larger, ranging between 200 and 400 residues in length, on average. While the basic physics of folding and binding hold true for larger proteins, there are additional considerations that complicate their studies. In larger proteins, the number of misfolding traps is expected to be proportionally higher with nonlocal effects dominating the landscape. They also exist in multiple quarternary structural states apart from their predisposition to bind

various agonists and antagonists, thus confounding the structure-function relationships. Last, large proteins have independently folded domains that interact in a complex manner, and a simple spectroscopic measure (circular dichroism or fluorescence) will not do justice to the expected intricate unfolding process. True to this expectation, misfolding is common in large proteins with multiple domains, but the extent to which this happens diminishes with increasing sequence divergence of adjacent domains.[162,163] These works provide strong evidence for an additional level of complexity that is positively selected for during protein evolution.

A majority of experimental studies on large proteins have exploited the power of NMR, H/D exchange, mass spectrometry, and single-molecule fluorescence resonance energy transfer combined with mutational perturbations to study various processes including ligand binding, allostery, and structural transitions. Energy landscape theory and its associated structure-based models have also been successful in capturing numerous features associated with large enzymes and membrane and multidomain proteins.[164,165] Intensive MD simulations are increasingly providing vivid insights into structural changes and allosteric transitions occurring in ion channels. However, evidence is accumulating for the role of membranes and their lipid composition in determining and fine-tuning signals, aspects that were earlier not explored in detail in simulations.[166,167] Since the time scale of folding scales with protein length, larger proteins will fold and unfold significantly slower; in other words, the associated structural transitions could be in the order of milliseconds or slower, a time scale that is still out of reach for simulations.

Finally, there are not many models or computational approaches that can consistently predict protein structures from sequences, sequences that fold to specific structures, their stabilities, and time scales of various local and global structural transitions. Similarly, many diseases have their origins in small changes in amino acid sequences that can have a dramatic effect on solubility, protein-ligand (partner protein, DNA, or small molecules) interactions, and misfolding tendencies. All of these are intimately related to the energetics of interactions within and between proteins, solvent effects, and their relative magnitudes that can be tuned by the environment. Protein folding-stability-function studies from experimental approaches and computational treatments of the same are thus at crossroads where the onus is on continually improving the resolution (spatial and temporal) at which structural events are observed, not just for small, singe-domain proteins but for large functional units or enzymes, and for a detailed quantitative characterization of folding-function landscapes. An engaging crosstalk between experiments and simulations will constantly challenge these approaches while pushing the frontiers of what is possible. Importantly, they can potentially provide elusive solutions to diseases associated with protein (mis)folding that have their origins in missense mutations or environmental conditions that manifest as surprisingly small changes in atomic-level interaction energy terms.

## Acknowledgments

This chapter is built on decades of work by students, postdoctoral researchers, colleagues, and researchers across the globe. I apologize if I have missed out on obvious references, given the vast literature on protein folding and dynamics spanning varied approaches. A.N.N is a Wellcome Trust/DBT India Alliance Intermediate fellow.

## References

1. Ramachandran GN, Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain configurations. *J Mol Biol* 1963;**7**:95–9.
2. Daquino JA, Gomez J, Hilser VJ, Lee KH, Amzel LM, Freire E. The magnitude of the backbone conformational entropy change in protein folding. *Proteins* 1996;**25**(2):143–56.
3. Bryngelson JD, Onuchic JN, Socci ND, Wolynes PG. Funnels, pathways, and the energy landscape of protein-folding - a synthesis. *Proteins* 1995;**21**(3):167–95.
4. Veitshans T, Klimov D, Thirumalai D. Protein folding kinetics: timescales, pathways and energy landscapes in terms of sequence-dependent properties. *Fold Des* 1997;**2**(1):1–22.
5. Sali A, Shakhnovich EI, Karplus M. How does a protein fold? *Nature.* 1994;**369**:248–51.
6. Finkelstein AV, Shakhnovich EI. Theory of cooperative transitions in protein molecules. II. Phase diagram for a protein molecule in solution. *Biopolymers* 1989;**28**:1681–9.
7. Akmal A, Muñoz V. The nature of the free energy barriers to two-state folding. *Proteins* 2004;**57**(1):142–52.
8. Ferreiro DU, Komives EA, Wolynes PG. Frustration in biomolecules. *Q Rev Biophys* 2014;**47**:285–363.
9. Socci ND, Onuchic JN, Wolynes PG. Protein folding mechanisms and the multidimensional folding funnel. *Proteins* 1998;**32**(2):136–58.
10. Socci ND, Onuchic JN, Wolynes PG. Diffusive dynamics of the reaction coordinate for protein folding funnels. *J Chem Phys* 1996;**104**(15):5860–8.
11. Chan HS, Dill KA. Protein folding in the landscape perspective: Chevron plots and non-Arrhenius kinetics. *Proteins* 1998;**30**(1):2–33.
12. Chan HS, Shimizu S, Kaya H. Cooperativity principles in protein folding. *Methods Enzymol* 2004;**380**:350–79.
13. Abkevich VI, Gutin AM, Shakhnovich E. Impact of local and non-local interactions on thermodynamics and kinetics of protein folding. *J Mol Biol* 1995;**252**:460–71.
14. Klimov DK, Thirumalai D. Cooperativity in protein folding: from lattice models with sidechains to real proteins. *Fold Des* 1998;**3**(2):127–39.
15. Baldwin RL. Energetics of protein folding. *J Mol Biol* 2007;**371**(2):283–301.
16. Rohl CA, Strauss CEM, Misura KMS, Baker D. Protein structure prediction using rosetta. *Methods Enzymol* 2004;**383**:66–93.
17. Rocklin GJ, Chidyausiku TM, Goreshnik I, et al. Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science* 2017;**357**(6347):168–75.
18. Szilagyi A, Zavodszky P. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure* 2000;**8**(5):493–504.
19. Risso VA, Gavira JA, Mejia-Carmona DF, Gaucher EA, Sanchez-Ruiz JM. Hyperstability and substrate promiscuity in laboratory resurrections of Precambrian beta-lactamases. *J Am Chem Soc* 2013;**135**(8):2899–902.
20. Zou T, Risso VA, Gavira JA, Sanchez-Ruiz JM, Ozkan SB. Evolution of conformational dynamics determines the conversion of a promiscuous generalist into a specialist enzyme. *Mol Biol Evol* 2015;**32**(1):132–43.
21. Basak S, Nobrega RP, Tavella D, et al. Networks of electrostatic and hydrophobic interactions modulate the complex folding free energy surface of a designed betaalpha protein. *Proc Natl Acad Sci USA* 2019;**116**(14):6806–11.

22. Walters AL, Deka P, Corrent C, et al. The highly cooperative folding of small naturally occurring proteins is likely the result of natural selection. *Cell* 2007;**128**:613–24.
23. Sadqi M, de Alba E, Perez-Jimenez R, Sanchez-Ruiz JM, Muñoz V. A designed protein as experimental model of primordial folding. *Proc Natl Acad Sci USA* 2009;**106**:4127–32.
24. Jones CM, Henry ER, Hu Y, et al. Fast events in protein-folding initiated by nanosecond laser photolysis. *Proc Natl Acad Sci USA* 1993;**90**(24):11860–4.
25. Munoz V, Cerminara M. When fast is better: protein folding fundamentals and mechanisms from ultrafast approaches. *Biochem J* 2016;**473**(17):2545–59.
26. Ibarra-Molero B, Naganathan AN, Sanchez-Ruiz JM, Muñoz V. Modern analysis of protein folding by differential scanning calorimetry. *Methods Enzymol* 2016;**567**:281–318.
27. Sanchez-Ruiz JM. Probing free-energy surfaces with differential scanning calorimetry. *Ann Rev Phys Chem* 2011;**62**:231–55.
28. Neudecker P, Lundstrom P, Kay LE. Relaxation dispersion NMR spectroscopy as a tool for detailed studies of protein folding. *Biophys J* 2009;**18**:2045–54.
29. Tzeng SR, Kalodimos CG. Protein dynamics and allostery: an NMR view. *Curr Opin Struct Biol* 2011;**21**(1):62–7.
30. Wand AJ. The dark energy of proteins comes to light: conformational entropy and its role in protein function revealed by NMR relaxation. *Curr Opin Struct Biol* 2013;**23**(1):75–81.
31. Igumenova TI, Frederick KK, Wand AJ. Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution. *Chem Rev* 2006;**106**(5):1672–99.
32. Englander SW, Mayne L, Krishna MMG. Protein folding and misfolding: mechanism and principles. *Q Rev Biophys* 2007;**40**(4):287–326.
33. Englander SW. Protein folding intermediates and pathways studied by hydrogen exchange. *Ann Rev Biophys Biomol Struct* 2000;**29**:213–38.
34. Schuler B, Hofmann H. Single-molecule spectroscopy of protein folding dynamics-expanding scope and timescales. *Curr Opin Struct Biol* 2013;**23**:36–47.
35. Chung HS, Eaton WA. Protein folding transition path times from single molecule FRET. *Curr Opin Struct Biol* 2018;**48**:30–9.
36. Zoldak G, Rief M. Force as a single molecule probe of multidimensional protein energy landscapes. *Curr Opin Struct Biol* 2013;**23**(1):48–57.
37. Alegre-Cebollada J, Perez-Jimenez R, Kosuri P, Fernandez JM. Single-molecule force spectroscopy approach to enzyme catalysis. *J Biol Chem* 2010;**285**(25):18961–6.
38. Takahashi S, Kamagata K, Oikawa H. Where the complex things are: single molecule and ensemble spectroscopic investigations of protein folding dynamics. *Curr Opin Struct Biol* 2016;**36**:1–9.
39. Lumry R, Biltonen RL, Brandts JF. Validity of the "two-state" hypothesis for conformational transitions of proteins. *Biopolymers* 1966;**4**:917–44.
40. Naganathan AN, Perez-Jimenez R, Sanchez-Ruiz JM, Muñoz V. Robustness of downhill folding: guidelines for the analysis of equilibrium folding experiments on small proteins. *Biochemistry* 2005;**44**(20):7435–49.
41. Wako H, Saito N. Statistical mechanical theory of protein conformation. 1. General considerations and application to homopolymers. *J Phys Soc Japan* 1978;**44**(6):1931–8.
42. Ikegami A. Statistical thermodynamics of proteins and protein denaturation. *Adv Chem Phys* 1981;**46**:363–413.
43. Abe H, Go N. Noninteracting local-structure model of folding and unfolding transition in globular proteins. II. Application to two-dimensional lattice proteins. *Biopolymers* 1981;**20**(5):1013–31.
44. Best RB. Atomistic molecular simulations of protein folding. *Curr Opin Struct Biol* 2012;**22**: 52–61.
45. Piana S, Klepeis JL, Shaw DE. Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. *Curr Opin Struct Biol* 2014;**24**:98–105.
46. Leone V, Marinelli F, Carloni P, Parrinello M. Targeting biomolecular flexibility with metadynamics. *Curr Opin Struct Biol* 2010;**20**:148–54.
47. Doshi U, Hamelberg D. Towards fast, rigorous and efficient conformational sampling of biomolecules: advances in accelerated molecular dynamics. *Biochim Biophys Acta* 2015;**1850**:878–88.

48. Perez A, Morrone JA, Simmerling C, Dill KA. Advances in free-energy-based simulations of protein folding and ligand binding. *Curr Opin Struct Biol* 2016;**36**:25–31.
49. Chodera JD, Noe F. Markov state models of biomolecular conformational dynamics. *Curr Opin Struct Biol* 2014;**25**:135–44.
50. Hyeon C, Thirumalai D. Capturing the essence of folding and functions of biomolecules using coarse-grained models. *Nat Commun* 2011;**2**:487.
51. Chan HS, Zhang Z, Wallin S, Liu Z. Cooperativity, Local-Nonlocal Coupling, and Nonnative Interactions: Principles of Protein Folding from Coarse-Grained Models. In: Leone SR, Cremer PS, Groves JT, Johnson MA, eds. Ann. Rev. Phys. Chem. Vol 622011:301-326.
52. Naganathan AN. Coarse-grained models of protein folding as detailed tools to connect with experiments. *WIREs Comput Mol Sci* 2013;**3**:504–14.
53. Kmiecik S, Gront D, Kolinski M, Wieteska L, Dawid AE, Kolinski A. Coarse-grained protein models and their applications. *Chem Rev* 2016;**116**(14):7898–936.
54. Wako H, Saito N. Statistical mechanical theory of protein conformation. 2. Folding pathway for protein. *J Phys Soc Japan* 1978;**44**(6):1939–45.
55. Muñoz V, Eaton WA. A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc Natl Acad Sci USA* 1999;**96**(20):11311–6.
56. Naganathan AN. Predictions from an ising-like statistical mechanical model on the dynamic and thermodynamic effects of protein surface electrostatics. *J Chem Theory Comput* 2012;**8**(11):4646–56.
57. Inanami T, Terada TP, Sasai M. Folding pathway of a multidomain protein depends on its topology of domain connectivity. *Proc Natl Acad Sci USA* 2014;**111**(45):15969–74.
58. Naganathan AN. Predictive modeling of protein folding thermodynamics, mutational effects and free-energy landscapes. *Proc Indian Natl Sci Acad* 2016;**82**:1211–28.
59. Taketomi H, Ueda Y, Go N. Studies on protein folding, unfolding and fluctuations by computer-simulation. 1. Effect of specific amino-acid sequence represented by specific inter-unit interactions. *Inter J Prot Pep Res* 1975;**7**(6):445–59.
60. Best RB, Hummer G, Eaton WA. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc Natl Acad Sci USA* 2013;**110**(44):17874–9.
61. Truong HH, Kim BL, Schafer NP, Wolynes PG. Predictive energy landscapes for folding membrane protein assemblies. *J Chem Phys* 2015;**143**(24):243101.
62. Li W, Wolynes PG, Takada S. Frustration, specific sequence dependence, and nonlinearity in large-amplitude fluctuations of allosteric proteins. *Proc Natl Acad Sci USA* 2011;**108**(9):3504–9.
63. Ferreiro DU, Walczak AM, Komives EA, Wolynes PG. The energy landscapes of repeat-containing proteins: topology, cooperativity, and the folding funnels of one-dimensional architectures. *PLOS Comp Biol* 2008;**4**(5):e1000070.
64. Oliveberg M, Wolynes PG. The experimental survey of protein-folding energy landscapes. *Q Rev Biophys* 2005;**38**(3):245–88.
65. Giri Rao VH, Gosavi S. Using the folding landscapes of proteins to understand protein function. *Curr Opin Struct Biol* 2016;**36**:67–74.
66. Levy Y, Onuchic JN. Mechanisms of protein assembly: lessons from minimalist models. *Acc Chem Res* 2006;**39**:135–42.
67. Whitford PC, Onuchic JN. What protein folding teaches us about biological function and molecular machines. *Curr Opin Struct Biol* 2015;**30**:57–62.
68. Henry ER, Eaton WA. Combinatorial modeling of protein folding kinetics: free energy profiles and rates. *Chem Phys* 2004;**307**:163–85.
69. Naganathan AN, Sanchez-Ruiz JM, Munshi S, Suresh S. Are protein folding intermediates the evolutionary consequence of functional constraints? *J Phys Chem B* 2015;**119**:1323–33.
70. Munshi S, Naganathan AN. Imprints of function on the folding landscape: functional role for an intermediate in a conserved eukaryotic binding protein. *Phys Chem Chem Phys* 2015;**17**(16):11042–52.
71. Freiberger MI, Guzovsky AB, Wolynes PG, Parra RG, Ferreiro DU. Local frustration around enzyme active sites. *Proc Natl Acad Sci USA* 2019;**116**(10):4037–43.
72. Jahn TR, Parker MJ, Homans SW, Radford SE. Amyloid formation under physiological conditions proceeds via a native-like folding intermediate. *Nat Struct Mol Biol* 2006;**13**(3):195–201.
73. De Simone A, Dhulesia A, Soldi G, et al. Experimental free energy surfaces reveal the mechanisms of maintenance of protein solubility. *Proc Natl Acad Sci USA* 2011;**108**(52):21057–62.

74. Sengupta I, Udgaonkar J. Monitoring site-specific conformational changes in real-time reveals a misfolding mechanism of the prion protein. *eLife* 2019;**8**:e44698.
75. Sivanandan S, Naganathan AN. A disorder-induced domino-like destabilization mechanism governs the folding and functional dynamics of the repeat protein IκBα. *PLOS Comput Biol* 2013;**9**:e1003403.
76. Aviram HY, Pirchi M, Barak Y, Riven I, Haran G. Two states or not two states: single-molecule folding studies of protein L. *J Chem Phys* 2018;**148**(12):123303.
77. Naganathan AN, Orozco M. The conformational landscape of an intrinsically disordered DNA-binding domain of a transcription regulator. *J Phys Chem B* 2013;**117**:13842–50.
78. Naganathan AN. Modulation of allosteric coupling by mutations: from protein dynamics and packing to altered native ensembles and function. *Curr Opin Struct Biol* 2019;**54**:1–9.
79. Guarnera E, Berezovsky IN. On the perturbation nature of allostery: sites, mutations, and signal modulation. *Curr Opin Struct Biol* 2019;**56**:18–27.
80. Tokuriki N, Tawfik DS. Stability effects of mutations and protein evolvability. *Curr Opin Struct Biol* 2009;**19**(5):596–604.
81. Sikosek T, Chan HS. Biophysics of protein evolution and evolutionary protein biophysics. *J R Soc Interface* 2014;**11**(100):20140419.
82. Risso VA, Sanchez-Ruiz JM, Ozkan SB. Biotechnological and protein-engineering implications of ancestral protein resurrection. *Curr Opin Struct Biol* 2018;**51**:106–15.
83. Petrovic D, Risso VA, Kamerlin SCL, Sanchez-Ruiz JM. Conformational dynamics and enzyme evolution. *J R Soc Interface* 2018;**15**(144).
84. Gupta K, Varadarajan R. Insights into protein structure, stability and function from saturation mutagenesis. *Curr Opin Struct Biol* 2018;**50**:117–25.
85. Bershtein S, Serohijos AW, Shakhnovich EI. Bridging the physical scales in evolutionary biology: from protein sequence space to fitness of organisms and populations. *Curr Opin Struct Biol* 2017;**42**:31–40.
86. Serohijos AWR, Shakhnovich EI. Merging molecular mechanism and evolution: theory and computation at the interface of biophysics and evolutionary population genetics. *Curr Opin Struct Biol* 2014;**26**:84–91.
87. Tokuriki N, Oldfield CJ, Uversky VN, Berezovksy IN, Tawfik DS. Do viral proteins possess unique biophysical features? *Trends Biochem Sci* 2008;**34**:53–9.
88. Horovitz A, Fleisher RC, Mondal T. Double-mutant cycles: new directions and applications. *Curr Opin Struct Biol* 2019;**58**:10–7.
89. Goldsmith M, Tawfik DS. Directed enzyme evolution: beyond the low-hanging fruit. *Curr Opin Struct Biol* 2012;**22**(4):406–12.
90. Rollins NJ, Brock KP, Poelwijk FJ, et al. Inferring protein 3D structure from deep mutation scans. *Nat Genet* 2019;**51**(7):1170–6.
91. Schmiedel JM, Lehner B. Determining protein structures using deep mutagenesis. *Nat Genet* 2019;**51**(7):1177–86.
92. Rajasekaran N, Sekhar A, Naganathan AN. A universal pattern in the percolation and dissipation of protein structural perturbations. *J Phys Chem Lett* 2017;**8**(19):4779–84.
93. Roche J, Caro JA, Dellarole M, et al. Structural, energetic, and dynamic responses of the native state ensemble of staphylococcal nuclease to cavity-creating mutations. *Proteins* 2013;**81**(6):1069–80.
94. Whitley MJ, Zhang J, Lee AL. Hydrophobic core mutations in CI2 globally perturb fast side-chain dynamics similarly without regard to position. *Biochemistry* 2008;**47**(33):8566–76.
95. Ben-David M, Huang H, Sun MGF, et al. Allosteric modulation of binding specificity by alternative packing of protein cores. *J Mol Biol* 2019;**431**(2):336–50.
96. Rajasekaran N, Suresh S, Gopi S, Raman K, Naganathan AN. A general mechanism for the propagation of mutational effects in proteins. *Biochemistry* 2017;**56**:294–305.
97. Haririnia A, Verma R, Purohit N, et al. Mutations in the hydrophobic core of ubiquitin differentially affect its recognition by receptor proteins. *J Mol Biol* 2008;**375**(4):979–96.
98. Bouvignies G, Vallurupalli P, Hansen DF, et al. Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* 2011;**477**(7362):111–4.
99. Sarkar SS, Udgaonkar JB, Krishnamoorthy G. Unfolding of a small protein proceeds via dry and wet globules and a solvated transition state. *Biophys J* 2013;**105**:2392–402.
100. Halskau O, Perez-Jimenez R, Ibarra-Molero B, et al. Large-scale modulation of thermodynamic protein folding barriers linked to electrostatics. *Proc Natl Acad Sci USA* 2008;**105**(25):8625–30.

101. Narayan A, Campos LA, Bhatia S, Fushman D, Naganathan AN. Graded structural polymorphism in a bacterial thermosensor protein. *J Am Chem Soc* 2017;**139**:792–802.
102. Narayan A, Naganathan AN. Switching protein conformational substates by protonation and mutation. *J Phys Chem B* 2018;**122**:11039–47.
103. Bah A, Vernon RM, Siddiqui Z, et al. Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* 2015;**519**(7541):106–U240.
104. Gopi S, Rajasekaran N, Singh A, Ranu S, Naganathan AN. Energetic and topological determinants of a phosphorylation-induced disorder-to-order protein conformational switch. *Phys Chem Chem Phys* 2015;**17**:27264–9.
105. Munshi S, Rajendran D, Naganathan AN. Entropic control of an excited folded-like conformation in a disordered protein ensemble. *J Mol Biol* 2018;**430**(17):2688–94.
106. Uversky VN. What does it mean to be natively unfolded? *Eur J Biochem* 2002;**269**(1):2–12.
107. Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins: Struct Funct Genet* 2000;**41**(3):415–27.
108. van der Lee R, Buljan M, Lang B, et al. Classification of intrinsically disordered regions and proteins. *Chem Rev* 2014;**114**:6589–631.
109. Babu MM, van der Lee R, de Groot NS, Gsponer J. Intrinsically disordered proteins: regulation and disease. *Curr Opin Struct Biol* 2011;**21**:432–40.
110. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nature Rev Mol Cell Biol* 2005;**6**(3):197–208.
111. Bah A, Forman-Kay JD. Modulation of intrinsically disordered protein function by post-translational modifications. *J Biol Chem* 2016;**291**(13):6696–705.
112. Uversky VN. A decade and a half of protein intrinsic disorder: biology still waits for physics. *Protein Sci* 2013;**22**:693–724.
113. Peran I, Holehouse AS, Carrico IS, Pappu RV, Bilsel O, Raleigh DP. Unfolded states under folding conditions accommodate sequence-specific conformational preferences with random coil-like dimensions. *Proc Natl Acad Sci USA* 2019;**116**(25):12301–10.
114. Meng WL, Lyle N, Luan BW, Raleigh DP, Pappu RV. Experiments and simulations show how long-range contacts can form in expanded unfolded proteins with negligible secondary structure. *Proc Natl Acad Sci USA* 2013;**110**(6):2123–8.
115. Meng W, Luan B, Lyle N, Pappu RV, Raleigh DP. The denatured state ensemble contains significant local and long-range structure under native conditions: analysis of the N-terminal domain of ribosomal protein L9. *Biochemistry* 2013;**52**(15):2662–71.
116. Breydo L, Wu JW, Uversky VN. Alpha-synuclein misfolding and Parkinson's disease. *Biochimica et biophysica acta* 2012;**1822**(2):261–85.
117. Uversky VN, Eliezer D. Biophysics of Parkinson's disease: structure and aggregation of alpha-synuclein. *Curr Protein Peptide Sci* 2009;**10**(5):483–99.
118. Abedini A, Raleigh DP. A critical assessment of the role of helical intermediates in amyloid formation by natively unfolded proteins and polypeptides. *Protein Eng Des Sel* 2009;**22**(8):453–9.
119. Uversky VN, Li J, Fink AL. Evidence for a partially folded intermediate in alpha-synuclein fibril formation. *J Biol Chem* 2001;**276**(14):10737–44.
120. Li J, Uversky VN, Fink AL. Conformational behavior of human alpha-synuclein is modulated by familial Parkinson's disease point mutations A30P and A53T. *Neurotoxicology* 2002;**23**(4-5):553–67.
121. Matos CA, Almeida LP, Nobrega C. Proteolytic cleavage of polyglutamine disease-causing proteins: revisiting the toxic fragment hypothesis. *Curr Pharm Des* 2017;**23**(5):753–75.
122. Hoffner G, Djian P. Polyglutamine aggregation in huntington disease: does structure determine toxicity? *Mol Neurobiol* 2015;**52**(3):1297–314.
123. Das RK, Ruff KM, Pappu RV. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr Opin Struct Biol* 2015;**32**:102–12.
124. Das RK, Pappu RV. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc Natl Acad Sci USA* 2013;**110**(33):13392–7.
125. Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc Natl Acad Sci USA* 2010;**107**:8183–8.

126. Keul ND, Oruganty K, Schaper Bergman ET, et al. The entropic force generated by intrinsically disordered segments tunes protein function. *Nature* 2018;**563**(7732):584–8.
127. Motlagh HN, Wrabl JO, Li J, Hilser VJ. The ensemble nature of allostery. *Nature* 2014;**508**(7496):331–9.
128. Hilser VJ, Thompson EB. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proc Natl Acad Sci USA* 2007;**104**(20):8311–5.
129. Rajasekaran N, Gopi S, Narayan A, Naganathan AN. Quantifying protein disorder through measures of excess conformational entropy. *J Phys Chem B* 2016;**120**:4341–50.
130. Paul F, Noe F, Weikl TR. Identifying conformational-selection and induced-fit aspects in the binding-induced folding of PMI from Markov state modeling of atomistic simulations. *J Phys Chem B* 2018;**122**(21):5649–56.
131. Mollica L, Bessa LM, Hanoulle X, Jensen MR, Blackledge M, Schneider R. Binding mechanisms of intrinsically disordered proteins: theory, simulation, and experiment. *Front Mol Biosci* 2016;**3**:52.
132. Vuzman D, Levy Y. Intrinsically disordered regions as affinity tuners in protein-DNA interactions. *Mol Biosystems* 2012;**8**(1):47–57.
133. Marcovitz A, Levy Y. Frustration in protein-DNA binding influences conformational switching and target search kinetics. *Proc Natl Acad Sci USA* 2011;**108**(44):17957–62.
134. Itoh Y, Murata A, Takahashi S, Kamagata K. Intrinsically disordered domain of tumor suppressor p53 facilitates target search by ultrafast transfer between different DNA strands. *Nucleic Acids Res* 2018;**46**(14):7261–9.
135. Subekti DRG, Murata A, Itoh Y, et al. The Disordered Linker in p53 participates in nonspecific binding to and one-dimensional sliding along DNA revealed by single-molecule fluorescence measurements. *Biochemistry* 2017;**56**(32):4134–44.
136. Sherry KP, Das RK, Pappu RV, Barrick D. Control of transcriptional activity by design of charge patterning in the intrinsically disordered RAM region of the Notch receptor. *Proc Natl Acad Sci USA* 2017;**114**(44):E9243–52.
137. Stenzoski NE, Luan B, Holehouse AS, Raleigh DP. The unfolded state of the C-terminal domain of L9 expands at low but not at elevated temperatures. *Biophys J* 2018;**115**(4):655–63.
138. Schuler B, Soranno A, Hofmann H, Nettels D. Single-molecule FRET spectroscopy and the polymer physics of unfolded and intrinsically disordered proteins. *Annu Rev Biophys* 2016;**45**:207–31.
139. Borgia A, Zheng W, Buholzer K, et al. Consistent view of polypeptide chain expansion in chemical denaturants from multiple experimental methods. *J Am Chem Soc* 2016;**138**(36):11714–26.
140. Riback JA, Bowman MA, Zmyslowski AM, et al. Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. *Science* 2017;**358**(6360):238–41.
141. Song J, Gomes GN, Shi T, Gradinaru CC, Chan HS. Conformational heterogeneity and FRET data interpretation for dimensions of unfolded proteins. *Biophys J* 2017;**113**(5):1012–24.
142. Thirumalai D, Samanta HS, Maity H, Reddy G. Universal nature of collapsibility in the context of protein folding and evolution. *Trends Biochem Sci* 2019;**44**(8):675–87.
143. Henriques J, Cragnell C, Skepo M. Molecular dynamics simulations of intrinsically disordered proteins: force field evaluation and comparison with experiment. *J Chem Theory Comput* 2015;**11**(7):3420–31.
144. Narayan A, Bhattacharjee K, Naganathan AN. Thermally versus chemically denatured protein states. *Biochemistry* 2019;**58**(21):2519–23.
145. Davis CM, Gruebele M, Sukenik S. How does solvation in the cell affect protein folding and binding? *Curr Opin Struct Biol* 2018;**48**:23–9.
146. Sukenik S, Salam M, Wang Y, Gruebele M. In-cell titration of small solutes controls protein stability and aggregation. *J Am Chem Soc* 2018;**140**(33):10497–503.
147. Feng R, Gruebele M, Davis CM. Quantifying protein dynamics and stability in a living organism. *Nat Commun* 2019;**10**(1):1179.
148. Guseman AJ, Perez Goncalves GM, Speer SL, Young GB, Pielak GJ. Protein shape modulates crowding effects. *Proc Natl Acad Sci USA* 2018;**115**(43):10965–70.
149. Monteith WB, Cohen RD, Smith AE, Guzman-Cisneros E, Pielak GJ. Quinary structure modulates protein stability in cells. *Proc Natl Acad Sci USA* 2015;**112**(6):1739–42.
150. Smith AE, Zhou LZ, Gorensek AH, Senske M, Pielak GJ. In-cell thermodynamics and a new role for protein surfaces. *Proc Natl Acad Sci USA* 2016;**113**(7):1725–30.

151. Sukenik S, Ren P, Gruebele M. Weak protein-protein interactions in live cells are quantified by cell-volume modulation. *Proc Natl Acad Sci USA* 2017;**114**(26):6776–81.
152. Cohen RD, Pielak GJ. A cell is more than the sum of its (dilute) parts: a brief history of quinary structure. *Protein Sci* 2017;**26**(3):403–13.
153. Cheng K, Wu Q, Zhang Z, Pielak GJ, Liu M, Li C. Crowding and confinement can oppositely affect protein stability. *Chemphyschem* 2018;**19**(24):3350–5.
154. Smith AE, Zhang Z, Pielak GJ, Li C. NMR studies of protein folding and binding in cells and cell-like environments. *Curr Opin Struct Biol* 2015;**30**:7–16.
155. Guseman AJ, Speer SL, Perez Goncalves GM, Pielak GJ. Surface charge modulates protein-protein interactions in physiologically relevant environments. *Biochemistry* 2018;**57**(11):1681–4.
156. Wirth AJ, Platkov M, Gruebele M. Temporal variation of a protein folding energy landscape in the cell. *J Am Chem Soc* 2013;**135**:19215–21.
157. Dhar A, Girdhar K, Singh D, Gelman H, Ebbinghaus S, Gruebele M. Protein stability and folding kinetics in the nucleus and endoplasmic reticulum of eucaryotic cells. *Biophys J* 2011;**101**(2):421–30.
158. Dhar A, Ebbinghaus S, Shen Z, Mishra T, Gruebele M. The diffusion coefficient for PGK folding in eukaryotic cells. *Biophys J* 2010;**99**(9):L69–71.
159. Mu X, Choi S, Lang L, et al. Physicochemical code for quinary protein interactions in *Escherichia coli*. *Proc Natl Acad Sci USA* 2017;**114**(23):E4556–e4563.
160. Munshi S, Gopi S, Asampille G, et al. Tunable order-disorder continuum in protein-DNA interactions. *Nucleic Acids Res* 2018;**46**(17):8700–9.
161. Allred BE, Gebala M, Herschlag D. Determination of ion atmosphere effects on the nucleic acid electrostatic potential and ligand association using AH(+).C wobble formation in double-stranded DNA. *J Am Chem Soc* 2017;**139**(22):7540–8.
162. Borgia A, Kemplen KR, Borgia MB, et al. Transient misfolding dominates multidomain protein folding. *Nat Commun* 2015;**6**:8861.
163. Borgia MB, Borgia A, Best RB, et al. Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. *Nature* 2011;**474**(7353):662–5.
164. Lu W, Schafer NP, Wolynes PG. Energy landscape underlying spontaneous insertion and folding of an alpha-helical transmembrane protein into a bilayer. *Nat Commun* 2018;**9**(1):4949.
165. Kim BL, Schafer NP, Wolynes PG. Predictive energy landscapes for folding alpha-helical transmembrane proteins. *Proc Natl Acad Sci USA* 2014;**111**(30):11031–6.
166. Chavent M, Duncan AL, Sansom MS. Molecular dynamics simulations of membrane proteins and their interactions: from nanoscale to mesoscale. *Curr Opin Struct Biol* 2016;**40**:8–16.
167. Pliotas C, Naismith JH. Spectator no more, the role of the membrane in regulating ion channel function. *Curr Opin Struct Biol* 2017;**45**:59–66.