Review

Beyond the Native State: From Protein Structure to Function Through Energy Landscapes and Conformational Ensembles

Au-delà de l'état natif : de la structure protéique à la fonction à travers les paysages énergétiques et les ensembles conformationnels

Ishaan S. Goswami[1†*], Isra F. Omar[1†]

1. University of Ottawa, Ottawa, ON, Canada
†Co-Authors
*Corresponding author. Email: igosw085@uottawa.ca

Abstract | Résumé

Rather than viewing protein folding as the formation of a single native structure, modern biophysics describes proteins as statistical ensembles of interconverting conformations whose populations are determined by thermodynamics, kinetics, and the energy landscape imposed by molecular interactions. Within this framework, folding is not considered a discrete event, but a dynamic process. When placed in a cellular context, these dynamic conformational ensembles are coupled to protein function and are directly influenced by the intracellular environment and interacting proteins. Protein folding and function still follow biophysical laws, albeit with varying significances, and can be described using kinetic models of intra- and intermolecular interactions. This review explores the biophysical, biochemical, and cellular determinants of protein folding, specifically highlighting that protein behaviour is a property of ensemble dynamics and the intracellular environment.

———

Plutôt que de considérer le repliement des protéines comme la formation d'une structure native unique, la biophysique moderne décrit les protéines comme des ensembles statistiques de conformations interconvertissantes dont les populations sont déterminées par la thermodynamique, la cinétique et le paysage énergétique imposé par les interactions moléculaires. Dans ce cadre, le pliage n'est pas considéré comme un événement discret, mais comme un processus dynamique. Lorsqu'ils sont placés dans un contexte cellulaire, ces ensembles conformationnels dynamiques sont couplés à la fonction des protéines et sont directement influencés par l'environnement intracellulaire et les protéines interagissant. Le repliement et la fonction des protéines suivent toujours les lois biophysiques, bien que leurs significations varient, et peuvent être décrits à l'aide de modèles cinétiques d'interactions intra- et intermoléculaires. Cette revue explore les déterminants biophysiques, biochimiques et cellulaires du repliement des protéines, en soulignant spécifiquement que le comportement des protéines est une propriété de la dynamique d'ensemble et de l'environnement intracellulaire.

Keywords: Protein folding; Energy landscapes; Conformational ensembles; Statistical mechanics; Allostery; Intrinsically disordered proteins; Molecular chaperones; Conformational selection

Introduction

Protein folding is a dynamic process governed by the principles of thermodynamics, kinetics, energetics, and statistical mechanics that is undertaken in unique biochemical and cellular contexts. This allows for the emergence of a dynamic conformation ensemble that is statistically distributed across an underlying energetic landscape, contributing to diverse biochemical functionality. This paradigm is opposed to the notion of protein folding as the deterministic acquisition of a single native structure.

The Biophysical Determinants of Protein Folding

In this section, the determinants of protein folding are highlighted from the angles of thermodynamics, molecular biophysics, and statistical mechanics (1–6).

Classical thermodynamics and the folding problem

The classical thermodynamic hypothesis of protein folding, first articulated by Anfinsen, postulates that the native structure corresponds to the global free-energy minimum under physiological conditions (2, 4–7). Evidence suggests that the "instructions" for folding a polypeptide into a native globular protein structure are contained within its primary sequence. This is known as Anfinsen's dogma, or the potential for self-assembly. However, this is not the full picture, as it portrays protein folding as deterministic, wherein one sequence will invariably produce only one structure through the same folding pathway.

Proteins are not a simple dichotomy between their amino acid sequence and the final native state. Rather, in folding, proteins populate several states, some of which are more favourable than others (2, 4–6).

As proteins can fold spontaneously in physiological conditions, protein folding must be a favourable energetic reaction. As such, it would have a negative Gibbs Free Energy (as defined by the following equation:

$$\Delta G_{folding} = \Delta H_{folding} - T\Delta S_{folding} \quad \text{(Eq. 1)}$$

Where:

$$\Delta G_{folding} < 0 \quad \text{(Eq. 2)}$$

And where enthalpy $(\Delta H_{folding})$ is defined as:

$$\Delta H_{folding} = \sum H_{folded} - \sum H_{unfolded} \quad \text{(Eq. 3)}$$

Gibbs Free Energy is a function of enthalpy $(\Delta H_{folding})$, entropy $(\Delta S_{folding})$, and temperature (T). The enthalpy of folding is due to contributions from intrachain non-covalent bonds, which is an exothermic process. This folding takes place through three mechanisms: charge-charge interactions, internal hydrogen (H)-bonding, and van der Waals interactions. Charge-charge interactions occur between cationic and anionic side chains at physiological pH. Internal H-bonding occurs between H-bond donors, the amide nitrogen of the peptide backbone, and acceptors, the carbonyl oxygen of the backbone. Van der Waals interactions are the result of the dense packing of non-polar groups in the protein core, which increases the strength of dipole-induced dipole interactions. All of these individual interactions provide low contributions, however when all intrachain bonds are summed, a large value of exothermic enthalpy is found (2, 4–6).

The entropy of folding is defined as its conformational entropy. Folding reduces the mobility, or degrees of freedom, of the polypeptide, resulting in negative entropy. However, this is overcome by contributions from solvation, or the entropy of solvation. When a hydrophobic solute is exposed to water, H-bonds in the solvent are broken. To restore the energy lost, water forms clathrates during the dissolution of said hydrophobic solute, which decreases the entropy of the solvent. With protein folding, hydrophobic residues tend to pack to the core of the protein, which minimizes the ordering of water molecules and the disruption of their H-bonds and thus restores degrees of freedom in the solvent. This is known as the hydrophobic collapse, and it can overcome the loss of peptide entropy and favour binding. The hydrophobic effect can harness entropy to create an apparent increase in order by coupling it to a greater increase in disorder among a class of smaller, more numerous objects like water molecules (2, 4–6). Therefore,

$$\Delta S_{folding} = \Delta S_{conformational} + \Delta S_{solvent} \quad \text{(Eq. 4)}$$

Combining the enthalpic and entropic contributions gives:

$$\Delta G_{folding} = \sum H_{folded} - \sum H_{unfolded} - T\left(\Delta S_{conformational} + \Delta S_{solvent}\right) \quad \text{(Eq. 5)}$$

This means that protein folding is the balance of competing energetic terms; it is a thermodynamic balance that allows for an understanding of which potential protein states are favourable. Protein folding is thus not simply proteins seeking the lowest energy, but rather, balancing competing energetic contributions (2, 4–6).

It is important to note that contributions from enthalpy and entropy also result in a small net stability of a protein. This marginal stability allows proteins to be on “the edge” of unfolding, such that they can remain flexible for function (2, 4–6).

Kinetics and folding pathways

Thermodynamics gives an idea of how favourable folding may be. However, even if a reaction or process is favourable, it may be kinetically limited. Therefore, the kinetics of protein folding must also be examined in order to diagnose which states are reachable and on what timescales (2, 4–6).

The most famous thought experiment, or paradox, of protein folding is Levinthal’s Paradox (8). The Phi-Psi (φ-ψ) dihedral angles that dictate the steric orientation between subsequent amino acids have preferred states, as visualized via the Ramachandran plot, to minimize any steric clashes between the side chains of the amino acids. As such, each residue samples a limited number of sterically permitted regions of conformational space. Assuming 10^{15} conformations can be sampled per second, the amount of time to sample all possible conformations of any protein of an arbitrary size would be absurdly large —larger than the lifetime of the universe, let alone the time biological life has existed on Earth. This is because Levinthal’s Paradox makes a false assumption that folding is a random process that requires the sampling of all possible conformations before reaching the final native state. Rather, the working kinetic “pathway” model of protein folding outlines that the primary sequence dictates the native fold which undergoes a hydrophobic collapse into molten globule intermediates in many proteins. The molten globule is a compact, native-like structure with secondary structures defined by H-bonding and backbone topology, but without defined-tertiary native structure. This pathway of protein folding necessarily eliminates the need for the primary structure to exhaustively sample all possible conformations. Instead, it follows a biased stochastic search that will lead to the native fold. The classical pathway model entails an overall decrease in ΔG as the protein folds to its native fold. Each step in the pathway is temporarily “trapped” in its conformational state as the final favourable ΔH is not yet achieved but is overcome via the loss of ΔS with folding (2, 4–6).

This challenges the notion of the two-state folding model as a universal model for protein folding, wherein two states of the protein can exist: the denatured primary sequence and the fully

folded native structure. Thus, even if the native state is thermodynamically favourable, the path taken to it matters. Proteins must traverse an energy landscape, "sampling" different conformations to find the optimal fold (2, 4–6).

Energy landscapes and force fields

The classical thermodynamic and kinetic models of protein folding paint a deterministic model of folding. However, modern biophysics portrays protein folding as a more complex and dynamic process, wherein the "sampling funnel" of classical kinetics is maintained but is not a singular path of gradient descent towards a universal ΔG minimum. Rather, it is a rugged energy landscape with constantly competing interactions dictated by energy force fields. The implication of this is that a sequence does not correspond to a single structure but to an ensemble of conformations with different probabilities (1, 3, 9–11).

Moving from macroscopic thermodynamic descriptions to microscopic conformational probabilities requires turning to statistical mechanics. Force fields are a model of molecular interactions that determine the microscopic energy of a particular protein conformation. These force fields thus determine the shape of the energetic landscape that is sampled during protein folding. Here, bonded interactions such as bond stretching, angle bending, and torsion, non-bonded terms such as van der Waals/Lennard-Jones interactions and electrostatics, and solvent-mediated effects are used to describe the energy of a certain conformation. The bonded terms constrain local geometry, torsion governs backbone and side-chain geometry and flexibility, Lennard-Jones accounts for packing and steric interactions, electrostatics captures charge interactions, and solvation accounts for water exposure and burial. All of these interactions together determine which conformations are low or high in energy (1, 3, 9–11). The following equations describe the molecular mechanics, statistical thermodynamics, and free-energy aspects of protein folding, with Table 1 containing a summary of the symbols and variables used.

$$E(R) = E_{bonded} + E_{nonbonded} + E_{solvation} \quad \text{(Eq. 6)}$$

Where is the vector, or full set of atomic coordinates of the protein.

Expanded out, the energy equation is:

$$E(R) = E_{Bond\ Stretching} + E_{Angle\ Bending} + E_{Dihedral\ rotation} + E_{Lennard-Jones} + E_{electrostatics} + E_{solvation} \quad \text{(Eq. 7)}$$

Or:

$$E(R) = \sum_{bonds\ i} k_{b,i}(b_i - b_{0,i})^2 + \sum_{angles\ i} k_{\theta,i}(\theta_i - \theta_{0,i})^2 + \sum_{dihedrals\ i} k_{\phi,i}[1 + \cos(n_i\phi_i - \delta_i)] + \sum_{i<j} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{i<j} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} + E_{solvation} \quad \text{(Eq. 8)}$$

Where:

$$E_{Bond\ Stretching} = \sum_{bonds\ i} k_{b,i}(b_i - b_{0,i})^2 \quad \text{(Eq. 9)}$$

Which is a harmonic spring approximation, wherein the bonds resist being stretched away from their equilibrium length b_0.

$$E_{Angle\ Bending} = \sum_{angles\ i} k_{\theta,i}(\theta_i - \theta_{0,i})^2 \quad \text{(Eq. 10)}$$

Which is another harmonic restoring force, which enforces the local geometry of angles.

Table 1. Definitions of symbols and variables used in the molecular mechanics, statistical thermodynamics, and free-energy descriptions of protein folding. The notation includes atomic coordinates, force-field parameters, thermodynamic quantities, and reaction-coordinate variables used throughout the derivation

Symbol	Meaning
R	Vector of all atomic coordinates
P	Vector of all atomic momenta
r_{ij}	Distance between atoms i and j
b_i	Length of bond i
$b_{0,i}$	Equilibrium bond length
θ_i	Bond angle
$\theta_{0,i}$	Equilibrium bond angle
ϕ_i	Dihedral angle
$k_{b,i}$	Bond force constant
$k_{\theta,i}$	Angular force constant
$k_{\phi,i}$	Torsional force constant
ε_{ij}	Lennard-Jones well depth
σ_{ij}	Distance at which Lennard-Jones potential is zero
q_i	Charge on atom i
m_i	Mass of atom i
k_B	Boltzmann constant
T	Absolute temperature (K)
Z	Partition function
Q	Folding reaction coordinate
γ	Solvent friction coefficient
$\Gamma(t)$	Random thermal force

$$E_{Dihedral\ rotation} = \sum_{dihedrals\ i} k_{\phi,i}[1 + \cos(n_i\phi_i - \delta_i)] \quad \text{(Eq. 11)}$$

Which determines the allowed backbone φ and ψ conformations and therefore secondary structure. Dihedral rotation describes the energetic cost of rotation around a bond. Rotation is periodic, so the energy repeats in a cosine pattern rather than as a harmonic operation.

$$E_{Lennard-Jones} = \sum_{i<j} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \quad \text{(Eq. 12)}$$

Where the r^{-12} term is the short-range steric repulsion from Pauli's exclusion, and the r^{-6} term is the attractive London dispersion forces, which allow for tight core packing and exclude impossible conformations.

$$E_{electrostatics} = \sum_{i<j} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \quad \text{(Eq. 13)}$$

Which accounts for electrostatic attractive and repulsive forces (1, 3, 9–11).

The force field gives the equation for one exact conformation. The force field itself does not create the "folding funnel." Rather, the set of all force fields for all conformations creates an energy landscape for every point in conformational space:

$$R \mapsto E(R) \quad \text{(Eq. 14)}$$

This mapping creates a remarkably high-dimensional landscape which includes all possible contributions of bond angles, torsions, sidechain rotamers, backbone motions, and so on (1, 3, 9–11).

Hamiltonian and microscopic states

To formally connect these energetic descriptions to observable conformational populations, the framework of classical statistical mechanics is required. In classical mechanics, a complete microscopic state of the protein needs the position of all atoms, R, and the momenta of all atoms, P (1, 3). Thus, in order to represent the total energy of that microstate, the Hamiltonian is required, which is defined as:

$$H(R,P) = T(P) + E(R) \quad \text{(Eq. 15)}$$

Where:

$T(P) = the\ kinetic\ energy\ term$

$E(R) = the\ potential\ energy\ term$

T(P) is defined as:

$$T(P) = \sum_i \frac{p_i^2}{2m_i} \quad \text{(Eq. 16)}$$

Which is the sum of all kinetic energies of all microstates in terms of momentum.

Hence, the expanded Hamiltonian is:

$$H(R,P) = \sum_i \frac{p_i^2}{2m_i} + \sum_{bonds\ i} k_{b,i}(b_i - b_{0,i})^2 + \sum_{angles\ i} k_{\theta,i}(\theta_i - \theta_{0,i})^2 + \sum_{dihedrals\ i} k_{\phi,i}[1 + \cos(n_i\phi_i - \delta_i)] + \sum_{i<j} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_{i<j} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} + E_{solvation} \quad \text{(Eq. 17)}$$

Thus, force fields are not simply a description of protein energetics. Rather, they define the probability distribution over conformational space and determine the ensemble from which function emerges (1, 3).

Langevin dynamics

While the Hamiltonian defines the total energy of a microstate, the actual trajectory of a protein across this landscape is a stochastic process. In a cellular environment, the protein is not an isolated system but is in constant contact with a solvent "heat bath." This motion is described by the Langevin Equation (2):

$$m\frac{d^2R}{dt^2} = -\nabla E(R) - \gamma\frac{dR}{dt} + \Gamma(t) \quad \text{(Eq. 18)}$$

Here, $-\nabla E(R)$ represents the force derived from the molecular force field, $\gamma\frac{dR}{dt}$ accounts for the viscous drag of the solvent, and $\Gamma(t)$ represents the random thermal "kicks" from water molecules. Instead of simple gradient descent down the energetic funnel, the protein undergoes Brownian motion. This stochastic term is mathematically vital; it provides the energy for the protein to escape the local kinetic traps and "rugged" minima of the energy landscape, eventually allowing the system to sample the full conformational ensemble. Without this thermal noise, a protein would get stuck in the first local minimum it reached and would never fold into the native state (2).

This stochastic sampling is temperature-dependent. Below the dynamical "glass" transition (~200 K), the thermal noise, , is insufficient to overcome potential barriers, effectively "freezing" the ensemble into a single microstate and severely restricting conformational dynamics requisite for biological function (9–11).

The Ergodic Hypothesis posits that, given enough time, a single molecule governed by stochastic Langevin dynamics is assumed to visit every microstate in its conformational ensemble with a frequency proportional to that state's Boltzmann weight over sufficiently long timescales. This is the basis of the assumption that a time average of a single protein's trajectory is equivalent to an ensemble average. This allows the partition function to become a physical prediction of how a single protein may structurally fluctuate (9–11).

Boltzmann distribution and partition function

At thermal equilibrium, the probability of a microstate is proportional to its Boltzmann weight (1, 3):

$$P(R,P) \propto e^{\frac{-H(R,P)}{k_B T}} \quad \text{(Eq. 19)}$$

In order to become a true probability, the Boltzmann must be normalized by the partition function:

$$P(R,P) = \frac{e^{\frac{-H(R,P)}{k_B T}}}{Z} \quad \text{(Eq. 20)}$$

Where:

$$Z = \int e^{-\frac{H(R,P)}{k_B T}} dR dP \quad \text{(Eq. 21)}$$

Now, the Hamiltonian may be substituted into the partition function:

$$Z = \int e^{-\frac{[T(P)+E(R)]}{k_B T}} dR dP \quad \text{(Eq. 22)}$$

Via the exponential identity:

$$e^{-\frac{[T(P)+E(R)]}{k_B T}} = e^{\frac{-T(P)}{k_B T}} e^{\frac{-E(R)}{k_B T}} \quad \text{(Eq. 23)}$$

Thus,

$$Z = \int e^{\frac{-T(P)}{k_B T}} e^{\frac{-E(R)}{k_B T}} dR dP \quad \text{(Eq. 24)}$$

$$Z = \left[\int e^{\frac{-T(P)}{k_B T}} dP\right]\left[\int e^{\frac{-E(R)}{k_B T}} dR\right] \quad \text{(Eq. 25)}$$

The integral may be separated since one factor depends only on P and the other only on R.

When integrating over momenta, the Gaussian integral over all momenta is obtained which depends only on masses and temperature, not on the conformation R. This means that when comparing conformations, the whole momentum contribution is just a constant pre-factor. This allows for the simplification wherein P(R, P) may be treated as P(R) for the purposes of protein folding. This separation thus implies that conformational probabilities only depend on the potential energy surface E(R), allowing for reduction from phase space to configurational space (1, 3).

These conformations are populated statistically, according to a Boltzmann distribution, with the force fields giving E(R) (1, 3).

$$P(R) = \frac{e^{-\frac{E(R)}{k_B T}}}{Z_{conf}} \quad \text{(Eq. 26)}$$

With the partition function specific for configurational probability:

$$Z_{conf} = \int e^{-\frac{E(R)}{k_B T}} dR \quad \text{(Eq. 27)}$$

Which allow for the normalization over all possible conformations.

As this follows a Boltzmann distribution, lower-energy conformations are more greatly populated, but are not uniquely occupied. Native-like structures tend to satisfy multiple favorable interactions simultaneously, including hydrophobic burial, hydrogen-bond stabilization, electrostatic complementarity, steric packing, and reduced solvent penalty. Thus, these conformations occupy a lower free energy than most unfolded states and are thus more stable. However, there are still many local minima, so the landscape is rugged, not a smooth absolute global minimum. Consequently, the native state is better understood as a basin of closely related low free-energy conformations instead of a single rigid structure (1, 3).

Projection onto a reaction coordinate

The distribution of conformations is then projected onto a low-dimensional reaction coordinate that may have different ways of measuring folding progress, such as the fraction of native contacts, root-mean-squared distance from native state, or radius of hydration. This creates a distribution of all states on this simplified coordinate diagram (1, 3):

$$F(Q) = -k_B T \ln P(Q) \quad \text{(Eq. 28)}$$

Where:

$Q = a\ folding\ coordinate$

$P(Q) = the\ probability\ of\ finding\ the\ protein\ at\ that\ value\ of\ Q$

$F(Q) = the\ free - energy\ profile\ at\ that\ coodinate$

Because the full configurational space is too high-dimensional to visualize directly, the probability distribution is often projected onto a reduced reaction coordinate (1, 3).

And *P(Q)* is defined as:

$$P(Q) = \sum_{R \in Q} P(R) \quad \text{(Eq. 29)}$$

As moves toward more native-like conformations, tends to decrease on average, accounting for “bumps” from intermediates and kinetic conformational traps. defines the free-energy profile whose topology gives rise to the folding funnel, which is a visualization of the statistical distribution of all conformations. The resulting free-energy landscape can be visualized as a rugged

folding funnel (Figure 1) in which the top of the folding funnel is broad, not because all unfolded states are necessarily of similar energies, but due to the high degeneracy of these structures. The bottom of the funnel has many closely related native-like conformations which are dynamic, not a perfectly rigid structure (1, 3).

Force-field terms generate the microscopic energetic surface over conformational space. When this surface is viewed through the lens of statistical mechanics and projected onto suitable coordinates, it yields these rugged free-energy funnels to describe protein folding (1, 3).

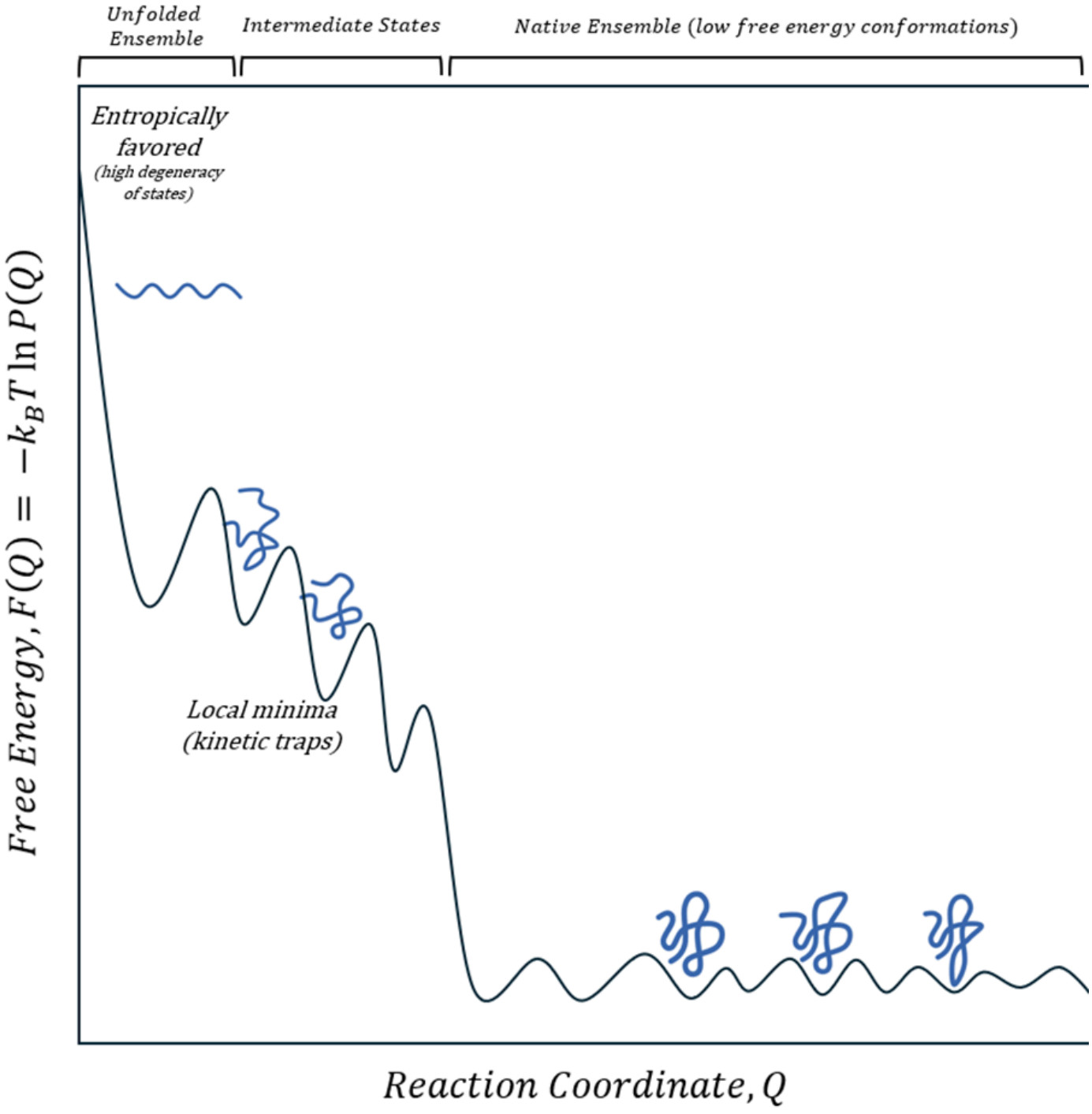

Figure 1. Rugged free-energy landscape of protein folding. The free energy F(Q) is plotted as a function of a reaction coordinate Q describing folding progress. The unfolded ensemble occupies a high-entropy region with a large degeneracy of accessible conformations. Intermediate states correspond to local minima and kinetic traps arising from competing interactions. The native state is represented as a basin of low free-energy conformations, reflecting a dynamic conformational ensemble rather than a single structure. The overall funnel shape indicates a thermodynamic bias toward the native basin, while the ruggedness reflects the complexity of the underlying energy landscape.

Critically, the folding funnel represents a free-energy landscape rather than a pure potential-energy landscape. Although unfolded conformations often possess higher potential energy, they are entropically favored because of their enormous number (1, 3).

$$F(Q) = E(Q) - TS(Q) \quad \text{(Eq. 30)}$$

Thus, the logic of this derivation may be understood as follows:

$$E(R), T(P) \rightarrow H(R,P) = T(P) + E(R) \rightarrow P(R,P) = \frac{e^{\frac{-H(R,P)}{k_B T}}}{Z} \rightarrow P(R) = \frac{e^{-\frac{E(R)}{k_B T}}}{Z_{conf}}$$
$$\rightarrow P(Q) = \sum_{R \in Q} P(R) \rightarrow F(Q) = -k_B T \ln P(Q) \quad \text{(Eq. 31)}$$

This has the profound implication that a sequence does not dictate one singular protein conformation; it dictates a dynamic ensemble of conformations (1, 3).

Proteins as Dynamic Ensembles

Protein function is not encoded in a single idealized native structure, but in the statistical distribution of conformational states and their relative populations within the native landscape, as shown above (1, 3). This may be modeled as:

$$\langle f \rangle = \sum_{I} P_i f_i \quad \text{(Eq. 32)}$$

Where:

$P_i = the\ probability\ of\ state\ i$

$f_i = the\ value\ of\ a\ measurable\ property\ in\ the\ state\ i$

Thus, if one conformation binds a ligand strongly, another binds weakly, and another is completely inactive, the observed behaviour of the protein will be the weighted average of all three. This directly upsets the traditional structure-function dogma, wherein the function is the weighted sum of all constituent components of the ensemble, rather than a clean one-to-one mapping (1, 3).

The Biochemical Determinants of Protein Folding

The biophysical frameworks outlined above provide a quantitative framework for understanding how proteins navigate energy landscapes. However, proteins exist in a cellular context, where folding does not occur in isolation, nor where it acts as an endpoint to ensure biological function. In a biochemical context, protein folding is coupled to function, with conformations modulated by ligands, the intracellular environment, the intrinsic disorder within proteins, and chaperones.

Conformational selection within ensembles

Of the many ways protein activity can be regulated, allostery is particularly relevant in the context of conformational selection. An allosteric effect occurs when the binding of a molecule to a non-orthosteric site (i.e., a site distinct from the primary, active site) triggers a structural shift in the protein and alters the protein's function. Allostery is thus the ability of a protein to transduce signals from an allosteric site to its (often distant) orthosteric site and influence its activity and thus the biological outcome. This can occur in both enzymes and proteins without catalytic ability (12). Two "textbook" models of allostery are frequently referenced: the Monod-Wyman-Changeux (MWC) or concerted model, and the Koshland-Nemethy-Filmer (KNF) or sequential model (Figure 2). The former describes allostery as a cooperative conformational transition of protein oligomers, while the latter describes allostery as progressive conformational transitions of individual/distinct domains within a protein (12–14).

Taking hemoglobin (which has four subunits) binding oxygen as an example, the MWC model assumes there is an equilibrium between two protein states, relaxed (R) and tense (T), where all subunits are simultaneously shifting between the R-state (which has higher affinity for oxygen) or the T-state (which has lower affinity for oxygen). Oxygen preferentially binds R-state subunits, and when binding of oxygen to all four subunits occurs, a shift in the R-T equilibrium causes a concerted conversion of all subunits from the T- to the R-state, creating more favourable binding sites for subsequent oxygen molecules. Allosteric effects are thus a result of an equilibrium-shift between distinct states in the MWC model, where the conformation of each subunit is constrained by its association with the other three subunits (12–14).

On the other hand, the KNF model describes an induced-fit mechanism; ligand binding induces structural changes. It assumes that binding of oxygen to one T-state subunit induces a conformational change only in this subunit, which then shifts the conformation and affinity of its neighbouring subunits. The KNF model therefore includes intermediate states for hemoglobin's structure, with the conformation and binding to each subunit being distinct, yet cooperative (12, 15).

Taking into account dynamics and thermodynamics, however, both the MWC and KNF models for allostery are oversimplified. In 1992, Parsegian et al. identified that the transition from hemoglobin's deoxygenated T-state to its oxygenated R-state was also characterized by the binding of 60 additional water molecules (16). Moreover, within the same year, Arnone et al. determined a third quaternary structure for hemoglobin (i.e., the "R2" state) using X-ray crystallography (17). Although the MWC model, especially, may already hint at conformational selection, both MWC and KNF models assume that protein states are well-defined structures (instead of continuous conformational distributions) and they fail to consider the protein's environment.

Parsegian et al.'s discovery showed that hemoglobin's transition between states is not only a result of the protein's structural rearrangement but also coupled to solvent changes. Of course, the activity of only one solution component cannot be varied with all

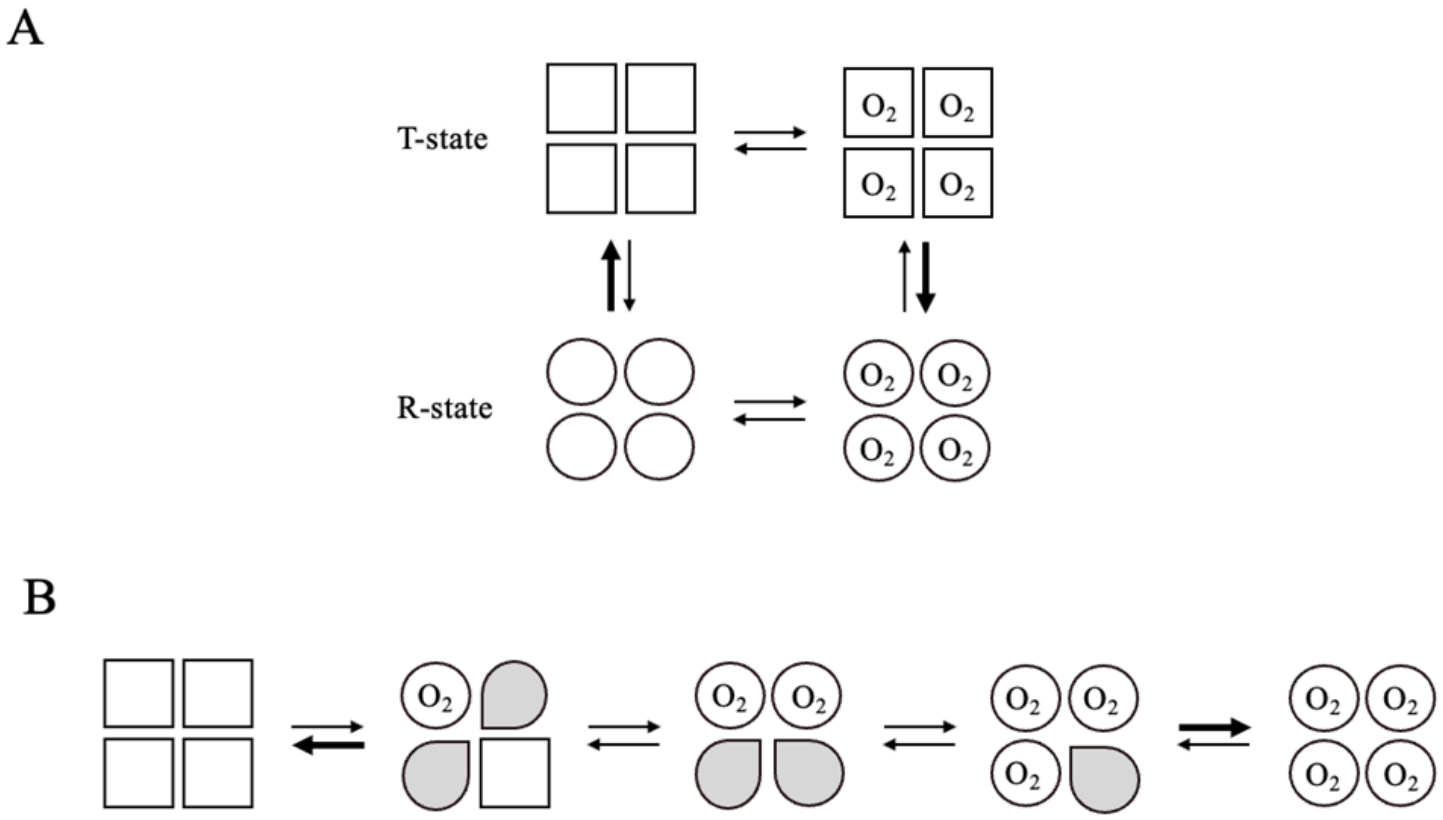

Figure 2. Models of allostery. Both panels use hemoglobin's four subunit structure as an example, with squares representing the tense (T) state, circles the relaxed (R) state, and pointed circles as an intermediate state with higher affinity for oxygen (O_2). **A. Monod-Wyman-Changeux (MWC) model.** In this concerted model, all subunits transition between the T and R states. Without oxygen, the R-T equilibrium favours the T-state. On the other hand, oxygen preferentially binds the R state due to higher affinity. Once oxygen binds to all subunits, the equilibrium shifts to favour the R-state. All subunits are thus either in the T- or R-state conformations, depending on substrate binding. **B. Koshland-Nemethy-Filmer (KNF) model.** In this sequential model, binding of oxygen to one subunit causes a conformational shift and a change in affinity in the neighbouring subunits. This conformational shift is to an intermediate state other than the T- or R-states considered within the MWC model. Panel adapted from Figure 1 in Monstserrat-Canals, Cordara, and Krengel (2025) (12).

others held constant; changing the concentration of one solute/ligand will directly affect the activity and distribution of water too. In a physiological medium, the binding of these 60 water molecules is "osmotic work" and requires 0.2 kcal/mol of energy. Water therefore acts thermodynamically as an allosteric ligand and contributes energetically to a protein's conformational selection and functional regulation (16). Next, Arnone et al.'s discovery redefined the entire "T- versus R-state" depiction of hemoglobin. Intermediate conformations do exist and are energetically accessible structures (17), demonstrating how protein conformations are continuous and how alternative conformations have functional relevance. As such, both findings support an ensemble view for protein conformations.

Because an ensemble view for protein conformations describes proteins as dynamic, heterogenous populations of different conformational states that constantly interconvert (as depicted in Equation 32), ligands are not seen as binding to a single static state. Instead, they interact with a population of states and stabilize specific conformations (this depicts conformational selection) and induce population shifts towards certain states within the ensemble (18, 19). Both conformational changes and binding/unbinding events in proteins require the crossing of free-energy barriers since they are thermally activated processes. However, the transition time needed to cross a free-energy barrier is shorter than the "dwell times" in conformational states before/after energy barrier crossing. This means that transition times are typically poorly resolved in experiments, with conformational changes instead appearing as sudden "jumps" between established conformational states (19).

With this in mind, conformational selection and the KNF model become more similar. In the KNF model, conformational changes occur, or are induced, after a ligand binds to the unbound ground-state conformation of a protein. On the other hand, in conformational selection, the conformational change occurs before ligand binding, and then the ligand will select a given conformation for binding and stabilize it. Nevertheless, a conformational selection-based mechanism (i.e., conformational excitation from a ground-state, low-energy confirmation to a

higher-energy conformation) essentially becomes an induced change-based mechanism (i.e., conformational relaxation from an excited-stated to a lower-energy, ground-state conformation) when binding/unbinding direction is reversed (Figure 3) (19). The MWC and KNF models are thus not invalidated, but redefined and broadened within a higher-dimensional energy landscape view of protein allostery and function, as described by conformational selection.

Functional consequences of conformational ensembles: Intrinsically disordered proteins

Interestingly, function may emerge entirely from ensembles. This is the case for intrinsically disordered proteins (IDPs). IDPs are functional proteins that lack a fixed three-dimensional structure, often characterized instead by their amino acid sequence and simplicity. Although IDPs are unable to fold spontaneously into unique, stable, globular structures, their undefined structures continuously interconvert between a continuum of conformations, and they are still thermodynamically stable (20, 21). This means that their highly entropic disorder (i.e., the equilibrium of their distinct conformations) is lower in free energy than a single conformation or a select set of conformations (21). Unlike many proteins, IDPs satisfy the "disorder-function paradigm," whereby proteins may still perform cellular functions without ever achieving a stable, physiological, three-dimensional structure. It is this conformational flexibility that allows IDPs to be involved in cell signalling, regulation, and binding (22).

IDPs have significant conformational entropy due to their ability to sample many conformations. A large amount of intra- and intermolecular interactions would conversely result in structure and limit sampling, thus restricting conformational entropy. Resultantly, the same biophysical forces that apply to normal proteins also apply to IDPs, but with varying relative importances.

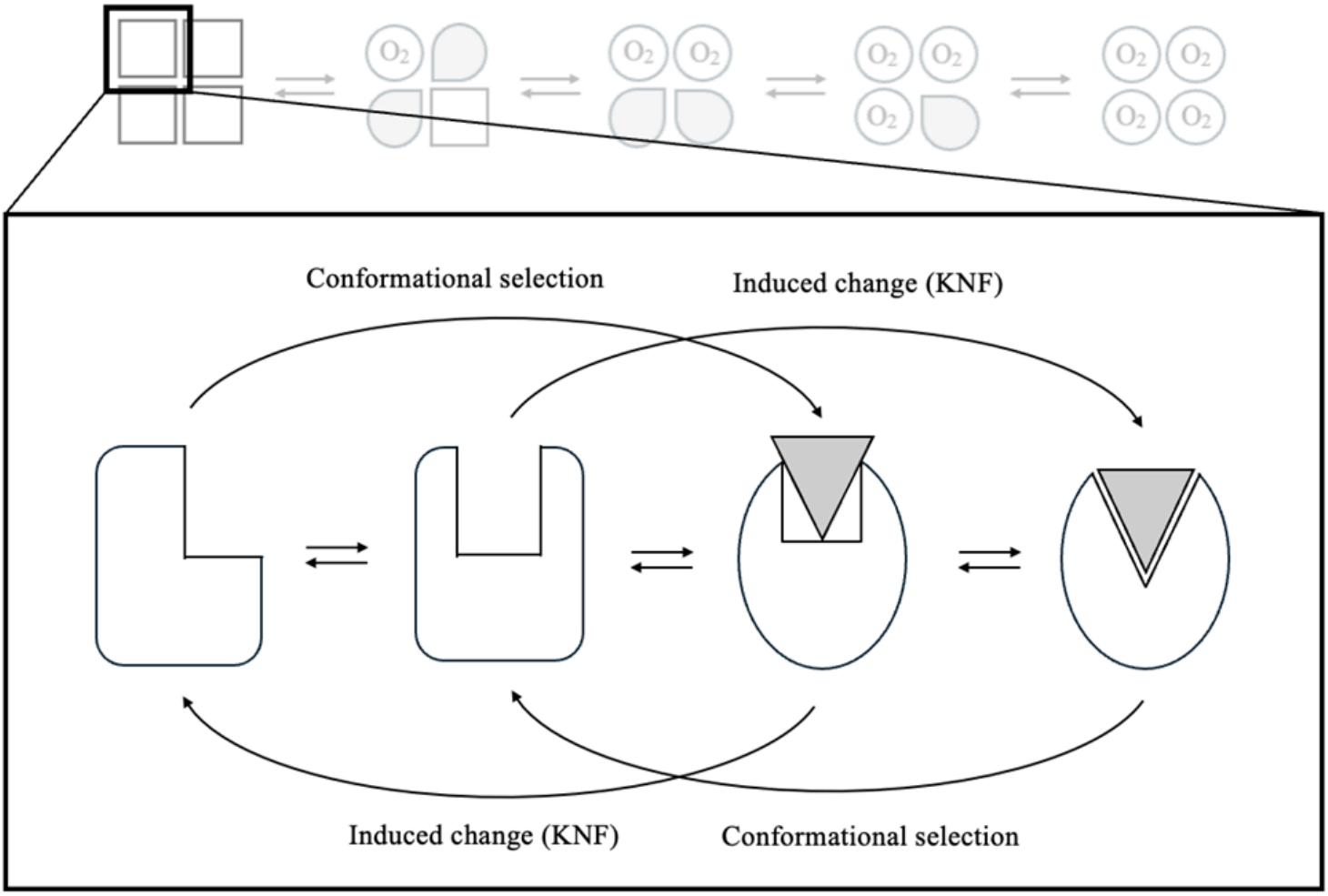

Figure 3. The relationship between the KNF and conformational selection dynamic ensemble models. This figure looks at one T-state subunit (shown in Figure 2's KNF model) in the context of dynamic ensembles as an example. In the dynamic ensembles model, this subunit will continuously interconvert between conformations, which in this figure, are the ground (T) or excited (R) states. Both the KNF and conformational selection models represent mechanisms where ligand binding is coupled to changes in conformation. These models appear equivalent when the reaction is considered in forward versus reverse directions. In the forward direction, induced fit (KNF) occurs when ligand binding precedes and drives the conformational change, while conformational selection occurs when the ligand binds a pre-existing conformation. In the reverse direction, a conformation formed after binding becomes a pre-existing state prior to unbinding, and a ligand-stabilized conformation undergoes conformational change after ligand release.

For example, IDP sequences are rich in charged and polar residues and glycine and proline, but they lack hydrophobic residues. The uncommon presence of hydrophobic residues is instead usually found within motifs that allow IDPs to recognize binding partners. As such, the hydrophobic effect is not a driving force for IDP dynamics, function, or structure. Functionally, this is advantageous since IDPs contain motifs for recognition by enzymes that carry out post-translational modifications, and the disorder and accessibility of these motifs allows IDPs to be rich in post-translational modifications and this, combined with their ability to bind many targets, facilitates IDPs' ability to act as protein hubs (i.e., central nodes within protein-protein interaction networks that accommodate many binding partners) (20, 21).

Furthermore, dynamic conformational sampling also facilitates the averaging of electrostatic fields within IDPs, so their binding or structural properties are more dependent on net charge or charge distributions. Notably, their high proportion of charged and polar residues enhances IDP solubility and even makes it so that some IDPs can function as proteinaceous detergents, depending on their sequence distribution (21, 23).

Moreover, conformational selection within IDPs does not only impact their function, but also their implication in health and disease. IDPs are involved in many signalling pathways, including the regulation of transcription, translation, and the cell cycle (20), and the involvement of disorder within these processes makes it so that IDPs are associated with many diseases, particularly those characterized by a loss of biological regulation (e.g., cancer), and those characterized by the formation of protein aggregates (e.g., Alzheimer's and Parkinson's diseases) (21). These pathological effects may arise from different imbalances within IDPs, with one being missense mutations that result in disorder-to-order transitions (21, 24). In this case, it becomes evident that disorder and a lack of native structure is beneficial for IDPs.

IDPs thereby serve as an example that principles governing protein folding and function are not dependent on a well-defined three-dimensional structure. In the case where no native fold exists, the biophysical principles still exist but are altered in their relative importance and driving force capabilities. Even with following the same dynamic laws but having no set three-dimensional structure, IDPs favour continuous interconversion between a spectrum of conformations within an ensemble, which provides them with their many functions, roles, and pathological risks. IDPs thus represent how biological function is not only encoded in structure, but also in conformational selection.

Non-spontaneous protein folding

Of course, the laws that govern spontaneous protein folding cannot be considered independently without also considering the cellular environment. Protein folding is a balance between thermodynamic stability and conformational flexibility. In the cell, where proteins tend to be marginally stable, this means that protein folding is in constant competition with protein aggregation. As aforementioned, aggregation (and other proteostatic processes) contributes to the development of disease, so molecular chaperones are essential in ensuring proper protein folding and the undertaking of certain native states (25, 26). In vivo, protein folding is complicated by its coupling to translation, the need for many newly translated proteins to be transported into subcellular compartments, and a crowded cellular environment (25).

In vivo folding after synthesis by the ribosome is energetically restricted. The ribosome's exit tunnel prevents large-scale folding and only allows for the formation of smaller tertiary structural elements. This makes it so that C-terminal amino acids cannot participate in more distal interactions essential for cooperative domain folding. Productive protein folding is thus delayed until entire protein domains are translated. This sequential exit and then folding of domains within a protein prevents non-native interactions between simultaneously folding domains, thus preventing the formation of unproductive structural intermediates. However, under stress conditions (e.g., heat shock, oxidative stress), proteins become destabilized (25). As aforementioned, under normal conditions, the free-energy surface that must be navigated by a folding protein is also rugged, creating kinetic traps that can become populated by partially folded states. These trapped intermediates tend to undergo hydrophobic collapse into disorganized globules or become "misfolded states" (if stabilized by non-native interactions), which tend to aggregate in a concentration-dependent manner (26).

Molecular chaperones, proteins that interact with and help in the folding/refolding or assembly of other proteins without being present in the final structure, are thus crucial interactors. Chaperones can be ATP-dependent, like the larger heat shock proteins Hsp70s or Hsp90s, or ATP-independent, like smaller heat shock proteins, depicting how chaperone-assisted folding is not purely equilibrium-driven like conformational ensembles, but active energetic processes (25, 26).

In brief, chaperones recognize non-native protein states after binding to hydrophobic segments. Binding to chaperones prevents aggregation and reduces the amount of freely folding intermediates, although transient release of the hydrophobic regions is necessary for folding to continue. Successful folding is achieved if the rate of folding is greater than the rate of chaperone rebinding (and thus the rate of aggregation). If folding is slower than either process, then the protein may be transferred to a different chaperone system or to degradation machinery (25). Overall, while the dynamics and biophysics of folding are still applicable to proteins in vivo, it can be more accurately described as a kinetically partitioned process, where environmental surroundings and interactors like chaperones bias the folding outcome towards a native state fold.

Conclusion

Both biophysical and biochemical frameworks discussed within this review highlight that protein folding and function cannot be accurately described as transitions between a small set of structures. Instead, they are a product of conformational selection from a continuum of states, which arise from dynamic conformational ensembles, governed by thermodynamics, entropy, and kinetics. When extended to cellular contexts, this further elucidates how biological function is not simple encoded in a single three-dimensional structure, but in population shifts and averages amongst a heterogenous pool of conformations. Overall, when taken together, these biophysical and biochemical insights connect folding and function, making them both consequences of the same continuous exploration process of an energy landscape within conformational ensembles.

Editorial Conflict of Interest Statement

Ishaan S. Goswami and Isra F. Omar are members of the OSURJ editorial team. Both authors were fully recused from all aspects of the editorial process for this manuscript, including reviewer selection, peer review, and final decision-making. The manuscript was handled independently by other members of the editorial board.

References

1. M. L. Boas, Mathematical Methods in the Physical Sciences (Wiley, Hoboken, NJ, 2006).
2. R. Phillips, J. Kondev, J. Theriot, H. G. Garcia, N. Orme, Physical Biology of the Cell (Garland Science, London; New York, NY, 2013), pp. 187-236, 311-354.
3. D. Schroeder, Introduction to Thermal Physics. (Oxford Univ Press, 2020), pp. 220-256.
4. J. Kuriyan, Boyana Konforti, D. Wemmer, The Molecules of Life: Physical and Chemical Principles (Garland Science, Taylor & Francis Group, New York, 2013), pp. 191-238, 220-256.
5. P. C. Nelson, D. S. Goodsell, K. Chen, S. Bromberg, Biological Physics: Energy, Information, Life (Chiliagon Science, Philadelphia, PA, 2020) pp. 184-376.
6. B. Alberts, R. Heald, A. Johnson, D. Morgan, M. Raff, K. Roberts, P. Walter, Molecular Biology of the Cell (W. W. Norton & Company, New York, ed. 7, 2022), pp. 115-182.
7. N. Kresge, R. D. Simoni, R. L. Hill, The Thermodynamic Hypothesis of Protein Folding: the Work of Christian Anfinsen. J. Biol. Chem. 281, e11–e13 (2006). 10.1016/S0021-9258(19)56522-X
8. R. Zwanzig, A. Szabo, B. Bagchi, Levinthal's paradox. Proc. Natl. Acad. Sci. U.S.A. 89, 20–22 (1992).
9. S. S. Plotkin, J. Wang, P. G. Wolynes, Statistical mechanics of a correlated energy landscape model for protein folding funnels. J. Chem. Phys. 106, 2932–2948 (1997).
10. J. D. Bryngelson, J. N. Onuchic, N. D. Socci, P. G. Wolynes, Funnels, pathways, and the energy landscape of protein folding: A synthesis. Proteins. 21, 167–195 (1995).
11. N. Onuchic, H. Nymeyer, A. E. García, J. Chahine, N. D. Socci, "The energy landscape theory of protein folding: Insights into folding mechanisms and scenarios" in Protein folding mechanisms (Academic Press, 2000; vol. 53, pp. 87–152.
12. M. Montserrat-Canals, G. Cordara, U. Krengel, Allostery. Q. Rev. Biophys. 58, e5 (2025). 10.1017/S0033583524000209
13. L. Zhang, M. Li, Z. Liu, A comprehensive ensemble model for comparing the allosteric effect of ordered and disordered proteins. PLoS Comput. Biol. 14, e1006393 (2018). 10.1371/journal.pcbi.1006393
14. E. R. Henry, C. M. Jones, J. Hofrichter, W. A. Eaton, Can a two-state MWC allosteric model explain hemoglobin kinetics? Biochemistry 36, 6511–6528 (1997).
15. T. Yonetani, M. Laberge, Protein dynamics explain the allosteric behaviors of hemoglobin. Biochim. Biophys. Acta Proteins Proteom. 1784, 1146–1158 (2008).
16. M. F. Colombo, D. C. Rau, V. A. Parsegian, Protein solvation in allosteric regulation: A water effect on hemoglobin. Science256, 655–659 (1992).
17. M. M. Silva, P. H. Rogers, A. Arnone, A third quaternary structure of human hemoglobin A at 1.7-Å resolution. Journal of Biological Chemistry267, 17248–17256 (1992).
18. H. Frauenfelder, S. G. Sligar, P. G. Wolynes, The energy landscapes and motions of proteins. Science (1979). 254, 1598–1603 (1991).
19. T. R. Weikl, F. Paul, Conformational selection in protein binding and function. Protein Science 23, 1508–1518 (2014).
20. P. E. Wright, H. J. Dyson, Intrinsically disordered proteins in cellular signalling and regulation. Nat. Rev. Mol. Cell Biol. 16, 18–29 (2015).
21. J. D. Forman-Kay, T. Mittag, From sequence and forces to structure, function, and evolution of intrinsically disordered proteins. Structure 21, 1492–1499 (2013).
22. R. Trivedi, H. A. Nagarajaram, Intrinsically Disordered Proteins: An Overview. Int. J. Mol. Sci. 23, e14050 (2022). 10.3390/ijms232214050
23. R. W. Bailey, A. K. Dunker, C. J. Brown, E. C. Garner, M. D. Griswold, Clusterin, a binding protein with a molten globule-like region. Biochemistry40, 11828–11840 (2001).
24. V. Vacic, P. R. L. Markwick, C. J. Oldfield, X. Zhao, C. Haynes, V. N. Uversky, L. M. Iakoucheva, Disease-Associated Mutations Disrupt Functionally Important Regions of Intrinsic Protein Disorder. PLoS Comput. Biol. 8, e1002709 (2012). 10.1371/journal.pcbi.1002709
25. Y. E. Kim, M. S. Hipp, A. Bracher, M. Hayer-Hartl, F. Ulrich Hartl, Molecular chaperone functions in protein folding and proteostasis. Annu. Rev. Biochem.82, 323–355 (2013).
26. F. U. Hartl, A. Bracher, M. Hayer-Hartl, Molecular chaperones in protein folding and proteostasis. Nature 475, 324–332 (2011).