

**Statistical Physics of Biological Self-Organization:
Case Studies in Protein Folding and Adaptation to
Time-Varying Fields**

by

Nikolay Perunov

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2016

© Massachusetts Institute of Technology 2016. All rights reserved.


Author

Signature redacted

Department of Physics
January 8, 2016

Certified by

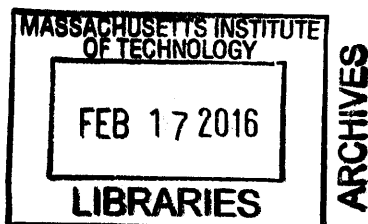
Signature redacted


Jeremy England
Assistant Professor
Thesis Supervisor

Accepted by

Signature redacted

Nergis Mavalvala
Associate Department Head for Education



Statistical Physics of Biological Self-Organization: Case Studies in Protein Folding and Adaptation to Time-Varying Fields

by

Nikolay Perunov

Submitted to the Department of Physics
on January 8, 2016, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Physics

Abstract

In this thesis, we use the methods of statistical physics to provide quantitative insights into the behavior of biological systems. In the first half of the thesis, we use equilibrium statistical physics to develop a phenomenological model of how the hydrophobic effect impacts the structure of proteins, and in the second half, we study the phenomenon of adaptation and Darwinian selection from the standpoint of non-equilibrium statistical physics.

It has been known for a long time that the hydrophobic effect plays a major role in driving protein folding. However, it has been challenging to translate this understanding into a predictive, quantitative theory of how the full pattern of sequence hydrophobicity in a protein helps to determine its structure. Here, we develop and apply a phenomenological theory of the sequence-structure relationship in globular protein domains. In an effort to optimize parameters for the model, we first analyze the patterns of backbone burial found in single-domain crystal structures and discover that classic hydrophobicity scales derived from bulk physicochemical properties of amino acids are already nearly optimal for prediction of burial using the model. Subsequently, we apply the model to studying structural fluctuations in proteins and establish a means of identifying ligand-binding and protein-protein interaction sites using this approach.

In the second half of the thesis, we undertake to address the question of adaptation from the standpoint of physics. Building on past fundamental results in non-equilibrium statistical mechanics, we demonstrate a generalization of the Helmholtz free energy for the finite-time stochastic evolution of driven Newtonian matter. By analyzing this expression, we show a general tendency in a broad class of driven many-particle systems toward self-organization into states formed through reliable absorption and dissipation of work energy from the surrounding environment. We demonstrate how this tendency plays out in the familiar example of Darwinian competition between two exponentially growing self-replicators. Subsequently, we illus-

trate the more general mechanism by which extra dissipation drives adaptation by analyzing the process of random hopping in driven energy landscapes.

Thesis Supervisor: Jeremy England
Title: Assistant Professor

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction to protein structure | 15 |
| 1.1 | The building blocks | 15 |
| 1.2 | Physical interactions that determine the properties of proteins | 17 |
| 1.3 | Anfinsen's experiments and thermodynamic principle | 20 |
| 1.4 | Hydrophobicity at small and large length scales | 22 |
| 1.5 | The current state in protein structure prediction | 27 |
| 2 | Quantitative theory of hydrophobic effect as a driving force of protein structure | 33 |
| 2.1 | Introduction | 33 |
| 2.2 | Results | 35 |
| 2.2.1 | Burial mode model | 35 |
| 2.2.2 | Parameter optimization | 38 |
| 2.2.3 | Sequence diversity in globins | 43 |
| 2.2.4 | Binding and mutation as triggers of conformational change . . | 45 |
| 2.3 | Discussion | 50 |
| 2.4 | Materials and methods | 54 |
| 3 | Overview of non-equilibrium statistical mechanics | 57 |
| 3.1 | Foundations of non-equilibrium statistical mechanics | 58 |
| 3.2 | Microscopic reversibility relation | 62 |
| 3.3 | Fluctuation theorems | 68 |
| 3.4 | Macroscopic irreversibility | 73 |

| | | |
|----------|---|------------|
| 3.5 | Generalization of Helmholtz free energy | 75 |
| 4 | Thermodynamics of adaptation | 81 |
| 4.1 | Toy model of self-replication | 81 |
| 4.2 | Drift and diffusion in driven energy landscapes | 85 |
| 4.3 | Discussion | 91 |
| 5 | Outlook | 95 |
| 5.1 | Phenomenological models of protein folding | 95 |
| 5.2 | Thermodynamics of self-replication | 96 |
| A | Burial Mode Model (BMM) | 99 |
| A.1 | Basic Assumptions and Approximations of BMM | 99 |
| A.2 | BMM Model Parameters | 101 |
| A.3 | Brute Force Search of 4-letter Hydrophobicity Scale | 102 |
| A.4 | Conformational changes (statistical significance) | 104 |
| A.5 | Comparison of BMM and Regression Models | 107 |
| B | Thermodynamics of self-replication | 111 |
| B.1 | Chemical reaction $X \rightleftharpoons 2X$ | 111 |
| | B.1.1 Deterministic solution and steady-state distribution | 112 |
| | B.1.2 Forward probability | 112 |
| | B.1.3 Return probability | 113 |
| | B.1.4 Entropy production | 115 |
| B.2 | Two chemical reactions $A \rightleftharpoons 2A, B \rightleftharpoons 2B$ | 116 |
| | B.2.1 Surface of fixed return probability | 116 |

List of Figures

| | | |
|-----|---|----|
| 1-1 | Three amino acids joined by two peptide bonds form a polypeptide. Peptide bonds are shown in red. ϕ, ψ , and ω are torsion and improper torsion angles. | 16 |
| 1-2 | Solvation free energy, ΔG , for a spherical hydrophobic particle as a function of particle size, R . For large particles, the solvation free energy per unit area approaches the liquid-vapour surface tension, γ | 25 |
| 2-1 | Basic assumptions of the burial mode model. A. The protein backbone is represented as a linear chain (solid red line) with residues indexed by the number s and that have position $\vec{r}(s)$ relative to the center of the globule. The solid black line shows the maximum size of the globule, while the black dashed line shows the radius of gyration R_g . The hydrophathy of each residue $\varphi(s)$ is determined by the type of the residue. Neighbouring residues are connected by harmonic springs of stiffness κ . Blue and red residues represent hydrophilic and hydrophobic amino acids, respectively. The plot in the bottom right corner shows the contribution of different residues to the system energy as a function of the distance to the center of the globule. B. Burial traces computed using the model (blue lines) and from the crystal structures (red lines) of sperm whale myoglobin (1BZP) and sialoadhesin (1OD7). The PCC between the model and the structure is 0.6 for myoglobin and -0.1 for sioloadhesin. | 36 |

| | | |
|-----|---|----|
| 2-2 | Extracting a model hydrophobicity scale from a set of proteins with known structures. | 41 |
|-----|---|----|

2-2 (Continued)

A. For a given protein, one can compute the burial trace (right panel) corresponding to its 3D structure (left panel). Then, one can count how many times a residue of type i (leucine (L) in the figure) is closer to the center of the globule than residue of type j (lysine (K) in the figure) given that they are the nearest neighbors on the chain.

B. Repeating the procedure described above for all proteins from the set, one can compute the matrix of relative positions M_{ij} (left panel). On the right, comparison of the hydrophobicity scale (α -rpm) calculated from the matrix of relative positions M_{ij} with Kyte-Doolittle and Wimley-White hydrophobicity scales. The matrix M_{ij} was constructed using α domains, with unique sequences of length between 100 and 300 a.a. from the SCOP database (970, in total). To compute this matrix, we used only the residues that are far from the center of a domain ($|\vec{r}(s)|^2 > 0.5R^2$).

C. Distribution of PCC between the burial traces predicted by the model using KD and α -rpm scales and the burial traces computed from the crystal structures for α -helical and β -stranded proteins from SCOP.

2-3 Inter-domain interaction in hemoglobin.

A. Distribution of PCC between the burial traces predicted from the sequence using the KD scale and the burial traces computed from the crystal structures for the family of globins (SCOP ID a.1.1.2).

B. Burial traces of α and β chains of hemoglobin (1Y4V) computed from crystal structures (black lines) and using the model (red and green lines). Gray bars correspond to inter-domain contacts, which were determined by the distance between C_α atoms with the threshold 6.5 Å.

C. PCC between the burial traces extracted from crystal structures of α and β chains of hemoglobin (1Y4V) and the burial traces computed using the model when one of the residues is pinned to the surface of the globule. Thin solid black lines correspond to the same procedure for the random sequence. The dashed horizontal lines correspond to PCC without pinning (0.28 for α -chain, and 0.10 for β -chain), whereas solid black lines correspond to the random sequence.

D. Hydrophobicity profiles of myoglobin (blue line) and hemoglobin (red and green lines) calculated using a sliding window of 10 residues. 46

2-4 Conformational changes in sperm whale myoglobin (1BZP), H-Ras (3K8Y), and chymotrypsinogen (1PYT, D).

A. On the bottom panel, the solid black line corresponds to the burial trace of myoglobin computed from the crystal structure, while red lines correspond the burial traces of low-energy excited states ($\Delta E = 4 k_B T$). On the top panel, structural variability $\text{var}[r^2(s)]$ is computed from these burial traces. The gray bars on both sub-plots correspond to heme-binding sites (residues 65 and 94).

B. The crystal structure of myoglobin is colored according to the structural variability $\text{var}[r^2(s)]$. A heme molecule is shown in red.

C. Conformational changes in H-Ras (3K8Y). On the top, burial traces of low-energy excited states of H-Ras are depicted. On the bottom, the structural variability is both plotted and colored on the crystal structure for H-Ras, as computed for burial traces of $\Delta E = 4 k_B T$. GTP binding sites are shown as gray bars, while GTP is shown in red. . . . 48

2-4 (Continued)

D. Structural variability of chymotrypsinogen (1PYT, D). Here, green lines correspond to the burial traces and structural variability computed for the uniprot sequence (before the signal peptide of chymotrypsinogen is cut), while red lines were computed for the chymotrypsinogen sequence taken from the PDB file (before the activation peptide is cleaved). Catalytic sites (H74, D121, and S216), signal and activation peptides are shown in gray.

On all subplots, the structural variability $\text{var}[r(s)]$ is shown in arbitrary units. 49

| | | |
|-----|---|----|
| 2-5 | Conformational change triggered by mutation. | |
| | A. 3D structures of the 2KDM and 2KDL proteins show that mutation L45Y leads to the transformation of a $3\text{-}\alpha$ fold into a $4\beta+\alpha$ fold. Structural variability, plotted in red and green, was computed from the burial traces of the low-energy excited states ($\Delta E = 4 k_B T$). The positions of other mutations are shown as the gray bars on the plot. | |
| | B. Response matrices $\delta r^2(s)/\delta\varphi(s')$ of the 2KDM and 2KDL proteins. The plots on the bottom were obtained by taking the sum of the absolute values along the rows of the response matrices. In both proteins, the residues near the termini and residues 43-47 are the most sensitive to changes in amino acid hydrophobicity. | 51 |
| 3-1 | An example of forward and reversed trajectories for a discrete time Markov process. | 67 |
| 3-2 | An example of periodic driving protocol. This protocol is symmetric between t_a and any time separated by an integer number of periods $t_a + nT$ | 72 |
| 4-1 | Variation of total entropy production on the surfaces of fixed return probability. $g_A = 1$, $g_B = 2$, $\delta = e^{-60}$, $\tau = 10$. $N = e^{20}$ | 84 |
| 4-2 | Three microstates in the absence of drive have the same energy $E = 0$, and the heights of the barriers between adjacent microstates is $B_{ij} = \Delta E$. In the presence of the drive, the barrier and state energies are varied in such a way that the transition rates $r_{3\rightarrow 2}$, $r_{2\rightarrow 3}$, and $r_{1\rightarrow 2}$ are not changed, while the transition rate from state 2 to 1 is $r_{2\rightarrow 1} = r \exp[-\beta\Delta E(1 - \cos[\omega t]/2)]$ | 87 |
| 4-3 | An example of system where the drift is suppressed by fluctuations Φ . | 89 |

| | | |
|-----|---|-----|
| A-1 | The mean and standard deviation of the distributions of PCC for 200 α -helical proteins when amino acids are divided into groups according to KD scale. The model number is computed as follows: $m = ind_1 + N_1 \cdot ind_2 + (N_1 N_2) \cdot ind_3 + (N_1 N_2 N_3) \cdot ind_4$, where N_i is the number of values along i -th axis. | 102 |
| A-2 | The mean and the standard deviation of the distributions of PCC when amino acids are divided into random groups. | 102 |
| A-3 | 4-letter hydrophobicity scales that provide the best and worst performance of the model on a large group of proteins, when amino acids are divided into groups according to their hydrophobicity indices in the KD scale. | 103 |
| A-4 | 4-letter hydrophobicity scales that provide the best and worst performance of the model on large group of proteins, when amino acids were divided into random groups. | 103 |
| A-5 | Comparison of performance of the LRM and BMM for a group of myoglobin (SCOP unid: 46469). Red points on the rightmost panel correspond to the training set (20 proteins). | 108 |
| A-6 | Comparison of performance of the LRM and BMM for a group of hemoglobin A (SCOP unid: 46486). Red points on the rightmost panel correspond to the training set (20 proteins). | 108 |
| A-7 | Comparison of performance of the LRM and BMM for a group of hemoglobin B (SCOP unid: 46500). Red points on the rightmost panel correspond to the training set (20 proteins). | 109 |
| A-8 | Comparison of performance of the LRM and BMM for a group of immunoglobulins (SCOP: b.1.1.1). Red points on the rightmost panel correspond to the training set (20 proteins). | 109 |
| B-1 | A sketch of a monotonic destruction path, $\mathcal{P}(t_4, t_3, t_2, t_1)$ | 114 |

List of Tables

| | | |
|-----|--|-----|
| 1.1 | Summary of interactions in biological systems. | 18 |
| 2.1 | Comparison of the model performance with different hydrophobicity scales for different classes of proteins from the SCOP database. Each column shows the mean and the standard deviation of the distributions of PCC between the burial traces computed from sequences using different hydrophobicity scales and the burial traces extracted from protein structures for a given SCOP class. | 39 |
| A.1 | Myoglobin (1BZP). $N = 153$. $dE = 5 k_B T$. Binding sites are determined using a distance cutoff. | 105 |
| A.2 | Myoglobin (1BZP). $N = 153$. $dE = 5 k_B T$. Binding sites are determined to be nearest neighbors to the residues, located closer than 3 \AA to the ligand. | 105 |
| A.3 | H-Ras (3K8Y). $N = 166$. $dE = 5 k_B T$. Binding sites are determined using a distance cutoff. | 105 |
| A.4 | H-Ras (3K8Y). $N = 166$. $dE = 5 k_B T$. Binding sites are determined to be the nearest neighbors to the residues, located closer than 3 \AA to the ligand. | 106 |
| A.5 | H-Ras (3K8Y). $N = 166$. $dE = 5 k_B T$. Binding sites are taken from UNIPROT. | 106 |
| A.6 | Chymotrypsinogen (1PYT). $N = 251$. $dE = 5 k_B T$. Binding sites are taken from UNIPROT. | 106 |

Chapter 1

Introduction to protein structure

Proteins are molecules of life — they are the building blocks of every living organism, and they participate in nearly every biochemical process, starting with regulation of gene expression inside one cell and ending with transduction of signals between organs. The functions performed by proteins depend on their shape and ability to change shape during interaction with the environment. Therefore, in order to design any drug or enzyme to promote or inhibit a specific reaction, it is crucial to understand why proteins fold in particular shapes. Over the past thirty years, the central question in the study of the protein folding has been the following: How does the amino acid sequence determine the three-dimensional structure and, therefore, function of a protein?

To answer this question, a large number of experimental, computational, and theoretical methods has been developed. In this chapter, we are going to review these methods and build the foundation for introducing the phenomenological models of protein folding presented in later chapters.

1.1 The building blocks

Despite the vast range of biological functions, proteins are a relatively homogeneous class of macromolecules. All proteins consist of one or more linear polymers built of various combinations of 20 amino acids. Figure 1-1 illustrates how amino acids

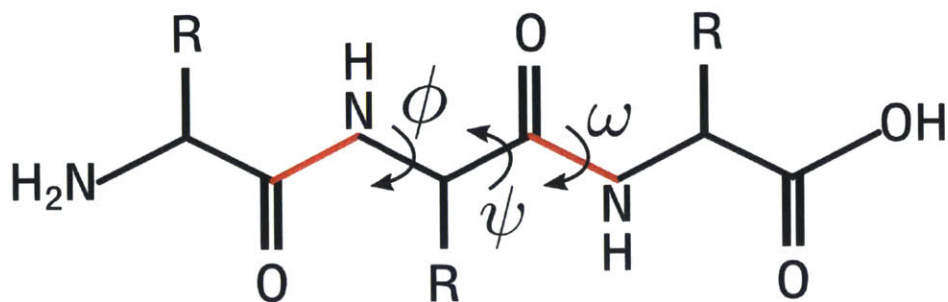


Figure 1-1: Three amino acids joined by two peptide bonds form a polypeptide. Peptide bonds are shown in red. ϕ , ψ , and ω are torsion and improper torsion angles.

are linked by peptide bonds to form a polypeptide chain. The particular sequence of amino acids in the chain is called the primary structure, the different parts of which might form local regular secondary structures, such as alpha helices and beta sheets. The overall shape of a single chain is called a tertiary structure and is formed by packing together secondary structure elements. Commonly, proteins contain more than one linear chain, in which case different polypeptide chains are arranged in the quaternary structure. The functional properties of a protein depend on its three-dimensional shape and amino acid sequence. Therefore, the functional diversity of proteins comes simply from the diversity of chemical properties of the amino acids and the astronomical number of primary sequences.

The first three-dimensional structures of proteins were obtained using X-ray crystallography in 1958 [1]. In 1984, the nuclear magnetic resonance (NMR) framework for protein structure determination, which does not require protein crystallization, was developed [2, 3]. Since then, the three-dimensional structures for thousands of proteins have been determined. This has allowed for classification of proteins by their structures and functions [4, 5]. Although X-ray crystallography and NMR spectroscopy significantly improved our understanding of the relationship between protein structure and function, these methods do not explain how the structure of the protein arises from its sequence. To get closer to answering this question, one first needs to understand the nature and variety of interactions in biological systems.

1.2 Physical interactions that determine the properties of proteins

Even though the majority of forces in living systems are manifestations of electromagnetic interaction, it is often convenient to group these forces by their strengths and by the types and charges of interacting particles. Table 1.1 shows that the strength of interactions also varies considerably depending on their nature. Because the energies of quantum mechanical bound states, such as covalent bonds and bond angles, are much greater in magnitude than the energies of other interactions, including thermal fluctuations, one can approximate interactions of quantum mechanical nature by classical potential energy:

- Covalent bonds are formed when two atoms share a pair of electrons. The energy of breaking of covalent bonds is around $30 k_B T$. The latter allows one to assume that the lengths of these bonds is close to optimal and to describe the energy cost of stretching the bond by classical harmonic potential.
- When three atoms are connected by two covalent bonds, the angle between these bonds tends to be close to the value corresponding to the ground state of two atoms. The energy required to significantly change the bond angle is of the same order of magnitude as the energy of breaking a covalent bond. Therefore, in proteins, one can take into account only small variations in the bond angle around optimal value, “bond bending”, which for practical purposes can be described by classical potential.
- When four atoms are connected by three covalent bonds, two side bonds prefer to be in certain orientation relative to each other. This orientation is described by torsion angle, the angle between two half planes that contain side chains and have middle covalent bond in common. Figure 1-1 shows the definition of torsion angles, ϕ and ψ , and of an improper torsion angle, ω , for a polypeptide backbone. The improper torsion angle is close to 180° in all proteins, and small deviations from this value can be characterized by harmonic potential. However, the ϕ and ψ torsion angle interactions are different from bond stretching and

| Interaction | Nature | Range, Å | Strength, $k_B T$ |
|----------------|---------------|-------------------------|-------------------|
| Covalent bonds | quantum | 1 – 2 | 30 – 100 |
| Torsion angle | quantum | NA | 5 – 10 |
| Electrostatic | entropic | solvent dependent | 3 – 5 |
| Hydrogen bonds | electrostatic | < 4 | 3 – 10 |
| Van der Waals | quantum | < 10 | 0.3 – 1 |
| Hydrophobic | entropic | particle size dependent | ~ 1 |

Table 1.1: Summary of interactions in biological systems.

bending interactions in two ways: First, they are periodic in torsion angle, and, second, more importantly, the rotation barriers are only about $3 k_B T$, which is comparable to the energy of thermal motion. Therefore, rotations around bonds are treated as Fourier series of lengths, providing a good agreement with experimental torsion surfaces.

To conclude, the degrees of freedom associated with bond lengths and bond angles are effectively frozen at room temperature, so the local flexibility and structural variation of the protein backbone are due to the ϕ , ψ torsion angles.

Because strong quantum mechanical interactions determine the local geometry of a protein, it is the cooperativity of weaker interactions, such as hydrogen bonds and van der Waals interaction, that control the structure of the entire protein. It is practical to group weaker forces by the types and charges of interacting particles:

- The long-range electrostatic interactions between two ions, which can be on the order of $100 k_B T$ in a vacuum, are screened by counter-ions in physiological solvents and should be taken into account at distances of less than several Debye length, $\lambda_D = 1/\sqrt{\lambda_B n_{ions}}$, where $\lambda_B = e^2/(\epsilon k_B T) \approx 7 \text{ Å}$ is Bjerrum length of the solvent, and n_{ions} is the concentration of the counter-ions.
- Hydrogen bonds form as a result of competition of two electronegative atoms for the same hydrogen atom. That is, the main component of a hydrogen bond is an electrostatic interaction between the dipole of donor-hydrogen pair (D^-H^+) and electronegative acceptor atom (A^-). In proteins, hydrogen bonds most frequently form between carboxyl CO and ammonium NH groups. Typ-

ically, the distance between nitrogen and oxygen atoms in a hydrogen bond ($\text{N-H}\dots\text{O} = C$) is about 3 Å, and the energy of this bond is about $3 k_B T$. Although these bonds are relatively weak, they can provide extra stability for the protein structure. For example, it has experimentally shown that the salt bridges — hydrogen bonds between anionic carboxylate (RCOO^-) of aspartic acid (Asp) or glutamic acid (Glu) and the cationic ammonium (RNH_3^+) of lysine (Lys) or guanidinium ($\text{RNHC}(\text{NH}_2)_2^+$) of arginine (Arg) — stabilize the native structure of lysozyme [6].

- Van der Waals interaction is the result of correlations in fluctuations of polarization of non-charged particles. This interaction is relatively weak (about $1 k_B T$), but it cannot be ignored because van der Waals forces are acting between all nearby particles, and the energies of these interactions might add up to a significant contribution to protein energy.

One of the most popular functional forms of effective potential energy of a protein, which combines contributions of all interactions, is given by [7, 8, 9]:

$$\begin{aligned}
 U = & \sum_{\text{all bonds}} \kappa_b (b - b_0)^2 + \sum_{\text{all angles}} \kappa_\theta (\theta - \theta_0)^2 + \sum_{\text{Urey-Bradley}} \kappa_{UB} (u - u_0)^2 \\
 & + \sum_{\text{all torsion angles}} \kappa_{\phi,n} [1 + \cos(n\phi - \gamma_n)] + \sum_{\text{all improper angles}} \kappa_\omega (\omega - \omega_0)^2 + \\
 & + \sum_{\text{non-bonded atoms}} \epsilon \left[\left(\frac{r_0}{r} \right)^{12} - \left(\frac{r_0}{r} \right)^6 \right] + \sum_{\text{all charges}} q_i q_j / r \quad (1.1)
 \end{aligned}$$

where κ_b , κ_θ , κ_{UB} , $\kappa_{\phi,n}$, and κ_ω are the bond, angle, Urey-Bradley¹, torsion angle, and improper torsion angle force constants respectively; b , θ , u , ϕ , and ω are the bond length, angle, Urey-Bradley distance, torsion angles and improper torsion angle, respectively; and the zero subscript corresponds to optimal values of distances and angles. The last two terms in this equation describe van der Waals interaction calculated with 12-6 Lennard-Jones potential and the electrostatic energy with a Coulomb

¹Urey-Bradley interaction is a harmonic term in distance between atoms 1 and 3 separated by two covalent bonds (1–2–3). Not all force fields consider Urey-Bradley interaction because it is a correction to bond bending and stretching terms.

potential. The interaction parameters of the energy function (1.1) are usually determined by fitting interaction energies and geometries of protein building blocks to experimental gas-phase geometries of these blocks, vibrational spectra, and torsion energy surfaces [10]. Importantly, the sums in the last two terms of energy function (1.1) include both intrachain and chain-solvent interactions.

Exact treatment of chain-solvent interaction is difficult because of the dependence of induced charges on each other and the complex organization of solvent molecules. For example, the molecules of the typical biological solvent — water — tend to form an extensive hydrogen bond network. Disruption of this network by protein backbone and side chains leads to local rearrangement of water molecules, which is associated with a change in free energy. This effect lies in the origin of hydrophobic interaction — the tendency of non-polar molecules to interact more strongly with each other rather than water molecules. The contribution of different interactions to potential energy of proteins has been studied both numerically and experimentally. The most influential experiments were done by Anfinsen in early 1960s. In the next chapter, we are going to review these experiments and demonstrate how Anfinsen came to a conclusion that hydrophobic interactions play the dominant role in determining the structure of globular proteins.

1.3 Anfinsen's experiments and thermodynamic principle

It is an indisputable fact that life is a far-from-equilibrium process. Living systems constantly absorb energy from the surrounding environment and use this energy to initiate a multiplicity of chemical reactions and to form regular structures that, from the standpoint of biology, perform particular biological functions. Because of this, the application of equilibrium thermodynamics to most biological systems is groundless. In early 1960s, Anfinsen conducted a series of experiments on kinetics of renaturation of globular proteins and observed that globular proteins can fold into

native structure without help from the cellular machinery [11, 12, 13]. These observations demonstrated that the process of protein folding does evolve toward equilibrium and motivated formulation of the "thermodynamic hypothesis" which states that the three-dimensional structure of a native protein in physiological environment is the one in which the Gibbs free energy of the whole system is the lowest — that is, the native conformation is determined by the sum of inter-atomic interactions and hence by the amino acid sequence.

In the work on renaturation of ribonuclease,² Anfinsen and colleagues used urea to disrupt non-covalent bonds and to unfold the protein [11]. Then, after removing urea, they let the system oxidize again and measured the restoration of protein activity as a function of time. After 10–20 hours, they observed the complete restoration of protein activity. It was known before this experiment that ribonuclease is a 124-residue protein with 8 cysteine residues forming 4 disulfide bonds that cross-link and stabilize its native structure. The lag period (about 1 hour) prior to the appearance of ribonuclease activity demonstrated that during the process of folding, disulfide bonds are formed at random and then are rearranged until the native configuration of protein is reached. It should be noted that this experiment was conducted *in vitro*, and there were no enzymes added that facilitate the formation of disulfide bonds. When such enzymes were added to the solvent, however, the folding occurred much more quickly, approximately 2 minutes, which is consistent with *in vivo* folding times. The results of this experiment strongly supported the idea that folding of ribonuclease is driven by thermodynamic forces toward the conformation with the minimal free energy, native ribonuclease.

Later, in the study of proteins that are not cross-linked, Anfinsen and colleagues identified the major physical forces that govern the process of protein folding [12]. Because the proteins that are not cross-linked undergo renaturation in a few seconds, Anfinsen and colleagues used stop-flow experiments to study the kinetics of folding of staphylococcal nuclease. They found that the fluorescence of nuclease, which depends on nuclease's proximity to its native state, as a function of renaturation time can be

²A protein that catalyzes the degradation of RNA into smaller parts.

described by combination of two processes: fast (55 ms) and slow (350 ms). Moreover, they investigated how the time scales of these processes change under different environmental conditions:

- Ionic strength and pH had no significant effect on folding time scales, which suggested that electrostatic forces do not affect rates of folding.
- The rate of fast process is independent of temperature, whereas slow process was almost four times faster at higher temperature ($\tau_{slow}(13^\circ\text{C}) = 600\text{ ms}$, $\tau_{slow}(38^\circ\text{C}) = 150\text{ ms}$). The latter indicated the importance of hydrophobic interactions in protein folding.

Anfinsen interpreted observed kinetics of renaturation as a sequence of two processes: Fast nucleation followed by slow rearrangement of hydrophobic interactions. It should be noted that not all proteins exhibit two-step renaturation [14]. However, numerous studies of renaturation of proteins under different environmental conditions confirm that hydrophobic interactions play a major role in the process of protein folding.

Contributions of hydrophobic interactions to protein folding thermodynamics have been studied extensively in the past 30 years [15, 16, 17]. However, the quantitative results of these studies vary substantially depending on the model of hydrophobic interactions. To understand the origin of these discrepancies, it is helpful to review basic models of hydrophobic interactions at different length scales.

1.4 Hydrophobicity at small and large length scales

Hydrophobic effect is usually understood as the preference of non-polar molecules for non-aqueous environments. The origin of hydrophobic interactions is well understood qualitatively — in the process of transferring a non-polar solute molecule into liquid, the unfavourable change in free energy is a result of creation of a suitably sized cavity for the solute molecule and associated rearrangement of the surrounding liquid molecules. However, the microscopic physical interpretation of hydrophobic effect is not straightforward.

The properties of liquids including water vary considerably depending on the

length scale at which they are probed. The exclusion of molecules of liquid around a large non-polar solute particle (about 1 μm in diameter) is associated with formation of a “liquid-solute” interface very similar to the liquid-vapor interface [18]; that is, the cost of formation of the interface is enthalpic and is due the fact the liquid molecules at the interface form fewer bonds than do the molecules in the bulk. For particles that are comparable in size to water molecules (about 3 \AA), the cost of forming of a cavity is mostly entropic. An extreme example of increased order of water molecules is formation of clathrates at low temperatures. In this example, all water molecules have the same number of hydrogen bonds on average, but the orientation of these bonds is fixed.

To estimate at what length scale the transition between bulk and microscopic regime happens, consider a system composed of two large parallel hydrophobic plates. At sufficiently large separations, the liquid bound between the plates forms two “liquid-plate” interfaces, which have unfavorable energy proportional to the total area of the interfaces, $\Delta E_{\text{surface}} = 2\gamma A$, where γ is the surface tension and A is the interface area. On the other hand, the favourable bulk free energy of liquid is proportional to the number of liquid molecules between the plates, $\Delta E_{\text{bulk}} = n_l |\mu_l - \mu_g| AD$, where n_l is the density of liquid, $\mu_{l,g}$ is the chemical potential of water/gas phase, and D is the distance between the plates. From the arguments above follows that at distances shorter than the critical distance,

$$D_c \sim \frac{2\gamma}{n_l |\mu_l - \mu_g|} \quad (1.2)$$

the surface energy dominates favourable bulk energy, and evaporation of liquid occurs. For water, the drying phenomenon happens at distances $D_c \approx 100$ nm, and at distances comparable to the size of biological macromolecules (about 10 nm), the phenomenon is affected by the short-range correlations in water structure.

The correlations in local structure of water are commonly described using radial distribution function, $g(r)$, which is the average density of water molecules at a given distance r from a reference particle. The radial density function depends on the size of

the reference solute particle: for particles larger than 10 nm, there is a depletion region where $g(r)$ goes to zero; for particles with radii less than 1 nm, there is an increase in the local density of water molecules. In other words, radial distribution function has a peak near the solute particle, which is often referred to as first hydration shell. This property of water motivated early models of solvation, where the solvation free energy of a particle comes from the first hydration shell is proportional to the volume of the hydration shell that is accessible to the solvent [19, 20]. Therefore, in these models, the total free energy of hydration of a solute macromolecule is given by

$$\Delta G_h = \sum_i g_i A_i, \quad (1.3)$$

where the sum is taken over all functional groups of solute (such as aromatic and carboxyl groups), g_i ³ is empirical hydration energy of the group i , and A_i is conformation-dependent accessible area of group i . This form of solvation free energy assumes the additivity of contributions of different groups and consistency of effective hydration energies g_i . Even though this method of calculating free energy can be extended to calculation of enthalpy and heat capacity of solvation, there exists evidence that the assumption that solvation free energy is proportional to the solvent-accessible surface area breaks down for polar groups [21].

A different microscopic theory of hydrophobic effect, which takes into account local correlations in density of water, was developed by Pratt and Chandler [22, 23, 24, 25]. This theory focuses on a component of fluid density that varies slowly in space and provides a description of hydrophobic effect at both small and large distances. At the heart of this theory lies the assumption that solvation energy is related to the probability of creation of a cavity in liquid that is large enough to fit a solute particle:

$$\Delta G_s = -k_B T \ln P_{cavity} = -k_B T \ln P_V(0), \quad (1.4)$$

where $P_V(0)$ is probability of observing zero liquid molecules in volume, V . Assuming

³The typical value of g_i is on the order of $10^{-2} k_B T / \text{\AA}^2$.

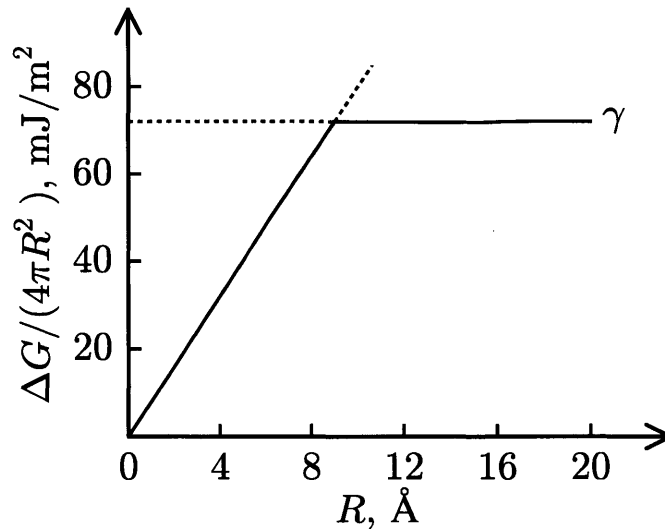


Figure 1-2: Solvation free energy, ΔG , for a spherical hydrophobic particle as a function of particle size, R . For large particles, the solvation free energy per unit area approaches the liquid-vapour surface tension, γ .

that local fluctuation of density in water are Gaussian (i.e. the probability of observing N particles in volume V is $P_V(N) = \mathcal{N}(\rho V, \sigma_V^2)$, where ρ is the density of bulk water), Chandler and coworkers found free energy of solvation and the variance in particle number σ_V^2 can be computed as

$$\Delta G_s = k_B T \frac{\rho^2 V^2}{2\sigma_V^2} + \frac{1}{2} k_B T \ln(2\pi\sigma_V^2) \quad (1.5)$$

$$\sigma_V^2 = \rho V + \rho^2 \int d\mathbf{x} d\mathbf{y} (g(|\mathbf{x} - \mathbf{y}|) - 1). \quad (1.6)$$

Within the framework of this model, they also computed the solvation free energy for a spherical particle in water as a function of particle size. The sketch of their result is shown on Figure 1-2. It is noteworthy that, for small particles ($R < 8 \text{ \AA}$), the free energy of solvation grows linear with the volume of a particle rather than the surface area, as is commonly assumed in hydration shell models. This result demonstrates that at small length scales, hydrophobic forces are not additive and affirms that hydrophobic effect is a collective phenomenon.

Experimentally, the magnitude of hydrophobic interactions is usually measured

by the free energy transfer of non-polar molecules from gas or non-polar liquid to water:

$$\Delta G_{tr} = -k_B T \ln \frac{n_{np}}{n_w}, \quad (1.7)$$

where n_{np} and n_w are equilibrium concentrations of non-polar molecules in non-polar environment and water [26]. The quantity defined by the equation (1.7) is usually called hydrophobicity, and the relative hydrophobicities of amino acids is called the hydrophobicity scale. From experimental measurements of hydrophobicities of individual amino acids, one can determine the effective hydration energies for hydration shell models, and using these values together with protein structural information, one can estimate the contribution of hydrophobic interactions to the free energy of the protein's native state by computing the difference in solvation free energy between unfolded and folded states [15, 16]. It is worth mentioning that there also exist a large number of empirical hydrophobicity scales derived using statistical analysis of the distributions of residues between the surface and the core in proteins with known structures. Some of these empirical scales we will discuss in more detail in Chapter 2.

Hydrophobicity scales have also been used to predict secondary structure of proteins, antigen binding sites, and partitioning of residues between the core and the exterior of a protein [27, 28, 29, 30, 31]. All of these methods are based on hydrophobicity profiles, moving average of sequence hydrophobicity. First, by analysing three-dimensional X-ray structures of globular proteins, Rose noticed a correspondence between chain turns and the place where hydrophobicity reaches a local minimum [27]. Later, he observed that protein packing density is also strongly correlated with hydrophobicity profile [28]. From these observations, Rose concluded that densely packed regions are formed primarily by hydrophobic parts of the chain, while chain turns and solvent-exposed parts of helices have low hydrophobicity. In addition, he suggested a hierarchical condensation mechanism for protein folding: Formation of dense hydrophobic clusters is followed by association of these clusters into native conformation. Later, Kyte and Doolittle argued that the interior of globular proteins

has high density and, thus, hydrophobicity profiles can be used to differentiate the buried parts of the chain from solvent-exposed parts based just on a sequence of a protein [31]. Likewise, Kyte and Doolittle used hydrophobicity profiles to identify parts of transmembrane proteins that are located within a lipid bilayer. A similar approach for transmembrane proteins was developed by Engelman [29]. In addition to being helpful for predicting buried regions of a protein, hydrophobicity profiles can also be used to locate protein antigenic determinants, parts of the antigen recognized by immune system. For example, Hopp and Woods argued that because antibody-binding interfaces are often located on the surface of the protein and contain charged, hydrophilic side chains, their positions can be predicted by finding regions with the highest local average hydrophilicity [30]. Despite descriptiveness and simplicity, hydrophobicity profiles have a limited practical use. These limitations come from the fact that the hydrophobicity profiles depend on the smoothing procedure and that the optimal window length for computing the average is not known beforehand. When the window length is short, the resulting hydrophobicity profile is noisy and contains a large number of extra of peaks and troughs, whereas averaging with large window length leads to the loss of resolution and artifacts such as inverting of initial data.

To conclude, because of the multifaceted nature of the hydrophobic effect, it is difficult to develop a theory that explains all features of this phenomenon. Therefore, the most reliable simulations of protein dynamics, calculations of energy surfaces, and methods of protein structure prediction are still performed with an explicit treatment of water.

1.5 The current state in protein structure prediction

After Anfinsen's experiments showed that the proteins can fold without help of cellular machinery and suggested that the information contained in the amino acid sequence of a globular protein is enough to determine its structure, the focus of theoretical research was to explain how proteins reach a single native state and how they do it in a reasonable amount of time. The first polymer physics models interpreted protein

folding first as coil-to-globule transition and then a native state transition [32, 33]. In addition, these models explained Anfinsen’s “all-or-none” observation as a first order phase transition and constructed phase diagrams of proteins. Later works on heteropolymers [34, 35, 36] considered energies of different protein conformations as independent random variables and associated the folded frozen phase of the Random Energy Model (REM) with the native structure of the protein. These studies demonstrated that a single native state of the protein can be thermodynamically dominant, provided that it has much lower energy than disordered states. In addition, the presence of a large energy gap in the spectrum of protein conformations sheds light on the kinetics of folding process [37]. Even though the theoretical models mentioned above provide deep insight into the physics of protein folding, they were not developed to predict the structure of real proteins.

The most successful approaches to protein structure prediction have been molecular dynamics (MD) and Monte Carlo (MC) simulations and statistical studies of homologous proteins. These methods were developed in the mid 1990s and take advantage of rapid/dramatic growth of computational power and advances in protein sequencing. MD simulations are widely used to analyse protein motion at atomic resolution, but they have limited time scales, usually shorter than those of protein folding and biologically important conformational changes. Limited computational power still impedes the use of MD simulations for *de novo* structure prediction of large proteins (greater than 100-residues), but some studies report that, using a special-purpose machine, they were able to run 1-millisecond simulations⁴ of folding of short polypeptides (about 40-residues) starting from random coil state [38]. In particular, Shaw and colleagues simulated folding dynamics of WW domain (FiP35) and bovine pancreatic trypsin inhibitor (BPTI). For FiP35, they ran two independent 100- μ s simulations, and in each of these simulations, FiP35 underwent several folding and unfolding events. These simulations allowed them to determine the typical folding pathway of the FiP35 and folding time $10 \pm 3 \mu$ s (comparable to experimental data 14 μ s). A 1-millisecond simulation of the folded protein BPTI helped them understand

⁴These simulations took several months of supercomputer time.

the native-state dynamics of this protein and to confirm separation of time scales of conformational changes: the relaxation of side-chains happens on the 10-ns time scales, whereas the time scale of backbone motion is on the order of 10 μs . The agreement between the results of MD simulations and experimental measurements validates the use of empirical force fields for describing dynamics of proteins over biologically important time scales.

A promising approach to interpretation of MD simulations and to propagation of dynamics to times far beyond of those directly simulated was proposed by Pande and colleagues [39]. In their work, Vaidyanathan and Pande ran a massive parallel simulation of the dynamics of the villin headpiece protein. The total simulation time of all trajectories was close to 500 μs , but unlike the simulation done later by Shaw and colleagues, the duration of single trajectory never exceeded 50 ns. Using trajectories that started from unfolded states, Vaidyanathan and Pande constructed a Markovian state model (MSM) that corresponds to observed dynamics. First, all conformations of the protein observed in these simulations were clustered/coarse grained using k -means clustering algorithm. Then, given transition time Δt , transition probabilities $P_{i \leftarrow j}$ between coarse grained states i and j were computed as the fraction of trajectories started in coarse grained state j at time $t = 0$ and finished in state i at time $t = \Delta t$. Finally, the real deterministic dynamics of the protein was approximated by probabilistic MSM dynamics. This approach allowed Vaidyanathan and Pande to compute the mean folding time, 9 μs , which is in good agreement with experimental measurements, 10 μs . In conclusion, it should be noted that the probabilistic approach proposed by Vaidyanathan and Pande provides a more descriptive picture of the dynamics than simple analysis of the multiple trajectories, but approximation of deterministic dynamics by MSM dynamics remains unproven.

Another powerful method of *de novo* protein structure prediction is built on the idea that the native structure of the protein corresponds to the lowest energy conformation, [40, 41].⁵ In order to find the minimal energy configuration Rosetta performs

⁵This method is called Rosetta and is a result of collaboration of multiple research groups. The code for the Rosetta project is available freely for academic purpose at (<http://www.rosettacommons.org/>).

a multistep MC search in conformational space of the protein. The energy function used by Rosetta changes during the search process — individual interactions are taken into account progressively. The first step in the search of global minima involves finding a large number for local minima in lower-resolution potential. At this step, Rosetta searches structure space by replacing torsion angles of 9-residue fragments with torsion angles from known structure fragments. The second step in energy minimization process starts from local minima found in the low-resolution search and adds more interactions, such as pairwise Lennard-Jones interaction, interaction with solvent, and hydrogen bonds. The predictive power of Rosetta was tested in the Critical Assessment of Techniques for Protein Structure Prediction (CASP) competition, and the high-resolution structures obtained with Rosetta agrees with crystal structure with less than 2 Å accuracy. Despite an exceptional ability to compute the native structure of the protein, the Rosetta program should be thought of as statistical/computational method rather than as a physical model for two important reasons: First, MC simulation is a great technique for obtaining ensemble averages for thermodynamic quantities, but its kinetic interpretation is unclear; second, the empirical energy function that Rosetta is minimizing changes during the simulation.

The recent progress in protein sequencing allowed for development of statistical methods for predicting protein structure from sequence variation [42]. This method is based on the idea that proteins that have similar sequences have similar shapes and perform similar functions. In addition, this method assumes that amino acids in positions important/critical for protein function are less likely to mutate in the process of evolution. Therefore, by studying covariation of amino acids in collections of homologous sequences, one can extract information about positions in linear sequence of protein that evolve together. The latter implies that residues at these positions are in contact in the three-dimensional structure. The workflow/algorithm of this method can be summarized as:

- sequence alignment
- prediction of contacts using maximum entropy methods
- generation of approximate three-dimensional structure, with contacts deter-

mined at the previous step

- refinement of three-dimensional structure using MD simulation

It should be noted that the refinement of three-dimensional structure in this method does not have to be done with MD simulations, it can also be done with the Rosetta energy minimization algorithm [43]. The major limitations of predicting a protein's structure using evolutionary information are the lack of ability to correctly predict non-local contacts and the unbiased choice of homologous sequences for alignment procedure. Current accuracy of contact prediction is only 50 % –70 % for the top 20 contacts.

Chapter 2

Quantitative theory of hydrophobic effect as a driving force of protein structure¹

2.1 Introduction

Since the experiments of Anfinsen [13], the field of structural biology has been motivated by the idea that the shape of a protein is completely determined by its sequence. Increasingly, however, it has been assumed that this mapping from sequence to structure is affected by such a diverse combination of physical interactions that a detailed simulation framework must be necessary to make accurate predictions about real proteins. Advances in hardware and simulation methods have led to various breakthroughs in the computer simulation of protein folding with all-atom resolution: The massive parallelization of trajectories for heavy sampling [39], the optimization of supercomputing on the millisecond timescale [38], and the improvement of algorithms for searching the energy landscapes of macromolecules have brought many structure-prediction and design goals within reach [41].

However, even considering the success of such computational methods in shedding

¹This chapter is a paper published in *Protein Science Journal* [44].

new light on macromolecular structure and function, their high computational cost [38, 39] and dependence on numerous modelling parameters raise the possibility that complementary insights might still be gained using a more more theoretically and computationally simple approach. Such a method would potentially have at least two advantages: that on a fixed computational budget it could be applied to a much larger corpus of protein sequences or used to sample a wider diversity of low-energy structures; and, that the small number of modelling assumptions would make it easier to determine where the model is expected to succeed, as well as where it might fail.

In the search for a simple physical principle to incorporate into the assumptions of such a model, the hydrophobic effect is a highly attractive choice. Various studies suggest that the hydrophobic effect plays a major role in the folding of proteins [45, 46, 25]. However, although the hydrophobic effect is well understood at the level of individual amino acids — non-polar amino acid residues tend to be buried in the core of the protein, and the polar residues are more likely to be on the surface — a quantitative theory of how the hydrophobic effect impacts structure as a whole in real globular proteins is difficult to construct. The lattice HP models, in which a protein is represented as a sequence of nonpolar (H) and polar (P) residues with attractive interaction between H residues, quite often do not give unique native structures, so that the predictions of these models cannot be translated to real protein structures [47, 48]. Hydrophobicity profiles, which are constructed by averaging sequence hydrophobicity, are known to correlate reasonably well with the burial of amino acid residues in globular proteins [49, 28]. However, the methods that use hydrophobicity profiles to predict burial generally do not include nontrivial effects of the polypeptide chain and do not account for the limited space in the core of a protein domain, which limits application of these methods.

Previously, England introduced a model of protein folding, termed here the “burial mode model”, that considers the hydrophobic effect, steric repulsion, and the polymeric constraints of the protein backbone to be the driving forces of protein structure [50]. Using only the amino acid sequence of a protein, this model allows one to compute not only the minimum energy conformational state of a protein but also an

ensemble of low-energy excited states. Knowledge of these states has in turn been demonstrated to be useful for studying coupled motion of different parts of a protein in allosteric motion.

For a 100–300 residue protein, it takes less than a second to use the burial mode model to compute tertiary structural information on a single CPU. Thus, it might eventually be appealing to apply the model to studying the large collections of sequence homologs that became available with high-throughput genomic sequencing. However, before doing so, one must clearly understand the model’s domain of applicability and which input parameters make it most successful in capturing the structural physics of protein domains.

In this study, we first examine whether our approach can be improved by choosing a better set of parameters. To accomplish this, we undertake to compute a new amino acid hydrophobicity scale from a large set of known protein structures, and we compare this performance of this scale to those of known hydrophobicity scales. Having identified a suitable set of parameters, we then undertake to explore the confounding effects of inter-domain interactions on the model’s ability to predict burial in protein monomers. By doing so, we discover a new application for the model in the analysis of conformational fluctuations related to ligand-binding and mutation.

2.2 Results

2.2.1 Burial mode model

In the burial mode model a globular protein domain is represented as a linear chain of N residues that are indexed by the number s and have position $\vec{r}(s) = [x(s), y(s), z(s)]$ relative to the center of mass of the globule (Figure 2-1). The polymeric bonds and the hydrophobic effect are incorporated into the system energy:

$$\mathcal{E} = \sum_{s=1}^{N-1} \kappa |\vec{r}(s+1) - \vec{r}(s)|^2 + \sum_{s=1}^N \varphi(s) |\vec{r}(s)|^2, \quad (2.1)$$

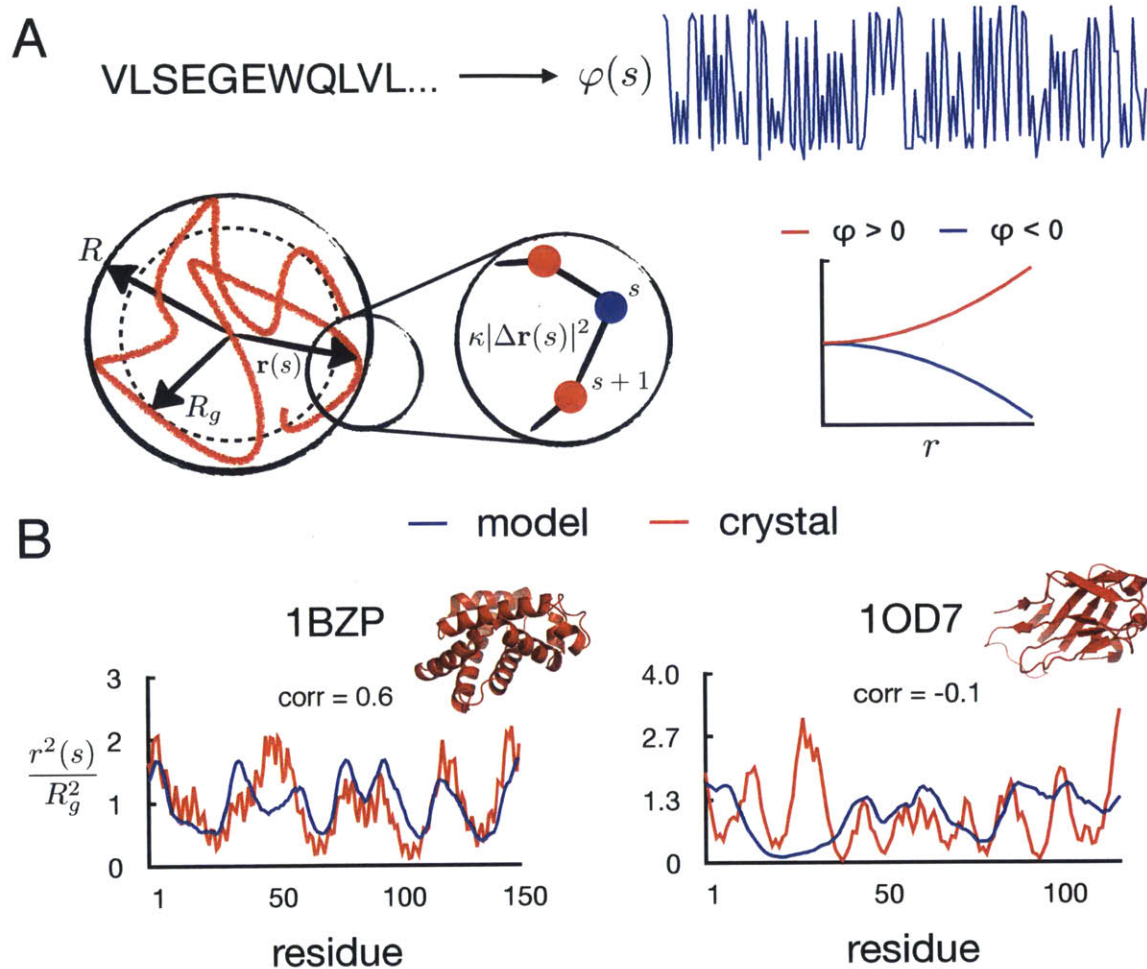


Figure 2-1: Basic assumptions of the burial mode model.

A. The protein backbone is represented as a linear chain (solid red line) with residues indexed by the number s and that have position $\vec{r}(s)$ relative to the center of the globule. The solid black line shows the maximum size of the globule, while the black dashed line shows the radius of gyration R_g . The hydrophathy of each residue $\varphi(s)$ is determined by the type of the residue. Neighbouring residues are connected by harmonic springs of stiffness κ . Blue and red residues represent hydrophilic and hydrophobic amino acids, respectively. The plot in the bottom right corner shows the contribution of different residues to the system energy as a function of the distance to the center of the globule.

B. Burial traces computed using the model (blue lines) and from the crystal structures (red lines) of sperm whale myoglobin (1BZP) and sialoadhesin (1OD7). The PCC between the model and the structure is 0.6 for myoglobin and -0.1 for sialoadhesin.

The bond stiffness κ determines the strength of “harmonic spring-like” attraction between adjacent monomers along the chain, which sets the overall elastic extensibility of the polymer. The relative hydrophathy $\varphi(s)$ reflects the tendency of each different amino acid in the chain to be exposed on the globule’s surface or buried in its core and is obtained by converting amino acid sequence into numbers using the standard Kyte-Doolittle (KD) hydrophobicity scale [31]. It should be noted that quadratic form of the hydrophobic contribution to the energy was chosen for two reasons: First, it allows the model to be analytically tractable; second, it has a physical intuition that a force acting on the residue near the surface is larger than in the core because on the surface the amino acid is more likely to have larger area exposed to the solvent. The steric repulsion between different parts of a chain is taken into account as a global constraint on the ratio α of the gyration radius squared to the maximum distance to the center of mass squared R^2

$$R_g^2 = \frac{1}{N} \sum_{s=1}^N |\vec{r}(s)|^2 = \alpha R^2. \quad (2.2)$$

The goal of this constraint is to prevent residues from collapsing into the center of the globule and, thus, to account for the limited space in the packed globular core.

To compute the lowest energy conformation of the protein, one should minimize the system energy (2.1) subject to constraint (2.2). As was shown in previous published work, this procedure can be reduced to an exactly solvable linear programming problem [50]. The optimized outcome of the linear program is given in the form of an energy-minimizing “burial trace,” that is, the squared distance $|\vec{r}(s)|^2$ from each residue to the center of mass.

To quantify the performance of the model on a given protein, one may compute Pearson’s correlation coefficient (PCC) between the burial trace computed from the sequence using the model and the burial trace generated from the known structure of the protein using coordinates of C_α atoms. (Note: To compute burial traces, one can also use coordinates of C_β atoms or side chain centroids, but this does not change burial traces significantly.) Examples of proteins for which the model gives different

PCC values are shown on Figure 2-1B. As one can see from this figure, for the proteins with high PCC (> 0.4), the resemblance between burial traces is striking, whereas for the proteins with low PCC (< 0.1), the model correctly predicts only positions of a few local extrema of the burial trace.

In globular protein domains, burial traces show what parts of the protein are buried in the core and what parts are exposed to water. In this regard, burial traces are similar to hydrophobicity profiles or window-averaged sequence hydrophobicities $\varphi(s)$, which are widely used to find out information about the secondary and the tertiary structure of proteins from their sequences [28, 30]. However, unlike hydrophobicity profiles, which do not contain any explicit information about conformational changes, the burial mode model allows one to compute the ensemble of burial traces for low-energy excited states of the chain and, thus, provides a framework for studying conformational fluctuations in proteins. Previously, this framework has been successfully employed to explain allosteric motion in a panel of test proteins for which the PCC between the burial traces from the sequence and from the structure was greater than 0.4 [50].

The mapping of the sequence-structure relationship that is affected in the burial mode model simplifies and thus accelerates the calculation so that it becomes an attractive tool for studying large collections of proteins. However, to use the model as a reliable method for analysis of conformational fluctuations, one should identify the set of physical parameters that makes the model applicable to the broadest set of proteins. Thus, motivated by the previous successes of the model in explaining allosteric motion for the proteins with high PCC, we attempted to improve the model's power to predict burial traces by optimizing its input parameters.

2.2.2 Parameter optimization

There are 21 independent parameters in the burial mode model: the bond stiffness (κ), the ratio of the squared gyration radius to the squared maximum radius of the protein (α), and 19 relative hydrophobicities of amino acid residues. However, not all parameters can be changed given the model's assumptions. First of all, the bond stiff-

| Hydrophobicity scale | Protein class | | | |
|----------------------|-----------------|-----------------|------------------|-----------------|
| | α | β | $\alpha + \beta$ | α/β |
| Kyte-Doolittle | 0.25 ± 0.22 | 0.22 ± 0.18 | 0.25 ± 0.18 | 0.23 ± 0.20 |
| Wimley-White | 0.24 ± 0.23 | 0.21 ± 0.19 | 0.21 ± 0.18 | 0.21 ± 0.19 |
| Janin | 0.22 ± 0.23 | 0.18 ± 0.19 | 0.23 ± 0.18 | 0.20 ± 0.19 |

Table 2.1: Comparison of the model performance with different hydrophobicity scales for different classes of proteins from the SCOP database. Each column shows the mean and the standard deviation of the distributions of PCC between the burial traces computed from sequences using different hydrophobicity scales and the burial traces extracted from protein structures for a given SCOP class.

ness κ fixes the units of length and must be chosen so that corresponding mean-square distance between neighbouring C_α atoms is equal to one; the parameter α ranges from 0.4 to 0.6 in real proteins and is set to $3/5$, which is the value that would hold for a globular protein that was spherical and had uniform density. The maximum radius of the protein, meanwhile, is estimated from the number of monomers in the chain and is given by $R = (3N/4\pi\rho_0)^{1/3}$, where ρ_0 is the density of monomers estimated from the crystal structure of the TIM barrel fold (PDB ID 2VXN). Thus, it is the amino acid hydrophobicity scale that offers some remaining parametric flexibility and could perhaps be optimized to improve the model’s burial trace prediction.

We first investigated how the burial mode model’s performance changes when we use different standard hydrophobicity scales. Based on the methods by which they were developed, hydrophobicity scales can be divided into two groups: experimental scales, which are based on the measurements of the free energy of solvation of single amino acids or short peptides in water and ethanol [31, 51, 52], and numerical scales, which are derived from the partition of amino acid residues between the core and the surface in proteins with known 3D structures [53, 45]. In our previous study, the relative hydrophobicities of amino acid residues were taken from the Kyte-Doolittle (KD) scale and standardized so that the energy change associated with transfer of glutamine from surface to the core of the globule is equal to $0.5 k_B T$. To compare the performance of the model with different hydrophobicity scales, we normalized all scales so that the difference between the maximum and the minimum hydrophobicities was the same as in the KD scale. Table 2.1 shows the mean and the variance of the

distributions of PCC for different classes of proteins from the Structural Classification of Proteins (SCOP) database [54]. Interestingly, despite the different origins of the hydrophobicity scales, none of the scales significantly altered the performance of the model on this large set of proteins (SCOP class).

Next, we did a brute-force search for a better hydrophobicity scale. For large groups of proteins (SCOP classes/folds), it is computationally costly to fit burial traces using a 20-letter amino acid alphabet, so we elected to use a reduced-size amino acid alphabet for these searches. We first split amino acids into four groups according to their hydrophobicity indices in the KD scale: (R, K, D, E, Q, N, H), (P, Y, W, S, T, G), (A, M, C, F), and (L, V, I). Because this is a somewhat arbitrary way to split amino acids into groups, as a control, we also divided amino acids into random groups. Then we generated a 4-dimensional rectangular grid with 10 nodes along each axis. The range of hydrophobicity indices was set between -9 and 9 , twice the minimum and maximum values of KD scale, respectively. In the case when amino acids were divided into groups at random, we found that the distributions of PCC for α -helical proteins were always broad (st.dev. ≈ 0.2) and their mean was never greater than 0.2 (the data are shown in Appendix A); whereas when amino acids were grouped according to the KD scale, the mean of the distribution of PCC never exceeded 0.3 and the standard deviation was about 0.2 . It should be noted that out of 10^4 different hydrophobicity scales we examined, only 2% had the mean of the distribution of PCC higher than 0.25 , the mean PCC for the KD scale. Furthermore, the hydrophobicity scales that provided high values of the mean PCC were in good agreement with the KD scale (Figure 2 of Appendix A). Taking into account the data in Table 2.1 and the results of the exhaustive search, one can conclude that one cannot achieve a significantly better performance for the model on large groups of proteins using 4-letter hydrophobicity scales.

To investigate whether the model’s power to predict burial traces can be improved with a 20-letter amino acid alphabet, we developed a method to derive a hydrophobicity scale from real protein structures, using physical assumptions in line with those of the model. In particular, we noted that any two amino acids of any two given

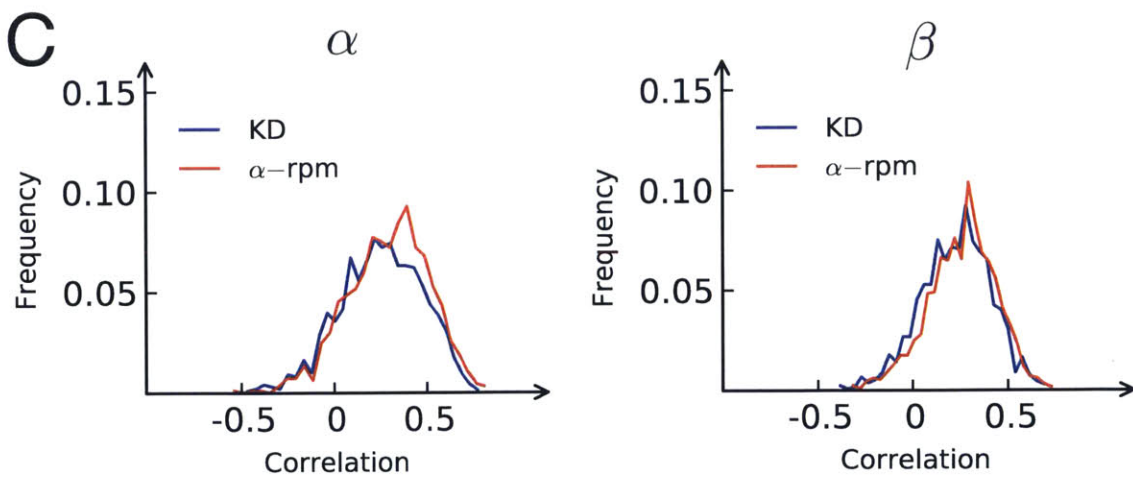
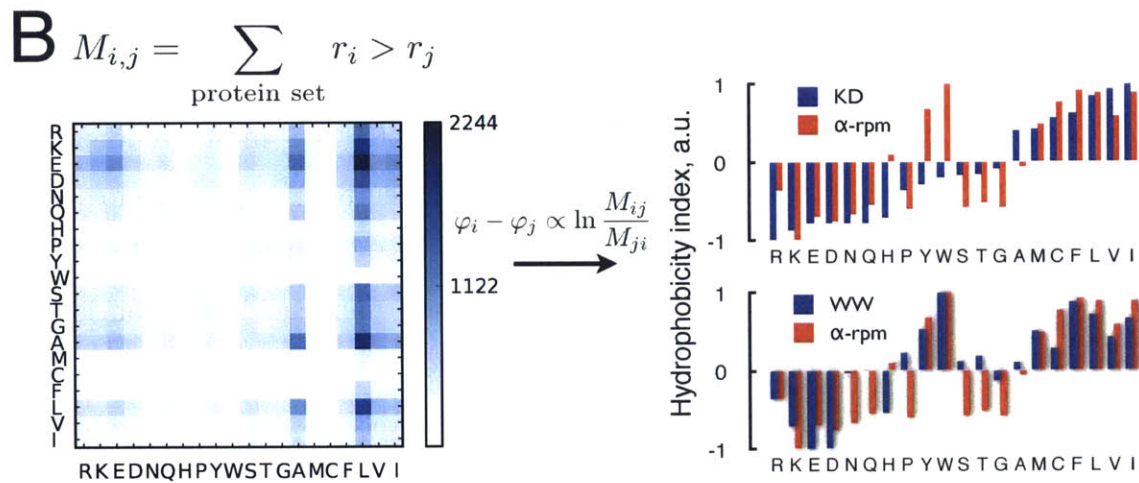
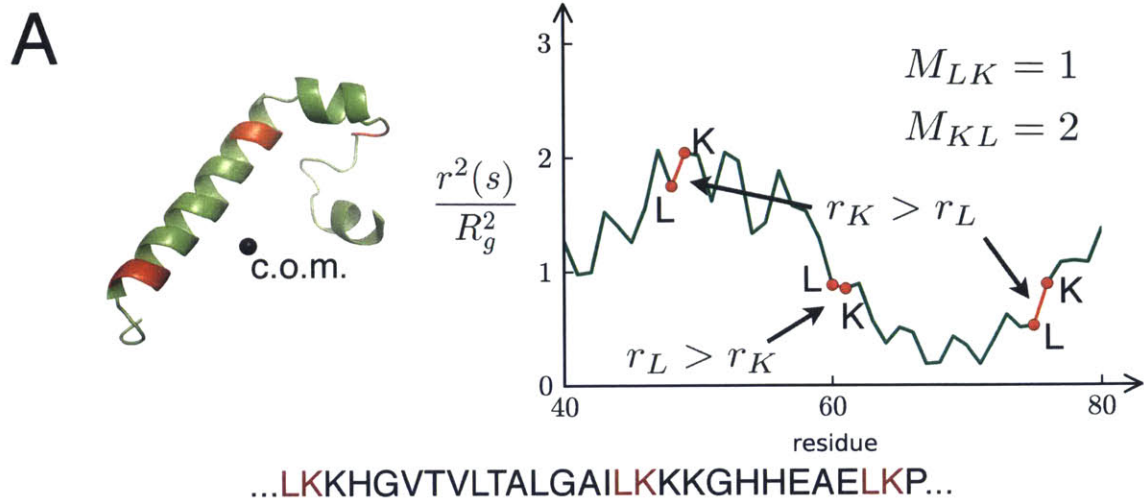


Figure 2-2: Extracting a model hydrophobicity scale from a set of proteins with known structures.

Figure 2-2: (Continued)

A. For a given protein, one can compute the burial trace (right panel) corresponding to its 3D structure (left panel). Then, one can count how many times a residue of type i (leucine (L) in the figure) is closer to the center of the globule than residue of type j (lysine (K) in the figure) given that they are the nearest neighbors on the chain.

B. Repeating the procedure described above for all proteins from the set, one can compute the matrix of relative positions M_{ij} (left panel). On the right, comparison of the hydrophobicity scale (α -rpm) calculated from the matrix of relative positions M_{ij} with Kyte-Doolittle and Wimley-White hydrophobicity scales. The matrix M_{ij} was constructed using α domains, with unique sequences of length between 100 and 300 a.a. from the SCOP database (970, in total). To compute this matrix, we used only the residues that are far from the center of a domain ($|\vec{r}(s)|^2 > 0.5R^2$).

C. Distribution of PCC between the burial traces predicted by the model using KD and α -rpm scales and the burial traces computed from the crystal structures for α -helical and β -stranded proteins from SCOP.

types in adjacent positions on a protein chain are forced to “live” in nearly identical environments. Because of this, one might suppose that their relative position in space with respect to the center of the protein globule in a crystal structure could provide an all-things-equal comparison of the tendencies of each amino acid to be buried in the globular core. Put another way, a relatively greater tendency of one amino acid in such a pair to be buried might be indicative of a relatively greater hydrophobicity.

In order to pursue this idea, we treated a large collection of proteins with known 3D structures as an ensemble of amino acid pairs, in which the relative burial of neighboring amino acids is determined only by their relative hydrophobicity. We examined the distribution of amino acid positions inside globular protein domains with unique sequences and constructed a matrix M_{ij} , each element of which was defined to be the number of times that a residue of type i is further from the center of the globule than residue of type j , given that these residues are the nearest neighbors on a chain (Figure 2-2). By positing that the probability of amino acid of type i being closer to the center of the globule than amino acid of type j is given by a Boltzmann weight, we find that the relative hydrophobicity $\Delta\varphi_{ij}$ of these amino acids is given by

$$\Delta\varphi_{ij} = \varphi_i - \varphi_j \propto \ln \frac{M_{ij}}{M_{ji}}.$$

Repeating this procedure for every pair of amino acids provides 190 relative hydrophobicities $\Delta\varphi_{ij}$. Thus, to compute 19 hydrophobicity indices φ_i of single amino acids, we did a least squares optimization. Figure 2-2B shows the matrix of relative positions of amino acid residues M_{ij} and the hydrophobicity indices φ_i computed for a set of α -helical protein domains with unique sequences of length between 100 and 300 a.a. from the SCOP database (970, in total). To compute this matrix, we used only the residues that are far from the center of a domain ($|\vec{r}(s)|^2 > 0.5R^2$). Strikingly, this new hydrophobicity scale (called “ α -rpm”) that we computed from burial information in real crystal structures turned out to agree quite well with the both the KD scale and with the Wimley-White (WW) scale (Figure 2-2B). Thus, by devising a new procedure to quantify the empirical relative statistical force on adjacent amino acids on a protein chain, we seem to have somewhat surprisingly discovered that classic hydrophobicity scales determined decades ago from bulk physicochemical measurements on amino acids already constitute a nearly optimal model of how the hydrophobic effect drives burial trends of adjacent amino acids.

To confirm this, we tested how the model works with the new hydrophobicity scale. As one can see from Figure 2-2C, the new parameters only slightly improve performance on a large set of proteins compared to the KD scale — roughly one quarter of all domains have PCC greater than 0.4. This finding, along with the results of our earlier searches of parameter space, suggests that there is no hydrophobicity scale that works significantly better than the KD scale, and there will always be many proteins whose structural physics cannot be captured by this simple model. Therefore, we sought next to understand better what other factors might limit the model’s domain of applicability.

2.2.3 Sequence diversity in globins

In search of systematic blind spots for the burial mode approach, we elected to look at a specific group of similar proteins for which the model’s performance showed a wide range of outcomes. The rationale in taking this approach was to reduce the number of sequence and structural differences among the proteins being compared,

so that it would be easier to correlate the remaining differences in these factors with resulting divergences in predicted burial trace.

An ideal group to consider for this purpose was the SCOP family of globins (SCOP ID a.1.1.2). The proteins in this family consist of eight α -helices forming a compact globule, which is appealing because the burial mode model does not account for non-local hydrogen bonding that is required for the formation of β -sheets. In light of the exceptionally good performance of the model in the case of myoglobin (PCC=0.56), we at first expected that the calculation should work just as well for all globins. However, examining more closely the full distribution of PCC for non-redundant proteins in this family, we found that the mean PCC is only 0.40 and that there are three separate peaks. Because the family of globins consists of two protein domains, myoglobin (a monomer) and hemoglobin (a heterotetramer), we decided to check whether the peaks in the distribution of PCC corresponded to these proteins. As one can see from Figure 2-3A, we indeed found that the model predicts burial traces significantly better for single domain myoglobins than for their multidomain hemoglobin cousins. For both chains of hemoglobin, Figure 2-3B shows that the model mistakenly predicts that the region 110-130, which corresponds to an inter-domain interface in the tetramer, is buried. These results suggested to us that inter-domain interaction, which is not included in the model, might change the amino acid propensity to burial by allowing hydrophobic residues to be a part of inter-domain interfaces on the surfaces of single domains.

To account for inter-domain interactions in hemoglobin we introduced a perturbation to the original burial mode model. In particular, we generated ensembles of burial traces where each residue of the chain was successively pinned to the surface of the globule by setting its hydrophobicity index to a large negative number. The PCC between these burial traces and the burial traces computed from the structures of α and β chains of hemoglobin as a function of pinning position is shown on Figure 2-3C. The idea behind this approach was that pinning the hydrophobic residues that are parts of inter-domain interfaces to the surface would push a protein into the correct shape by changing the amount of room in the protein core, and as one can see from

Figure 2-3C, the model indeed predicted the burial traces better when regions corresponding to inter-domain interfaces (residues 35–40, 110–130, and C-terminus) were forced to be on the surface. However, the highest PCC was achieved when residues 75–85 were pinned to the surface.

To understand why pinning this region, which is not a part of inter-domain interface, improves the performance of the model, we compared the hydrophobicity profiles of myoglobin and hemoglobin (Figure 2-3D). As one can see from the hydrophobicity profiles, the regions of hemoglobin corresponding to inter-domain interfaces are more hydrophobic than the same regions in myoglobin, but the largest differences in hydrophobicity occur in regions 62–72 and 75–85. The first region is more hydrophobic in myoglobin and is in close contact with a heme molecule [55], whereas the second region contains more hydrophobic residues in hemoglobin and can bind to 2,3-bisphosphoglyceric acid in the deoxy state of hemoglobin [56, 57]. Because of these differences in hydrophobicities, burying region 62–72 and exposing region 75–85 of hemoglobin is energetically less favourable in the framework of the original burial mode model. Therefore, by pinning residues 75–85 to the surface, we just restored the propensity of this region to exposure. To summarize, from the family of globins, we have learned that the tendency of amino acid residues to be buried or exposed might be determined not only by their hydrophobicity and the available space in the core but also by whether the residues are potential sites of interaction.

2.2.4 Binding and mutation as triggers of conformational change

The realization that regions involved in interactions have marginal propensities to be buried gave us the idea to look at conformational fluctuations, which we would expect the burial model to predict in regions least able to “decide” whether to be buried or exposed. Continuing to study the family of globins, we generated an ensemble of burial traces with energy $\Delta E = 1 - 5k_B T$ above the ground state energy for the sequence of sperm whale myoglobin (PDB ID: 1BZP) [50], and then from these burial traces we computed the variance of squared radial distance $\text{var}[r^2(s)]$ as a function of residue position along the chain. This function indicates the ability of each part of

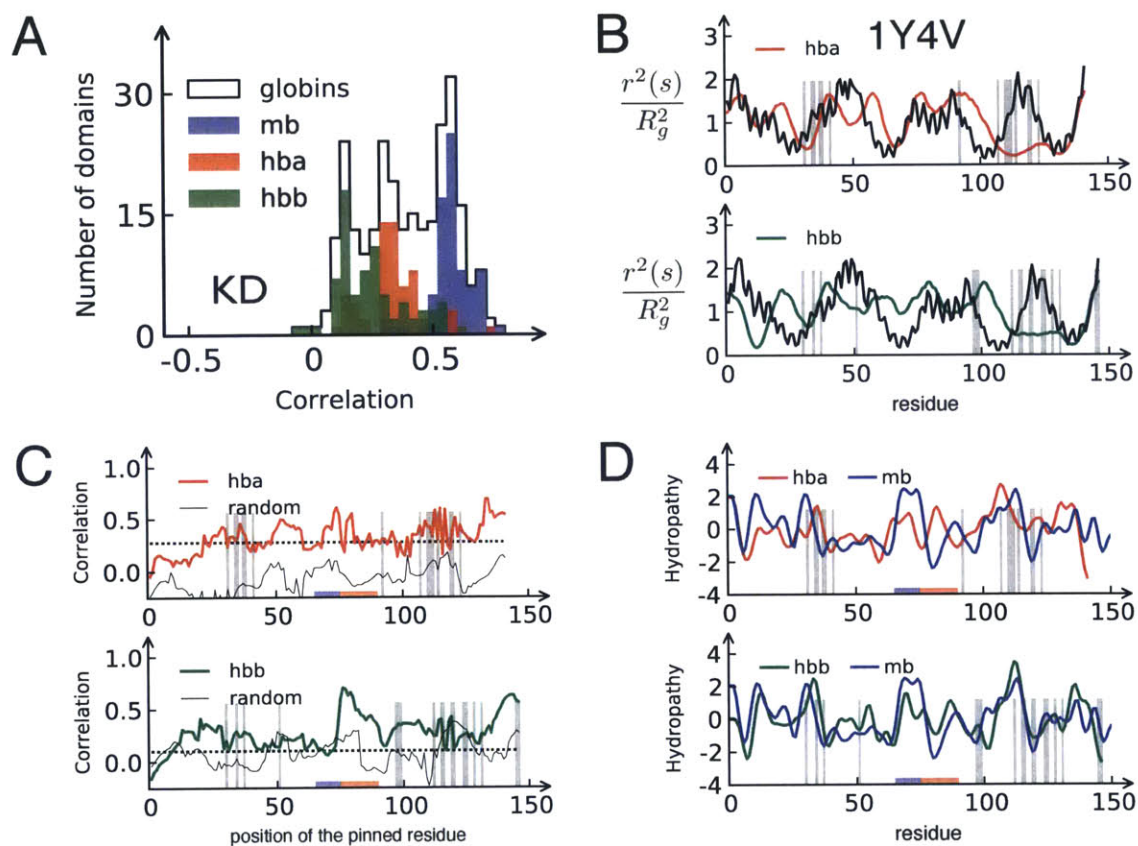


Figure 2-3: Inter-domain interaction in hemoglobin.

A. Distribution of PCC between the burial traces predicted from the sequence using the KD scale and the burial traces computed from the crystal structures for the family of globins (SCOP ID a.1.1.2).

B. Burial traces of α and β chains of hemoglobin (1Y4V) computed from crystal structures (black lines) and using the model (red and green lines). Gray bars correspond to inter-domain contacts, which were determined by the distance between C_{α} atoms with the threshold 6.5 Å.

C. PCC between the burial traces extracted from crystal structures of α and β chains of hemoglobin (1Y4V) and the burial traces computed using the model when one of the residues is pinned to the surface of the globule. Thin solid black lines correspond to the same procedure for the random sequence. The dashed horizontal lines correspond to PCC without pinning (0.28 for α -chain, and 0.10 for β -chain), whereas solid black lines correspond to the random sequence.

D. Hydrophobicity profiles of myoglobin (blue line) and hemoglobin (red and green lines) calculated using a sliding window of 10 residues.

the chain to change its shape. Figure 2-4 shows the structural variability $\text{var}[r^2(s)]$ and the 3D structure of the myoglobin colored according to this function. Strikingly, the most variable region of myoglobin corresponds the location of histidine 93, which chelates the protein's heme co-factor [55]. This result is consistent with our initial idea that the regions that can freely shift from core to surface are located close to interaction sites.

We decided to look at other proteins and to check whether our method of fluctuation analysis can be used to provide analogous insight into function in a broader range of cases. We selected two proteins in which the relation between function and conformational motion is understood and for which the model succeeds in predicting ground state burial traces: H-Ras protein (3K8Y, PCC=0.42) and chymotrypsinogen (1PYT, D chain, PCC=0.49). H-Ras is an intracellular protein that is involved in cell division regulation, while chymotrypsinogen is a secreted protein that possesses serine protease activity. H-Ras acts as a switch in a signal transduction from membrane to the cell nucleus. In its active state, H-Ras binds to GTP and converts it to GDP by cleaving the phosphate group. Figure 2-4C shows the 3D structure of H-Ras bound to GTP and the structural variability of H-Ras computed using burial mode analysis method. As one can see from this figure, the GTP binding sites of the H-ras protein (10–17, 57–61, 116–119) are located in highly fluctuating/variable regions [58].

Figure 2-5B shows the results of similar analysis performed for chymotrypsinogen and chymotrypsin (the active form of chymotrypsinogen). The conversion of chymotrypsinogen into its active form occurs in several steps: First, chymotrypsinogen is secreted and the signal peptide (residues 1–16) is cut; then, the activation peptide (residue 17–29) is removed by trypsin. The active form of chymotrypsin (residues 30–268) has catalytic activity [59]. As one can see from Figure 2-4D, both the activation peptide and the catalytic sites of chymotrypsin have high structural variability. These findings increase our confidence that the model correctly explains structural rearrangements in proteins where the burial trace prediction matches well with the known structure.

Structural variability may indeed be an important physical mechanism for bio-

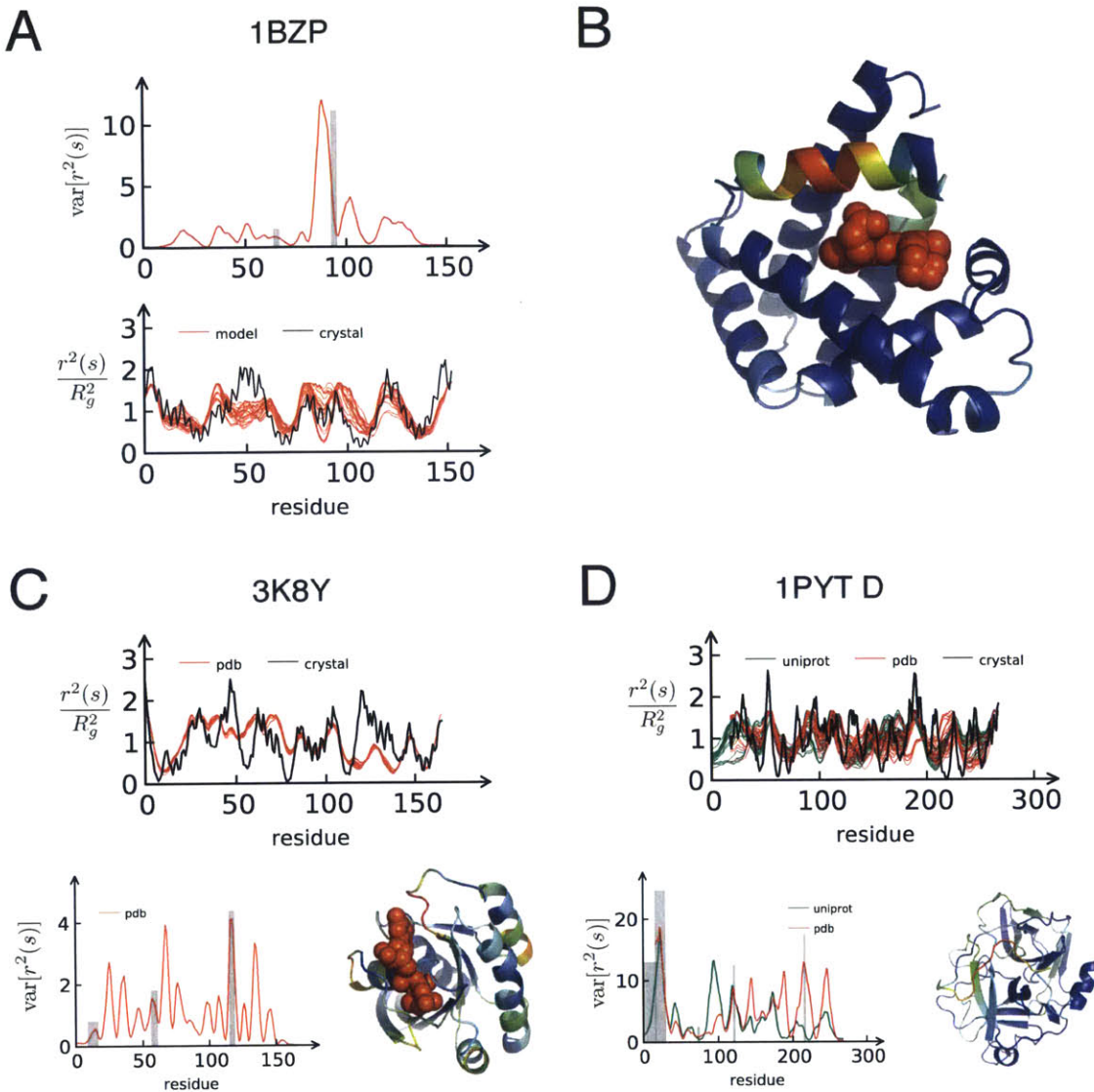


Figure 2-4: Conformational changes in sperm whale myoglobin (1BZP), H-Ras (3K8Y), and chymotrypsinogen (1PYT, D).

A. On the bottom panel, the solid black line corresponds to the burial trace of myoglobin computed from the crystal structure, while red lines correspond the burial traces of low-energy excited states ($\Delta E = 4 k_B T$). On the top panel, structural variability $\text{var}[r^2(s)]$ is computed from these burial traces. The gray bars on both sub-plots correspond to heme-binding sites (residues 65 and 94).

B. The crystal structure of myoglobin is colored according to the structural variability $\text{var}[r^2(s)]$. A heme molecule is shown in red.

C. Conformational changes in H-Ras (3K8Y). On the top, burial traces of low-energy excited states of H-Ras are depicted. On the bottom, the structural variability is both plotted and colored on the crystal structure for H-Ras, as computed for burial traces of $\Delta E = 4 k_B T$. GTP binding sites are shown as gray bars, while GTP is shown in red.

Figure 2-4: (Continued)

D. Structural variability of chymotrypsinogen (1PYT, D). Here, green lines correspond to the burial traces and structural variability computed for the uniprot sequence (before the signal peptide of chymotrypsinogen is cut), while red lines were computed for the chymotrypsinogen sequence taken from the PDB file (before the activation peptide is cleaved). Catalytic sites (H74, D121, and S216), signal and activation peptides are shown in gray.

On all subplots, the structural variability $\text{var}[r(s)]$ is shown in arbitrary units.

logical function in many proteins, but there are also situations where one would not expect to see a signature of conformational change in this metric. It is possible that a protein’s native fold might be well-structured, but that it could exhibit strong sensitivity to small changes in its sequence. For example, in a recent study [60], Alexander and colleagues demonstrated that it is possible to design a version of the streptococcal protein G such that a single point mutation (L45Y) leads to switching from 3α to $4\beta+\alpha$ fold. Furthermore, they obtained high-resolution NMR structures of two proteins (2KDL, 2KDM) that differ by three mutations (L20A, I30F, L45Y). These structures and the corresponding burial traces are shown at the top panel of Figure 2-5A. While the L20A and I30F mutants do not lead to a conformational rearrangement in the protein, the L45Y mutation does, and it is clear that the map of structural variability does not reflect the corresponding pattern of mutational sensitivity.

However, we also analyzed the sensitivity of both structures to changes in sequence hydrophobicity pattern. Using the burial mode model, we constructed the response matrix

$$\chi_{s,s'} = \frac{\delta r^2(s)}{\delta \varphi(s')},$$

where $\delta r^2(s)$ is the change in predicted optimal burial trace at position s following a small change in hydrophobicity $\delta \varphi(s')$ at position s' along the chain. The rows of this matrix show how sensitive the optimal structure of the protein is to mutations. The bottom panel of Figure 2-5B depicts the response matrices computed from the sequences of 2KDL and 2KDM proteins. It should be noted, that for both proteins, small changes in hydrophobicity in the region 43–47 lead to large changes in predicted burial trace. This result is strikingly consistent with the experimental fact that muta-

tion L45Y triggers a complete change of fold in the protein. Thus, the physical model of conformational energetics provided by the burial mode picture enables a diverse set of approaches to analyzing structural phenomena in globular protein domains.

2.3 Discussion

The problem of protein structure prediction from amino acid sequence has a long history. The most reliable approach to this problem so far — all-atom simulation — is computationally costly because it explicitly keeps track of the multitude of interactions among all atoms inside a protein. In this study, we set out to characterize a model of protein folding that sacrifices atomic details and that considers only backbone stretching, steric repulsion, and the hydrophobic effect to explain conformational preference in proteins. The advantages of this approach in studying the sequence-structure relationship are its high speed and the simplicity of interpreting results. However, a stumbling block preventing us from using the model to study large collections of proteins was a lack of clear understanding of the model’s limitations.

The parameter space of the burial mode model is defined by the hydrophobicity scale by which the amino acid sequence is mapped into a quantified string of relative burial tendencies. Thus, to improve the predictive power of the model, we searched for a better hydrophobicity scale. Having not found another standard hydrophobicity scale that works significantly better than KD scale, we did a brute force search for a new hydrophobicity scale with a reduced amino acid alphabet. Because this approach was not more effective than using KD scale, we devised a method to infer relative hydrophobicities of amino acid residues from analysis of known protein structures. This method is based on the idea that two amino acid residues that are the nearest neighbours on the chain are essentially in the same environment, and their tendency to burial is determined only by their relative hydrophobicity. It should be noted that using statistics of amino acid contacts and distances to infer amino acid interactions has been widely used before [61, 62]. However, our method is fundamentally different from Miyazawa-Jernigan and Sippl’s statistical potentials because it considers only

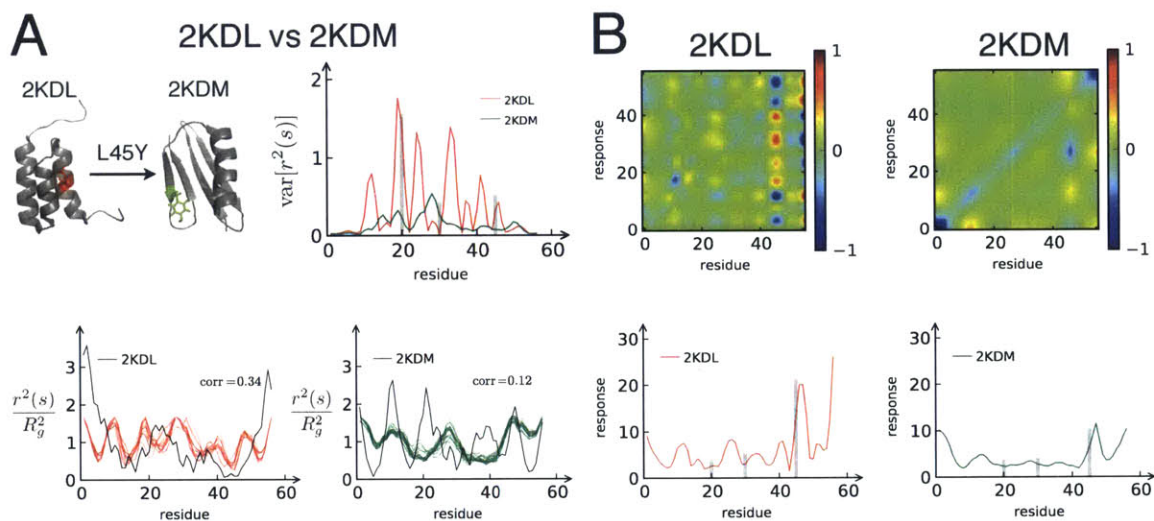


Figure 2-5: Conformational change triggered by mutation.

A. 3D structures of the 2KDM and 2KDL proteins show that mutation L45Y leads to the transformation of a $3\text{-}\alpha$ fold into a $4\beta+\alpha$ fold. Structural variability, plotted in red and green, was computed from the burial traces of the low-energy excited states ($\Delta E = 4 k_B T$). The positions of other mutations are shown as the gray bars on the plot.

B. Response matrices $\delta r^2(s)/\delta\varphi(s')$ of the 2KDM and 2KDL proteins. The plots on the bottom were obtained by taking the sum of the absolute values along the rows of the response matrices. In both proteins, the residues near the termini and residues 43-47 are the most sensitive to changes in amino acid hydrophobicity.

local interactions affecting relative burial of adjacent residues along the chain and focuses on the relative positions of amino acid residues with respect to the center of mass of the protein rather than pairwise distances.

Strikingly, the hydrophobicity scale computed with our method was in good agreement with the experimentally measured scales. This fact supports the idea that a large collection of proteins can be treated as a statistical ensemble of sequences and that our model of folding is based on sound physical assumptions about the forces driving native structure. Testing the model with the new scale, we found that performance on a large set of proteins was not improved; apparently, the model has limitations that may come from neglecting other intra-chain and/or inter-domain interactions that may be important to protein structure in any given case. Indeed, it is not surprising that the hydrophobic effect is not sufficient to explain the tertiary structures of globular proteins in all cases. Long-range hydrogen bonding interactions (such as in beta sheets), disulfide linkages, salt bridges, and dihedral angle constraints all are forces not included in the burial mode model that might play a definitive role in selecting a particular native structure in the case of a given protein. In this light, it is easy to understand why the alpha-rich globins proved such a fertile testing ground for the model.

Nonetheless, it should also be noted that the matrix of relative positions M_{ij} that we used to compute our new hydrophobicity scale contains more information about amino acid residues than a simple hydrophobicity scale, because it treats each pair of letters as having a unique local interaction. Thus, there are 190 parameters in this matrix that correspond to relative burial tendencies of different pairs, and an exciting future avenue of research will be to develop a model similar to the burial trace model that exploits all of the information in this statistical potential to predict the conformational physics of proteins. For example, it may eventually be possible using this information to develop better criteria for distinguishing between sequence trends that promote burial in the globular core and sequence trends that facilitate surface interaction with a hydrophobic ligand or protein-protein interface. Although both such trends might correspond to elevated hydrophobicity on the KD scale, one

type of sequence composition could well be distinguishable from the other with a more detailed model of the non-transitive relative burial tendency in each amino acid pair.

Having found that the burial mode model could not be substantially improved simply through parametric optimization, we set out to explore the origins of the model’s limitations. In particular, we looked at the family of globins, where the model performs exceptionally well with myoglobin and does not succeed with hemoglobin. From the comparison of these two proteins, we learned that the propensity of amino acid residues to burial might depend not only on their hydrophobicity but also on the interactions with molecules external to the monomeric protein chain, which are not included in the model. This realization gave us the idea to study conformational fluctuations in order to identify potential sites of interactions. For various proteins with good burial trace agreement (myoglobin, H-Ras protein, and chymotrypsinogen), we demonstrated that ligand binding and catalytic sites are located in the regions of high structural variability.

This finding is consistent with the “conformational selection” paradigm that has been suggested previously in the study of binding events [63] — regions of proteins that have to accommodate ligands, whether small molecules or other proteins, benefit from being structurally variable because the free energy of interaction is improved when the protein can optimize its shape to accommodate the moieties of the ligand. This process is accompanied by large structural rearrangements if there is an energy exchange between protein regions with “discrete breathers” (localized excitations) [64, 65, 66, 67]. The conformational selection paradigm implies that “discrete breathers” should be located close to ligand binding sites. Although at first sight the conformational selection paradigm and the approach that we used in this study look different, the similarity between them becomes clear if we make an analogy between “discrete breathers” and the eigenmodes of the burial mode model energy function [50] — in both descriptions, ligand binding suppresses one mode and stimulates another, coupling large scale motions to the transduction of small forces. Furthermore, it should be noted in passing that, unlike methods that use the normal mode analy-

sis to compute structural variability and mechanical response [68, 69, 70, 71], burial mode analysis relies only on sequence information and is not limited to small perturbations about a local energy minimum in a particular conformational state. Thus, burial mode analysis may yet prove useful as a general tool for prediction of catalytic and ligand binding sites from primary sequence information.

To conclude, we presented a simplified model of protein folding that allows one to compute information about protein structure directly from its sequence. In our attempt to optimize the input parameters, we discovered that the KD hydrophobicity scale provides nearly optimal performance and the limitations of the model come in part from the interactions with external molecules that are not considered in the model. To predict potential sites of ligand interaction, we exploited the idea of conformational selection and demonstrated that the burial mode model captures functionally relevant conformational changes in several cases of good burial trace agreement. Finally, we showed that sometimes the requirement for good burial trace agreement can be relaxed and that the model can also be used to predict regions most sensitive to mutations. This information can potentially be used in drug design to identify target sites and in SNP genotyping to distinguish neutral and disease-causing mutations. The model can also provide auxiliary information for MD simulations that use burial traces to generate initial protein configurations [72]. In addition, because of the high speed, the model can be employed as a tool to study large collections of homologous sequences, which became available with high-throughput genomic sequencing, and to access structural information about different mutants that are not yet crystallized.

2.4 Materials and methods

Calculation of hydrophobicity scale from the matrix of relative positions

To calculate hydrophobicity scale of n -letter amino acid alphabet from the matrix of relative positions, we first constructed two matrices, $A_{n(n-1)/2 \times n}$ and $B_{n(n-1)/2 \times 1}$,

elements of which were computed as follows:

$$A_{mi} = 1, \quad A_{mj} = -1, \quad (2.3)$$

$$A_{mk} = 0, \quad \text{for } k \neq i, j, \quad (2.4)$$

$$B_{m1} = -\ln M_{ij}/M_{ji} = \Delta\varphi_{ij}, \quad (2.5)$$

where $i = [1, n-1]$, $j = [i+1, n]$, and $m = (i-1) \times (2n-i)/2 + j - i$. Then, we used the method of least squares to find approximate solution for overdetermined system of linear equations $A \cdot \varphi = B$, where φ is n -letter hydrophobicity scale.

Generation of the burial traces of near-native states

The burial traces in the model can be written in terms of the eigenmodes $\psi_k(s)$ of energy function (2.1) and coefficients c_k : $r^2(s) = \sum_k c_k \psi_k^2(s)$. Thus, to compute the burial trace of the lowest energy state, one should minimize

$$E = \sum_k c_k \varepsilon_k, \quad (2.6)$$

where ε_k are the eigenvalues of the model energy function (2.1), subject to the steric constraints:

$$\sum_k c_k = \alpha N R^2, \quad (2.7a)$$

$$0 \leq \sum_k c_k \psi_k^2(s) \leq R^2, \quad \text{for } s \in [1, N], \quad (2.7b)$$

$$c_k \geq 0, \quad \text{for all } k. \quad (2.7c)$$

These equations set an exactly solvable linear programming problem with variables c_k , objective function (2.6), and linear constraints (2.7). The solution to this problem provides the energy of the lowest energy state E_{min} and optimal coefficients c_k^{opt} . To find the burial traces of excited states with energy $E_{min} + \Delta E$, we generated a set of coefficients c_k , which are the solution of another linear programming problem with constraints (2.7) and $\sum_k c_k \varepsilon_k = E_{min} + \Delta E$, and objective function $\sum_k c_k r_k$, where

r_k are random numbers.

To compute the structural variability $\text{var}[r^2(s)]$, we first computed $n = 100$ burial traces of near-native states $r_i^2(s)$ ($i = 1, 2, \dots, n$), and then for every position s we calculated the variance of $r^2(s)$:

$$\text{var}[r^2(s)] = \frac{1}{n} \sum_i (r_i^2(s) - \text{mean}[r^2(s)])^2, \text{ where} \quad (2.8)$$

$$\text{mean}[r^2(s)] = \frac{1}{n} \sum_i r_i^2(s). \quad (2.9)$$

Chapter 3

Overview of non-equilibrium statistical mechanics

The majority of physical processes in living systems occur far from thermal equilibrium. Biological systems are intrinsically open systems — that is, they continuously absorb mechanical and chemical work from the surrounding environment and use it to fuel their vital activities. For example, on an organism level, plants in the process of photosynthesis convert light energy into chemical energy which is stored in the adenosine triphosphate (ATP), or “energy currency of a cell”; on a single cell level, bacteria digest the nutrients in the environment and use the chemical energy extracted from the food to power the process of cell division, chemotaxis, etc; on a molecular scale, protein complexes such as molecular motors or polymerases use the free energy released in the hydrolysis of ATP to perform mechanical work of transporting a cargo along microtubules or unzipping of DNA in the process of transcription. Importantly, biological organisms are adaptive and can adjust their behavior to the changes in the environmental conditions. In the aforementioned examples, deciduous plants shed leaves with the onset of dry or cold weather, and single-cell organisms in the prolonged periods of starvation form colonies or switch into a dormant state in which normal activities are suspended. From the biological point of view, the phenomenon of adaptation is generally understood as the ability of an organism to change its qualities and behavior in a manner that allows it to survive and reproduce. However,

from the physics standpoint, there is no clear understanding of the general principles governing such adaptation.

In this chapter, we present a theory that might shed light on the physical mechanisms of adaptation and provide a unifying framework for quantitative comparison of the effectiveness of adaptation between different classes of living systems. This theory is based on the application of non-equilibrium fluctuation relations to the dynamics of living systems and allows us to compute the relative probability of the outcomes of finite-time stochastic dynamics. In the first section, we review the general approaches to modelling far-from-equilibrium systems; in the next two sections, we demonstrate that fluctuation theorems stem from the time-reversal symmetry; and finally, we derive and analyse the expression for a generalization of the Helmholtz free energy for finite-time stochastic evolution of driven Newtonian matter.

3.1 Foundations of non-equilibrium statistical mechanics

In recent years, the development of experimental techniques that allow us to manipulate individual molecules and measure tiny changes in their energy under non-equilibrium conditions, together with the extensive use of computer simulations of dynamics of molecular systems, have boosted the interest of theoretical studies in thermodynamics of small systems. These studies resolved the caveats of the second law of thermodynamics:¹ that is, of how microscopic equations of motion that are symmetric under time reversal can result in macroscopic behavior that does not possess this symmetry. To do so, one can interpret the second law as a constraint on the average value of macroscopic observable over multiple realizations of the physical process rather than their values measured for individual microscopic trajectories. Consequently, as one measures thermodynamic properties of systems with a small number of degrees of freedom, the statistical/thermal fluctuations around the aver-

¹Here, we refer to the formulation of the second law, which claims that the total entropy of the isolated system must not decrease, $dS_{tot} \geq 0$.

age become more important and the “violation” of the second law of thermodynamics is more likely to be observed.

The general framework of non-equilibrium thermodynamics assumes that a finite, classical system of interest is held in contact with a heat bath at constant inverse temperature $\beta \equiv 1/T$, and that there is an external parameter $\lambda(t)$ that allows us to control certain degrees of freedom of the system (e.g. the extension of the DNA molecule in the stretching experiments). The classical Hamiltonian of the whole set-up (system+bath) can be written as

$$\mathcal{H}_{tot}(\mathbf{x}, \mathbf{y}, \lambda(t)) = H_{sys}(\mathbf{x}, \lambda(t)) + H_{bath}(\mathbf{y}) + h_{int}(\mathbf{x}, \mathbf{y}), \quad (3.1)$$

where \mathbf{x} and \mathbf{y} represent degrees of freedom of the system and the bath, respectively; $H_{sys}(\mathbf{x}, \lambda(t))$ defines the Hamiltonian of the system of interest, including its interaction with the external time-varying field $\lambda(t)$, $H_{bath}(\mathbf{y})$ is the Hamiltonian of the heat bath, and $h_{int}(\mathbf{x}, \mathbf{y})$ describes the interaction between the system and the bath. Without loss of generality, one can assume that h_{int} is small, so that this term merely allows the exchange of energy between the system and the bath and may otherwise be ignored.

Now we can imagine a process in which at time $t = 0$, the system is prepared to be at some particular point of the phase space $\mathbf{x}(0)$, and then from time $t = 0$ to time $t = \tau$ is driven according to some protocol $\lambda(t)$. Crucially, the dynamics of the whole set-up (system+bath) is completely deterministic; however, when we focus only the system degrees of freedom, the observed dynamics appear to be stochastic due to the variation of the initial configurations of the heat bath. Therefore, when we model such non-equilibrium process, we generally posit that there is some probability distribution $\pi_\tau[\mathbf{x}(t)|\mathbf{x}(0), \lambda(t)]$ that expresses how likely it is that one would observe a particular micro-trajectory $\mathbf{x}(t)$ over time τ . Just as in equilibrium thermodynamics, we can compute the amount of work done on the system when the parameter λ is displaced by $\delta\lambda$ as the change of system energy due to this displacement $\delta W = \delta\lambda \frac{\partial H_{sys}}{\partial \lambda}(\mathbf{x}, \lambda)$,

so the work performed over the entire process is

$$W[\mathbf{x}(t)] = \int \delta W = \int_0^\tau dt \dot{\lambda} \frac{\partial H_{sys}}{\partial \lambda}(\mathbf{x}(t), \lambda(t)) \quad (3.2)$$

From the conservation of energy ($\delta E = \delta W - \delta Q$), one can also find that the heat dissipated from the system into the heat bath is

$$\Delta Q[\mathbf{x}(t)] = \int \delta Q = W[\mathbf{x}(t)] - H_{sys}(\mathbf{x}(\tau), \lambda(\tau)) + H_{sys}(\mathbf{x}(0), \lambda(0)). \quad (3.3)$$

One compelling avenue that we can pursue at this point is calculation of ensemble averages of macroscopic observables, such as work or dissipated heat. In particular, by considering an infinitely large number of repetitions of the same driving process, we can obtain a statistical ensemble of trajectories $\mathbf{x}_1(t), \mathbf{x}_2(t), \dots$ and compute any function of these trajectories $\mathcal{F}[\mathbf{x}_1(t)], \mathcal{F}[\mathbf{x}_2(t)], \dots$. If $\pi_\tau[\mathbf{x}(t)]$ is the probability distribution of individual trajectories, then the expected value of the functional $\mathcal{F}[\mathbf{x}(t)]$ is

$$\begin{aligned} \langle \mathcal{F}[\mathbf{x}(t)] \rangle &= \sum_{\mathbf{x}(t)} \pi_\tau[\mathbf{x}(t)] \mathcal{F}[\mathbf{x}(t)] = \sum_f f \cdot \sum_{\mathbf{x}(t)} \pi_\tau[\mathbf{x}(t)] \delta(\mathcal{F}[\mathbf{x}(t)] - f) \\ &= \sum_f f \cdot \rho_\tau(f), \end{aligned} \quad (3.4)$$

where $\rho_\tau(f)$ is the distribution function of $f[\mathbf{x}(t)]$. A groundbreaking result was obtained by Jarzynski [73], who considered the driving process in which at time $t = 0$ the system of interest is in equilibrium with the heat bath at $\lambda = A$, and from $t = 0$ to $t = \tau$, the external field is varied from $\lambda = A$ to $\lambda = B$. He demonstrated that for this driving protocol, the expected value of the functional $\mathcal{F}[\mathbf{x}(t)] = \exp(-W[\mathbf{x}(t)])$ is related to change of the equilibrium free energy, $\Delta F = F_B - F_A$,² as

$$\langle e^{-\beta W} \rangle = e^{-\beta \Delta F} \quad (3.5)$$

²Here, the equilibrium free energy is defined in a standard way: $F_\lambda = -k_B T \ln Z_\lambda$, where $Z_\lambda = \sum_{\mathbf{x}} e^{-H_{sys}(\mathbf{x}, \lambda)}$.

and does not depend on specific details of the driving protocol. This result has been derived using various approaches [74, 75, 76] and has been experimentally verified in RNA stretching [77], trapped colloidal particles [78], and torsion pendulum studies [79]. A simple way to derive Jarzynski work fluctuation theorem (3.5) is to assume that during the time interval $(0, \tau)$, the system is not connected to the bath. In this case, the dynamics of the system is deterministic, and for each repetition of the process there is one-to-one correspondence between the initial $\mathbf{x}(0)$ and final $\mathbf{x}(\tau)$ points in the phase space of the system. In addition, the work done in this process is equal to the change in system energy $W = H_{sys}(\mathbf{x}(\tau), \lambda(\tau) = B) - H_{sys}(\mathbf{x}(0), \lambda(0) = A)$. The only source of randomness in this simple scenario comes from the initial configuration of the heat bath (i.e. the probability of observing a particular trajectory $\mathbf{x}(t)$ is given by the probability of selecting state $\mathbf{x}(0)$ from the Boltzmann distribution, $\pi_\tau[\mathbf{x}(t)|\mathbf{x}(0), \lambda(t)] = p_{eq}(\mathbf{x}(0)) = e^{-H_{sys}(\mathbf{x}(0), \lambda=A)}/Z_A$). Combining these facts, we obtain

$$\begin{aligned}
\langle e^{-\beta W} \rangle &= \sum_{\mathbf{x}(t)} \pi_\tau[\mathbf{x}(t)|\mathbf{x}(0), \lambda(t)] e^{-W[\mathbf{x}(t)]} = \\
&= \sum_{\mathbf{x}(0)} \frac{e^{-H_{sys}(\mathbf{x}(0), \lambda(0)=A)}}{Z_A} e^{H_{sys}(\mathbf{x}(\tau), \lambda(\tau)=B) - H_{sys}(\mathbf{x}(0), \lambda(0)=A)} = \\
&= \sum_{\mathbf{x}(0)} \frac{e^{-H_{sys}(\mathbf{x}(\tau), \lambda(\tau)=B)}}{Z_A} = \sum_{\mathbf{x}(\tau)} \frac{e^{-H_{sys}(\mathbf{x}(\tau), \lambda(\tau)=B)}}{Z_A} = \frac{Z_B}{Z_A} = e^{-\beta \Delta F}. \quad (3.6)
\end{aligned}$$

An instructive interpretation of Jarzynski work fluctuation theorem can be obtained if we introduce another functional of the system's trajectory: dissipated work, $W_d[\mathbf{x}(t)] = W[\mathbf{x}(t)] - \Delta F$. The fluctuation theorem for dissipated work takes form

$$\langle e^{-W_d} \rangle = 1. \quad (3.7)$$

From Jensen inequality ($\langle e^x \rangle \geq e^{\langle x \rangle}$) immediately follows that $\langle W_d \rangle \geq 0$. We know that in an infinitely slow, reversible process the work done on the system is equal to the change in Helmholtz free energy, $W_d = 0$. Therefore, in driven systems, the dissipated work measures how irreversible a particular realization of the process was,

and equation (3.7) tells us that, on average, this quantity is greater or equal to zero. Another interpretation of the Jarzynski result that points to connection of non-equilibrium statistical mechanics and information theory comes from the relation between dissipated work and entropy production. In macroscopic theory, equilibrium free energy is defined as $F = E - TS$, and the work done on the system is related to dissipated heat through the first law $W = \Delta E + \Delta Q$. Thus, we find that the dissipated work is equal to

$$\beta W_d = \beta(W - \Delta F) = \beta(\Delta E + Q) - \beta\Delta(E - TS) = \beta Q + \Delta S = \Delta S_{tot},$$

where the total entropy change ΔS_{tot} is the sum of the entropy change of the system ΔS and the bath βQ . If this definition of total entropy production is extended to microscopic trajectories, we will recover the familiar statement of the second law $\langle S_{tot} \rangle \geq 0$.

It should be emphasized that the significance of Jarzynski work fluctuation theorem is not only in reconciling the inconsistency between the second law of thermodynamics and time-reversal symmetry, but also in setting bounds on the fluctuations of macroscopic observables. Indeed, any macroscopic observable σ that satisfies integral fluctuation theorem $\langle e^{-\sigma} \rangle = 1$ has a bound on the probability of negative values $\text{Prob}[\sigma < -\epsilon] \geq e^{-\epsilon}$.³ In the next section, we will show that probability distribution $\rho(\sigma)$ satisfies even stronger constraints given that the underlying microscopic equations of motion satisfy microscopic reversibility relation.

3.2 Microscopic reversibility relation

In this section, we describe a stochastic dynamics approach to modelling far-from-equilibrium thermodynamic processes. The reason for switching from deterministic to stochastic framework is purely pedagogical: the two approaches are equivalent, but derivation of fluctuation relations in stochastic framework is less technical, and

³This fact follows from the fact that for any $\epsilon > 0$, $1 = \langle e^{-\sigma} \rangle \geq \int_{-\infty}^{-\epsilon} d\sigma \rho(\sigma) e^{-\sigma} \geq e^{\epsilon} \text{Prob}[\sigma < -\epsilon]$.

it provides a more descriptive picture of the physical principles underlying these relationships.

One of the standard methods to describe the time evolution of the finite-size system in stochastic process is solving a set of master equations:

$$p_i(t + \delta t) = \left(1 - \sum_{j \neq i} W_{ij} \delta t\right) p_i(t) + \sum_{j \neq i} W_{ij} \delta t p_j(t), \quad (3.8)$$

where $p_i(t)$ is the probability that the system is at state i at time t , and $W_{ij} \delta t$ is the probability that in time δt the system makes a transition from state j to state i . As an example, we can consider a molecular motor walking along a microtubule: The state of the motor $i = (x, \sigma)$ in this case is represented by its position on the microtubule x and its internal conformation σ ; the transition rates W_{ij} describe how fast the motor can switch between internal states and walk along the microtubule. These rates can be controlled externally by changing the properties of the surrounding environment, such as concentration of ATP, or by applying force to the cargo attached to the motor. As one can see from the above example, the transition rates W_{ij} have to satisfy certain conditions in order for the master equation to describe thermodynamics of physical systems. To understand what these conditions are, one can first require that in the long time limit stochastic equations (3.8) are consistent with equilibrium thermodynamics.

For a large number of physical systems, one can assume that the stochastic dynamics is Markovian (i.e. the transition rates do not depend on the previous history of the system). In addition, for the systems in thermal equilibrium the rates should not change with time. In this case, the transition matrix W_{ij} satisfies the Perron-Frobenius theorem, which tells that master equation (3.8) has a unique steady state solution, p_i^{ss} . At equilibrium, this steady state probability distribution should be the same as Boltzmann distribution, $p_i^{ss} = p_i^{eq} \propto e^{-\beta E_i}$, where E_i is the energy of the system in state i , and the system satisfies detailed balance condition (i.e. the probability current between any pair of states (i, j) is zero, $J_{ij} = p_j W_{ij} - p_i W_{ji} = 0$.) These two

conditions imply that for any i, j

$$W_{ij}e^{-\beta E_j} = W_{ji}e^{-\beta E_i}, \quad (3.9)$$

which in turn means that transition rates can be written in Arrhenius law form

$$W_{ij} = r_0 e^{\beta E_j - \beta B_{ij}}, \quad W_{ji} = r_0 e^{\beta E_i - \beta B_{ji}}, \quad (3.10)$$

where r_0 is some constant rate that sets the time scale in the system, and $B_{ij} = B_{ij}$ can be interpreted as the energy of the barrier between states i and j . Interestingly, this form of transition rates demonstrates that the interaction between the heat bath and the system has potentially multiple effects on the dynamics of the system: Not only does it set the temperature of the system, but it also affects the equilibration time by determining the heights of the barriers B_{ij} .

The two conditions that transition rates satisfy in equilibrium point to two distinct ways to drive the system out of equilibrium: First, one can imagine scenario where the transition rates vary over time, but at any moment of time t , they still obey detailed balance and, thus, can be expressed using time-dependent Arrhenius law:

$$W_{ij}(t) = r_0 e^{\beta E_j(t) - \beta B_{ij}(t)}, \quad B_{ij}(t) = B_{ji}(t). \quad (3.11)$$

The systems that fall into this driving scenario include, but are not limited to, the system of colloidal particles in time-varying magnetic or electric field and driven reversible chemical reactions, where the reaction rates are modulated by changing solvent properties. In the second scenario, one can assume that transition rates stay constant in time but that they no longer satisfy detailed balance conditions (i.e. in steady state, there exist states for which $J_{ij} = p_j^{ss}W_{ij} - p_i^{ss}W_{ji} \neq 0$). This scenario can describe the irreversible chemical reactions similar to ATP hydrolysis that results in directed motion of molecular motors along cytoskeleton filaments. In the later example, still assume that the transition rates take form (3.11); however, there is no longer a single barrier energy, $B_{ij} \neq B_{ji}$, because the free energy released in ATP

hydrolysis lowers only the barrier of forward transitions B_{ij} and does not affect the barrier of reverse B_{ji} . Clearly, one can also imagine a driven system in which the two scenarios described above are combined, but these scenarios demonstrate two fundamentally different ways in which the driving field λ , introduced in the previous section, can couple to the system degrees of freedom.

Independent of the specific details of the drive, we can define thermodynamics observables such as work done on the system δW and heat dissipated from the system into the environment δQ :

$$\delta W = E[i(t + \delta t), \lambda(t + \delta t)] - E[i(t + \delta t), \lambda(t)] \quad (3.12)$$

$$\delta Q = E[i(t), \lambda(t)] - E[i(t + \delta t), \lambda(t)], \quad (3.13)$$

where $E[i(t), \lambda(t)]$ is the energy of the state $i(t)$ in driving field $\lambda(t)$ at time t . From these definitions follows that by driving the system, we perform work during time intervals when the system stays in the same state, and heat is released into environment every time the system jumps from one state to another. To compute the amount of heat that goes into the bath during the transition from state j to state i , it is useful to consider two driving scenarios described above. In the first scenario, during the jump process, the heat bath does not do any chemical work on the system, so the total dissipated heat $q_{ij}(t)$ is equal to the change of system energy δQ . Combining this result with the fact the instantaneous transitions rates are given by Arrhenius law, we find that the heat released into bath is

$$q_{ij}(t) = E_j[\lambda(t)] - E_i[\lambda(t)] = k_B T \ln \frac{W_{ij}(t)}{W_{ji}(t)}. \quad (3.14)$$

In the second scenario, when the transition rates are constant but do not satisfy detailed balance, the heat q_{ij} has two contributions: one from the change in energy of the system, $E_j - E_i$, and another from the work done to change the barrier energy $-(B_{ij} - B_{ji})$. Again, combining these two contributions with equation (3.10), we

obtain the same result

$$q_{ij}(t) = E_j[\lambda(t)] - E_i[\lambda(t)] - B_{ij}[\lambda(t)] + B_{ji}[\lambda(t)] = k_B T \ln \frac{W_{ij}(t)}{W_{ji}(t)}. \quad (3.15)$$

Having derived the above result for a single transition, we now are in a position to construct a measure of irreversibility of stochastic dynamics.

In classical mechanics, equations of motion are symmetric with respect to time reversal. The meaning of this statement becomes clear once we consider the evolution of a Hamiltonian system on a finite time interval $[0, \tau]$. Let trajectory $\mathbf{x}(t) = \{\mathbf{q}(t), \mathbf{p}(t)\}$, where \mathbf{q} and \mathbf{p} are generalized positions and momenta of the system, be a solution to Hamiltonian equations of motion in which the driving protocol λ is some function of time, $\lambda = \lambda(t)$. The time reversal operator \hat{T} acts on time, generalized coordinates, and driving protocol as follows: $\hat{T}t = -t$, $\hat{T}\mathbf{q} = \mathbf{q}$, $\hat{T}\mathbf{p} = -\mathbf{p}$, and $\hat{T}\lambda = \lambda$. If we imagine that at time τ we apply time reversal operator to Hamiltonian equations of motion in which the driving protocol is time independent, we will find that $\tilde{\mathbf{x}}(t) = \{\tilde{\mathbf{q}}(t), \tilde{\mathbf{p}}(t)\}$, where generalized coordinates are given by

$$\tilde{\mathbf{q}}(t) = \mathbf{q}(\tau - t) \text{ and } \tilde{\mathbf{p}}(t) = -\mathbf{p}(\tau - t) \quad (3.16)$$

is a solution of time-reversed equations of motion. The trajectory $\tilde{\mathbf{q}}(t)$ that satisfies equation (3.16) is called the reversed trajectory. If we apply the time reversal operator when the driving protocol is not constant, then in order to observe the reversed trajectory one also needs to run the protocol backwards in time, $\tilde{\lambda}(\tilde{t}) = \lambda(\tau - \tilde{t})$. The driving protocol $\tilde{\lambda}(\tilde{t})$ is usually called the reversed protocol.

In stochastic dynamics, it is absurd to assume that after we apply the time reversal operator, the system will follow the reversed trajectory. What one can do instead is compare the probability of the forward trajectory $i(t)$ under the forward driving protocol to the probability of the reversed trajectory $\tilde{i}(t) = i(\tau - t)$ under the reversed protocol. It is easier to do so if instead of a continuous time, we consider a discrete time Markov process. The example of such a process for $\tau = 4$ is shown on Figure 3-1. It should be noted that one can always approximate a continuous

time Markov process by a discrete time Markov process if a time step δt in discrete process is small enough so that the transition rates $W_{ij}(t)$ do not significantly change during this time and transition probabilities are small, $W_{ij}(t)\delta t \ll 1$. Thus, the results obtained here for a discrete time process are also applicable to continuous time processes. The discrete time forward trajectory $i(t)$ can be parametrized by speci-

$$\begin{aligned} i(0) &\xrightarrow{\lambda(0)} i(1) \xrightarrow{\lambda(1)} i(2) \xrightarrow{\lambda(2)} i(3) \xrightarrow{\lambda(3)} i(4) \\ \tilde{i}(4) &\xleftarrow{\tilde{\lambda}(3)} \tilde{i}(3) \xleftarrow{\tilde{\lambda}(2)} \tilde{i}(2) \xleftarrow{\tilde{\lambda}(1)} \tilde{i}(1) \xleftarrow{\tilde{\lambda}(0)} \tilde{i}(0) \end{aligned}$$

Figure 3-1: An example of forward and reversed trajectories for a discrete time Markov process.

fying the system state and the moment time when the system arrives in this state: $i(t) = \{(i_0, \tau_0), (i_1, \tau_1), \dots, (i_N, \tau_N)\}$, where N is the number of transitions in time interval $[0, \tau]$. The probability of the forward process can be written as

$$\mathcal{P}_F[i(t)] = \left(\prod_{n=0}^N \prod_{t=\tau_n}^{\tau_{n+1}-1} [1 - \sum_j W_{ji_n}(t)\delta t] W_{i_{n+1}i_n}(\tau_{n+1} - 1)\delta t \right) \times \prod_{t=\tau_N}^{\tau-1} [1 - \sum_j W_{ji_N}(t)\delta t], \quad (3.17)$$

where the terms in the form $[1 - \sum_j W_{ji_n}(t)\delta t]$ describe the probability that the system will not make a transition in time interval $[t, t + \delta t)$, and terms in the form $W_{i_{n+1}i_n}(\tau_{n+1} - 1)\delta t$ describe the probability of making a transition from state i_n to state i_{n+1} during times step $[\tau_{n+1} - 1, \tau_{n+1})$. Similarly, the reversed trajectory can be represented as $\tilde{i}(t) = i(\tau - t) = \{(\tilde{i}_0, \tilde{\tau}_0), (\tilde{i}_1, \tilde{\tau}_1), \dots, (\tilde{i}_N, \tilde{\tau}_N)\}$, and its probability under the reversed driving protocol is

$$\mathcal{P}_R[\tilde{i}(t)] = \left(\prod_{n=0}^N \prod_{t=\tilde{\tau}_n}^{\tilde{\tau}_{n+1}-1} [1 - \sum_j \tilde{W}_{j\tilde{i}_n}(t)\delta t] \tilde{W}_{\tilde{i}_{n+1}\tilde{i}_n}(\tilde{\tau}_{n+1} - 1)\delta t \right) \times \prod_{t=\tilde{\tau}_N}^{\tau-1} [1 - \sum_j \tilde{W}_{j\tilde{i}_N}(t)\delta t]. \quad (3.18)$$

Using the fact that $\tilde{i}_n = i_{N-n}$, $\tilde{\tau}_n = \tau - \tau_n + 1$, and $\tilde{W}_{ij}(t) = W_{ij}(\tau - t)$, we find the ratio of the probability of the forward process to the probability of the reversed

process

$$\frac{\mathcal{P}_F[i(t)]}{\mathcal{P}_R[\tilde{i}(t)]} = \prod_{n=0}^{N-1} \frac{W_{i_{n+1}i_n}(\tau_{n+1}-1)}{W_{i_n i_{n+1}}(\tau_{n+1}-1)}. \quad (3.19)$$

Taking into account equation (3.14), which relates the heat dissipated into the bath during a single transition, we obtain

$$\frac{\mathcal{P}_F[i(t)]}{\mathcal{P}_R[\tilde{i}(t)]} = \prod_{n=0}^{N-1} e^{\beta q_{i_{n+1}i_n}(\tau_{n+1}-1)} = e^{\beta Q[i(t)]} = e^{\Delta S_{bath}}, \quad (3.20)$$

where $Q[i(t)] = \sum_n q_{i_{n+1}i_n}(\tau_{n+1}-1)$ is the heat dissipated into the bath along the forward trajectory, and $\Delta S_{bath} = \beta Q[i(t)]$ is the change in entropy of the bath. Equation (3.20) is called Crooks microscopic reversibility relation, and it represents the essence of all fluctuation theorems. This relation can be generalized to the case of multiple baths by replacing ΔS_{bath} with the total entropy change of the baths connected to the system ΔS_{baths} . Microscopic reversibility has been derived in various ways [80, 81], using Markovian dynamics (similar to how they are presented in this section) [82], using Hamiltonian dynamics [83], and using Langevin dynamics [84, 76].

In the next two sections, we will first use microscopic reversibility relation to obtain work and entropy production fluctuation theorems for a broad class of driven systems [73, 74, 85, 86], and then derive a similar relation for transitions between macrostates [87].

3.3 Fluctuation theorems

There are two types of relations that are commonly referred to as fluctuation theorems in the literature: detailed fluctuation theorems, which relate the probability distribution of some thermodynamic observable $\mathcal{F}[x(t)]$ in a forward process and those in a reversed process, and the integral fluctuation theorems, which set a constraint on the exponential average of thermodynamic observable:

$$\frac{\rho_F(\mathcal{F})}{\rho_R(\mathcal{F})} = e^{-\mathcal{F}}, \text{ and } \langle e^{-\mathcal{F}} \rangle = 1. \quad (3.21)$$

It should be noted that the detailed fluctuation theorem implies the integral fluctuation theorem; however, the opposite is not true.

In the previous section, we found the ratio of probabilities of forward and reversed trajectories given that the initial state of the system is known. More generally, at time $t = 0$, a thermodynamic system is described by some distribution over state $p_i(0)$, so probabilities of observing forward $x(t)$ and reversed $\tilde{x}(t)$ trajectories are $\pi_\tau[x(t)] = p_{x(0)}(0)\mathcal{P}_F[x(t)]$, and $\pi_\tau[\tilde{x}(t)] = p_{\tilde{x}(0)}(0)\mathcal{P}_R[\tilde{x}(t)]$. Using the fact that $p_{\tilde{x}(0)}(0) = p_{x(\tau)}(\tau)$, we find that the ratio of probabilities of forward and reversed trajectories is given by

$$\frac{\pi_\tau[x(t)]}{\pi_\tau[\tilde{x}(t)]} = \frac{p_{x(0)}\mathcal{P}_F[x(t)]}{p_{\tilde{x}(0)}\mathcal{P}_R[\tilde{x}(t)]} = e^{\ln p_{x(0)}(0) - \ln p_{x(\tau)}(\tau) + \beta Q[x(t)]}. \quad (3.22)$$

If, following the information theory, we define the entropy of a microstate i of the system as a logarithm of the probability of being in this state, $s_{int}(i) = -\ln p_i$, then we can interpret the expression in the exponent in equation (3.22) as the total entropy production along the trajectory $x(t)$

$$\Delta S_{tot}[x(t)] = \ln p_{x(0)}(0) - \ln p_{x(\tau)}(\tau) + \beta Q[x(t)] = \Delta s_{int} + \Delta S_{bath}. \quad (3.23)$$

From this definition and equation (3.22), we immediately find that for any driving protocol, and for any initial conditions, the total entropy production satisfies the integral fluctuation theorem

$$\langle e^{-\Delta S_{tot}} \rangle = \sum_{x(t)} \pi_\tau[x(t)] e^{-\Delta S_{tot}[x(t)]} = \sum_{\tilde{x}(t)} \pi_\tau[\tilde{x}(t)] = 1. \quad (3.24)$$

Because in general case $p_{x(0)}(0) \neq p_{\tilde{x}(\tau)}(\tau)$, there is no simple relation between the entropy production along forward and reversed trajectories, $\Delta S_{tot}[x(t)] \neq -\Delta S_{tot}[\tilde{x}(t)]$, and the total entropy production ΔS_{tot} does not necessarily satisfy a detailed fluctuation theorem.

For the rest of this section, unless otherwise explicitly stated, we will focus on the

systems and driving protocols for which

$$p_{\tilde{x}(0)}(0) = p_{x(\tau)}(\tau) \text{ and } p_{x(0)}(0) = p_{\tilde{x}(\tau)}(\tau). \quad (3.25)$$

We will show that, in this case, the total entropy production satisfies a detailed fluctuation theorem, and we will give examples of two broad groups of physical systems for which this theorem is relevant. When the conditions (3.25) are satisfied, the total entropy production is anti-symmetric under time reversal, and we obtain that its probability distribution satisfies

$$\begin{aligned} \rho_F(\omega) &= \sum_{x(t)} \pi_\tau[x(t)] \delta(\omega - \Delta S_{tot}[x(t)]) = \sum_{x(t)} \pi_\tau[\tilde{x}(t)] e^{\Delta S_{tot}[x(t)]} \delta(\omega - \Delta S_{tot}[x(t)]) = \\ &= e^\omega \sum_{\tilde{x}(t)} \pi_\tau[\tilde{x}(t)] \delta(\omega + \Delta S_{tot}[\tilde{x}(t)]) = e^\omega \rho_R(-\omega). \end{aligned} \quad (3.26)$$

Motivated by Jarzynski work fluctuation theorem, we will first focus on the group of the systems that from time $t = -\infty$ to time $t = 0$ are in equilibrium with the heat bath at some constant external field $\lambda = A$; then, during finite time interval $[0, \tau]$, they are driven arbitrary far from equilibrium according to protocol $\lambda(t)$; finally, after time τ , the protocol is kept constant and the systems relax toward new equilibrium $\lambda(\tau) = B$. Because under the reversed protocol at the start the system is in equilibrium at $\lambda = B$ and in the end is in equilibrium at $\lambda = A$, the system satisfies conditions (3.25), and the total entropy production along the forward trajectory equals the dissipated work

$$\begin{aligned} \Delta S_{tot}[x(t)] &= \ln p_{x(-\infty)}^{eq}(\lambda = A) - \ln p_{x(\infty)}^{eq}(\lambda = B) + \beta Q[x(t)] = \\ &= \ln Z_B - \ln Z_A + \beta(E_{x(\infty)} - E_{x(-\infty)}) + \beta Q[x(t)] = -\beta \Delta F + W[x(t)] = W_d[x(t)]. \end{aligned} \quad (3.27)$$

From this follows that the dissipated work satisfies detailed and integral fluctuation theorems, and we recover Jarzynski's result, $\langle e^{-W_d} \rangle = 1$. Moreover, because the change in free energy is anti-symmetric under time reversal and does not depend on the system's trajectory $x(t)$, the probability distribution functions for work is

$\rho(W) = \rho(W_d - \Delta F)$ and satisfies

$$\frac{\rho_F(W)}{\rho_R(-W)} = e^{\beta W - \beta \Delta F}. \quad (3.28)$$

This relation is known as the Crooks work fluctuation theorem [74], and it is valid even if the system is driven far from equilibrium and has been experimentally verified [88].

To derive the fluctuation relations (3.5, 3.28), we restricted ourselves to the group of systems that are initially in equilibrium. However, this is not the only group of systems for which fluctuation theorems are applicable. We can also consider a group of system that are driven symmetrically into non-equilibrium steady state (NESS). For example, a Markovian system discussed in the previous section, in which transition rates are driven periodically and symmetrically in time according to time-dependent Arrhenius law (3.11), will eventually reach a periodic steady state. Between the moments of time with respect to which the driving protocol is symmetric, the entire system is invariant under time reversal⁴ (i.e. the forward and reversed dynamics are the same). Figure 3-2 shows an example of periodic driving protocol, and the points in time with respect to which the protocol is symmetric and conditions (3.25) are met. It should be emphasized that for systems in NESS, a detailed fluctuation theorem is valid only for total entropy production ΔS_{tot} . The changes in internal ΔS_{sys} and heat bath entropy ΔS_{bath} do not individually satisfy detailed or integral fluctuation theorems. To complete this discussion, it is informative to list a few real systems for which these fluctuation theorems are relevant: nanoscale machines such as molecular rotors [89], driven chemical reactions, self-replicators, and fluid or suspension of colloidal particles under the constant shear.

Since the late 1990s, there have been many theoretical works that split the total entropy production into contributions that satisfy detailed or integral fluctuation theorems in NESS. Hatano and Sasa used Langevin dynamics framework and the concepts of houskeeping Q_{hk} and excess $Q_{ex} = Q_{bath} - Q_{hk}$ heat introduced in phe-

⁴In this example, we assume that the system has discrete states and, therefore, no momenta.

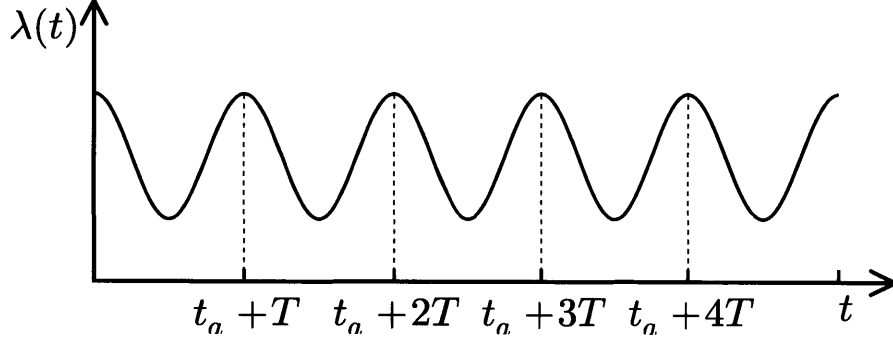


Figure 3-2: An example of periodic driving protocol. This protocol is symmetric between t_a and any time separated by an integer number of periods $t_a + nT$.

nomenological steady-state thermodynamics (SST) [90] to show that integral fluctuation theorem holds for $Q_{ex} + \Delta S_{sys}$ [85]. Jarzynski and colleagues demonstrated that this quantity also satisfies a detailed fluctuation theorem [91]. At the same time, Speck and Seifert obtained an integral fluctuation theorem for housekeeping heat Q_{hk} [92], which is defined as

$$Q_{hk}[x(t)] \equiv \gamma \int_0^\tau dt \frac{J^{ss}(x(t), \lambda(t))}{p^{ss}(\lambda(t))} \dot{x}(t), \quad (3.29)$$

where J^{ss} is a steady state probability current and γ is a drag coefficient. In Markovian dynamics framework, the housekeeping heat is defined as

$$Q_{hk}[x(t)] = \sum_i \delta Q_{hk}(x_{i+1}, x_i, \lambda_i) = \sum_i \ln \frac{W_{x_{i+1}, x_i}(\lambda_i) p_{x_i}^{ss}(\lambda_i)}{W_{x_i, x_{i+1}}(\lambda_i) p_{x_{i+1}}^{ss}(\lambda_i)} \quad (3.30)$$

and measures the degree to which detailed balance is violated. Recently, Esposito and Broeck divided the total entropy production along a trajectory into the adiabatic ΔS_a and nonadiabatic parts ΔS_{na} and demonstrated that all three quantities satisfy detailed fluctuation theorems [86]⁵. Finally, Hatano and Sasa's relation has been tested in experiments on optically trapped beads dragged through viscous fluid [93].

⁵To derive these theorems, Esposito and Broeck defined ΔS_a using equation (3.30) generalized to multiple baths.

3.4 Macroscopic irreversibility

In the previous section, we demonstrated how microscopic reversibility relation sets constraints on the ensemble averages and probability distribution functions of thermodynamic quantities such as dissipated work and housekeeping heat. Here, we will use this relation to determine constraints on the transition probabilities between coarse-grained states.

We may define a macrostate of the system as a set of microstates that share some observable property \mathbf{I} (e.g, the number of bacteria in the Petri dish, or the magnetization of a collection of spins). We may also assume that there is some reproducible experimental procedure that allows us to prepare the system in macrostate \mathbf{I} , and in this macrostate, the probability density function over microstates is $p(x|\mathbf{I})d\mathbf{x} \equiv p_i(\mathbf{x})d\mathbf{x}$. If the system that was prepared to be in macrostate \mathbf{I} is driven according to some protocol $\lambda(t)$, then after finite time τ with probability $\pi_\tau[\mathbf{I} \rightarrow \mathbf{II}; \lambda(t)]$, we may observe the system to be in macrostate \mathbf{II} , where the microstates are distributed according to a new density function $p(\mathbf{x}|\mathbf{II}, \mathbf{I}; \lambda(t); \tau)d\mathbf{x} \equiv p_f(\mathbf{x})d\mathbf{x}$. In addition, we can define $\pi_\tau^{\mathbf{I}}[\mathbf{II}^\dagger \rightarrow \mathbf{I}^\dagger; \lambda(\tau - t)]$ to be the probability of returning back to macrostate \mathbf{I} under reversed protocol $\lambda(\tau - t)$, given that all generalized momenta change signs ($\mathbf{x}^\dagger = \{(q_1, -p_1), \dots\}$). In his recent work [87], England demonstrated that the transition probabilities introduced above satisfies

$$\frac{\pi_\tau^{\mathbf{I}}[\mathbf{II}^\dagger \rightarrow \mathbf{I}^\dagger; \lambda(\tau - t)]}{\pi_\tau[\mathbf{I} \rightarrow \mathbf{II}; \lambda(t)]} = \langle e^{-\Delta S_{tot}} \rangle_{\mathbf{I} \rightarrow \mathbf{II}}, \quad (3.31)$$

where the total entropy production along trajectory $\mathbf{x}(t)$ is defined in the usual way $\Delta S_{tot}[\mathbf{x}(t)] = \ln p_i(\mathbf{x}(0))/p_f(\mathbf{x}(\tau)) + \Delta Q[\mathbf{x}(t)]$, and averaging is done over all paths that start at some microstate $i \in \mathbf{I}$ and end in microstate $j \in \mathbf{II}$. More general derivation of this result, which is valid for quantum systems, has been done by Ruelle [94, 95]. To understand the structure of expression (3.31), it is instructional to reproduce the original derivation.

To emphasize the use of microscopic reversibility relation, we first derive an expression similar to (3.31) for a pair of microstates: $\mathbf{x}(0) = i$, $\mathbf{x}(\tau) = j$. In the

equation below, we will use the fact that for forward process $\delta_{\mathbf{x}(\tau),j} \delta_{\mathbf{x}(0),i} \frac{\pi_\tau[\mathbf{x}(t)]}{\pi_\tau[i \rightarrow j; \lambda(t)]} = \pi_\tau[\mathbf{x}(t)|i, j, \lambda(t)]$, conditional probability of trajectory $\mathbf{x}(t)$ given that it started at i and finished at j is

$$\begin{aligned} \pi_\tau[j^\dagger \rightarrow i^\dagger; \lambda(\tau - t)] &= \sum_{\mathbf{x}^\dagger(t)} \delta_{\mathbf{x}^\dagger(0),j^\dagger} \delta_{\mathbf{x}^\dagger(\tau),i^\dagger} \pi_\tau[\mathbf{x}^\dagger(t)] = \sum_{\mathbf{x}(t)} \delta_{\mathbf{x}(\tau),j} \delta_{\mathbf{x}(0),i} \pi_\tau[\mathbf{x}(t)] e^{-\beta \Delta Q[x(t)]} = \\ &= \pi_\tau[i \rightarrow j; \lambda(t)] \sum_{\mathbf{x}(t)} \delta_{\mathbf{x}(\tau),j} \delta_{\mathbf{x}(0),i} \frac{\pi_\tau[\mathbf{x}(t)]}{\pi_\tau[i \rightarrow j; \lambda(t)]} e^{-\beta \Delta Q[x(t)]} = \\ &= \pi_\tau[i \rightarrow j; \lambda(t)] \langle e^{-\Delta S_{bath}} \rangle_{i \rightarrow j}. \end{aligned} \quad (3.32)$$

Having obtained the relation for two microstates, we may now express transition probabilities between macrostates \mathbf{I} and \mathbf{II} as

$$\begin{aligned} \pi_\tau[\mathbf{I} \rightarrow \mathbf{II}; \lambda(t)] &= \int_{\mathbf{II}} dj \int_{\mathbf{I}} di p(i|\mathbf{I}) \pi_\tau[i \rightarrow j; \lambda(t)] \\ \pi_\tau^\dagger[\mathbf{II}^\dagger \rightarrow \mathbf{I}^\dagger; \lambda(\tau - t)] &= \int_{\mathbf{II}^\dagger} dj \int_{\mathbf{I}^\dagger} di p(j|\mathbf{II}^\dagger) \pi_\tau[j \rightarrow i; \lambda(\tau - t)]. \end{aligned} \quad (3.33)$$

Using the fact that $p(j^\dagger|\mathbf{II}^\dagger) = p(j|\mathbf{II}) = p_f(j)$ and $\frac{p(i|\mathbf{I}) \pi_\tau[\mathbf{I} \rightarrow \mathbf{II}; \lambda(t)]}{\pi_\tau[\mathbf{I} \rightarrow \mathbf{II}; \lambda(t)]}$ equals to conditional probability of transition from microstate i to microstate j given that $i \in \mathbf{I}$ and $j \in \mathbf{II}$, we obtain

$$\begin{aligned} \frac{\pi_\tau^\dagger[\mathbf{II}^\dagger \rightarrow \mathbf{I}^\dagger; \lambda(\tau - t)]}{\pi_\tau[\mathbf{I} \rightarrow \mathbf{II}; \lambda(t)]} &= \frac{1}{\pi_\tau[\mathbf{I} \rightarrow \mathbf{II}; \lambda(t)]} \int_{\mathbf{II}^\dagger} dj \int_{\mathbf{I}^\dagger} di \frac{p_f(j)}{p_i(i)} p_i(i) \pi_\tau[j \rightarrow i; \lambda(\tau - t)] = \\ &= \int_{\mathbf{II}} dj \int_{\mathbf{I}} di p(i|\mathbf{I}) \frac{p(i|\mathbf{I}) \pi_\tau[\mathbf{I} \rightarrow \mathbf{II}; \lambda(t)]}{\pi_\tau[\mathbf{I} \rightarrow \mathbf{II}; \lambda(t)]} e^{\ln p_f(j)/p_i(i)} \langle e^{-\Delta S_{bath}} \rangle_{i \rightarrow j} = \\ &= \langle e^{\ln p_f(j)/p_i(i)} \langle e^{-\Delta S_{bath}} \rangle_{i \rightarrow j} \rangle_{\mathbf{I} \rightarrow \mathbf{II}} = \langle e^{-(\Delta S_{int} + \Delta S_{bath})} \rangle_{\mathbf{I} \rightarrow \mathbf{II}}. \end{aligned} \quad (3.34)$$

Now, if we define the Shannon entropy as $\mathcal{S} = -\sum_i p_i \ln p_i = \langle S_{int} \rangle$, take the logarithm of the expression above, and use Jensen inequality $\langle e^x \rangle \geq e^{\langle x \rangle}$, we find lower bound on the average of the total entropy production

$$\langle \Delta S_{tot} \rangle_{\mathbf{I} \rightarrow \mathbf{II}} = \Delta \mathcal{S} + \beta \langle \Delta Q \rangle_{\mathbf{I} \rightarrow \mathbf{II}} \geq \ln \frac{\pi_\tau[\mathbf{I} \rightarrow \mathbf{II}; \lambda(t)]}{\pi_\tau^\dagger[\mathbf{II}^\dagger \rightarrow \mathbf{I}^\dagger; \lambda(\tau - t)]}. \quad (3.35)$$

When dynamics of the system is dominated by diffusive motion, and the external protocol is time-symmetric, the bound on total entropy production significantly simplifies

$$\langle \Delta S_{tot} \rangle_{\mathbf{I} \rightarrow \mathbf{II}} = \Delta \mathcal{S} + \beta \langle \Delta Q \rangle_{\mathbf{I} \rightarrow \mathbf{II}} \geq \ln \frac{\pi_\tau[\mathbf{I} \rightarrow \mathbf{II}]}{\pi_\tau[\mathbf{II} \rightarrow \mathbf{I}]}.$$
 (3.36)

As an example of the phenomenon that falls into framework of equation (3.36), England considered biological self-replication. In particular, he demonstrated the effect of growth rate and durability of replicators on the heat dissipated into the environment in the processes of DNA and RNA hydrolysis and bacterial cell division.

3.5 Generalization of Helmholtz free energy

In this section, we will use macroscopic irreversibility relation (3.31) to derive a generalization of the Helmholtz free energy for a finite time stochastic evolution of Newtonian matter. We will analyze this expression term by term and show that the relative probability of observing two macrostates is strongly affected by the amount of dissipated work during the process of entering these macrostates. Finally, we will argue that many structures formed far from equilibrium may appear to be selected for their ability to absorb work from the environment.⁶

Here we will consider the same setup as described in the previous section. However, instead of deriving the generalized second law of thermodynamics and constraints obeyed by self-replicators, we will focus on the thermodynamics of driven stochastic evolution. Let us assume that the system is initially prepared to be in some macroscopic state \mathbf{I} , and that macrostates \mathbf{II} and \mathbf{III} are possible macroscopic arrangements of the system at time τ after it was driven according to protocol $\lambda(t)$. It should be noted that this framework is quite general and covers even ridiculous possibilities, such as an *E.coli* cell in rich medium spontaneously degrading into individual atoms instead of making a copy of itself. Independently of specific details of the system, we will try to answer the question of what makes some outcomes more likely than others. By dividing macroscopic irreversibility relations (3.31) for two possible macrostates

⁶In this section we will closely follow the logic of our manuscript [96].

and taking the logarithm of both sites, we can write

$$\ln \left[\frac{\pi_\tau[\mathbf{I} \rightarrow \mathbf{II}; \lambda(t)]}{\pi_\tau[\mathbf{I} \rightarrow \mathbf{III}; \lambda(t)]} \right] = \ln \left[\frac{\pi_\tau^\dagger[\mathbf{II}^\dagger \rightarrow \mathbf{I}^\dagger; \lambda(\tau - t)]}{\pi_\tau^\dagger[\mathbf{III}^\dagger \rightarrow \mathbf{I}^\dagger; \lambda(\tau - t)]} \right] - \ln \left[\frac{\langle \exp(-\Delta S_{tot}) \rangle_{\mathbf{I} \rightarrow \mathbf{II}}}{\langle \exp(-\Delta S_{tot}) \rangle_{\mathbf{I} \rightarrow \mathbf{III}}} \right] \quad (3.37)$$

If we now assume that the system is driven time-symmetrically and that its dynamics is diffusive (i.e. the particle momenta are not relevant), then we can drop \dagger operators and replace reversed transition probabilities with return probabilities in forward process. Furthermore, if the system is driven for a long time, it is reasonable to assume that there is no correlation between initial $\mathbf{x}(0)$ and final $\mathbf{x}(\tau)$ microstates of the system. Using equation (3.27), which states that when the system starts in equilibrium at $\lambda(0)$ and finishes at equilibrium $\Delta S_{tot}[x(t)] = W_d[x(t)]$, we can find that for arbitrary initial and final conditions, the total entropy production is

$$\Delta S_{tot}[x(t)] = \ln \frac{p_i(\mathbf{x}(0))}{p_{eq}^{\lambda(0)}(\mathbf{x}(0))} - \ln \frac{p_i(\mathbf{x}(\tau))}{p_{eq}^{\lambda(\tau)}(\mathbf{x}(\tau))} + W_d[x(t)], \quad (3.38)$$

where p_{eq}^λ is Boltzmann distribution at constant external field λ . Because dissipated work $W_d[x(t)]$ depends on the entire trajectory, when the system is driven for time τ longer than the typical time needed to lose the memory of the initial distribution over states, we can neglect any correlation between W_d , p_f , and p_i and write the logarithm of the ratio of transition probabilities as a sum of three terms

$$\ln \left[\frac{\pi_\tau(\mathbf{I} \rightarrow \mathbf{II})}{\pi_\tau(\mathbf{I} \rightarrow \mathbf{III})} \right] = - \ln \left[\frac{\left\langle \frac{p_f^\mathbf{II}}{p_{eq}^\mathbf{II}} \right\rangle_{\mathbf{II}}}{\left\langle \frac{p_f^\mathbf{III}}{p_{eq}^\mathbf{III}} \right\rangle_{\mathbf{III}}} \right] + \ln \left[\frac{\pi_\tau(\mathbf{II} \rightarrow \mathbf{I})}{\pi_\tau(\mathbf{III} \rightarrow \mathbf{I})} \right] - \ln \left[\frac{\langle e^{-\beta W_d} \rangle_{\mathbf{I} \rightarrow \mathbf{II}}}{\langle e^{-\beta W_d} \rangle_{\mathbf{I} \rightarrow \mathbf{III}}} \right]. \quad (3.39)$$

The first term in equation (3.39) has a clear intuitive meaning: Since for any distribution p_f functional $-\ln\langle p_f/p_{eq} \rangle \leq 0$, and zero is reached only when $p_f \equiv p_{eq}$, this term measures the proximity of non-equilibrium distribution p_f to equilibrium distribution p_{eq} . If all other terms in equation (3.39) are equal, then the macrostate that is closer to equilibrium is more likely to be observed. To conclude, the first term merely indicates the general tendency of systems to evolve toward thermal equilibrium

and contains no information about the effect of the drive on the system's dynamics.

The interpretation of the remaining terms and the practical use of equation (3.39) seem to present two major difficulties. Our initial goal was to make a statement about the probabilities of forward transition between macrostates, and yet almost tautologically, in order to do this, we need to know the reversal probabilities, the calculation of which requires knowledge of all of the microscopic transition rates in the system. The second problem involves a well-known difficulty of calculating exponential averages of dissipated work that have dominant contribution from rare trajectories with extremely low values of dissipated work. This means that to answer our original question, we need to know the dissipation in stochastic events that are immeasurably unlikely to occur [97].

In the remaining part of this section, we will try to resolve both difficulties by narrowing down the set of *all* possible trajectories to a subset of *typical* trajectories that carry the most probability current from one macrostate to another. Importantly, all calculations that we have done so far to derive equation (3.39) are also applicable for the case when the averages are computed using any subset of possible paths. When we do the averaging over the typical forward paths, forward transition probabilities between the macrostates are almost not affected (i.e. $\pi_{\mathbf{I} \rightarrow \mathbf{II}}^{fwd} \simeq \pi_{\mathbf{I} \rightarrow \mathbf{II}}$), but reversal probabilities and exponential averages of dissipated work might change significantly. Indeed, the average $\langle e^{-\beta W_d} \rangle^{fwd}$ taken over the typical paths does not have a contribution from rare events, which means that the reversal probability might be $\pi_{\mathbf{II} \rightarrow \mathbf{I}}^{rev} \ll \pi_{\mathbf{II} \rightarrow \mathbf{I}}$. These simplifying assumptions allow us to rewrite equation (3.39) in a more compact form

$$\ln \left[\frac{\pi_{\mathbf{I} \rightarrow \mathbf{II}}^{fwd}}{\pi_{\mathbf{I} \rightarrow \mathbf{III}}^{fwd}} \right] = -\Delta \ln \left\langle \frac{p_f}{p_{eq}} \right\rangle_{\mathbf{II}, \mathbf{III}} + \ln \left[\frac{\pi_{\mathbf{II} \rightarrow \mathbf{I}}^{rev}}{\pi_{\mathbf{III} \rightarrow \mathbf{I}}^{rev}} \right] - \ln \left[\frac{\langle \exp(-\beta W_d) \rangle_{\mathbf{I} \rightarrow \mathbf{II}}^{fwd}}{\langle \exp(-\beta W_d) \rangle_{\mathbf{I} \rightarrow \mathbf{III}}^{fwd}} \right]. \quad (3.40)$$

If we further notice that $-\ln \langle e^{-\beta W_d} \rangle$ is a cumulant generating function for dissipated work and expand it, $-\ln \langle e^{-\beta W_d} \rangle = \beta \langle W_d \rangle - \frac{\beta^2}{2} \sigma_{W_d}^2 + \frac{\beta^3}{6} \langle W_d \rangle_c^3 \dots \equiv \Psi - \Phi$, we can introduce two new quantities: the average of dissipated work $\Psi = \beta \langle W_d \rangle$, and $\Phi = \ln \langle e^{-\beta W_d} \rangle + \Psi$, which accounts for all fluctuations around the average. It

should be noted that Φ is a non-negative $\Phi \geq 0$, because $e^{-\Psi+\Phi} = \langle \exp(-\beta W_d) \rangle \geq \exp(-\beta \langle W_d \rangle) = e^{-\Psi}$. The intuitive meaning of Φ can be acquired from the systems in the non-equilibrium steady state. In these systems, $\langle e^{-\Delta S_{tot}} \rangle = 1$, and $\Psi = \Phi$. Because in steady state the average rate of dissipation can take any value but is always counter-balanced by fluctuations associated with cyclical motion, the presence of Φ in the equation for relative probabilities points to the idea that contribution to dissipation coming from cyclical motion should be ignored. This idea is closely related to the concept of housekeeping heat introduced previously [90, 85]. By incorporating these two new quantities into equation (3.39), we obtain

$$\ln \left[\frac{\pi_{\mathbf{I} \rightarrow \mathbf{II}}^{fwd}}{\pi_{\mathbf{I} \rightarrow \mathbf{III}}^{fwd}} \right] = -\Delta \ln \left\langle \frac{p_f}{p_{eq}} \right\rangle_{\mathbf{II}, \mathbf{III}} + \ln \left[\frac{\pi_{\mathbf{II} \rightarrow \mathbf{I}}^{rev}}{\pi_{\mathbf{III} \rightarrow \mathbf{I}}^{rev}} \right] + \Delta \Psi_{\mathbf{II}, \mathbf{III}}^{\mathbf{I}, fwd} - \Delta \Phi_{\mathbf{II}, \mathbf{III}}^{\mathbf{I}, fwd}. \quad (3.41)$$

This equation can be thought of as a generalization of Helmholtz free energy for finite time stochastic evolution of driven systems. In the next chapter, we will demonstrate how this relation can be used to provide a physical explanation of adaptation phenomenon. In particular, we will look at the toy models of thermodynamics of competing self-replicators and the random hopping of a single particle in time-varying energy landscapes.

When we derived equation (3.41), we restricted calculation of average to typical paths. However, we did not explain how these paths can be identified. One plausible approach might be to consider a procedure in which some finite number $\mathcal{N} \gg 1$ of paths are randomly drawn from the distribution of forward paths leading from one macrostate to another. Such a procedure implicitly excludes “abnormal” types of events unlikely to occur in \mathcal{N} experiments. It should be noted that the distribution of work obtained in such a way may still be non-Gaussian and may have a significant contribution of higher-order cumulants to $\langle e^{-\beta W_d} \rangle$.

To conclude, it is informative to go thorough the physical principles that lie at the core of equation (3.41) and to give a brief interpretation of the individual terms in this equation. Like many fluctuation theorems for far-from-equilibrium systems, equation (3.41) is based on the conservation of energy and microscopic reversibility.

The first term in this equation describes the general tendency of physical system to equilibrium, the second term reflects the effects of kinetics, and the last two terms express the effect of the drive on stochastic dynamics: Reliable dissipation described by Ψ makes forward transitions more likely, while fluctuations in dissipation due to cyclical motion Φ have the opposite effect.

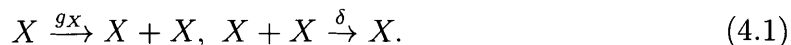
Chapter 4

Thermodynamics of adaptation

In Chapter 3, we derived fluctuation relations (3.37, 3.39, 3.41) that can be thought of as a generalization of the Helmholtz free energy for finite-time stochastic evolution of driven systems. In this chapter, we will use these relations to provide a potential explanation of adaptation in far-from-equilibrium systems. We will start in a familiar setting and consider a toy model of self-replication. Then we will illustrate how average dissipated work and its fluctuations affect the random motion of a single particle in a time-varying energy landscape. Finally, we will discuss the implications of our physical model of adaptation for biological systems.

4.1 Toy model of self-replication

Let us consider a system consisting of two populations of self-replicating particles, A and B . The coarse-grained state of this system can be specified by an ordered pair (N_A, N_B) , which counts the number of particles of each type. We will assume that the particles of both types can spontaneously produce copies of themselves and undergo a reversal process when two particles of the same type combine into one particle. In other words, the self-replication rules are described by two chemical reactions:



In principle, the rates of these reactions g and δ might depend on the positions and momenta of the particles. However, we will neglect by this dependence, assuming that concentration of particles is low and that the system is well mixed. In addition, we will assume that the rate of reversal reaction δ is the same for both types, and without loss of generality set $g_B > g_A$. It should be noted that the replication rules (4.1) that we have chosen and time-reversibility imply that the minimum number of particles of each type in the system is one. In order to have zero particles in the system, we need to introduce processes of spontaneous particle creation from a vacuum and single-particle annihilation. At the end of this section, we will discuss the impact of these processes on the thermodynamics of the system.

Because particles of different types do not directly interact with each other, the probability of finding N particles of type X in the system obeys the master equation

$$\dot{p}_{N_X}(t) = - \left(gN_X + \delta \frac{N_X(N_X - 1)}{2} \right) p_{N_X}(t) + g(N_X - 1)p_{N_X - 1}(t) + \delta \frac{(N_X + 1)N_X}{2} p_{N_X + 1}(t). \quad (4.2)$$

In the calculations that follow, we will be interested in the limit that occurs when reversal reactions are extremely rare $\delta \rightarrow 0$ but δ remains non-zero so that the entropy production in replication event $N_X - 1 \rightarrow N_X$ remains finite

$$\Delta S_X = - \ln \frac{p[N_X \rightarrow N_X - 1]}{p[N_X - 1 \rightarrow N_X]} = \ln \left[\frac{2g_X}{\delta} \right] - \ln N_X. \quad (4.3)$$

Furthermore, we will consider the regime in which the average number particles grows exponentially (i.e. $1 \ll N_X \sim e^{g\tau} \ll N_X^* = 2g/\delta$, where N_X^* is the average number of particles of types X in the steady state).¹ Our goal is to demonstrate that, in this regime, the nonequilibrium dynamics of the system can be predicted by comparing the total entropy production within ensembles of macrostates that have the same value of reversal probability.

The probability of forward process in which the system initially is in state $(1, 1)$ and at time τ is in state (N_A, N_B) can be found analytically from equation (4.2) if

¹The solution of deterministic equation associated with equation (4.2) and the steady-state distribution are presented in Appendix B.

we neglect by terms that are proportional to δ

$$p[(1, 1) \rightarrow (N_A, N_B), \tau] = e^{-g_A\tau}(1 - e^{-g_A\tau})^{N_A-1}e^{-g_B\tau}(1 - e^{-g_B\tau})^{N_B-1}. \quad (4.4)$$

In a sense, this approximation is equivalent to restricting our attention to typical paths along which the number of particles grows monotonically in time. The next logical step is to compute the probability of the reversed process. This can be done in two different ways. Conceptually, the most straightforward way is to take the sum of probabilities of all paths along which the number of particles monotonically decreases (Appendix B). However, this approach is mathematically challenging, and one can obtain the same result by exploiting Crooks-type relation (3.32)

$$\ln \frac{p[(1, 1) \rightarrow (N_A, N_B), \tau]}{p[(N_A, N_B) \rightarrow (1, 1), \tau]} = \Delta S(1 \rightarrow N_A) + \Delta S(1 \rightarrow N_B) \quad (4.5)$$

where the total entropy production $\Delta S(1 \rightarrow N_X)$ can be found by adding together all contributions from individual transitions² (4.3)

$$\Delta S(1 \rightarrow N_X) = (N_X - 1) \ln \left[\frac{2g_X}{\delta} \right] - \ln N_X!. \quad (4.6)$$

Combining these expressions with the one for $p[(1, 1) \rightarrow (N_A, N_B), \tau]$, we find that the probability of the reversed process occurring is

$$p[(N_A, N_B) \rightarrow (1, 1), \tau] = \prod_{X=\{A,B\}} e^{-g_X\tau}(1 - e^{-g_X\tau})^{N_X-1}e^{-g_X\tau}(1 - e^{-g_X\tau})^{N_X-1}N_X! \quad (4.7)$$

In principle, the space of all possible outcomes of the self-replication process described above consists of all ordered pairs ($N_A \geq 1, N_B \geq 1$) and the probability of outcomes spans a large range of values. However, in this space, there exist a certain region of macrostates that have the same reversal probability. In the limit of small δ ($\ln g_X/[N_X\delta]$, $N_X \geq 1$) and comparable order-of-magnitude growth

²Here, we used the fact that total entropy production is the same for all paths that have the same starting and ending points because the transition rates g, δ are constant, and all paths are topologically identical to $1 \rightarrow 2 \rightarrow \dots \rightarrow n-1 \rightarrow n$.

rates $g_A \sim g_B \sim \frac{1}{\tau} \ln N_X$, these regions can be well approximated by the lines $N_A + N_B = \text{const}$. On these lines, entropy production takes the simple form

$$\Delta S \approx C'' + N_B (e^{-g_A \tau} - e^{-g_B \tau}), \quad (4.8)$$

where C'' is a constant independent of N_A and N_B .³ Figure 4-1 shows the variation of entropy production along these lines for $g_B > g_A$. Because in this case the expression in parentheses in the equation above is always positive, the entropy production increases linearly with N_B along each line. Therefore, the relative probability of states on lines of fixed reversal probability grows exponentially with N_B , and we recover an intuitive notion that in the system where the total number of particles is bounded from below, the most likely outcomes are those where $N_B \gg N_A$.

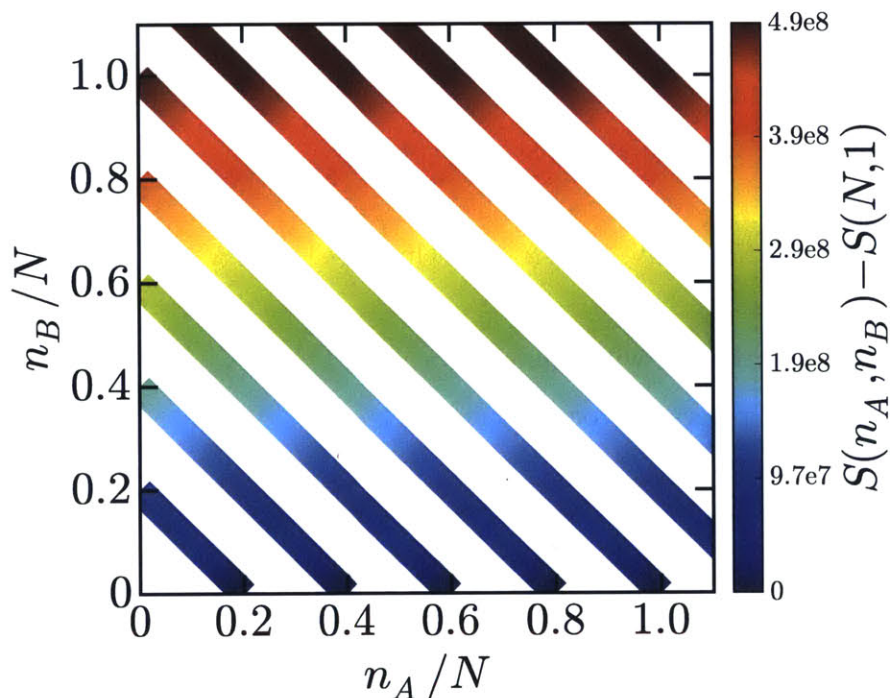


Figure 4-1: Variation of total entropy production on the surfaces of fixed return probability. $g_A = 1$, $g_B = 2$, $\delta = e^{-60}$, $\tau = 10$. $N = e^{20}$.

Certainly, the result that we obtained is not surprising — without doing any cal-

³The derivation of this result and the relation between N_A and N_B on the surfaces of fixed return probability is presented in Appendix B.

culations, one can immediately say that in the competition between two populations where the particles are almost never destroyed, the particles that divide more quickly are expected to be more numerous after a long enough time. However, our goal here was to provide the explanation of this result using a thermodynamic approach that can be applicable in other situations, where the stochastic dynamics of the system is more complicated.

Finally, let us consider an extra replication rule that allows the spontaneous birth and death of particles ($X \leftrightarrow 0$). Let g' and δ' be the rates of these processes. One might assume that the addition of this rule will complicate the calculation of the exponential average $\langle e^{-\Delta S_{tot}} \rangle$ because of the diversity of paths with different values of entropy production associated with a transition from $(1, 1)$ to (N_A, N_B) . However, in the limit of $g'_X \simeq \delta' \ll g_X$, this calculation simplifies significantly — we can take the sum over the *typical* paths where spontaneous birth and death processes in forward direction are ignored, so in calculation of reversal probability, we do not have to consider these processes. Therefore, as long as the spontaneous birth and death of particles are rare, we recover previous results (4-1). To conclude, this scenario clearly demonstrates the advantages of summing only over the typical paths in equation (3.41).

4.2 Drift and diffusion in driven energy landscapes

In this section, we will consider a simple, stochastic model of a single particle hopping in a discrete landscape of energy states. Following the Markovian dynamics framework introduced in Section 3.2, we assume that each state i had an energy E_i , and each pair of states i and j is separated by an activation barrier $B_{ij} = B_{ji}$. To introduce the notion of the heat bath and the temperature $T \equiv 1/\beta$, we will further assume that at any moment of time, the transition rate $r_{i \rightarrow j}$ of jumping for state i to state j obeys time-dependent Arrhenius law (3.11)

$$r_{i \rightarrow j} = r_{ij}^0 e^{-\beta(B_{ij} - E_i)}, \quad (4.9)$$

where $r_{ij}^0 = r_{ji}^0$ is a rate constant that stays constant when E_i , E_j , and B_{ij} are varied. It is worth noting that this form of transition rates implies that at any moment of time, the instantaneous steady-states distribution is a Boltzmann $p_{ss}(i) = p_{eq}(i) \propto e^{-\beta E_i}$ and that the system obeys detailed balance condition $J_{i \rightarrow j}^{ss} = p_{ss}(i)r_{i \rightarrow j} - p_{ss}(j)r_{j \rightarrow i} = 0$. In the presence of an external drive that couples to the barrier and state energies, the thermodynamic quantities, such as the work done on the system and heat dissipated into the surrounding bath, are defined in a standard way (3.12).

In principle, there are many physical phenomena that can be studied using this framework, but our primary interest in this section is to demonstrate how the generalized Helmholtz free energy (3.41) can be used to explain how fluctuations and dissipation affect the flow of probability in stochastic evolution. The differences in mean dissipated work Ψ and its fluctuations Φ determine the relative likelihood of two outcomes only if these outcomes have the same “distance” to equilibrium ($\ln\langle p_f/p_{eq} \rangle$) and return probability π^{rev} . In the case of the single particle that we are considering, the first term is always zero $-\ln\langle p_f/p_{eq} \rangle = 0$ because each microstate is assumed to have zero entropy.

Figure 4-2 shows an example of a system where the return probability can be tuned to be the same for certain states. This system consists of only three states arranged in a row so that $r_{12}^0 = r_{23}^0 = r$ and $r_{13}^0 = 0$. In the absence of the external driving field, all states have the same energy $E_i = 0$, and the barrier heights between adjacent states are equal $B_{12} = B_{23} = \Delta E > 0$. When at time $t = 0$ the particle is in state 2, then in the absence of an external drive, the probabilities of hopping to the right and to the left are the same for any time interval τ because of the symmetry. For the same reason, the probability of returning from 1 to 2 in time τ is the same as that of transitioning from 3 to 2. Now let us consider a system that is driven in such a way that $E_1(t) = -\Delta \cos(\omega t)/2$ and $B_{12} = \Delta E - \Delta E \cos(\omega t)/2$, and the energies of other states and activation states remain the same. In general, in order to compute forward and return probabilities, we need to solve a time-dependent master equation, but in the limit of fast drive and short times $\omega \gg 1/\tau \gg r$, this calculation significantly simplifies. On this time scale, the transition probability from one state to another

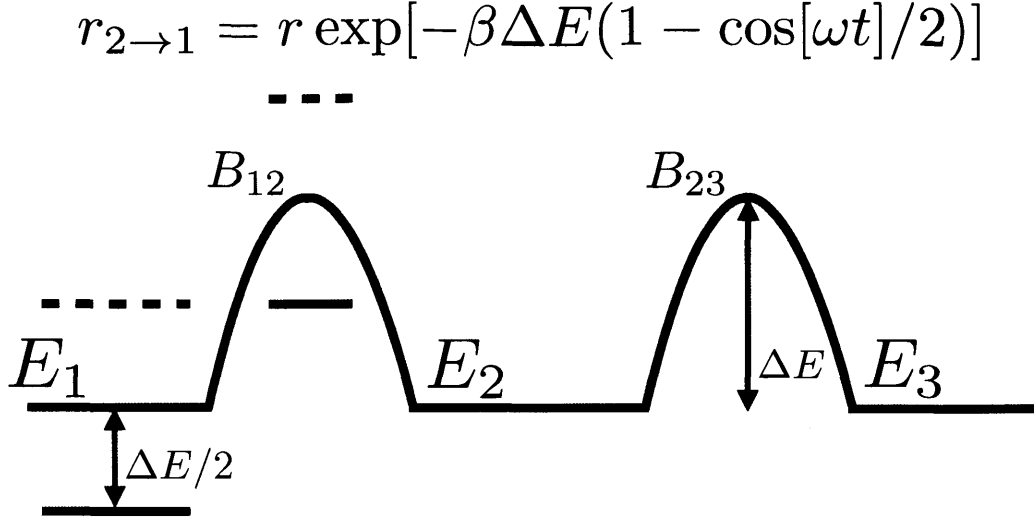


Figure 4-2: Three microstates in the absence of drive have the same energy $E = 0$, and the heights of the barriers between adjacent microstates is $B_{ij} = \Delta E$. In the presence of the drive, the barrier and state energies are varied in such a way that the transition rates $r_{3 \rightarrow 2}$, $r_{2 \rightarrow 3}$, and $r_{1 \rightarrow 2}$ are not changed, while the transition rate from state 2 to 1 is $r_{2 \rightarrow 1} = r \exp[-\beta \Delta E (1 - \cos[\omega t]/2)]$.

can be found as $\pi_\tau[i \rightarrow j] \simeq \int_0^\tau r_{i \rightarrow j}(t) dt$. Using $r_{1 \rightarrow 2}(t) = \exp[-\beta \Delta E] = r_{3 \rightarrow 2}(t)$, we find that in the limit of interest, the drive has no effect on return probabilities

$$\pi_\tau[1 \rightarrow 2] \approx \pi_\tau[3 \rightarrow 2] \approx r\tau e^{-\beta \Delta}.$$

In other words, for the driving protocol that we have chosen, the particle is still equally likely to return from states 1 and 3 to state 2.

The consequences of fixing the return probability and the impact of the drive on the systems dynamics can be recognized, if we consider the forward probabilities $\pi_\tau[2 \rightarrow 1]$ and $\pi_\tau[2 \rightarrow 3]$. It should first be noted that our analyses are valid only between points in time between which the driving protocol is time-symmetric (e.g, $t = 0$ and $t = \tau = 2\pi n/\omega$). The rate of hopping to the right does not change in time $r_{2 \rightarrow 3} = r e^{-\beta \Delta E}$, whereas the rate of leftward hops is periodically attenuated and amplified because of oscillations of the barrier energy $B_{12}(t)$

$$r_{2 \rightarrow 1} = r \exp[-\beta \Delta E (1 - \cos[\omega t]/2)]. \quad (4.10)$$

In the limit of strong drive $\beta\Delta E \gg 1$, the leftward transition is much more likely to happen when the barrier $B_{12}(t)$ reaches minimum, and in this limit, the ratio of leftward to rightward transition probabilities is

$$\frac{\pi_\tau[2 \rightarrow 1]}{\pi_\tau[2 \rightarrow 3]} \approx \frac{r_{2 \rightarrow 1}^{max}}{r_{2 \rightarrow 3}} = e^{\beta\Delta E/2} > 1. \quad (4.11)$$

A more general result that is valid for any strength of the drive can be obtained if we average transition rates over one period of drive and recall that $\pi_\tau[2 \rightarrow 1]/\pi_\tau[2 \rightarrow 3] = 1/\tau \int_0^\tau e^{\beta\Delta E \cos(\omega t)/2} dt = I_0(\beta\Delta E/2) > 1$, where I_0 is the modified Bessel function of the first kind.

During the moments of time when the particle is most likely to make a transition from state 2 to state 1, the energy of state 1 takes its minimum value $-\beta E/2$, so leftward transitions are typically accompanied by a positive entropy production $\Psi_{2 \rightarrow 1} = \beta \langle \Delta Q \rangle_{2 \rightarrow 1} \simeq \beta\Delta E/2$. In contrast, the rightward transitions occur with no dissipation because $E_2 = E_3$, so $\Psi_{2 \rightarrow 3} = 0$. In more general case, one can show that $\langle e^{-\Delta S_{tot}} \rangle_{2 \rightarrow 1} = 1/I_0(\beta\Delta E/2)$. Therefore, if we now look back on equation (3.41), we will immediately recover that

$$\frac{\pi_\tau[2 \rightarrow 1]}{\pi_\tau[2 \rightarrow 3]} \approx \frac{r_{2 \rightarrow 1}^{max}}{r_{2 \rightarrow 3}} = e^{\Delta\Psi}. \quad (4.12)$$

This equation can be interpreted as the general tendency of a particle to move with higher probability in the direction where it dissipates more heat into the environment (i.e, particle drift and the mean dissipated heat are two profoundly connected quantities).

A similar system shown on Figure 4-3 can be examined to demonstrate the effect of fluctuations Φ on drift. This system consists of states 1 and 2, which in the absence of drive have the same energy $E_1 = E_2 = 0$. Unlike the previous example, where there was only one possible way to make a jump from one state to another, here, we assume that there are two different hopping paths that connect states 1 and 2. There might be different physical mechanisms to realize this scenario, but the simplest one is connecting the system to two heat baths. We assume that the external driving

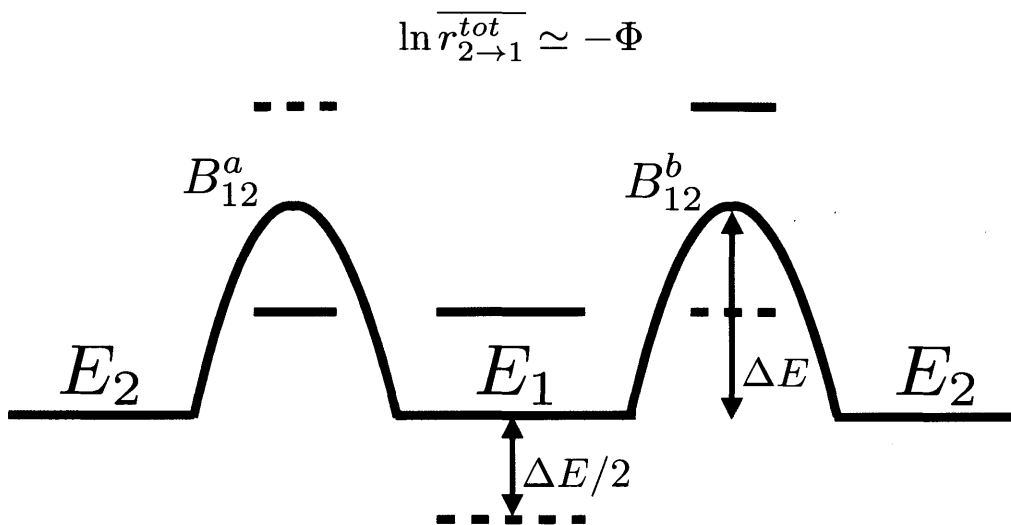


Figure 4-3: An example of system where the drift is suppressed by fluctuations Φ .

field couples to the system so that the barrier heights are changed over time in the following way

$$B_{12}^a(t) = \Delta E - A \cos(\omega t) \text{ and } B_{12}^b(t) = \Delta E + A \cos(\omega t),$$

where A is the amplitude of the drive, the exact value of which we will specify later. The energy of the second state is not affected by the drive, while the energy of the first state varies as $E_1(t) = A \cos(\omega t)$.

In the regime of strong $\beta A \gg 1$ and fast $r \ll 1/\tau \ll \omega$ drive, the transitions between states are rare and occur during the moments of time when transition rates have maximum value. First, we assume that at time $t = 0$, the system is in state 2, and we consider the forward process of hopping to state 1 in time τ through each path. Both transition rates $r_{2 \rightarrow 1}^a$ and $r_{2 \rightarrow 1}^b$ change over time, but the integral of this rate over integer number of driving cycles is the same. Thus, the forward transition probabilities over each path are

$$\pi_\tau[2 \xrightarrow{a} 1] = \pi_\tau[2 \xrightarrow{b} 1] \simeq r\tau e^{-\beta\Delta E} I_0(\beta A) \simeq r\tau e^{-\beta(\Delta E - A)}. \quad (4.13)$$

Next, we estimate the exponential average of dissipated heat for forward process. One

can notice that when a particle jumps to state 1 over the barrier a , the energy of the first state reaches its maximum value $E_1^{max} = A$; thus, this transition is associated with negative heat $\Delta Q^a = -A$ dissipated into the environment. In contrast, the transition through path b happens when the energy of the first state is minimal $E_1^{min} = -A$, and the system dissipated positive heat $\Delta Q^b = A$ into the bath. Combining ΔQ^a and ΔQ^b , we promptly find that the average total entropy production is zero, $\Psi = \beta(\Delta Q^a + \Delta Q^b)/2$, and that the fluctuations are $\Phi = 0 - (-\ln\langle e^{-\Delta S} \rangle) = \ln \cosh(\beta A) \approx \beta A$. Finally, we will determine the probability of the reversed process of coming back from state 1 to state 2. The transition rate $r_{1 \rightarrow 2}^a = r e^{-\beta \Delta E} = \text{const}$ is not affected by the drive, whereas the averaged transition rate through path b is amplified by the drive $\overline{r_{1 \rightarrow 2}^b} = r e^{-\beta \Delta E} I_0(2\beta A) \simeq r e^{-\beta(\Delta E - 2A)}$. The latter implies that we can ignore the return transitions over barrier a . Furthermore, if we set $A = \Delta E/2$, we find that the return probability $\pi_\tau[1 \rightarrow 2] \simeq r\tau$ is independent of drive. For this value of driving amplitude A , the fluctuations of total entropy production are

$$\Phi \approx \beta \Delta E/2 = -\ln \overline{r_{2 \rightarrow 1}^{tot}}. \quad (4.14)$$

This expression explicitly shows that fluctuations in entropy production decrease the forward transition rate when the return probability and average dissipation are held constant. The origin of this effect is in the diversity of possible ways to jump between the states: Not only can the drive accelerate forward transitions along some paths, but it also can speed up return transitions along other paths, resulting in dissipation that is not associated with directed motion. An extreme example of the drive that acts symmetrically on forward and return transition rates is no-pumping theorem [98], which claims that in the NESS, the current integrated over one driving period is non-zero if and only if both barrier and state energies are driven.

It should also be noted that, in a low dimensional system, in order to hold the return probability fixed, we had to change the transition rates in an extremely orchestrated way so that we had freedom in controlling other parameters. However, in a high dimensional system, we expect that tuning landscape parameters such as the

local drift velocity and diffusion coefficient would be much easier because one would be fixing small number quantities in a system with many degrees of freedom.

The examples considered in this section demonstrate the twofold effect of the drive on single-particle motion in time-varying landscapes. On one hand, the drive can cause the particle to drift in the direction associated with reliable dissipation. On the other hand, the drive can result in cyclic motion and futile dissipation. In the next section, we will discuss how these ideas can be generalized to other systems.

4.3 Discussion

The non-equilibrium systems described by equation (3.41) have a vast range of physical properties, but not all of these systems exhibit behavior that we call adaptation. In this chapter we analyzed the thermodynamics of self-replication and random motion in driven energy landscapes. The former example demonstrates the interplay between entropy production and reversal probability in determining outcomes of the stochastic evolution. In addition, the toy model of self-replication shows that adaptation through Darwinian selection can be thought of as a special case of physical mechanism of adaptation in which the system tends to evolve toward the regions of phase space where it can reliably absorb work from the drive and dissipate heat into the surrounding bath.

This interpretation of adaptation could have a broad application, but there are also certain subtle points related to it. Because equation (3.41) defines the probability of being in a particular state and using the dissipated work, in order to judge whether the system is well adapted, we need to know the initial state of the system and the history that brought it to the final state. Therefore, we expect this theory of adaptation to work only when, from the current configuration/structure of the system, one can infer how the system looked in the past. In a sense, this is analogous to Darwinian selection: When we see that species traits are adjusted well to the environment, we assume that the ancestors of this species had these traits as well.

Another example that supports our model of adaptation focuses on rearrangement

of a periodically driven spring network and its vibrational spectrum over time. In the NESS, this system adopts a dynamical structure, with the peak in the vibrational spectrum corresponding to the driving frequency. Furthermore, if the system is sufficiently large and is driven at multiple frequencies, then there might emerge structures with peaks in the spectrum that match each of driving frequencies. Physically, this means that the system can absorb more work from the drive when its natural frequency is close to the driving frequency. This behavior is in agreement with the recent experiments on silver nanorods that assemble into ring-like structures in the presence of light when their plasmon frequency matches the frequency of the laser beam [99].

An interesting observation consistent with the dissipative theory of adaptation has recently arisen in studies of self-organization in voltage-driven systems [100, 101]. In these experiments, conducting beads were placed in circular Petri dish filled with oil, and then high voltage was applied between two electrodes, one of which was located in the center of the dish and the other of which formed the walls of the dish. The viscosity of the oil was high so that on the time scale of the experiment, the diffusion of beads was negligible. In both experiments, the authors observed the formation of tree-like structures with statistically robust properties, such as number of termini or branch points. However, in the second experiment [101], the authors also tracked the dissipation rate in the system, and they found that the network of conducting beads evolves towards states of lower resistance and, thus, higher rates of entropy production. It is worth mentioning that the motion of the beads was collective, and regardless of the initial arrangement and boundary conditions, the final structure had lower resistance. Similar behavior was observed in the simulations of self-assembly of particles interacting through time-oscillatory potentials [102]. The simulated systems consisted of two types of charged particles whose charge was controlled by the pH of the surrounding solvent. When the period of pH oscillations was fast compared to particle diffusion, the formation of steady-state structures was observed. Depending on the particle density and driving protocol, self-assembled structures with disordered, dimer, fiber, honeycomb lattice, and square lattice morphologies were formed. Interestingly, some of these structures, such as fibers, were not formed in the absence

of time-varying potential. When the driving period was long, disordered oscillating structures were observed. The transition from a disordered to an ordered phase corresponded to the maximum energy dissipated per period. In addition, the dissipative theory of adaptation could potentially be used to explain spontaneous motion in mixtures of multiple proteins driven by ATP or GTP hydrolysis [103, 104] and to provide an alternative way to design structures with high yield [105, 106].

To conclude, it is instructive to discuss the connection between the generalization of Helmholtz free energy (3.41) and variational principles that set constraints on entropy production rate. In 1947, Prigogine formulated and proved the *minimum entropy production theorem* for irreversible processes in the linear response regime [107]. In particular, this theorem states that in a system with purely diffusive dynamics, the total entropy production rate is at a minimum in a non-equilibrium steady state with respect to the variations of the generalized forces compatible with boundary conditions. In the past 40 years, there was a lot of argument about this principle and its application to biological systems. Recently, Maes and Netocny used the large deviation theory approach to devise more specific criteria of when minimum entropy production principle is valid [108]. They demonstrated that this principle breaks for systems that are either non-linear or non-even under time reversal. The counterexample first proposed by Landauer clearly illustrates these scenarios [109]. In the electrical circuit where a resistor R and conductance L are connected in series and driven by applying a constant voltage V , the steady state current is $I = V/R$ and the entropy production rate is $\sigma = RI^2/T$, where T is the temperature of the surrounding bath. One can easily see that the entropy production is not minimized by current $I = V/R$. The important point in this example is that the generalized thermodynamic force, current I , is odd under time-reversal. A more general criticism of minimum entropy production principle and its validity for temperature-inhomogeneous systems was outlined by Jaynes [110]. It is worth emphasizing that, unlike Prigogine's minimum production principle, the generalization of the Helmholtz free energy (3.41) for finite time stochastic evolution is valid arbitrarily far from equilibrium and does not depend on the symmetry of generalized forces.

A different variational principle states that, in steady-state, the mean rate of entropy production is maximum. This principle is based on the application of Jaynes formalism of statistical mechanics [111] to path information entropy

$$S_I = - \sum_{\text{paths}} p_{\text{path}} \ln p_{\text{path}}.$$

By maximizing path information entropy, Dewar showed that the probability of a single path $x(t)$ is proportional to the exponential of the entropy production $p[x(t)] \propto \exp(\tau\sigma[x(t)])$ [112] (i.e, trajectories with higher mean entropy production rate are more probable). Similar results were obtained by others [113, 114]. However, the most legitimate derivation of this principle and discussion of how maximum and minimum entropy production principles reconcile with each other is due to Maes and colleagues [115]. In particular, they show that in linear systems, minimization of entropy production is associated with even under time reversal variables, and maximization is associated with odd. In some sense, the mean dissipated work Ψ in equation (3.41) resembles the maximum entropy production principle; however, in our framework of dissipative adaptation, we also have to consider the opposite effect of fluctuations Φ and the difference in kinetic accessibility of macrostates. Finally, it should be repeated that the central result of our theory of adaptation, equation (3.41), was derived for arbitrarily strongly driven systems.

Chapter 5

Outlook

In previous chapters, we used statistical physics to develop models of protein folding and adaptation to time-varying fields. Here, we will briefly discuss the limitations of these models and suggest future research directions.

5.1 Phenomenological models of protein folding

In Chapter 2, we presented a burial mode model of protein folding that allows us to study the sequence-structure relationship in globular protein domains. In particular, this model takes into account hydrophobic effect, polymeric bonds, and steric repulsion and allows us to compute the squared distance from each residue to the center of mass of the protein, called the “burial trace”. By analyzing fluctuations of burial traces in proteins, we establish a method of identifying ligand-binding and protein-protein interaction sites. Unfortunately, this analysis cannot be applied to a broad class of proteins because of the approximate nature of the model.

To reduce the minimization of Hamiltonian (A.1) that is subject to constraints (A.2, A.3) of a linear programming problem, we made a crucial approximation (A.7). To understand how this approximation affects the performance of the model, it might be interesting to perform Monte Carlo or molecular dynamics of Hamiltonian (A.1). Importantly, these simulations can provide an ultimate answer for why the model succeeds on some proteins and fails on others — the simulations will show us whether

the cause of failure is the lack of physical factors taken into account in burial mode model or whether the source of error is approximation (A.7).

In the process of parameter optimization of the burial mode model, we devised a new way to measure the hydrophobic effect: From the matrix of relative positions M_{ij} , an example of which is shown on Figure 2-2, one can obtain a set of 190 relative hydrophobicities of amino acids $\Delta\varphi_{ij}$. In Chapter 2, we used these numbers to obtain a new hydrophobicity scale ϕ_i for the burial mode model. However, in principle, one can use all 190 relative hydrophobicities to model the hydrophobic effect. For example, one can consider a class of models in which the energy of the protein conformation is

$$\mathcal{H} = \sum_s \left[\kappa |\vec{r}(s+1) - \vec{r}(s)|^2 + \Delta\varphi_{s,s+1} f(|\vec{r}(s+1) - \vec{r}(s)|) \right], \quad (5.1)$$

where $f(x)$ is some function of a scalar argument. It is worth noting that these models have the same physical assumptions as burial mode model, but because of the freedom in choosing $f(x)$, one might be able to compute the analytical expression for the partition function of Hamiltonian (5.1).

5.2 Thermodynamics of self-replication

In Chapters 3 and 4, we addressed the question of adaptation from the standpoint of physics. By employing the Crooks microscopic reversibility relation, we derived the expression for the relative probability of two macroscopic outcomes of stochastic evolution of driven Newtonian matter. By analyzing this expression term by term, we demonstrated the general tendency of the systems of self-replicators to evolve toward the regions of the phase space associated with the state formed through reliable absorption and dissipation of energy from the heat bath.

For a system of two non-interacting, self-replicating species, we showed that keeping the term that describes return probability in equation (3.41) constant corresponds to a simple constraint on the total number of particles in the system, $n_A + n_B \approx N$. An interesting avenue to pursue might be studying the class of self-replicating sys-

tems in which interactions between different species are allowed. These studies can provide a thermodynamic explanation of equations describing population growth (e.g, Lotka-Volterra or logistic equations). Furthermore, a deeper understanding of population dynamics can be obtained by comparing the theory of adaptation presented in Chapters 3 and 4 to Fisher's fundamental theorem of natural selection [116] and Price's equations [117].

Appendix A

Burial Mode Model (BMM)

A.1 Basic Assumptions and Approximations of BMM

In the burial mode model, a single domain protein is represented as a linear chain of N residues, which are indexed by number s and have position $\mathbf{r}(s)$ with respect to the center of mass of the protein. This model considers only three interactions that determine protein structure: polymeric bonds, hydrophobic interaction, and steric repulsion. The first two interactions are incorporated into the system energy directly, whereas steric repulsion is taken into account as a global constraint on the ratio α , between the radius of gyration of the protein chain, R_g , and its maximum size, R . Mathematically, these assumptions are expressed as:

- polymeric bonds and hydrophobic effect

$$\mathcal{H} = \int ds \left[\kappa \left| \frac{d\mathbf{r}(s)}{ds} \right|^2 + \varphi(s) |\mathbf{r}(s)|^2 \right], \quad (\text{A.1})$$

- steric repulsion and relation between R_g and R

$$\int ds |\mathbf{r}(s)|^2 = \alpha N R^2, \quad (\text{A.2})$$

- steric repulsion and constraint on maximum size

$$0 \leq |\mathbf{r}(s)|^2 \leq R^2. \quad (\text{A.3})$$

Because of the global constraints (A.2, A.3), it is difficult to obtain a closed form analytical expression for the partition function of the protein in BMM. Therefore, to compute the native state configuration of the protein, we have chosen to minimize the potential energy (A.1) that is subject to these constraints. The minimization procedure can be reduced to linear programming (LP) problem, if the following steps and approximations are made:

- diagonalize Hamiltonian (A.1)

$$\mathcal{H}\psi_k(s) = \epsilon_k\psi_k(s), \quad (\text{A.4})$$

- express $\mathbf{r}(s)$ in terms of eigenmodes $\psi_k(s)$

$$\mathbf{r}(s) = \left[\sum_k X_k\psi_k(s), \sum_k Y_k\psi_k(s), \sum_k Z_k\psi_k(s) \right], \quad (\text{A.5})$$

$$c_k = X_k^2 + Y_k^2 + Z_k^2, \quad (\text{A.6})$$

- neglect by cross-terms $\psi_i(s)\psi_j(s)$

$$|\mathbf{r}(s)|^2 = \sum_k c_k\psi_k^2(s) + \underline{\sum_{i,j} (X_iX_j + Y_iY_j + Z_iZ_j)\psi_i(s)\psi_j(s)}. \quad (\text{A.7})$$

The latter approximation is motivated by the fact that all protein conformations with given values of coefficients c_k have the same energy and radius of gyration and, thus, that these conformations can be represented by a subset of conformations for which cross-terms vanish.

Using the approximations described above, the minimization of the energy function (A.1) that is subject to constraints (A.2, A.3) can be written as:

$$\mathcal{H} = \sum_k c_k\epsilon_k \quad (\text{A.8})$$

$$\sum_k c_k = \alpha NR^2 \quad (\text{A.9})$$

$$0 \leq |\mathbf{r}(s)|^2 \approx \sum_k c_k\psi_k^2(s) \leq R^2 \quad (\text{A.10})$$

which is an LP problem with objective function (A.8) and constraints (A.9, A.10). The burial trace of the minimal energy configuration is

$$|\mathbf{r}(s)|^2 = \sum_k c_k^{opt} \psi_k^2(s), \quad (\text{A.11})$$

where c_{opt} is the solution of the LP problem.

A.2 BMM Model Parameters

The original parameter of BMM were set based on the following assumptions:

- Stiffness κ sets the length scale in the problem.

$$\langle |\mathbf{r}_s - \mathbf{r}_{s-1}|^2 \rangle = 3/2(\beta\kappa)^{-1} = 1 \quad (\text{A.12})$$

- The maximum size of the globule scales with the number of monomers N as

$$R = \left(\frac{3N}{4\pi\rho_o} \right)^{1/3} \quad (\text{A.13})$$

- The ratio of r_{rms}^2 to R^2 is set to 3/5 assuming uniform density of the protein.
- Kyte-Doolittle (KD) hydrophobicity scale (20 numbers), which is used to convert the primary sequence to the sequence of numbers $\varphi(s)$, is normalized so that $\varphi(\text{Gln})R^2 = 0.5 k_B T$.

A.3 Brute Force Search of 4-letter Hydrophobicity Scale

Figures A-1 and A-2 show the mean and the variance of the distribution of Pearson correlation coefficient (PCC) between the burial traces predicted by BMM with 4-letter alphabet and burial traces computed from crystal structures. In the former letter, the amino acids are grouped together according to KD hydrophobicity scale; in the latter, the groups are random. For random grouping, maximum Mean(PCC) does not exceed 0.2.

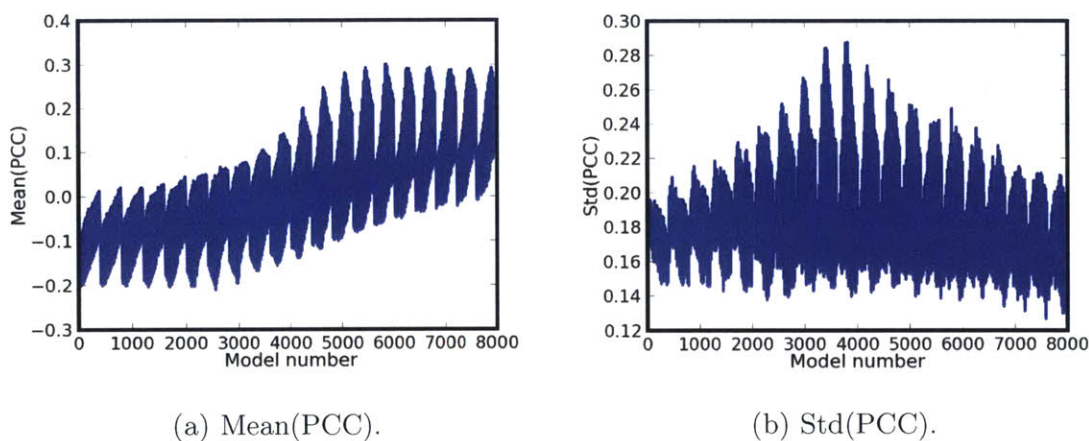


Figure A-1: The mean and standard deviation of the distributions of PCC for 200 α -helical proteins when amino acids are divided into groups according to KD scale. The model number is computed as follows: $m = ind_1 + N_1 \cdot ind_2 + (N_1 N_2) \cdot ind_3 + (N_1 N_2 N_3) \cdot ind_4$, where N_i is the number of values along i -th axis.

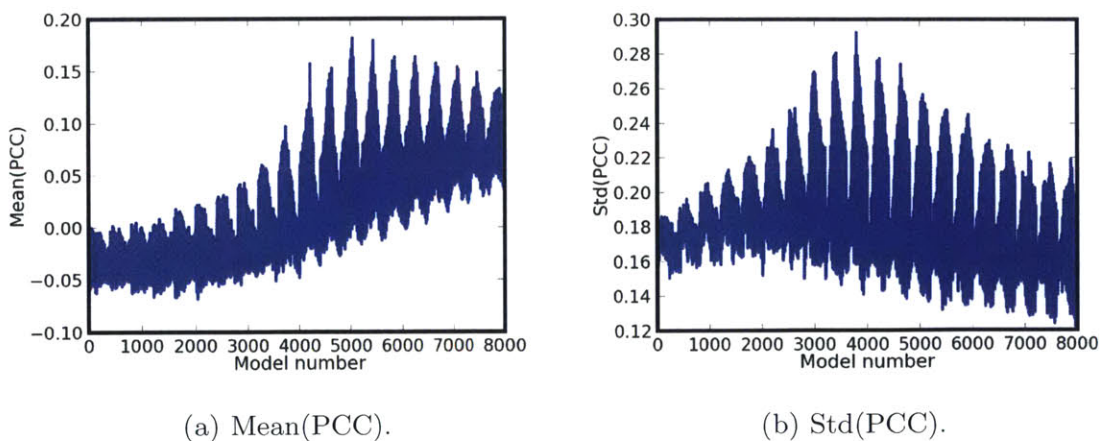
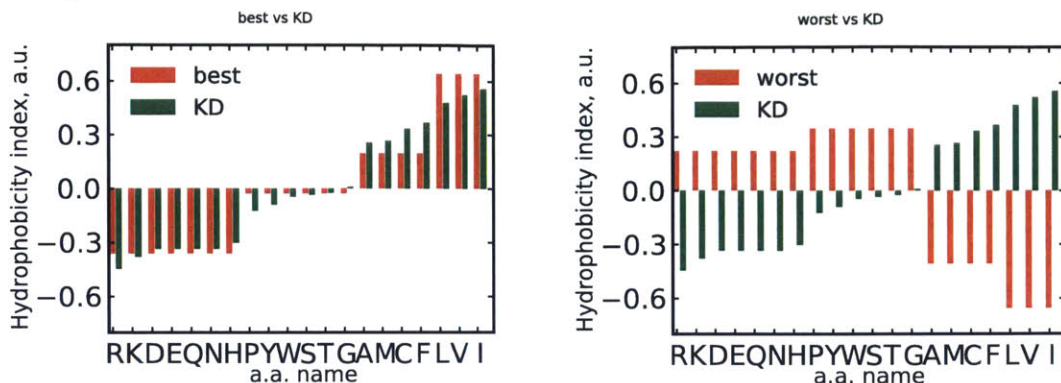


Figure A-2: The mean and the standard deviation of the distributions of PCC when amino acids are divided into random groups.

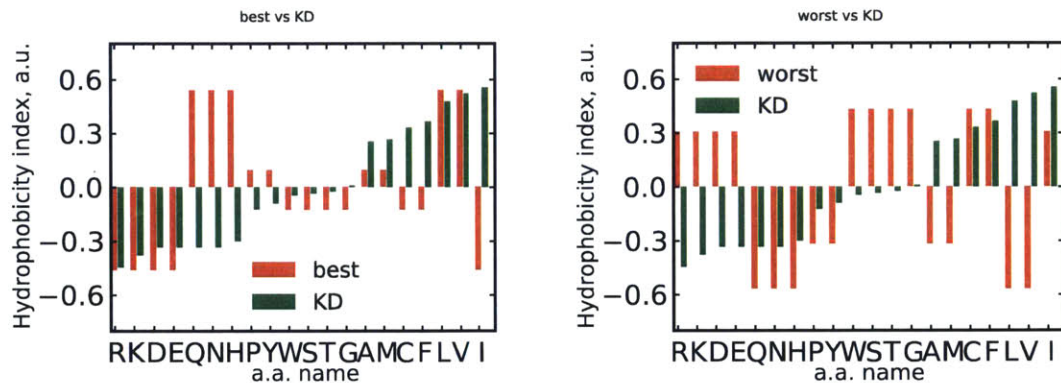
Figures A-3 and A-4 show 4-letter hydrophobicity scales that provide the best and worst performance of the model.



(a) Best vs KD.

(b) Worst vs KD.

Figure A-3: 4-letter hydrophobicity scales that provide the best and worst performance of the model on a large group of proteins, when amino acids are divided into groups according to their hydrophobicity indices in the KD scale.



(a) Best vs KD.

(b) Worst vs KD.

Figure A-4: 4-letter hydrophobicity scales that provide the best and worst performance of the model on large group of proteins, when amino acids were divided into random groups.

A.4 Conformational changes (statistical significance)

To estimate statistical significance of the results presented in the conformational changes section, we used the following test model. For the protein of length N , let F be a set of n residues responsible for performing specific function (ligand binding for the case of myoglobin (1BZP) and H-Ras (3K8Y), and serine protease activity for the case of chymotrypsinogen (1PYT D)); let V be a set of m residues with high value of structural variability computed using burial mode model, and let k be the size of the overlap $O = F \cap V$. Assuming that the set of residues performing specific function F is fixed and that the set of residues with high structural variability V is chosen at random, one can find that the probabilities of having an overlap O of size k and greater than k are given by:

$$\text{Prob}(|O| = k) = \binom{n}{k} \cdot \binom{N-n}{m-k} / \binom{N}{m}$$
$$\text{Prob}(|O| > k) = 1 - \sum_{i=0}^k \text{Prob}(|O| = i).$$

Thus, the p -value, defined as the probability of obtaining the overlap O of size at least k , is equal to $p = 1 - \text{Prob}(|O| \geq k)$.

The tables below show p -values for different parameters of the test model. For myoglobin and H-Ras, the residue was assumed to bind to the ligand if at least one of its atoms is located closer to the ligand than some distance d ; for chymotrypsinogen, the positions of residues that are involved in serine protease activity were taken from UNIPROT database (id: Q7M3E1); in some tests, we also treated first- and second-nearest neighbors as active sites. For all proteins, the set of residues V was determined using a cutoff on the normalized value of structural variability:

$$s \in V \text{ if } \text{var}[r^2(s)] / \langle \text{var}[r^2(s)] \rangle \geq \text{var}.$$

| $d, \text{\AA}$ | var | k | m | n | Prob($ O = k$) | Prob($ O > k$) | p-value |
|-----------------|-------|-----|-----|-----|-------------------|-------------------|----------------|
| 3.0 | 2.0 | 2 | 15 | 4 | 0.045 | 0.003 | 0.048 |
| | 2.5 | 2 | 9 | 4 | 0.017 | 0.001 | 0.017 |
| 4.0 | 2.0 | 4 | 15 | 17 | 0.052 | 0.014 | 0.066 |
| | 2.5 | 3 | 9 | 17 | 0.054 | 0.009 | 0.063 |
| 5.0 | 2.0 | 5 | 15 | 21 | 0.029 | 0.007 | 0.036 |
| | 2.5 | 4 | 9 | 21 | 0.019 | 0.003 | 0.021 |
| 6.0 | 2.0 | 5 | 15 | 28 | 0.079 | 0.034 | 0.112 |
| | 2.5 | 4 | 9 | 28 | 0.048 | 0.011 | 0.059 |

Table A.1: Myoglobin (1BZP). $N = 153$. $dE = 5 k_B T$. Binding sites are determined using a distance cutoff.

| sites | var | k | m | n | Prob($ O = k$) | Prob($ O > k$) | p-value |
|----------------------------|-------|-----|-----|-----|-------------------|-------------------|----------------|
| 44; 91; 92; 96 | 2.0 | 2 | 15 | 4 | 0.045 | 0.003 | 0.048 |
| | 2.5 | 2 | 9 | 4 | 0.017 | 0.001 | 0.017 |
| + first nearest neighbors | 2.0 | 3 | 15 | 10 | 0.051 | 0.009 | 0.060 |
| | 2.5 | 3 | 9 | 10 | 0.013 | 0.009 | 0.014 |
| + second nearest neighbors | 2.0 | 4 | 15 | 15 | 0.035 | 0.008 | 0.043 |
| | 2.5 | 4 | 9 | 15 | 0.005 | 0.000 | 0.006 |

Table A.2: Myoglobin (1BZP). $N = 153$. $dE = 5 k_B T$. Binding sites are determined to be nearest neighbors to the residues, located closer than 3\AA to the ligand.

| $d, \text{\AA}$ | var | k | m | n | Prob($ O = k$) | Prob($ O > k$) | p-value |
|-----------------|-------|-----|-----|-----|-------------------|-------------------|----------------|
| 3.0 | 2.0 | 2 | 22 | 10 | 0.262 | 0.131 | 0.392 |
| | 2.5 | 2 | 14 | 10 | 0.161 | 0.041 | 0.201 |
| 4.0 | 2.0 | 5 | 22 | 24 | 0.118 | 0.072 | 0.190 |
| | 2.5 | 4 | 14 | 24 | 0.090 | 0.034 | 0.123 |
| 5.0 | 2.0 | 9 | 22 | 31 | 0.006 | 0.002 | 0.008 |
| | 2.5 | 6 | 14 | 31 | 0.021 | 0.005 | 0.026 |
| 6.0 | 2.0 | 9 | 22 | 34 | 0.012 | 0.004 | 0.016 |
| | 2.5 | 6 | 14 | 34 | 0.031 | 0.010 | 0.041 |

Table A.3: H-Ras (3K8Y). $N = 166$. $dE = 5 k_B T$. Binding sites are determined using a distance cutoff.

| sites | <i>var</i> | <i>k</i> | <i>m</i> | <i>n</i> | Prob($ O = k$) | Prob($ O > k$) | p-value |
|---------------------------------|------------|----------|----------|----------|-------------------|-------------------|----------------|
| 14-17, 28, 29, 34, 59, 118, 145 | 2.0 | 2 | 22 | 10 | 0.262 | 0.131 | 0.392 |
| | 2.5 | 2 | 14 | 10 | 0.161 | 0.041 | 0.201 |
| + first nearest neighbors | 2.0 | 5 | 22 | 22 | 0.095 | 0.048 | 0.143 |
| | 2.5 | 4 | 14 | 22 | 0.071 | 0.023 | 0.094 |
| + second nearest neighbors | 2.0 | 7 | 22 | 34 | 0.080 | 0.050 | 0.130 |
| | 2.5 | 6 | 14 | 34 | 0.031 | 0.010 | 0.041 |

Table A.4: H-Ras (3K8Y). $N = 166$. $dE = 5 k_B T$. Binding sites are determined to be the nearest neighbors to the residues, located closer than 3 Å to the ligand.

| sites | <i>var</i> | <i>k</i> | <i>m</i> | <i>n</i> | Prob($ O = k$) | Prob($ O > k$) | p-value |
|----------------------------|------------|----------|----------|----------|-------------------|-------------------|----------------|
| 10-17, 57-61, 116-119 | 2.0 | 7 | 17 | 10 | 0.002 | 0.000 | 0.002 |
| | 2.5 | 6 | 14 | 17 | 0.001 | 0.000 | 0.001 |
| + first nearest neighbors | 2.0 | 9 | 22 | 23 | 0.001 | 0.000 | 0.001 |
| | 2.5 | 6 | 14 | 23 | 0.005 | 0.001 | 0.005 |
| + second nearest neighbors | 2.0 | 9 | 22 | 29 | 0.004 | 0.001 | 0.005 |
| | 2.5 | 6 | 14 | 29 | 0.015 | 0.004 | 0.019 |

Table A.5: H-Ras (3K8Y). $N = 166$. $dE = 5 k_B T$. Binding sites are taken from UNIPROT.

| sites | <i>var</i> | <i>k</i> | <i>m</i> | <i>n</i> | Prob($ O = k$) | Prob($ O > k$) | p-value |
|----------------------------|------------|----------|----------|----------|-------------------|-------------------|----------------|
| 74, 121, 216 | 2.0 | 1 | 30 | 3 | 0.280 | 0.038 | 0.319 |
| | 2.5 | 1 | 17 | 3 | 0.178 | 0.012 | 0.190 |
| + first nearest neighbors | 2.0 | 3 | 30 | 9 | 0.065 | 0.014 | 0.079 |
| | 2.5 | 3 | 17 | 9 | 0.015 | 0.001 | 0.017 |
| + second nearest neighbors | 2.0 | 5 | 30 | 15 | 0.018 | 0.004 | 0.022 |
| | 2.5 | 5 | 17 | 15 | 0.001 | 0.000 | 0.002 |

Table A.6: Chymotrypsinogen (1PYT). $N = 251$. $dE = 5 k_B T$. Binding sites are taken from UNIPROT.

A.5 Comparison of BMM and Regression Models

It is possible to construct a class of linear regression models (LRM) that are similar in spirit to BMM. Just as in BMM, one can represent the protein configuration by the distance from each amino acid to the center of mass of the protein and assume that each amino acid has a tendency to be at specific distance from the center of mass, $r_{opt}(\text{aa}) = r/R$. Moreover, one can assume that the effect of polypeptide bonds in LRM can be taken into account by averaging the preferred position $r_{opt}(\text{aa})$ over nearest neighbors. That is, the radial position of residue at position s in the sequence is

$$r(s) = \text{Mean}[r_{opt}[\text{aa}(s-n), r_{opt}[\text{aa}(s-n+1), \dots, r_{opt}[\text{aa}(s-n)]]], \quad (\text{A.14})$$

where $\text{aa}(s)$ is amino acid at position s , n is the number of nearest neighbors taken into account, and $\text{Mean}[x_1, x_2, \dots]$ is some weighted arithmetic or geometric mean function. Because parameters $r_{opt}(\text{aa})$ have physical meaning only when they are greater than zero, it is convenient to write these parameters as an exponent of “effective hydrophobicity”

$$r_{opt}(\text{aa}) = e^{-h_{opt}(\text{aa})}$$

and to compute radial positions as geometric mean

$$r(s) = \exp \left[-\frac{1}{2n+1} \sum_i h_{opt}[\text{aa}(i)] \right], \quad (\text{A.15})$$

where the sum is taken over nearest neighbors and n is the number of neighbors. Given this form of $r(s)$, one can easily find the parameters of the model $h_{opt}(\text{aa})$ using simple linear regression on $\ln r(s)$.

Figures A-5–A-8 show the results of the LRM described above for different groups of proteins. For each group, the optimal parameters were found from the training set consisting of 20 proteins, and then the model was tested as one whole group. In addition, the performance of the LRM on each group of proteins was compared to the

performance of BMM. Figure A-5 shows that LGM and BMM perform equally well on the group of myoglobin and that the optimal parameters $h_{opt}(aa)$ are similar to the KD hydrophobicity scale. For other groups of proteins (hemoglobin and immunoglobulin) the LRM works significantly better than BMM, but there is almost no correlation between optimal parameters of the LRM and KD hydrophobicity scales.

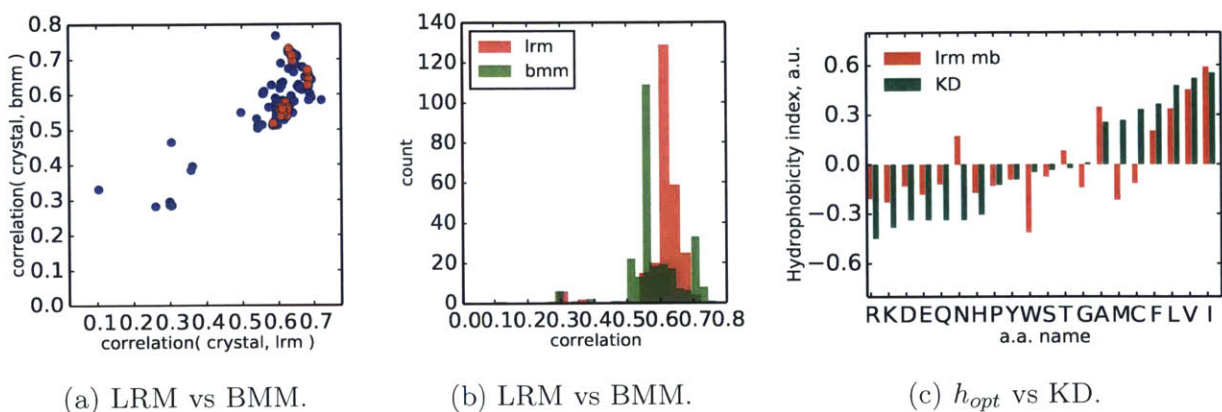


Figure A-5: Comparison of performance of the LRM and BMM for a group of myoglobin (SCOP unid: 46469). Red points on the rightmost panel correspond to the training set (20 proteins).

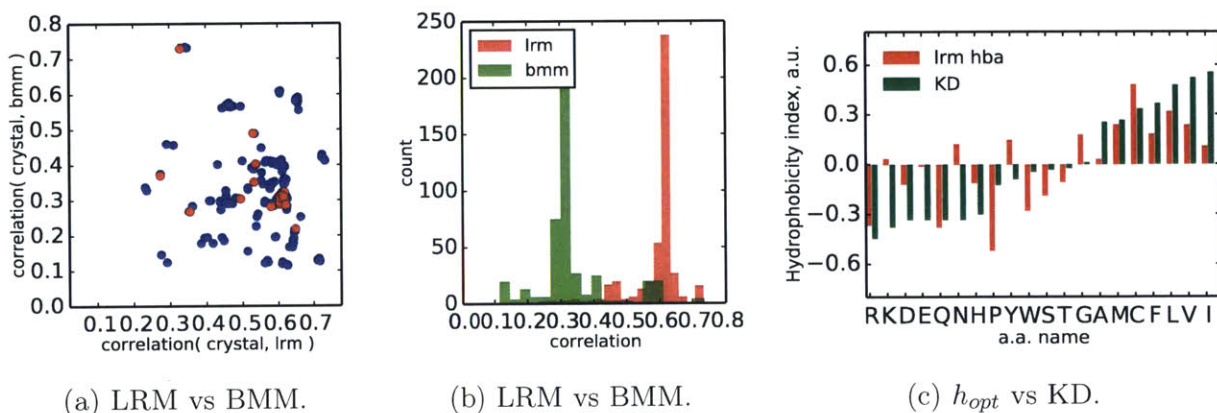


Figure A-6: Comparison of performance of the LRM and BMM for a group of hemoglobin A (SCOP unid: 46486). Red points on the rightmost panel correspond to the training set (20 proteins).

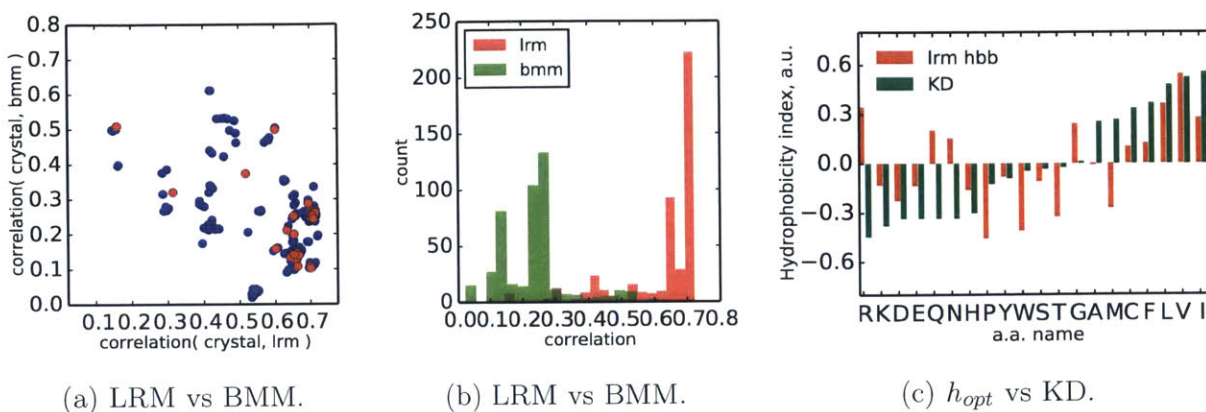


Figure A-7: Comparison of performance of the LRM and BMM for a group of hemoglobin B (SCOP unit: 46500). Red points on the rightmost panel correspond to the training set (20 proteins).

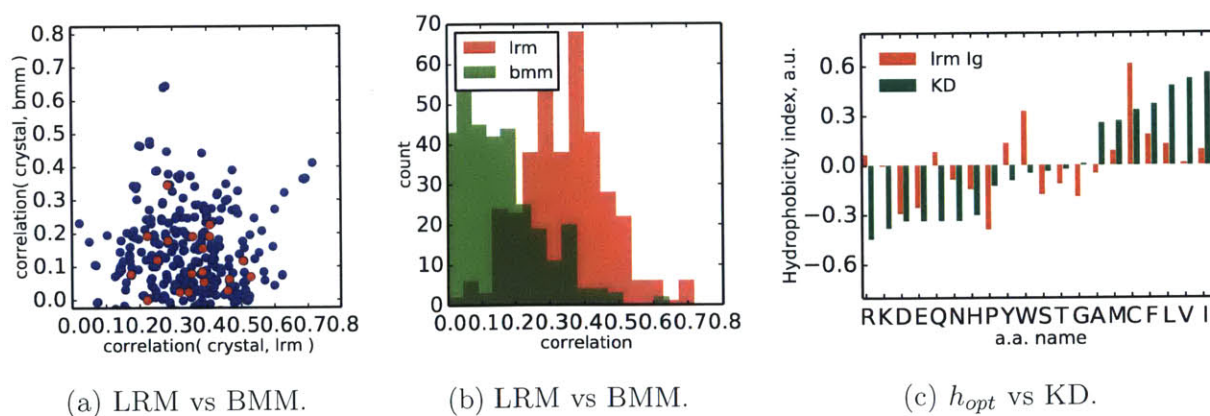


Figure A-8: Comparison of performance of the LRM and BMM for a group of immunoglobulins (SCOP: b.1.1.1). Red points on the rightmost panel correspond to the training set (20 proteins).

Appendix B

Thermodynamics of self-replication

B.1 Chemical reaction $X \rightleftharpoons 2X$

Let us consider two chemical reactions:



The first reaction describes the self-reproduction of molecule X at rate g , whereas the second reaction describes the annihilation at rate δ of one X molecule when two X molecules collide.

In principle, the state of a system with n molecules should be defined by specifying the positions and momenta of all molecules $\{(q_1, p_1), \dots, (q_n, p_n)\}$. However, for simplicity of calculation we are going to assume that all molecules have the same position and neglect their momenta, so the states of the system can be labelled by the number molecules in the system, n . The Master equation for the system can be written as

$$\dot{p}(n, t) = -(gn + \delta \binom{n}{2})p(n, t) + g(n-1)p(n-1, t) + \delta \binom{n+1}{2}p(n+1, t), \quad (\text{B.2})$$

where combinatorial factors n and $\binom{n}{2}$ account for the number of ways that first and the second reactions can happen.

B.1.1 Deterministic solution and steady-state distribution

The corresponding deterministic equation for this set of chemical reactions is

$$\dot{x} = gx - \delta x^2/2. \quad (\text{B.3})$$

The deterministic equation has a stable fixed point $x^* = 2g/\delta$ and the system approaches this point as

$$x(t) = \frac{x^* x_0 e^{gt}}{x^* + x_0 (e^{gt} - 1)}, \quad (\text{B.4})$$

where $x_0 = x(0)$.

Assuming that $x^* \gg 1$, we can rewrite Master equation (B.2) as a Fokker-Plank equation

$$\dot{p}(x, t) = -\frac{\partial}{\partial x} v(x) p(x, t) + \frac{\partial^2}{\partial x^2} D(x) p(x, t), \quad (\text{B.5})$$

where $v(x) = gx(1 - x/x^*)$ and $D(x) = gx(1 + x/x^*)/2$. The steady-state distribution is equal to

$$p_{ss}(x) = \frac{C}{D(x)} \exp\left(\int \frac{v(x)}{D(x)} dx\right) = \frac{e^{-2x}}{x} (1 + x/x^*)^{4x^* - 1} \quad (\text{B.6})$$

and is sharply peaked at $\bar{x} \approx x^*$, $\text{var}(x) \approx x^*$.

B.1.2 Forward probability

In this subsection, we are going to compute the time-evolution of the probability distribution $p(n, t|1, 0)$ for a system that at time $t = 0$ has only one molecule. We will consider the regime where the rate of particle annihilation δ is very small compared to growth rate g , $2g/\delta \gg 1$, and focus only on times when the average number of molecules in the system grows exponentially, that is $1 \ll gt \ll x^* = 2g/\delta$. In this regime, we can neglect terms corresponding to the annihilation of the molecules and rewrite Master equation (B.2) as

$$\dot{p}(n, t) = -gnp(n, t) + g(n-1)p(n-1, t). \quad (\text{B.7})$$

This equation can be easily solved by introducing the generation function $G(s, t) = \sum_{n=1}^{\infty} s^n p(n, t)$ and solving PDE for the generating function

$$G_t(s, t) = gs(s - 1)G_s(s, t).$$

The latter equation has a solution $G(s, t) = F(gt + \ln |\frac{s-1}{s}|)$, where $F(z)$ is arbitrary function that can be found from the initial conditions

$$p(n, t = 0) = \delta_{n, n_0}, \text{ or } G(s, t = 0) = s^{n_0}.$$

From $\frac{1-s}{s} = e^z$, we find that $F(z) = 1/(e^z + 1)^{n_0}$, i.e.

$$G(s, t) = \left(\frac{1}{e^{gt}(1-s)/s + 1} \right)^{n_0} = \left(\frac{1}{e^{gt} - 1} \right)^{n_0} \left(-1 + \frac{1}{1 - \beta s} \right)^{n_0}, \quad (\text{B.8})$$

where $\beta = 1 - e^{-gt}$. For $n_0 = 1$, we can simplify the result and obtain that

$$G(s, t) = e^{-gt}/\beta(-1 + 1 + \beta s + (\beta s)^2 + \dots) = e^{-gt}/\beta \sum_{n=1}^{\infty} (\beta s)^n.$$

Thus, we find that the probability of observing n molecules at time t given that at time $t = 0$ there was only one molecule is

$$P_F(n, t) = p(n, t|1, 0) = e^{-gt}\beta^{n-1} = e^{-gt}(1 - e^{-gt})^{n-1}. \quad (\text{B.9})$$

B.1.3 Return probability

In regime $1 \ll gt \ll x^* = 2g/\delta$, the most likely paths from a state with n molecules to a state with one molecule do not contain growing events because any growing event contributes the factor of $\delta/g \ll 1$ to the probability of the path. Thus, the probability of moving from a state with n molecules to a state with one molecule can be estimated as a sum of probabilities of all trajectories in which the number of

molecules monotonically decreases with time:

$$P_R = p(1, t | n, t = 0) = \sum_{\text{all } \downarrow \text{ paths}} \mathcal{P}_{\text{path}}(t_n, t_{n-1}, \dots, t_1).$$

It should be emphasized that this probability equals the probability of the process time reversed to the process $(1, t = 0) \rightarrow (n, t)$.

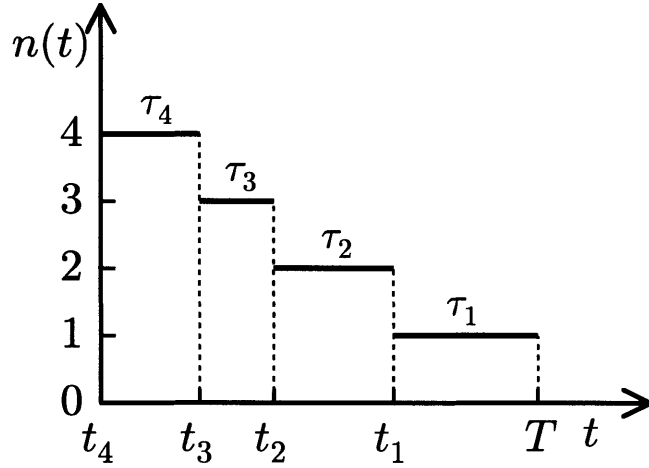


Figure B-1: A sketch of a monotonic destruction path, $\mathcal{P}(t_4, t_3, t_2, t_1)$.

A typical monotonically decaying path is shown in Figure B-1. The probability of such a path is given by the product of probabilities of having i particles on the time interval $[t_i, t_{i-1})$ and probabilities of annihilation of the particle in dt_{i-1}

$$\mathcal{P}_{\text{path}}(t_n, t_{n-1}, \dots, t_1) = e^{-g(T-t_1)} \prod_{i=n}^2 e^{-\alpha_i(t_{i-1}-t_i)} \binom{i}{2} \delta dt_{i-1},$$

where $\alpha_i = gi + \delta \binom{i}{2}$. Here, the extra term $e^{-g(T-t_1)}$ comes from the fact that the system should spend time $\tau_1 = T - t_1$ in a state with one molecule. Therefore, the probability of the reverse process is given by the integral over t_i , subject to the

constraint $0 = t_n \leq t_{n-1} \leq \dots t_1 \leq T$:

$$\begin{aligned}
P_R &= \int_{0=t_n \leq t_{n-1} \leq \dots t_1 \leq T} e^{-g(T-t_1)} \prod_{i=n}^2 e^{-\alpha_i(t_{i-1}-t_i)} \binom{i}{2} \delta dt_{i-1} \\
&= \left(\frac{\delta}{2}\right)^{n-1} n!(n-1)! \int_0^\infty \prod_{i=1}^n d\tau_i e^{-\alpha_i \tau_i} \delta\left(\sum_{i=1}^n \tau_i - T\right) = \\
&= \left(\frac{\delta}{2}\right)^{n-1} n!(n-1)! \frac{1}{2\pi i} \int dk e^{ikT} \int_0^\infty \prod_{i=1}^n d\tau_i e^{-(\alpha_i+ik)\tau_i} = \\
&= \left(\frac{\delta}{2}\right)^{n-1} n!(n-1)! \frac{1}{2\pi i} \int dk e^{ikT} \prod_{i=1}^n \frac{1}{\alpha_i + ik} = \\
&= \left(\frac{\delta}{2}\right)^{n-1} n!(n-1)! \sum_i e^{-\alpha_i T} \prod_{i \neq j} \frac{1}{\alpha_i - \alpha_j}. \tag{B.10}
\end{aligned}$$

Assuming that $n \sim e^{gt} \ll 2g/\delta$, we can simplify the expression above as

$$\begin{aligned}
P_R &= \left(\frac{\delta}{2g}\right)^{n-1} n!(n-1)! \sum_{i=1}^n \frac{(-1)^{i-1} e^{-igT}}{(n-i)!(i-1)!} = \\
&= \left(\frac{\delta}{2g}\right)^{n-1} n! e^{-gT} \sum_{i=0}^{n-1} \binom{n-1}{i} (-1)^i e^{-igt} = \\
&= \left(\frac{\delta}{2g}\right)^{n-1} n! e^{-gT} (1 - e^{-gt})^{n-1}. \tag{B.11}
\end{aligned}$$

B.1.4 Entropy production

Because the transition rates do not depend on time and to get from the state with one molecule to the state with n molecules, the system should go through all states,

$$1 \rightarrow 2 \rightarrow \dots \rightarrow n-1 \rightarrow n,$$

the entropy production is completely determined by the initial and final state of the system:

$$\begin{aligned}\Delta S(1, n) &= \sum_{i=1}^{n-1} \Delta s(i, i+1) = \sum_{i=1}^{n-1} \ln \frac{\mathcal{R}(i, i+1)}{\mathcal{R}(i+1, i)} = \sum_{i=1}^{n-1} \ln \frac{g^i}{\delta i(i-1)/2} = \\ &= (n-1) \ln \frac{2g}{\delta} - \ln n!\end{aligned}\tag{B.12}$$

Using equations (B.9, B.11, B.12), one can verify that our assumptions for P_F , P_R , and ΔS are self-consistent and that these quantities satisfy the Crooks relation:

$$\ln P_F - \ln P_R = \Delta S.$$

It should be noted that, in principle, it was not necessary to directly compute P_R through integration over monotonically decreasing paths. However, we decided to include this calculation just to demonstrate the consistency of our assumptions.

B.2 Two chemical reactions $A \rightleftharpoons 2A$, $B \rightleftharpoons 2B$

B.2.1 Surface of fixed return probability

If we consider the evolution of two different types of molecules A and B that do not interact between each other, we can write

$$\begin{aligned}\ln P_F(n_A, n_B, T) &= \ln P_F(n_A, T) + \ln P_F(n_B, t) = \\ &= -(g_A + g_B)T + (n_A - 1) \ln(1 - e^{-g_A T}) + (n_B - 1) \ln(1 - e^{-g_B T}) \\ \ln P_R(n_A, n_B, T) &= \ln P_R(n_A, T) + \ln P_R(n_B, t) = \\ &= -(g_A + g_B)T + (n_A - 1) \ln(1 - e^{-g_A T}) + (n_B - 1) \ln(1 - e^{-g_B T}) + \\ &\quad + \ln n_A! + \ln n_B! - (n_A - 1) \ln \frac{2g_A}{\delta_A} - (n_B - 1) \ln \frac{2g_B}{\delta_B} \\ \Delta S(n_A, n_B) &= (n_A - 1) \ln \frac{2g_A}{\delta_A} - (n_B - 1) \ln \frac{2g_B}{\delta_B} - \ln n_A! - \ln n_B!\end{aligned}\tag{B.13}$$

Assuming that $1 \ll n_{A/B} \ll g_{A/B}/\delta$, we can rewrite the probability of reverse process and the entropy production as:

$$\begin{aligned}\ln P_R(n_A, n_B, T) &= C(T) - (u + \delta u)n_A - (v + \delta v)n_B \\ \ln \Delta S(n_A, n_B, T) &= C'(T) + un_A + vn_B,\end{aligned}\tag{B.14}$$

where $u = \ln \frac{2g_A}{n_A\delta_A}$, $\delta u = -\ln(1 - e^{-g_A T}) = e^{-g_A T}$, $v = \ln \frac{2g_B}{n_B\delta_B}$, $\delta v = -\ln(1 - e^{-g_B T}) = e^{-g_B T}$. On the surface of fixed return probability $P_R(n_A, n_B) = \text{const}$, we can estimate the variation of entropy production ΔS as follows:

$$\begin{aligned}\Delta S(n_A, n_B) \Big|_{P_R=\text{const}} & - C''(T) = n_B \left(v \frac{\delta u}{u} - \delta v \right) \approx \\ & = n_B \left(e^{-g_A T} \frac{\ln n_B^*}{\ln n_A^*} - e^{-g_B T} \right) = n_B e^{-g_B T} \left(e^{(g_B - g_A)T} \frac{\ln n_B^*}{\ln n_A^*} - 1 \right),\end{aligned}\tag{B.15}$$

that is the molecules that grow more quickly dissipate more, and the outcomes with larger fraction of these molecules are more likely on the surface of fixed return probability.

If we assume that $g_B = 2g_A$ and $\delta_B = 2\delta_A$, then $n_A^* = n_B^*$ and

$$\Delta S(n_A, n_B) \Big|_{P_R=\text{const}} - C''(T) = n_B (e^{-g_A T} - e^{-g_B T}) = n_B (e^{-g_A T} - e^{-2g_A T}) \approx n_B e^{-g_A T}.$$

Taking into account that the average number molecules of type B is on the order of $n_B \sim e^{g_B t}$, we find that $S(n_B = 1) \sim e^{(g_B - g_A)T} \gg 1$, that is, the variation of entropy on the surface of fixed return probability can be dramatic.

Bibliography

- [1] J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips, "A three-dimensional model of the myoglobin molecule obtained by X-ray analysis," *Nature*, vol. 181, pp. 662–666, 03 1958.
- [2] G. Wagner and K. Wüthrich, "Sequential resonance assignments in protein 1H nuclear magnetic resonance spectra: Basic pancreatic trypsin inhibitor," *Journal of Molecular Biology*, vol. 155, pp. 347–366, 3 1982.
- [3] K. Wuthrich, "The way to NMR structures of proteins," *Nat Struct Mol Biol*, vol. 8, pp. 923–925, 11 2001.
- [4] T. J. P. Hubbard, B. Ailey, S. E. Brenner, A. G. Murzin, and C. Chothia, "SCOP: a structural classification of proteins database," *Nucleic Acids Research*, vol. 27, pp. 254–256, 01 1999.
- [5] C. Orengo, A. Michie, S. Jones, D. Jones, M. Swindells, and J. Thornton, "CATH — a hierarchic classification of protein domain structures," *Structure*, vol. 5, pp. 1093–1109, 8 1997.
- [6] D. E. Anderson, W. J. Becktel, and F. W. Dahlquist, "ph-induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of t4 lysozyme," *Biochemistry*, vol. 29, no. 9, pp. 2403–2408, 1990.
- [7] M. Karplus *et al.*, "Charmm: A program for macromolecular energy, minimization, and dynamics calculations," *J Comput Chem*, vol. 4, p. 187217, 1983.
- [8] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," *Journal of computational chemistry*, vol. 25, no. 9, pp. 1157–1174, 2004.
- [9] W. R. Scott, P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krüger, and W. F. van Gunsteren, "The gromos biomolecular simulation program package," *The Journal of Physical Chemistry A*, vol. 103, no. 19, pp. 3596–3607, 1999.
- [10] A. D. MacKerell, D. Bashford, M. Bellott, R. Dunbrack, J. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. a. Ha, *et al.*, "All-atom empirical potential for molecular modeling and dynamics studies of proteins," *The journal of physical chemistry B*, vol. 102, no. 18, pp. 3586–3616, 1998.

- [11] C. B. Anfinsen, E. Haber, M. Sela, and F. H. White, "The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 47, pp. 1309–1314, 09 1961.
- [12] H. F. Epstein, A. N. Schechter, R. F. Chen, and C. B. Anfinsen, "Folding of staphylococcal nuclease: Kinetic studies of two processes in acid renaturation," *Journal of Molecular Biology*, vol. 60, pp. 499–508, 9 1971.
- [13] C. B. Anfinsen, "Principles that govern the folding of protein chains," *Science*, vol. 181, pp. 223–230, 07 1973.
- [14] F. Pohl, "On the kinetics of structural transition i of some pancreatic proteins," *FEBS letters*, vol. 3, no. 1, pp. 60–64, 1969.
- [15] G. I. Makhatadze and P. L. Privalov, "Contribution of hydration to protein folding thermodynamics: I. the enthalpy of hydration," *Journal of molecular biology*, vol. 232, no. 2, pp. 639–659, 1993.
- [16] P. L. Privalov and G. I. Makhatadze, "Contribution of hydration to protein folding thermodynamics: Ii. the entropy and gibbs energy of hydration," *Journal of molecular biology*, vol. 232, no. 2, pp. 660–679, 1993.
- [17] T. Lazaridis and M. Karplus, "Thermodynamics of protein folding: a microscopic view," *Biophysical chemistry*, vol. 100, no. 1, pp. 367–395, 2002.
- [18] F. H. Stillinger, "Structure in aqueous solutions of nonpolar solutes from the standpoint of scaled-particle theory," *Journal of Solution Chemistry*, vol. 2, no. 2-3, pp. 141–158, 1973.
- [19] Y. K. Kang, G. Nemethy, and H. A. Scheraga, "Free energies of hydration of solute molecules. 1. improvement of the hydration shell model by exact computations of overlapping volumes," *Journal of Physical Chemistry*, vol. 91, no. 15, pp. 4105–4109, 1987.
- [20] T. Ooi, M. Oobatake, G. Nemethy, and H. A. Scheraga, "Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides," *Proceedings of the National Academy of Sciences*, vol. 84, no. 10, pp. 3086–3090, 1987.
- [21] T. Lazaridis, G. Archontis, and M. Karplus, "Enthalpic contribution to protein stability: insights from atom-based calculations and statistical mechanics," *Advances in protein chemistry*, vol. 47, pp. 231–306, 1995.
- [22] L. R. Pratt and D. Chandler, "Theory of the hydrophobic effect," *The Journal of Chemical Physics*, vol. 67, no. 8, pp. 3683–3704, 1977.

- [23] S. Garde, G. Hummer, A. E. García, M. E. Paulaitis, and L. R. Pratt, "Origin of entropy convergence in hydrophobic hydration and protein folding," *Physical Review Letters*, vol. 77, no. 24, p. 4966, 1996.
- [24] K. Lum, D. Chandler, and J. D. Weeks, "Hydrophobicity at small and large length scales," *The Journal of Physical Chemistry B*, vol. 103, no. 22, pp. 4570–4577, 1999.
- [25] D. Chandler, "Interfaces and the driving force of hydrophobic assembly," *Nature*, vol. 437, no. 7059, pp. 640–647, 2005.
- [26] R. Smith and C. Tanford, "Hydrophobicity of long chain n-alkyl carboxylic acids, as measured by their distribution between heptane and aqueous solutions," *Proceedings of the National Academy of Sciences*, vol. 70, no. 2, pp. 289–293, 1973.
- [27] G. D. Rose, "Prediction of chain turns in globular proteins on a hydrophobic basis," *Nature*, vol. 272, no. 5654, pp. 586–590, 1978.
- [28] G. D. Rose and S. RoY, "Hydrophobic basis of packing in globular proteins," *PNAS*, vol. 77, no. 8, pp. 4643–4647, 1980.
- [29] D. Engelman, T. Steitz, and A. Goldman, "Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins," *Annual review of biophysics and biophysical chemistry*, vol. 15, no. 1, pp. 321–353, 1986.
- [30] T. P. Hopp and K. R. Woods, "Prediction of protein antigenic determinants from amino acid sequences," *Proceedings of the National Academy of Sciences*, vol. 78, no. 6, pp. 3824–3828, 1981.
- [31] J. Kyte and R. F. Doolittle, "A simple method for displaying the hydrophobic character of a protein," *Journal of molecular biology*, vol. 157, no. 1, pp. 105–132, 1982.
- [32] O. Ptitsyn, A. Kron, Y. Eizner, *et al.*, "The models of the denaturation of globular proteins. i. theory of globula–coil transitions in macromolecules," in *Journal of Polymer Science Part C: Polymer Symposia*, vol. 16, pp. 3509–3517, Wiley Online Library, 1967.
- [33] I. Lifshitz, "Some problems of the statistical theory of biopolymers," *Sov. Phys. JETP*, vol. 28, no. 6, 1969.
- [34] J. D. Bryngelson and P. G. Wolynes, "Spin glasses and the statistical mechanics of protein folding," *Proceedings of the National Academy of Sciences*, vol. 84, no. 21, pp. 7524–7528, 1987.
- [35] E. Shakhnovich and A. Gutin, "Formation of unique structure in polypeptide chains: theoretical investigation with the aid of a replica approach," *Biophysical chemistry*, vol. 34, no. 3, pp. 187–199, 1989.

- [36] V. S. Pande, A. Y. Grosberg, and T. Tanaka, "Heteropolymer freezing and design: Towards physical models of protein folding," *Rev. Mod. Phys.*, vol. 72, pp. 259–314, Jan 2000.
- [37] J. N. Onuchic and P. G. Wolynes, "Navigating the folding routes," *Science*, vol. 267, no. 5204, pp. 1619–1620, 1995.
- [38] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, and W. Wriggers, "Atomic-level characterization of the structural dynamics of proteins," *Science*, vol. 330, pp. 341–346, 10 2010.
- [39] G. Jayachandran, V. Vishal, and V. S. Pande, "Using massively parallel simulation and Markovian models to study protein folding: Examining the dynamics of the villin headpiece," *The Journal of Chemical Physics*, vol. 124, no. 16, p. 164902, 2006.
- [40] C. A. Rohl, C. E. Strauss, K. M. Misura, and D. Baker, "Protein structure prediction using rosetta," *Methods in enzymology*, vol. 383, pp. 66–93, 2004.
- [41] R. Das and D. Baker, "Macromolecular modeling with rosetta," *Annual Review of Biochemistry*, vol. 77, pp. 363–382, 2015/10/20 2008.
- [42] D. S. Marks, T. A. Hopf, and C. Sander, "Protein structure prediction from sequence variation," *Nat Biotech*, vol. 30, pp. 1072–1080, 11 2012.
- [43] S. Ovchinnikov, L. Kinch, H. Park, Y. Liao, J. Pei, D. E. Kim, H. Kamisetty, N. V. Grishin, and D. Baker, "Large-scale determination of previously unsolved protein structures using evolutionary information," *eLife*, vol. 4, p. e09248, 2015.
- [44] N. Perunov and J. L. England, "Quantitative theory of hydrophobic effect as a driving force of protein structure," *Protein Science*, vol. 23, no. 4, pp. 387–399, 2014.
- [45] G. D. Rose, A. R. Geselowitz, G. J. Lesser, R. H. Lee, and M. H. Zehfus, "Hydrophobicity of amino acid residues in globular proteins," *Science*, vol. 229, no. 4716, pp. 834–838, 1985.
- [46] R. L. Baldwin, "Energetics of protein folding," *Journal of molecular biology*, vol. 371, no. 2, pp. 283–301, 2007.
- [47] K. F. Lau and K. A. Dill, "A lattice statistical mechanics model of the conformational and sequence spaces of proteins," *Macromolecules*, vol. 22, no. 10, pp. 3986–3997, 1989.
- [48] K. Yue, K. M. Fiebig, P. D. Thomas, H. S. Chan, E. I. Shakhnovich, and K. A. Dill, "A test of lattice protein folding algorithms," *Proceedings of the National Academy of Sciences*, vol. 92, no. 1, pp. 325–329, 1995.

- [49] C. Chothia, "The nature of the accessible and buried surfaces in proteins," *Journal of molecular biology*, vol. 105, no. 1, pp. 1–12, 1976.
- [50] J. L. England, "Allostery in protein domains reflects a balance of steric and hydrophobic effects," *Structure*, vol. 19, no. 7, pp. 967–975, 2011.
- [51] W. C. Wimley, T. P. Creamer, and S. H. White, "Solvation energies of amino acid side chains and backbone in a family of host-guest pentapeptides," *Biochemistry*, vol. 35, no. 16, pp. 5109–5124, 1996.
- [52] Y. Nozaki and C. Tanford, "The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions establishment of a hydrophobicity scale," *Journal of Biological Chemistry*, vol. 246, no. 7, pp. 2211–2217, 1971.
- [53] J. Janin, "Surface and inside volumes in globular proteins.," *Nature*, vol. 277, no. 5696, pp. 491–492, 1979.
- [54] A. G. Murzin, S. E. Brenner, T. Hubbard, and C. Chothia, "Scop: a structural classification of proteins database for the investigation of sequences and structures," *Journal of molecular biology*, vol. 247, no. 4, pp. 536–540, 1995.
- [55] T. Takano, "Structure of myoglobin refined at 2.0 Å resolution: I. crystallographic refinement of metmyoglobin from sperm whale," *Journal of molecular biology*, vol. 110, no. 3, pp. 537–568, 1977.
- [56] R. Benesch and R. E. Benesch, "The effect of organic phosphates from the human erythrocyte on the allosteric properties of hemoglobin," *Biochemical and biophysical research communications*, vol. 26, no. 2, pp. 162–167, 1967.
- [57] A. Arnone, "X-ray diffraction study of binding of 2, 3-diphosphoglycerate to human deoxyhaemoglobin.," *Nature*, vol. 237, no. 5351, pp. 146–149, 1972.
- [58] F. McCormick, B. Clark, T. La Cour, M. Kjeldgaard, L. Norskov-Lauritsen, and J. Nyborg, "A model for the tertiary structure of p21, the product of the ras oncogene," *Science*, vol. 230, no. 4721, pp. 78–82, 1985.
- [59] C. Branden, J. Tooze, *et al.*, *Introduction to protein structure*, vol. 2. Garland New York, 1991.
- [60] P. A. Alexander, Y. He, Y. Chen, J. Orban, and P. N. Bryan, "A minimal sequence code for switching protein structure and function," *Proceedings of the National Academy of Sciences*, vol. 106, no. 50, pp. 21149–21154, 2009.
- [61] S. Miyazawa and R. L. Jernigan, "Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation," *Macromolecules*, vol. 18, no. 3, pp. 534–552, 1985.

- [62] M. J. Sippl, “Calculation of conformational ensembles from potentials of mean force: an approach to the knowledge-based prediction of local structures in globular proteins,” *Journal of molecular biology*, vol. 213, no. 4, pp. 859–883, 1990.
- [63] P. Csermely, R. Palotai, and R. Nussinov, “Induced fit, conformational selection and independent dynamic segments: an extended view of binding events,” *Trends in biochemical sciences*, vol. 35, no. 10, pp. 539–546, 2010.
- [64] F. Piazza and Y.-H. Sanejouand, “Discrete breathers in protein structures,” *Physical biology*, vol. 5, no. 2, p. 026001, 2008.
- [65] F. Piazza and Y.-H. Sanejouand, “Long-range energy transfer in proteins,” *Physical biology*, vol. 6, no. 4, p. 046014, 2009.
- [66] G. Kopidakis and S. Aubry, “Intraband discrete breathers in disordered nonlinear systems. i. delocalization,” *Physica D: Nonlinear Phenomena*, vol. 130, no. 3, pp. 155–186, 1999.
- [67] G. Kopidakis, S. Aubry, and G. Tsironis, “Targeted energy transfer through discrete breathers in nonlinear systems,” *Physical Review Letters*, vol. 87, no. 16, p. 165501, 2001.
- [68] M. Bathe, “A finite element framework for computation of protein normal modes and mechanical response,” *Proteins: Structure, Function, and Bioinformatics*, vol. 70, no. 4, pp. 1595–1609, 2008.
- [69] D.-N. Kim, R. S. Sedeh, C. T. Nguyen, and M. Bathe, “Finite element framework for mechanics and dynamics of supramolecular protein assemblies,” in *ASME 2010 First Global Congress on NanoEngineering for Medicine and Biology*, pp. 315–316, American Society of Mechanical Engineers, 2010.
- [70] R. J. Hawkins and T. C. McLeish, “Coarse-grained model of entropic allostery,” *Physical Review Letters*, vol. 93, no. 9, p. 098104, 2004.
- [71] M. Levitt, C. Sander, and P. S. Stern, “Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme,” *Journal of molecular biology*, vol. 181, no. 3, pp. 423–447, 1985.
- [72] J. R. Rocha, M. G. van der Linden, D. C. Ferreira, P. H. Azevêdo, and A. F. P. de Araújo, “Information-theoretic analysis and prediction of protein atomic burials: on the search for an informational intermediate between sequence and structure,” *Bioinformatics*, vol. 28, no. 21, pp. 2755–2762, 2012.
- [73] C. Jarzynski, “Nonequilibrium equality for free energy differences,” *Physical Review Letters*, vol. 78, no. 14, p. 2690, 1997.

- [74] G. E. Crooks, “Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences,” *Physical Review E*, vol. 60, no. 3, p. 2721, 1999.
- [75] G. Hummer and A. Szabo, “Free energy reconstruction from nonequilibrium single-molecule pulling experiments,” *Proceedings of the National Academy of Sciences*, vol. 98, no. 7, pp. 3658–3661, 2001.
- [76] U. Seifert, “Entropy production along a stochastic trajectory and an integral fluctuation theorem,” *Physical Review Letters*, vol. 95, no. 4, p. 040602, 2005.
- [77] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco, and C. Bustamante, “Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski’s equality,” *Science*, vol. 296, no. 5574, pp. 1832–1835, 2002.
- [78] V. Blickle, T. Speck, L. Helden, U. Seifert, and C. Bechinger, “Thermodynamics of a colloidal particle in a time-dependent nonharmonic potential,” *Physical Review Letters*, vol. 96, no. 7, p. 070603, 2006.
- [79] F. Douarche, S. Ciliberto, A. Petrosyan, and I. Rabbiosi, “An experimental test of the Jarzynski equality in a mechanical experiment,” *EPL (Europhysics Letters)*, vol. 70, no. 5, p. 593, 2005.
- [80] R. E. Spinney and I. J. Ford, “Fluctuation relations: a pedagogical overview,” *arXiv preprint arXiv:1201.6381*, 2012.
- [81] U. Seifert, “Stochastic thermodynamics, fluctuation theorems and molecular machines,” *Reports on Progress in Physics*, vol. 75, no. 12, p. 126001, 2012.
- [82] G. E. Crooks, “Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems,” *Journal of Statistical Physics*, vol. 90, no. 5-6, pp. 1481–1487, 1998.
- [83] C. Jarzynski, “Hamiltonian derivation of a detailed fluctuation theorem,” *Journal of Statistical Physics*, vol. 98, no. 1-2, pp. 77–102, 2000.
- [84] C. Maes, F. Redig, and A. Van Moffaert, “On the definition of entropy production, via examples,” *Journal of mathematical physics*, vol. 41, no. 3, pp. 1528–1554, 2000.
- [85] T. Hatano and S.-i. Sasa, “Steady-state thermodynamics of langevin systems,” *Physical Review Letters*, vol. 86, no. 16, p. 3463, 2001.
- [86] M. Esposito and C. Van den Broeck, “Three detailed fluctuation theorems,” *Physical review letters*, vol. 104, no. 9, p. 090601, 2010.
- [87] J. L. England, “Statistical physics of self-replication,” *The Journal of chemical physics*, vol. 139, no. 12, p. 121923, 2013.

- [88] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco, and C. Bustamante, “Verification of the crooks fluctuation theorem and recovery of rna folding free energies,” *Nature*, vol. 437, no. 7056, pp. 231–234, 2005.
- [89] J. Michl and E. C. H. Sykes, “Molecular rotors and motors: recent advances and future challenges,” *ACS nano*, vol. 3, no. 5, pp. 1042–1048, 2009.
- [90] Y. Oono and M. Paniconi, “Steady state thermodynamics,” *Progress of Theoretical Physics Supplement*, vol. 130, pp. 29–44, 1998.
- [91] V. Y. Chernyak, M. Chertkov, and C. Jarzynski, “Path-integral analysis of fluctuation theorems for general Langevin processes,” *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2006, no. 08, p. P08001, 2006.
- [92] T. Speck and U. Seifert, “Integral fluctuation theorem for the housekeeping heat,” *Journal of Physics A: Mathematical and General*, vol. 38, no. 34, p. L581, 2005.
- [93] E. Trepagnier, C. Jarzynski, F. Ritort, G. E. Crooks, C. Bustamante, and J. Liphardt, “Experimental test of Hatano and Sasa’s nonequilibrium steady-state equality,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 42, pp. 15038–15041, 2004.
- [94] D. Ruelle, “Biology and nonequilibrium: remarks on a paper by JL England,” *arXiv preprint arXiv:1502.01600*, 2015.
- [95] D. Ruelle, “A generalized detailed balance relation,” *arXiv preprint arXiv:1510.08357*, 2015.
- [96] N. Perunov, R. Marsland, and J. England, “Statistical physics of adaptation,” *arXiv preprint arXiv:1412.1875*, 2014.
- [97] C. Jarzynski, “Rare events and the convergence of exponentially averaged work values,” *Physical Review E*, vol. 73, no. 4, p. 046105, 2006.
- [98] S. Rahav, J. Horowitz, and C. Jarzynski, “Directed flow in nonadiabatic stochastic pumps,” *Physical Review Letters*, vol. 101, no. 14, p. 140602, 2008.
- [99] S. Ito, H. Yamauchi, M. Tamura, S. Hidaka, H. Hattori, T. Hamada, K. Nishida, S. Tokonami, T. Itoh, H. Miyasaka, *et al.*, “Selective optical assembly of highly uniform nanoparticles by doughnut-shaped beams,” *Scientific reports*, vol. 3, 2013.
- [100] J. K. Jun and A. H. Hübler, “Formation and structure of ramified charge transportation networks in an electromechanical system,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 3, pp. 536–540, 2005.

- [101] D. Kondepudi, B. Kay, and J. Dixon, “End-directed evolution and the emergence of energy-seeking behavior in a complex system,” *Physical Review E*, vol. 91, no. 5, p. 050902, 2015.
- [102] M. Tagliazucchi, E. A. Weiss, and I. Szleifer, “Dissipative self-assembly of particles interacting through time-oscillatory potentials,” *Proceedings of the National Academy of Sciences*, p. 201406122, 2014.
- [103] V. Schaller, C. Weber, C. Semmrich, E. Frey, and A. R. Bausch, “Polar patterns of driven filaments,” *Nature*, vol. 467, no. 7311, pp. 73–77, 2010.
- [104] T. Sanchez, D. T. Chen, S. J. DeCamp, M. Heymann, and Z. Dogic, “Spontaneous motion in hierarchically assembled active matter,” *Nature*, vol. 491, no. 7424, pp. 431–434, 2012.
- [105] Z. Zeravcic, V. N. Manoharan, and M. P. Brenner, “Size limits of self-assembled colloidal structures made using specific interactions,” *Proceedings of the National Academy of Sciences*, vol. 111, no. 45, pp. 15918–15923, 2014.
- [106] C. Jäckel, P. Kast, and D. Hilvert, “Protein design by directed evolution,” *Annu. Rev. Biophys.*, vol. 37, pp. 153–173, 2008.
- [107] I. Prigogine and G. Nicolis, “Biological order, structure and instabilities,” *Quarterly Reviews of Biophysics*, vol. 4, no. 2-3, pp. 107–148, 1971.
- [108] C. Maes and K. Netočný, “Minimum entropy production principle from a dynamical fluctuation law,” *Journal of mathematical physics*, vol. 48, no. 5, p. 053306, 2007.
- [109] R. Landauer, “Inadequacy of entropy and entropy derivatives in characterizing the steady state,” *Physical Review A*, vol. 12, no. 2, p. 636, 1975.
- [110] E. T. Jaynes, “The minimum entropy production principle,” *Annual Review of Physical Chemistry*, vol. 31, no. 1, pp. 579–601, 1980.
- [111] E. T. Jaynes, “Information theory and statistical mechanics,” *Physical review*, vol. 106, no. 4, p. 620, 1957.
- [112] R. Dewar, “Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states,” *Journal of Physics A: Mathematical and General*, vol. 36, no. 3, p. 631, 2003.
- [113] L. Martyushev and V. Seleznev, “Maximum entropy production principle in physics, chemistry and biology,” *Physics reports*, vol. 426, no. 1, pp. 1–45, 2006.
- [114] L. M. Martyushev, “The maximum entropy production principle: two basic questions,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 365, no. 1545, pp. 1333–1334, 2010.

- [115] S. Bruers, C. Maes, and K. Netočný, “On the validity of entropy production principles for linear electrical circuits,” *Journal of statistical physics*, vol. 129, no. 4, pp. 725–740, 2007.
- [116] R. A. Fisher, *The genetical theory of natural selection: a complete variorum edition*. Oxford University Press, 1930.
- [117] G. R. Price *et al.*, “Selection and covariance.,” *Nature*, vol. 227, pp. 520–21, 1970.