Protein Actions: Principles and Modeling by I. Bahar, R.L. Jernigan, and K.A. Dill

**Garland Science, Taylor & Francis Group, 1st Edition,
Feb 09, 2017, ISBN: 9780815341772**

Brigita Urbanc^{1,2}

Published online: 10 October 2017
© Springer Science+Business Media B.V. 2017

When I started teaching a course on biophysics intended for upper-level undergraduates and graduate students, I decided to focus on the biophysics of proteins, which was at a time the main focus of the biophysics research group in our department. I chose *Protein Physics* by Alexei V. Finkelstein and Oleg B. Ptitsyn. As this textbook was out-of-print and the population of students attending the course requested a broader course focus, I started using *Physical Biology of the Cell* by Rob Phillips, Jane Kondev, and Julie Theriot, which is one of the most comprehensive biophysics textbooks on the market with numerous end-of-chapter problems, showcasing the power of statistical mechanics in developing our current understanding of molecular biophysics not only of proteins, but also DNA, lipids, and cells. But, how do you convey all complexity of molecular biophysics contained in *Physical Biology of the Cell* in 10 weeks of a single quarter? Then, I came across this most recent textbook on protein biophysics offered by Garland Science (Taylor & Francis Group), *Protein Actions* by I. Bahar, R.L. Jernigan, and K.A. Dill. There are 13 chapters in the book, each neatly formatted in a way that allows a reader to make notes, comments, and calculations on the side of the text (yes!). Each chapter is organized in sections and subsections, which makes it easy to follow while simultaneously focusing on a specific aspect under discussion. Specific examples of the more general concepts and approaches are worked out in blue boxes, which is another wonderful feature of the book. The style of writing is non-assuming. Very refreshing, indeed. Knowing the fundamentals of classical statistical physics is required, however. My personal favorite stylistic feature is a recurring phrase "Here's how you use (understand) ...", which is typically followed by a description of how a particular equation or concept can help you get the missing information from the observables. This is a smart pedagogical approach, which, I am sure, many students will benefit from.

✉ Brigita Urbanc
bu25@drexel.edu

¹ Drexel University, 3141 Chestnut St., Philadelphia, PA 19104, USA

² Faculty of Mathematics and Physics, Jadranska ulica 19, 1000, Ljubljana, Slovenia

Chapters 1 and 2 introduce protein structure and function. Chirality of individual amino acids, which has far-reaching consequences for the twisted supramolecular structure, is explained and so is the use of the right-hand rule. Planar peptide bonds and backbone dihedral angles, ϕ and ψ , which are essential determinants of the secondary structure and define the conformational space of individual amino acids, are introduced. Chapter 1 covers frequently used concepts, such as a contact map to quantify tertiary (and for multi-chain proteins also quaternary) structure and explains that proteins can be classified with respect to both their structure and evolutionary properties. The native state is described not as a rigid conformation but rather as fluctuating around the native structure, resulting in an ensemble of conformations. Chapter 2 tracks proteins from their “birth”, when genetic DNA code is transcribed into messenger RNA and translated into a protein sequence, to their “final moments” of degradation of ubiquitin-tagged “defective” proteins by proteasomes. Most of this chapter describes what proteins do during their lifetime, which brings us to the magical world of their versatile actions. Protein complexes, known as molecular chaperons, help proteins fold into their native structures by providing an enclosure that prevents other entities from interfering with the process. Proteins can act as enzymes, which catalyze the synthesis of other biomolecules, are involved in transcription and regulation of DNA expression (that is, help select genes to be expressed as specific proteins), convert energy produced by the cell into motion (so-called motor proteins, with a classical case of kinesin walking along a microtubule), are involved in transport of small molecules, ions, or proteins across the membrane, and act as transducers of signals across the membrane.

Chapters 3-6 provide a beautifully delivered theoretical basis of protein folding, binding, assembly, and folding kinetics. These chapters are based on a conjecture that a protein adopts a stable equilibrium state, independent of the kinetic process that lead to this state, and can thus be analyzed using tools of classical statistical thermodynamics. These chapters are impressive because they cover material, which students traditionally find difficult to comprehend, yet intuitive enough to bring a direct meaning into the statistical mechanics quantities, such as partition function and probabilistic description of quantities that can be observed experimentally. Whenever possible, theoretical predictions are compared to available experimental data. More involved calculations in support of the main text of these chapters are collected into seven Appendices.

Chapter 3 introduces Anfinsen’s hypothesis and then delves into protein unfolding by thermal and chemical denaturation. Protein unfolding as a function of temperature or concentration of a denaturant provides information about protein stability, whereby melting and denaturation temperatures can be experimentally determined by calorimetric measurements. The language of classical statistical mechanics, including the partition function and Boltzmann distribution are described, followed by an application to a hydrophobic-polar (HP) lattice model of protein, for which the melting temperature can be analytically derived. The free energy landscape of protein folding is discussed and the shape of the landscape is explained in terms of the density of states. Chapter 3 is indeed rich in content: In addition to a two-state model, which is critical for understanding folding-unfolding dynamics of short proteins, pK_a values of ionizable amino acid residues are introduced, followed by a description of a denatured state in terms of an end-to-end radius, a quantity that is borrowed from polymer physics. Using a random-flight polymer (i.e., self-avoiding random walk) model for a denatured protein, the scaling exponent of the end-to-end radius of denatured proteins is derived to be $3/5$, thus larger than both the Flory exponent of $1/3$ and random walk model exponent of $1/2$. The polymer is “swollen” relative to the random walk prediction as it has to avoid itself due to volume exclusion. Simple and effective. The Appendix of this chapter

analyzes a simple model of electrostatics to elucidate why proteins denature at low and high pH values.

Protein binding plays a central role in protein actions. A binding partner may be an ion, small molecule, or another protein. In many instances, binding does not affect the protein structure only at a local binding site but rather induces global conformational changes, phenomena known as allosteric effects. Chapter 4 delivers a key mathematical concept, a binding polynomial, which makes the analytical treatment of a variety of protein binding models easy to follow. The logical progression from a two-state binding to more complex models with increased cooperativity is given. A classical example of hemoglobin, a protein comprising four chains, each of which can bind oxygen, is used as an example of highly cooperative binding. Consequently, the binding curve is very sharp, whereby hemoglobin is either oxygen free below the critical oxygen concentration or fully bound by four oxygens above the critical concentration, while intermediate states are populated only close to the critical oxygen concentration. In this context, the Monod–Wyman–Changeaux (MWC) model is of a particular relevance, as it couples binding to allostery. The MWC model is general enough to be applicable to systems other than oxygen binding of hemoglobin. Another example is binding of neurotransmitters to ion channels, which causes a conformational switch between the open and closed ion channel states. Binding can also be coupled to a directed motion, such as in a Brownian ratchet, a model applicable to helicases, proteins that move along a double-stranded DNA.

Protein folding and aggregation are viewed as cooperative transitions in Chapter 5. Different models of a helix-to-coil transition are described, and finally the Schellman model, which describes this transition in terms of nucleation and propagation, is adopted and applied to folding of a three-helix bundle protein. The nucleation-propagation modeling is applied also to amyloid fibril formation, which predicts how the fibril length and fibril concentration depend on the protein concentration. Chapter 6 introduces the Levinthal paradox stating that protein folding cannot occur on the time scale of the universe due to an astronomically large number of conformations that need to be sampled in order to find the unique native state. Which is clearly wrong. But how does one study protein folding kinetics? Mass-action models are introduced and a two-state kinetics is discussed in more detail. Folding of larger proteins involves in addition to unfolded and folded states also kinetic intermediates. The two-state kinetics is then generalized and the corresponding formalism, which involves setting up and solving master equations, is worked out in Appendix 6A. Chemical denaturation strongly affects folding or unfolding rates or both. A chevron plot captures both folding and unfolding rate coefficients in dependence on the denaturant concentration. Finally, the Levinthal puzzle is solved by visualizing the free energy landscape of protein folding as a funnel, whereby a relatively modest funnel incline is sufficient to account for experimentally observed folding rates as mathematically formulated in Appendix 6B in the context of the Zwanzig–Szabo–Bagchi model.

Chapter 7 focuses on protein evolution, starting with Darwin's principle of descent with modification though random mutation and natural selection. Terms like homologs, sequence identity, similarity, and homology are introduced. After two protein sequences are optimally aligned, their similarity can be quantified by using the Hamming distance. The idea behind sequence comparison is that proteins with similar sequences not only perform similar functions but may share a common ancestor. Horizontal gene transfer is discussed, whereby a gene enters a foreign cell and becomes stably incorporated into its genome, which can accelerate evolution. This process complicates the Darwin's idea of descent as well as construction of a phylogenetic tree. The sequence-structure relationship with its redundancy

(many different sequences adopt the same native fold) is further illustrated using the lattice HP model. Directed evolution is discussed as a way of improving protein function and creating new proteins with desired properties in the laboratory. Chapter 8 introduces the principles of bioinformatics as a means to compare and analyze proteins or nucleic acids to learn about their function. Several on-line databases and software tools that help align and compare the sequences are referenced. Prior to comparison, two (or more) sequences need to be aligned and each alignment is given a score that depends on the scoring function. The tool of bioinformatics is Basic Local Alignment Search Tool (BLAST) method, which compares an unknown sequence against an entire database of known sequences with an objective to identify similar sequences needed to construct a phylogenetic tree. The Shannon entropy expresses the degree of residue conservation during evolution and a quantity referred to as the mutual information can be used to find correlations between mutated amino acids, which can help draw inferences about the structure of the unknown sequence.

Chapters 9 and 10 offer fundamental insights into atomistic description of protein conformation and interatomic potentials that need to be defined to study protein dynamics by computer simulations. Protein Data Bank (PDB) is a publicly available on-line collection of PDB-formatted files that contain three-dimensional coordinates of each atom of all structurally resolved proteins or protein complexes. Computer simulations of proteins are based on the force field, a collection of interatomic potentials and their parameterization. The water model is needed for explicit solvent simulations. To increase computational efficiency, some force fields utilize an implicit solvent approach, whereby interactions between charged protein atoms are "renormalized" to account for the presence of water using a mean-field generalized Born or Poisson–Boltzmann model. Two major simulation methods are described, molecular dynamics (MD) and Monte Carlo (MC). MD, which is arguably the lead computer simulation method for studies of protein dynamics, is based on empirical continuous interatomic potentials, and gives a complete classical description of protein dynamics. Metropolis MC, which can also be used in combination with square-well-type (discontinuous) potentials, is generally computationally more efficient than MD but is limited to conformational sampling and finding free energy minima. Some enhanced sampling methods are introduced in Appendix 10C. In nature, a protein sequence uniquely determines protein structure. Although all-atom MD simulations can predict the structures of some small proteins, the sequence to structure mapping is far from resolved. Chapter 11 addresses several state-of-the-art approaches, such as comparative modeling and knowledge-based potentials. Critical Assessment of Structure Prediction (CASP) is a yearly community-wide competition that compares and evaluates methods of protein structure prediction of many target sequences that are made publicly available.

Protein actions are not determined solely by their three-dimensional structure but can arise from protein motion. Chapter 12 discusses the importance of collective modes of protein motion, in which changes in atomic coordinates are correlated with each other and involve the entire protein. These soft modes (the term bears some analogy to the condensed-matter soft modes arising from symmetry breaking in the vicinity of a second-order phase transition) are often related to protein action, as exemplified by the walk of reverse transcriptase along RNA. Elastic network models represent powerful tools that can tease out the normal modes of protein motion. These models can also be applied to protein complexes, and protein assemblies, such as viruses. All you need is the adjacency matrix, which contains the number of contacts for each amino acid in the sequence. The formalism is worked out in six Appendices. Simple, beautiful, and, importantly, predicted modes fit experimental data!

Fundamental understanding of proteins represents one of the many aspects of human knowledge, which is one of the cornerstones of human civilization. Its prosperity is plagued by many still incurable diseases. Tackling these diseases and finding cures requires molecular-level understanding of underlying mechanisms of each disease, in which proteins are often key players. Chapter 13 delivers the basics of drug discovery, including specific terminology used in the field, multiple stages of discovery and specific conditions that a drug needs to satisfy. Ligand-based and target-based drug discovery approaches are discussed alongside docking algorithms. Drugs are classified into small molecules and biologics and pros and cons of each are elaborated on. This chapter is of value also because it contains references to several databases and publicly available software packages, which in principle allows any motivated individual to join the research.

I recommend this textbook to any graduate student who is considering a career in protein science or proteomics. I thoroughly enjoyed reading the book cover to cover. The book's format is close to the US letter format, which nicely fits with all its glossy 322 pages into my handbag so I could easily carry it around. The book fully survived my not-always-gentle handling while reading it in a coffee shop, mall, moving car (not while driving, mind you!), even at a dinner table. This is more than I can say about another textbook (to be kept anonymous), which fell apart leaf by leaf by the end of the quarter-term course. Thanks *Garland Science*!