

Preface

Protein folding, the process by which newly synthesized proteins fold into the specific three-dimensional structures defining their biologically active states, is an old scientific problem that can be dated back at least seven decades, namely to the experiments of Anson and Mirsky in the 1930s. It is also multifaceted, changing its definition according to the background and emphasis of the particular researcher. For the cell biologist-biochemist the *in vivo* protein folding problem consists of identifying, isolating and characterizing all components of the cellular machinery in charge of facilitating and catalyzing protein folding inside the cell. From a bioinformatics viewpoint, the folding problem could be seen as devising methods to predict with high accuracy the native three-dimensional structure of proteins from the amino acid sequence alone. Physics-inclined scientists phrase the folding problem as understanding the processes and mechanisms that control the self-organization of disordered protein molecules to form unique, exquisitely detailed structures, while avoiding their irreversible assembly into high-order aggregates. The scope of this book belongs to the last of these viewpoints.

To introduce the book it is useful to take a historical perspective, which illustrates how the prevailing views about the mechanisms of protein folding have closely followed the idiosyncrasies in the catalog of available proteins and experimental approaches. In the early days and for a long time after, folding was circumscribed to equilibrium denaturation experiments on a small group of complex proteins, such as hemoglobin, because they were readily available. A theoretical framework to interpret experiments was not available, and there was significant discussion as to whether simple models based on elementary chemical reactions could be applied to protein folding (incidentally a similar discussion has regained center stage in recent years). The development of techniques that exploited thiol chemistry to trap intermediary folding species in proteins containing disulfide bonds, together with folding coupled to prolyl-bond isomerization, opened the era of kinetic experiments. This led to the characterization of folding as a convoluted process involving multiple pathways, misfolded intermediates, and heterogeneous unfolded states. However,

the question was whether such heterogeneity was intrinsic to the folding reaction or induced by the trapping reactions, which involved the formation or breakage of covalent bonds. Later on, the combination of molecular biology and stopped-flow kinetic methods with millisecond resolution changed the landscape dramatically. Many single domain proteins with neither disulfide bonds nor *cis* prolines were studied, showing what appeared to be very simple behavior. In the absence of chemical trapping, folding of small proteins looked like a two-state process. Two-state implied the existence of a high free energy barrier separating the folded and unfolded states, which seemed to agree with the still slow (seconds to milliseconds) folding kinetics observed in these proteins. In a two-state regime the only mechanistic information accessible to experiment relies on mapping out the properties of the top of the folding barrier (*i.e.* the folding “transition-state”) from the effects that small free energy perturbations have on folding and unfolding rates. Combining this idea with structure-oriented site directed mutagenesis resulted in the protein engineering approach to protein folding, which was independently initiated in the labs of Goldenberg, Fersht, and Matthews, and then fully developed by the Fersht group. The approach quickly caught on among protein biochemists, who applied it to many two-state proteins. Theoreticians immediately saw this avalanche of new experimental results as an opportunity to test results from theory and computer simulations, leading to the first *de facto* connection between the worlds of experiment and theory in protein folding. At the same time research in protein misfolding and aggregation was starting to reach the status of quantitative science. This state of affairs has been portrayed in detail by several books that appeared in the 1990s and 2000, including “Protein Folding” edited by Creighton, “The Mechanisms of Protein Folding” edited by Pain, and “Protein Folding Mechanisms” edited by Richards, Eisenberg, and Kim.

However, in the last 10 years there have been important developments in the area of protein folding and aggregation that have not yet been discussed in a book of these characteristics. These advances have emerged from a close partnership between statistical theory, novel approaches that dramatically increase the temporal, structural and ensemble resolution of folding experiments, and the maturity of computer simulations, which are now capable of producing results directly comparable to experiments. Once again, these advances are changing our general perception of protein folding to one that emphasizes the stochastic nature of the process and the subtle energetic balance that eventually determines whether a protein folds, the mechanisms by which it does, and its propensity to aggregate. The aim of this book is to provide an account of these major advances as seen by some of the main contributors. The book is intended for graduate students and postdoctoral researchers actively involved in protein folding research, other scientists interested in the recent progress of the field, and instructors revamping the protein folding section of their biochemistry and biophysics courses. Chapters 1 and 2 focus on the α -helix, one of the basic structural elements found in proteins. The main attraction in investigating α -helix formation is that one encounters many of the features of protein folding but in their simplest version. These two chapters will introduce the reader to

conformational ensembles, partially cooperative unfolding processes, the connection between protein energetics and stereochemistry, detailed kinetic modeling, and simple examples of the application of statistical approaches to the analysis of experimental data. Chapter 3 explains the statistical theory that, even if just judged by the number of times it is cited throughout the book, provides the conceptual backbone for most of the subsequent experimental and computational developments. Chapters 4 through 7 discuss experimental approaches for the investigation of folding mechanisms. This selection is not intended to be comprehensive, but to include techniques that either probe or exploit the stochastic nature of protein folding: classical hydrogen-exchange techniques (Chapter 4), novel ensemble-based methods to estimate folding free energy surfaces from differential scanning calorimetry (Chapter 5), fast-folding kinetic experiments and their most important findings (Chapter 6), and the application of single molecule spectroscopy to protein folding (Chapter 7). Chapters 8 and 9 deal with the impressive recent developments in computational approaches; starting from atomistic simulations of complete folding (Chapter 8) and continuing with applications to *de novo* protein design (Chapter 9). Finally, Chapters 10 and 11 are devoted to the experimental and computational investigation of the other side of the problem, that of protein misfolding and aggregation.

