Available online at www.sciencedirect.com
ScienceDirect
Archives of Biochemistry and Biophysics 469 (2008) 4–19
ABB
www.elsevier.com/locate/yabbi

Review

Protein folding: Then and now

Yiwen Chen [1], Feng Ding [1], Huifen Nie [1], Adrian W. Serohijos [1], Shantanu Sharma [1], Kyle C. Wilcox [1], Shuangye Yin [1], Nikolay V. Dokholyan [*,1]

Department of Biochemistry and Biophysics, The University of North Carolina at Chapel Hill, School of Medicine, Chapel Hill, NC 27599, USA

Received 20 March 2007, and in revised form 11 May 2007
Available online 8 June 2007

Abstract

Over the past three decades the protein folding field has undergone monumental changes. Originally a purely academic question, how a protein folds has now become vital in understanding diseases and our abilities to rationally manipulate cellular life by engineering protein folding pathways. We review and contrast past and recent developments in the protein folding field. Specifically, we discuss the progress in our understanding of protein folding thermodynamics and kinetics, the properties of evasive intermediates, and unfolded states. We also discuss how some abnormalities in protein folding lead to protein aggregation and human diseases.

Keywords: Protein folding; Misfolding; Aggregation; Folding kinetics; Protein thermodynamics; Discrete molecular dynamics (DMD)

Protein folding refers to the process by which a protein assumes its characteristic structure, known as the native state. The most fundamental question of how an amino-acid sequence specifies both a native structure and the pathway to attain that state has defined the protein folding field. Over more than four decades the protein folding field has evolved (Fig. 1), as have the questions pertaining to it. This evolution can be divided into two predominant phases. During the first phase, research was focused on understanding the mechanisms of protein folding and uncovering the fundamental principles that govern the folding transition. While the first phase provided general answers to the protein folding question, new and no less ambitious questions arose: what are the mechanisms of protein folding in a context, such as under the influence of other biological molecules in the cellular environment? This next set of questions defined the second phase in protein folding field evolution.

The first phase is akin to a romantic stage of research, where the final goal of studies may not be directly applicable to a broader understanding, or exploitable in a relevant science. The final goal is to determine the basic principles that relate protein sequence and structure. The second phase is a more pragmatic stage of research, where the applications drive research in the field and the rational manipulation of derived knowledge allows engineering of tools for advancement of a relevant science. For example, understanding the functional intermediates that accompany the transition of a protein *en route* to its native state may allow rational manipulation of protein structure via protein design. This example not only relates protein sequence, structure and function, but also demonstrates the engineering aspect of the modern protein folding field.

Next, we survey the questions of the modern protein folding field. We attempt to describe a number of directions where understanding protein folding offers insights into more complex questions in molecular and cellular biology as well as medicine. We also describe new approaches and tools to address complexities associated with these new areas of research. We review studies of protein stability, folding kinetics, intermediate and unfolded states, and protein self-association and aggregation.

* Corresponding author. Fax: +1 919 966 2852.
E-mail address: dokh@med.unc.edu (N.V. Dokholyan).
[1] All authors contributed equally to this work.

0003-9861/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.abb.2007.05.014

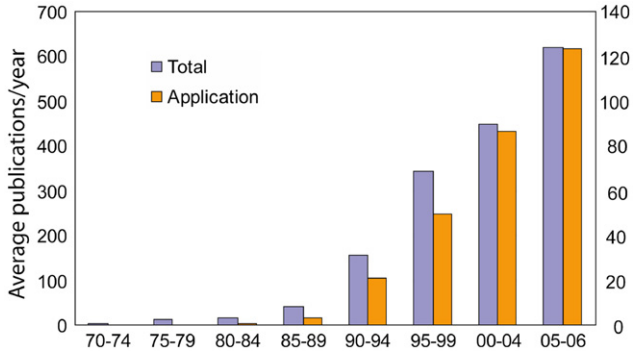

Fig. 1. Growth of the Protein Folding Field. The average number of publications per year in protein folding field (left *y* axis) and the average number of publications per year that are dedicated to application (right *y* axis) were plotted every five years between 1970 and 2004, and 2005–2006. The first dataset was generated by searching articles in PubMed that contain the keyword 'protein folding' or 'protein unfolding' in either title or abstract. The second dataset was extracted from the previous dataset by searching with the following additional keywords: 'engineering', 'design', 'misfolding', 'aggregation', 'amyloid' and 'amyloid disease'.

Studying protein folding

The protein folding field has witnessed significant changes and progress since the original work of Anfinsen showing that proteins can fold spontaneously [1,2]. Early *in vitro* studies showed that the folding process typically occurs on a milliseconds-to-seconds time scale, much faster than the rate estimated assuming that folding proceeds by a random search of all possible conformations. Based upon this observation, Levinthal then proposed that a random conformation search does not occur in folding and that proteins fold by specific 'folding pathways' [3]. On these pathways, the protein molecule passes through well-defined partially-structured intermediate states. Based on this view, numerous experiments and simulations were conducted to test the existence of transient folding intermediates [4,5]. It was expected that the determination of the structures and population of folding intermediates could help elucidate protein folding mechanisms. Earlier experimental studies on protein folding kinetics monitored the structural changes through relaxation of the protein's spectroscopic properties after exposing the protein to folding or unfolding conditions. The data obtained from such experiments exhibit single- or multiple-exponential time-decay: a single-exponential decay is interpreted as a signature of two-state kinetics between the native state and the denatured state, whereas models involving more than two states are required to explain multiple-exponential decay data. These experiments generally probe only the average behavior of proteins, and they are not able to provide information about the folding/unfolding process in atomic details.

The discovery of a class of simple, single-domain proteins which fold via two-state kinetics without any detectable intermediates in the early 1990s [6,7], the development of experimental techniques with improved spatial/temporal resolution [8–13], and the application of computer simulations using simplified lattice and off-lattice models [14,15] greatly enhanced our understanding of various aspects of the protein folding problem. Based on the nucleation theory [16–18], one of the early proposed mechanisms for protein folding, the nucleation-condensation model was formulated [19–21]. In this scenario, a small number of residues (folding nucleus) need to form their native contacts in order for the folding reaction to proceed fast into the native state. The cooperativity of the protein folding process is analogous to that exhibited in first-order phase transitions, which proceed via a nucleation and growth mechanism [22]. Because of these similarities, terminology used in studies of phase transitions, such as energy landscapes and nucleation, was introduced into the discussion of protein folding. The concepts of the nucleation and the free-energy landscape have promoted much of the recent progress in understanding the process of protein folding. Proteins are generally thought to have evolved to exhibit globally funneled energy landscapes [23–25] which allow proteins to fold to their native states through a stochastic process in which the free energy decreases spontaneously. The unfolded state, transition state, native state and possible intermediates correspond to local minima or saddle points in the free-energy landscape.

Advances in experimental techniques such as protein engineering, nuclear magnetic resonance (NMR)[2], mass spectrometry, hydrogen exchange, fluorescence resonance energy transfer (FRET), and atomic force microscopy (AFM), have made it possible to obtain detailed information about the different conformations occurring in the folding process [26,27]. At the same time, computational methods have been developed to better interpret experimental data by using simulations to obtain structural information about the states which are populated during the folding process. In Table 1, we list several advances in experimental and computational methodologies used for investigating the folding of model proteins.

All-atom protein models with explicit or implicit solvents were developed to study the folding thermodynamics and the unfolding dynamics of specific proteins. Technological advances in computation allowed folding simulations of small proteins and peptides at atomic detail [28–30]. However, due to the complexity and vast dimensionality of protein conformational space, all-atom MD simulations have severe limitations on the time and length scales that can be studied. Novel simulation protocols have been proposed to improve conformational sampling

[2] *Abbreviations used:* DMD, discrete molecular dynamics; NMR, nuclear magnetic resonance; FRET, fluorescence resonance energy transfer; AFM, atomic force microscopy; MC, Monte Carlo; GdHCl, guanidinium HCl; DFIRE, distance-scaled finite ideal-gas reference state; CFTR, cystic fibrosis transmembrane conductance regulator; FALS, familial amyloid sclerosis; TSE, transition state ensemble; HX, hydrogen exchange; FAT, focal adhesion targeting; FAK, focal adhesion kinase; SMD, steered-molecular dynamics; ECM, extracellular matrix.