Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Structural Biology

Combining experiment and simulation in protein folding: closing the gap for small model systems

R Dustin Schaeffer[1], Alan Fersht[2] and Valerie Daggett[1,3]

All-atom molecular dynamics (MD) simulations on increasingly powerful computers have been combined with experiments to characterize protein folding in detail over wider time ranges. The folding of small ultrafast folding proteins is being simulated on μs timescales, leading to improved structural predictions and folding rates. To what extent is 'closing the gap' between simulation and experiment for such systems providing insights into general mechanisms of protein folding?

Addresses
[1] Biomolecular Structure & Design Program, University of Washington, Seattle, WA 98195, USA
[2] MRC Center for Protein Engineering, MRC Centre, Hills Road, Cambridge CB2 2QH, UK
[3] Department of Bioengineering, University of Washington, Seattle, WA 98195-5061, USA

Corresponding author: Fersht, Alan (arf25@cam.ac.uk) and Daggett, Valerie (daggett@u.washington.edu)

Current Opinion in Structural Biology 2008, **18**:4–9

This review comes from a themed issue on
Folding and binding
Edited by Laura Itzhaki and Peter Wolynes

Available online 1st February 2008

0959-440X/$ – see front matter

DOI 10.1016/j.sbi.2007.11.007

Introduction

All-atom molecular dynamics (MD) simulations can generate atomic-resolution structural models for states that are only indirectly accessible by experiment [1,2]. Recently, increased computer power has been used to generate multiple microsecond (μs) simulations. Research groups using such simulations have reported refolding from unfolded states for a variety of small model proteins and peptides, using both explicit and implicit solvent models. We will discuss the implications of new long timescale simulations using these brute-force methods, as well as how simulation is being used to model poorly populated states. Parallel developments that use increasingly sophisticated models with simplifying features for capturing longer timescales in realistic detail [3–6] will not be discussed.

Partially folded states are heterogeneous, consisting of many rapidly exchanging conformations. Ensemble averaging from such states complicates the interpretation of the experimental data that would normally be readily analyzed for well-folded systems. Structural signals detected by ensemble methods such as nuclear magnetic resonance (NMR) or optical spectroscopy may originate from many different structures (rather than a single average structure) or from a small fraction of folded structure in a large ensemble of unfolded structures [7–9]. All-atom MD complements experimental approaches by providing a molecular framework for interpretation of experiment and fleshing out the protein folding process.

All-atom MD simulations provide a detailed representation of protein dynamics on picosecond (ps) to microsecond (μs) timescales. Robust treatment of protein–protein and protein–solvent interactions can be achieved with explicit solvent MD simulations. The effects of solvent can also be modeled implicitly to increase computational efficiency [10], but it comes at the cost of accuracy. For example, protein folding involves significant displacement of solvent that is not taken into account by such models [11]. Nonetheless, structures generated with implicit solvent simulation can be useful provided they have been experimentally validated.

Comparison of simulation and experiment is complicated by differences in accessible timescale. MD simulations typically run for tens to hundreds of nanoseconds (ns), up to 1–2 μs. Ultrafast folding proteins such as the engrailed homeodomain and the villin headpiece have folding times on the order of μs near room temperature. In general, refolding from extended states using explicit solvent has been out of reach at these timescales. Instead, methods that increase the probability of observing a transition between states are used [12–14]. Also, thermal or chemical denaturant unfolding simulations can be used to characterize events along the folding pathway. Alternatively, biasing potential terms can be applied to limit the time that simulations linger in local minima or to limit sampling of conformational space [2,15].

Where does protein folding start? Residual structure in denatured states

Chemically denatured states are generally highly expanded and obey a power law dependence between expansion, as measured by radius of gyration (Rg), and protein length [16]. But, denatured states generated by other methods typically display signals incompatible with sequence-independent residual structure [17]. Recently, Förster-resonance energy transfer (FRET) was used to measure protein expansion in chemical denaturants. Eaton and co-workers used single-molecule FRET to determine the radius of gyration of Protein L and Csp*Tm*

under a range of guanidinium chloride (GdmCl) concentrations [18]. The two proteins collapsed to different degrees at low denaturant concentrations, contrary to simple length dependence of the Rg. This contrasts with previous results showing no collapse for Protein L at low denaturant concentrations using time-resolved SAXS [19]. The dyes used for the FRET studies may be perturbing the denatured state. To investigate this further, both proteins were simulated in water/urea mixtures and the Rg increased with increasing urea; however, comparison with experiment is complicated by the different solvent conditions.

Protein unfolding: observation of reversible results

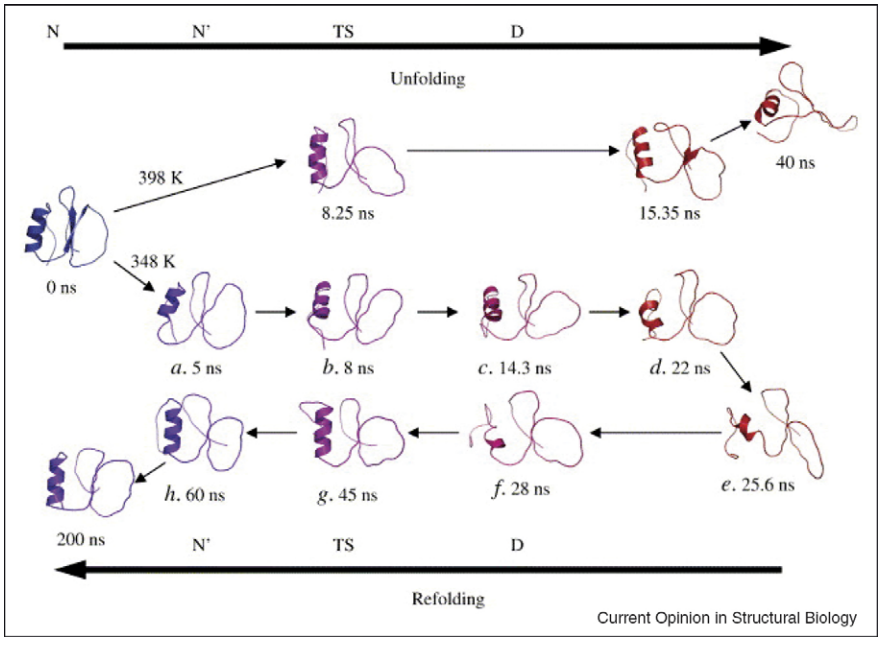
According to the principle of microscopic reversibility, the unfolding pathway should be the reverse of the folding pathway, and we expect unfolding simulations to shed light on folding. All-atom MD simulations of chymotrypsin inhibitor 2 (CI2) have demonstrated microscopic reversibility of the folding pathway [20]. A CI2 simulation conducted at the protein's T_m (348 K) crosses the transition state (TS) to the denatured state, remains there for 36 ns, and then refolds (Figure 1). The structures of both the starting and refolded states are very similar on the basis of physical properties, topology, and pairwise contact maps. Detailed analysis of the loss of interactions during unfolding demonstrates that the same contacts are gained in reverse during the folding process. Thus, this study demonstrates that folding and unfolding are microscopically reversible within a single trajectory. Additional demonstrations of microscopic reversibility are forthcoming. Such studies provide justification for studying protein folding by more computationally efficient unfolding simulations.

A common criticism of simulation is that there is the difficulty in determining the properties of an ensemble of conformations from a limited number of simulations. Day and Daggett showed that 5–10, 20-ns thermal unfolding simulations capture the properties of a larger set of 100 simulations [21]. Further, the average unfolding/folding pathway delineated by these properties accurately describes the nucleation–condensation folding process of CI2, with simultaneous formation of secondary and tertiary contacts, in agreement with experiment.

Protein refolding: repeated simulation of microsecond timescales

Protein refolding from extended structures using MD is hindered by the approximations necessary to access long timescales. In many cases hydrophobic collapse can be observed, but the precise criteria used to assess refolding vary, measures such as backbone Cα RMSD, or particular contacts in the hydrophobic core are used in an inconsistent fashion (see Table 1). Consider the landmark 1 μs simulation of Duan and Kollman in which hydrophobic collapse was achieved but the native state was not reached, as expected from the minimal 4 μs folding time [22,23]. The seminal refolding prediction of Trp-cage by

Figure 1



Chymotrypsin inhibitor 2 unfolds through its transition state to the denatured state, and then returns to a *N′* state through the same transition state at 348 K. This is the first observed case of microscopic reversibility in a simulation system.