MAR08-2007-002413

Abstract for an Invited Paper
for the MAR08 Meeting of
the American Physical Society

Folding peptides and proteins with all-atom physics: methods and applications

M. SCOTT SHELL, Department of Chemical Engineering, University of California Santa Barbara

Computational methods offer powerful tools for investigating proteins and peptides at the molecular-level; however, it has proven challenging to reproduce the long time scale folding processes of these molecules at a level that is both faithful to the atomic driving forces and attainable with modern commodity cluster computing. Alternatively, the past decade has seen significant progress in using bioinformatics-based approaches to infer the three dimensional native structures of proteins, drawing upon extensive knowledge databases of known protein structures [1]. These methods work remarkably well when a homologous protein can be found to provide a structural template for a candidate sequence. However, in cases where homology to database proteins is low, where the folding pathway is of interest, or where conformational flexibility is substantial—as in many emerging protein and peptide technologies—bioinformatics methods perform poorly. There is therefore great interest in seeing purely physics-based approaches succeed. We discuss a purely physics-based, database-free folding method, relying on proper thermal sampling (replica exchange molecular dynamics) and molecular potential energy functions. In order to surmount the tremendous computational demands of all-atom folding simulations, our approach implements a conformational search strategy based on a putative protein folding mechanism called zipping and assembly [2-4]. That is, we explicitly seek out potential folding pathways inferred from short simulations, and iteratively pursue all such routes by coaxing a polypeptide chain along them. The method is called the Zipping and Assembly Method (ZAM) and it works in two parts: (1) the full polypeptide chain is broken into small fragments that are first simulated independently and then successively re-assembled into larger segments with further sampling, and (2) consistently stable structure in fragments is detected and locked into place, in order to avoid re-sampling those degrees of freedom in subsequent steps. ZAM pursues all potential folding routes it finds, which may be mutually exclusive, and it ranks these by calculating free energies along the way. Importantly, it gives full conformational ensembles and folding pathways, features not captured by bioinformatics approaches. We also discuss ways in which the structural ensembles and folding pathways of ZAM can facilitate the rational design of peptide technologies. In particular, we examine the coupling of ZAM-produced structures with coarse-grained theories of transport and association, in order to model the interactions of peptides with membranes (for insertion processes), proteins (for binding processes), and other peptides (for aggregation processes). Importantly, this approach is able to capture highly sequence-specific effects due to the atomistic nature of the ZAM folding simulations, providing a predictive tool for targeted sequence mutations. 1. J. Moult, *A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction,* Curr. Opin. Struct. Biol. **15,** (2005). 2. K.M. Fiebig and K.A. Dill, *Protein core assembly processes,* J. Chem. Phys. **98,** (1993). 3. S.B. Ozkan, G.H.A. Wu, J.D. Chodera, and K.A. Dill, *Protein folding by zipping and assembly,* Proc. Natl. Acad. Sci. U. S. A. **104,** (2007). 4. M.S. Shell, S.B. Ozkan, V.A. Voelz, G.H.A. Wu, and K. Dill, *Can molecular physics predict the native structures of globular proteins?, under review*, (2007).