

Minimalist Representations and the Importance of Nearest Neighbor Effects in Protein Folding Simulations

Andrés Colubri^{1,2,3,5}, Abhishek Jha^{1,2,5}, Min-yi Shen⁴, Andrej Sali⁴, R. Stephen Berry^{1,5}, Karl F. Freed^{*1,5} and Tobin R. Sosnick^{2,3}

¹ Department of Chemistry, The University of Chicago, Chicago, IL 60637

² Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL 60637

³ Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL 60637

⁴ Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, and California Institute for Quantitative Biomedical Research, University of California, San Francisco, San Francisco, CA 94143

⁵ The James Franck Institute, The University of Chicago, Chicago, IL 60637

Abstract

In order to investigate the level of representation required to simulate folding and predict structure, we test the ability of a variety of reduced representations to identify native states in decoy libraries and recover the native structure given the advanced knowledge of the very broad native Ramachandran basin assignments. Simplifications include the removal of the entire side chain or the retention of only the C_{β} atom. Scoring functions are derived from an all-atom statistical potential that distinguishes between atoms on different residue types. Structures are obtained by minimizing the scoring function with a computationally rapid simulated annealing algorithm. Results are compared for simulations in which backbone conformations are sampled from a PDB-based backbone rotamer library generated by either ignoring or including a dependence on the identity and conformation of the neighboring residues. Only when the C_{β} atom and nearest neighbor effects are included do the lowest energy structures generally fall within 4 Å of the native backbone RMSD, despite the initial configuration being highly expanded with an average RMSD ≥ 10 Å. The side chains can be reinserted into the C_{β} models with minimal steric clash, demonstrating that the C_{β} -level simulations respect the packing of side groups. Therefore, the detailed, all-atom information lost in descending to a C_{β} -level representation can be recaptured to a large measure using backbone sampling that includes nearest neighbor effects and an appropriate scoring function.