426d Development of Rigorous Distance Bounds for Improved Protein Structure Prediction

Scott R. McAllister and Christodoulos A. Floudas

A multitude of approaches yielding significant advances have defined the field of protein structure prediction in recent years. The first major class uses comparative modeling techniques to predict structures by making comparisons with experimentally-determined homologs. A second category applies fold recognition methods that rely upon evolutionary distant homologs and the knowledge that protein structure is more conserved than sequence. The final class of protein structure prediction is made up of first principles approaches. Knowledge-based first principles methods incorporate distance constraints from known structures into statistical models. These approaches can be contrasted with physics-based first principles approaches, which try to predict protein structure based solely upon the primary sequence and the application of detailed force fields and energy models. The recent development of a novel physics-based first principles approach, ASTRO-FOLD, has shown tremendous promise in this area ([1]). A recent review by Floudas et al. (2005) provides a thorough overview of the methods for each class of protein structure prediction ([2]).

One area of critical interest is the need for quicker and more accurate methods for tertiary structure prediction, given secondary structure information and β-sheet topologies. This is especially important in physics-based first principles approaches, as a carefully restrained problem can significantly reduce the the feasible search space of the protein. The search space reduction can result in both an increase in the speed of the algorithm, as well as improved results through a more focused problem statement.

A multi-stage framework has been developed for the identification of tight, rigorous bounds on the inter-residue distances within a protein structure ([3]). The main components of this framework are (i) initial distance bound identification through distance geometry rules, (ii) bound tightening through a novel linear programming model, (iii) further distance bound refinement based on the generation of an ensemble of structures and the hypothesis of a well-packed protein. The initial distance bounds are derived from rigorously quantified observations for inter-residue distances within local secondary structure elements and loop residues, as well as between β-sheets, Cysteine disulfide bridges, and secondary structure to loop residues. By applying the idea of triangle inequalities to linear programming, these initial bounds can be iteratively tightened to satisfy both these triangle constraints and those based on residue contact monotonicity. After generating an ensemble of protein structures using DYANA ([4]), which employs simulated annealing with molecular dynamics in torsion angle space, a detailed statistical analysis can be followed to derive tighter upper bounds on the distances of the residues that pack within the hydrophobic core of the protein. Early work with this protocol has demonstrated the ability to rapidly predict the structure of Protein G (56 AA) with an RMSD of < 5 Å. A complete set of the computational results for a number of protein systems will be presented.

[1]-J. L. Klepeis and C. A. Floudas. ASTRO-FOLD: A Combinatorial and Global Optimization Framework for Ab Initio Prediction of Three-Dimensional Structures of Proteins from the Amino Acid Sequence. Biophysical Journal, 85 (2003): 2119-2146.

[2]-C. A. Floudas, H. K. Fung, S. R. McAllister, M. Mönningmann, and R. Rajgaria. Advances in Protein Structure Prediction and De Novo Protein Design: A Review. Chemical Engineering Science, (2005). in press.

[3]-S. R. McAllister and C. A. Floudas. Rigorous Distance Bounds for Protein Structure Prediction via Distance Geometry and Linear Programming. (2005). In preparation.

[4]-P. Güntert, C. Mumenthaler and K. Wüthrich. Torsion angle dynamics for NMR structure calculation with the new program DYANA. Journal of Molecular Biology, 273 (1997): 283-298.