Publications

The research conducted in Jean-Claude Latombe's group studies computational methods to help biologists explore the conformational spaces of individual molecules, as well as of groups of interacting molecules. In particular, it investigates such techniques as bounding-volume hierarchies to quickly detect steric clashes between atoms, probabilistic roadmaps to compute ensemble properties of molecular pathways (properties characterizing many pathways simultaneously), and nearest-neighbor techniques to quickly detect similarities in 3-D structures. Three new papers were recently submitted and accepted for publication in computational biology venues.

● Two graduate students, Mitul Saha and Rohit Singh, attended the Pacific Symposium on Biocomputing held in Lihue, Kauai, Hawaii, on January 3-7, 2003, where they presented a paper entitled IDENTIFYING STRUCTURAL MOTIFS IN PROTEINS.

In biological macromolecules, 3-D structural patterns (motifs) are often repeated across different molecules. Detection of these motifs in a new molecule can provide useful clues to its functional properties. Mitul and Rohit formulate the problem of identifying a given motif in a target protein and discuss the notion of complete matches vs. partial matches. They describe the error criterion that has to be minimized and propose different metrics for evaluating the quality of partial matches. They also present a new polynomial-time algorithm for the problem of matching a given motif in a target protein. Unlike other methods, the error minimized by the algorithm directly translates to the root mean square deviation (RMSD), the most commonly accepted metric for structure matching in biological macromolecules. The paper presents experiments exploring the quality of matches found by the algorithm in matching active sites in proteins. Details of this work can be found at http://robotics.stanford.edu/~mitul/psb03.ps.

● A paper by Serkan Apaydin, Doug Brutlag, Carlos Guestrin, David Zhu, Jean-Claude Latombe, and Chris Varma - entitled STOCHASTIC ROADMAP SIMULATION: AN EFFICIENT REPRESENTATION AND ALGORITHM FOR ANALYZING MOLECULAR MOTION - has been accepted for publication in the Journal of Computational Biology. Serkan and Carlos are PhD students in Latombe's and Prof. Koller's groups, respectively. In 2001-02, Chris was a Master student in Latombe's group, and is now a PhD student in the Harvard-MIT Bio-Medical program. Doug Brutlag is a professor in Biochemistry at Stanford. Until recently, David Hsu was a postdoctoral researcher in Jack Snoeyink's group at UNC, and is now a faculty member at the National University of Singapore.

In this paper, the authors describe a new technique to compute ensemble properties of molecular motion efficiently and accurately using a graph constructed by sampling a molecular conformation space at random. SRS is successfully applied to two biological problems: computing the probability of folding (Pfold), an important order parameter that measures the "kinetic distance" of a protein's conformation from its native state (see Figure 1); and estimating the expected time to escape from a ligand-protein binding site. The paper can be accessed at http://robotics.stanford.edu/~apaydin/JCB.pdf.

● A paper by Itay Lotan and Fabian Schwarzer, APPROXIMATION OF PROTEIN STRUCTURE FOR FAST SIMILARITY MEASURES, has been accepted for publication at the 7th Annual International Conference on Research in Computational Molecular Biology (RECOMB) that will be held in Berlin on April 10-13, 2003. Itay is a PhD student in Latombe's group. Until December 2002, Fabian was a postdoctoral researcher in this group.

In this paper, Itay and Fabian show that structural similarity between proteins can be decided well with much less information than the commonly used $C\alpha$ representation. Indeed, this representation contains redundant information due to the facts that proteins have a chain-like kinematic structure and limited compactness avoiding steric clashes. A wavelet analysis of random chains and proteins justifies approximating sub-chains by their centers of mass. For proteins, similarity measures that use this approximation are highly correlated to similarity measures using the full representation and are therefore useful, e.g., as fast filters. Experimental results with such simplified similarity measures in two applications, nearest neighbor search and automatic structural classification, show significant computational speed-up.
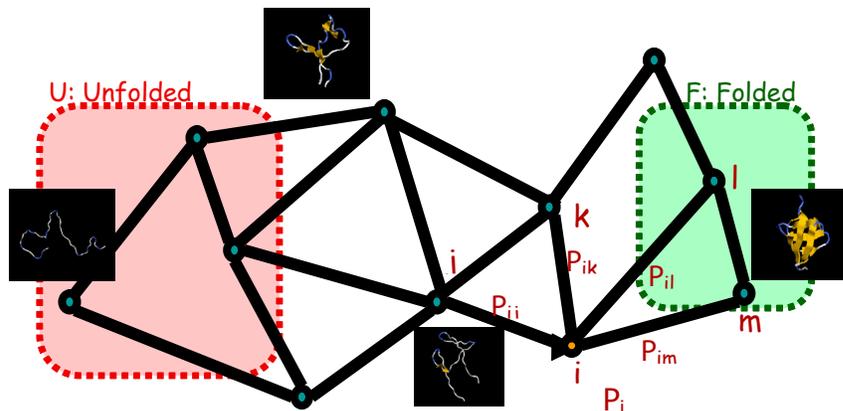

Figure 1: A simple probabilistic roadmap to compute the Pfold parameter

Student Profile: Andrew Leaver-Fay

Andrew Leaver-Fay is a second year PhD student at the University of North Carolina in Chapel Hill, working under the supervision of PI Jack Snoeyink. Andrew did his undergraduate work at the University of Virginia where he studied biology, chemistry, philosophy, neuroscience, and computer science.

Andrew began his research with the UNC biogeometry group in the fall of 2001. He joined the team to model the binding geometry of various small molecules to the active site of the human pregnane xenobiotic receptor (hPXR), a critical mediator in the pathway that breaks down many drugs. Matt Redinbo (UNC Chemistry) and his group crystallized hPXR and determined its structure. They observed that the binding pocket in hPXR is large, smooth, and predominantly hydrophobic, so they conjectured that the few hydrogen bonds donors and acceptors would determine the binding geometry with different ligands. The ITR team used a combinatorial search of hydrogen bond donors and acceptors between hPXR and the small molecule ligands to find pairs of bonds. Hydrogen bonds were modeled as opposing wedges

and sticks and binding positions predicted by non-linear optimization. The results were presented in the 2001 ITR project meeting and as a poster at RECOMB. Andrew in particular explored validation by comparing with existing docking programs; we would like validation from further work in crystallography, but this is difficult.

Andrew is a member of the first class of trainees in the new bioinformatics program, an interdepartmental certificate-granting program which serves to provide a solid background in the diverse issues surrounding computational biology. UNC Chapel Hill has an excellent, long-standing, interdisciplinary training program in biophysics (lead by Prof. Barry Lentz, Biochemistry). Although this program includes some computer science faculty, it has never had a computer science student because the program requirements do not align well with those of the computer science PhD. The new bioinformatics program (led by Prof. Alex Tropsha, Pharmacy) was constructed to allow students to fulfill degree requirements in their PhD programs, yet still give them an interdisciplinary experience through required courses, seminars, and research rotations.

This past fall Andrew did a rotation with Dr. Tropsha, who investigates four-body Delaunay statistical potentials and their application to analyze protein packing and to discriminate native protein structures from "decoys." For a sample of a thousand high-resolution protein structures with low sequence identity, Andrew examined Delaunay triangles and tetrahedra computed from representative points for each amino acid. Adding a distinction based on buriedness within a four-body potential would have produced too many

configurations to sample adequately, so Andrew analyzed a "three-body potential" by focusing on Delaunay facets instead of entire tetrahedra. Describing boundary facets and their vertices as "exposed" and then discriminating non-boundary facets on the number of exposed vertices they contain, the three body potential makes five discriminations between facets on the basis of their buriedness.

The three-body potential alone showed moderate success in decoy discrimination. In a weighted combination with the four-body potential, it improved Z-scores by about 15%. Three-body profile scores (the sum of the scores for each facet a given vertex participates in) identify well-formed cores in folded protein structures. The residues with high profile scores that were in contact in the folded structure of the CI2 domain were found to be in contact in post-transition state structures, but these contacts were not as buried as in the folded structure and were thus scored lower. This suggests that a folding algorithm that asks, "What should I try to bury?" would be able to take advantage of this potential. Andrew presented these results at the bioinformatics program poster session in January.

This semester, Andrew is working again with PI Snoeyink. He is adding a rotamer library to the MATLAB toolbox that is being developed at UNC, and learning the dezymer system of PI Homme Hellinga (Duke Biochemistry) with an eye toward protein design by inverse folding on malleable protein backbones. This is an exciting project, and one where his interdisciplinary training is very important.

- Profile by Jack Snoeyink

Workshop

At the end of January, Patrice Koehl, Rachel Kolodny, Andrea Mantler, and Jack Snoeyink participated in Ileana Streinu's workshop on Rigidity and Proteins at the Bellairs Research Institute.

From left, Patrice talks about protein sequence space; Jack talks about almost-Delaunary tetrahedra; and Patrice, Rachel, Jack and Andrea work on origami protein folding.