BioGeometry NEWS
May 2003
biogeometry.cs.duke.edu

Schedule for PI Meeting at Stanford

Thursday, June 5

8:00	Coffee & continental breakfast
8:45	Welcome by Michael Levitt (Stanford)
9:00	Weighted volume derivatives by Patrice Koehl (Stanford)
9:30	2-D Monte Carlo chain folding, based on expansive/contractive motions by Rachel Kolodny (Stanford)
10:00	Modeling and simulation of lymph node germinal center reaction by Ben Wong (Stanford)
10:30	Approximation of protein structures for fast similarity measures by Itay Lotan (Stanford)
11:00	Discussion
11:30	Lunch
2:00	Deformable spanners for proximity maintenance and collision detection of molecular chains by An Nguyen (Stanford)
2:30	A comparison of techniques for incrementally updating Delaunay and power diagrams across simulation time steps by Daniel Russell (Stanford)
3:00	The derivatives of backbone motion by Kimberly Noonan (UNC)
3:30	Speeding up REDUCE for placing hydrogen by Yuanxin Liu (UNC)
4:00	Coffee
4:30	PI Meeting & NSF Renewal Planning for PIs only
6:30	Dinner

Friday, June 6

8:00	Coffee & continental breakfast
9:00	Three-body potentials and molecular surfaces by Andrew Leaver-Fay (UNC)
9:30	The fine structure of protein-protein interfaces by Andrew Ban (Duke)
10:00	Protein-protein docking I: Exhaustive search by Johannes Rudolph (Duke)
10:30	Protein-protein docking II: Plans for local improvement by Vicky Choi (Duke)
11:00	Protein-protein docking III: Features marked by critical points by Yusu Wang (Duke)
11:30	Lunch
2:00	Discussion on docking and interfaces
2:30	Jacobi sets of two Morse functions by Alper Üngör (Duke)
2:45	Local squaring equations: real/reciprocal space duality by Jeff Roach (UNC)
3:30	Coffee
4:00	Concluding discussion and planning

People

Johannes Rudolph, Assistant Professor of Chemistry and Biochemistry at Duke, has been named a PI on the BioGeometry Project, effective June 1, 2003. He is interested in protein structures and the mechanisms of their interactions related to cell growth and cancer. He works with the rest of the Duke group on relevant geometric questions, such as the shape of interfaces and computational docking as a tool to predict interactions.

Patrice Koehl, Senior Research Associate in the Department of Structural Biology at Stanford, has also been named a PI on the project, effective February 1, 2003. Patrice takes an essentially computational approach to some of the central problems in structural biology, including protein design, protein folding and structure prediction. He interacts with the rest of the Stanford group but also collaborates beyond institutional bounds, thus strengthening the coherence of the entire project.

Research

Local Squaring Functions for Multi-atom Templates

The local squaring functions are intended to measure the likelihood that a particular molecular fragment occupies a given region of the unit cell. When the only fragments considered are single atoms, the local squaring functions take a simplified form where it is unnecessary to model fragment orientation. Unfortunately this simplification requires high resolution data to be effective. At resolutions typical of macromolecular crystallography, say 3.0 to 2.0 angstroms, single atom fragments must be accompanied by larger, more complex multi-atom fragments.

These multi-atom fragments encode information not only on the placement of atoms in the protein but also the interactions between them. This additional information necessarily complicates the situation. Not only is it necessary to determine some suitable parameterization of the orientation space, it is also necessary to determine which individual fragment among a large library of potential fragments is the best choice given the initial electron-density. By emphasizing density in regions of the unit cell likely to contain a particular fragment and dampening regions likely to contain no fragment at all, an iterative electron-density improvement regime can be established. Furthermore by maintaining a list of particularly well placed fragments, the process ends with an initial atomic model of the protein. Naturally this initial atomic model is unsuitable for atomic coordinate refinement, however, it can be used for the next cycle of iterative electron-density refinement.

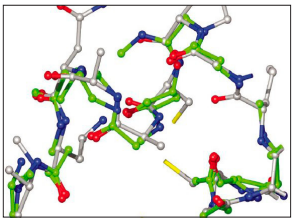


Figure 1: Green fragments placed by local squaring function method.

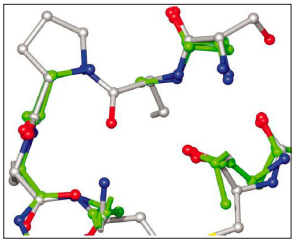


Figure 2: Note that even though some atoms are not covered by a green fragment, the protein backbone is still well represented,

Where high resolution data has been available, successive cycles of electron-density refinement and model building have resulted in high quality density maps using only single atom fragments. These maps, when submitted to coordinate refinement, produce very accurate structural solutions.

Coordinate refinement, being in essence a high dimensional, non-linear optimization technique, is susceptible to local extrema and over fitting. It is becoming increasingly clear that misuse of coordinate refinement can lead to incorrect structures that appear to be correct. The best way to remedy this situation is to couple strong validation techniques with an improved starting point. The local squaring functions intend to provide both.

Although some progress has been made in the transformation from single atom fragments to multi-atom fragments, much remains to be accomplished. Initial studies by PI Charlie Carter and postdoc Jeff Roach, both of UNC, on

sampling and interpolating within a particular parameterization of the orientation space have been improved somewhat by introducing a local coordinate framework. They are continuing research on the types and number of molecular fragments to compose our library. Aside from the simple "lego"-type fragments that they initially considered, they are exploring fragments from Kolodny, Koehl, Guibas and Levitt (2002) as well as fragments developed by Alex Tropsha. Finally their calculations have been greatly improved by the share they purchased in the IBM Blade Server. This distributed computing environment greatly increases the number of fragments they can routinely consider as well as improves their orientation sampling density.

Their initial results have been published in: Roach, J. & Carter, C. W. Jr. (2003). Local squaring functions for non-spherical templates. Acta Cryst. A59, 273--280.

Student Profile: Serkan Apaydin

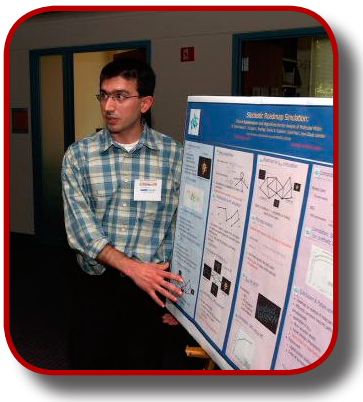
Mehmet Serkan Apaydin is a Ph.D. student in Electrical Engineering at Stanford. He works with Prof. Jean-Claude Latombe from Computer Science and Prof. Doug Brutlag from Biochemistry. He obtained his MSEE degree from Stanford in 1999, and before that he was an undergraduate at Bilkent University, EE department, Ankara, Turkey. Without prior background in biology (he was working on computer reliability), Serkan decided to change fields upon reading an interview with Prof. Don Knuth on what he would work on if he were to start research today (Don Knuth answered "robotics and biology"). But more importantly, a David L. Cheriton Stanford graduate fellowship (SGF) allowed him to pursue his interests, even before the BioGeometry project had started.

Since then, along with his collaborators, he has successfully developed a new simulation technique, Stochastic Roadmap Simulation (SRS), to efficiently compute ensemble properties of molecular motion. Ensemble properties are quantities that depend on a multitude of pathways. Examples include pfold, an order parameter in protein folding; or escape time, a quantity that measures the time to escape from a funnel of attraction of a potential binding site in ligand-protein binding. His work on SRS is a collaborative effort of people in three research groups at two (now three) universities. Working with Carlos Guestrin in Prof. Daphne Koller's group and David Hsu (formerly a member of the BioGeometry project at UNC, now a professor at the National University of Singapore), Serkan realized that the complicated conformational roadmaps we had been using to simulate protein folding pathways could be considered a Markov process. This realization dramatically simplified the computation allowing us to tackle even more complicated problems. This seminal result will have impact on nearly all protein structure simulations that currently involve Monte Carlo methods, as well as on other simulations (e.g., ligand-protein binding). Serkan presented his results at conferences in 2002, and a paper combining these results has been accepted for publication in the Journal of Computational Biology, to appear in summer 2003.

Serkan is now working on improving SRS applicability to higher dimensional problems, with more detailed protein representations and more complicated energy models. He has demonstrated (with Prof. David Hsu) that the stationary distribution convergence property (to Boltzmann distribution) of SRS can still be achieved when sampling non-uniformly. He is studying whether ensemble properties can also be accurately computed using techniques such as Gaussian sampling originated in robotics motion planning community. Finally, Serkan is also working on the release of the SRS software on the web.

Prof. Brutlag says the following of Serkan: "Serkan's mild manner and low key sense of humor belies a great intelligence and deep understanding of both biology and computer science. I am always impressed when I present him with another biological constraint on his problem and a few days later he has either discovered or created a computational implementation which embodies the constraint and simplifies the solution. As a biologist, I view biological complexities as challenges and Serkan sees them as opportunities to simplify the problem. I was very proud to be able to present Serkan's work as a keynote at the last European Conference on Computational Biology in Saarbruecken (October 2002)."

Serkan is also active in outreach activities. He mentored undergraduate student Julie Greenberg from Harvard University last summer, and has co-organized BCATS 2002 (Biomedical computation at Stanford), a 450-people symposium bringing together students from Stanford and the vicinity for a day of talks and poster presentations.



Publications:

[1] M.S. Apaydin, D.L. Brutlag, C. Guestrin, D. Hsu, and J.C. Latombe. Stochastic roadmap simulation: an efficient representation and algorithm for analyzing molecular motion. Proc. RECOMB 12-21, (2002)

[2] M.S Apaydin, C. Guestrin, C. Varma, D.L. Brutlag, and J.C. Latombe. Studying protein-ligand interactions with stochastic roadmap simulation. Bioinformatics, 18, 18-26 (2002).

[3] M.S. Apaydin, D.L. Brutlag, C. Guestrin, D. Hsu, J.C. Latombe. Stochastic Conformational roadmaps for computing ensemble properties of molecular motion. Workshop Alg. Found. Robotics (WAFR), 2002.

[4] M.S. Apaydin, D.L. Brutlag, C. Guestrin, D. Hsu, J.C. Latombe, and C. Varma. Stochastic roadmap simulation: an efficient representation and algorithm for analyzing molecular motion. J. Computational Biology 10 (2003), to appear.

- Profile by Jean-Claude Latombe

BioGeometry News is the monthly newsletter of the BioGeometry project.

For more information, please visit http://biogeometry.cs.duke.edu/newsletter.

The project is funded by the National Science Foundation under grant CCR-00-86013.