BioGeometry NEWS
August 2004
biogeometry.cs.duke.edu

Education

An important component of education in computational structural biology is the development and delivery of topical courses. We present a short description of one new BioGeometry course.

CS 273: Algorithms for Structure and Motion in Biology (Spring 2004; Batzoglou, Guibas, Latombe; Stanford)

Molecular biology is an exciting application area of computer science, because it presents new computational challenges that ultimately advance the field of CS. Proteins, for example, are long macromolecules that can be modeled as chains in three dimensions with rotational degrees of freedom, a given surface area, and kinematic properties. In that respect, computational methods from robot arm movement, computational geometry, and dynamic simulations, are relevant to the modeling of proteins. However, proteins are vastly more complicated than typical robot arms, and have a complex surface that interacts with the environment. Traditional computational methods have to be modified, or give way to new techniques that are appropriate for the novel challenges of modeling proteins.

This course covers algorithms for modeling structure and motion in biology. It is a course about exciting new algorithms that are motivated by the challenges we face in modeling macromolecules and predicting their properties in silico. We cover biology concepts as needed to present the algorithmic material in context. As this is a relatively new field, many of the methods are currently in development, and the structure of the course reflects that fact. The main reading materials are papers from recent literature, powerpoint slides, and lecture notes that students scribe during the class. Coursework consists of one scribed lecture, two problem sets, and a small project that involves using some existing software for analysis of biological data.

Student Summer Projects

- **Daniel Evans**, now a senior student in computer science at UNC Chapel Hill, spent June and July implementing a tool to help interact with molecular display programs projected on a whiteboard/screen. Through an NSF equipment grant, we have been equipping more offices (for faculty, staff, and students) with projectors. White board trackers such as mimeo allow you to point-and-click directly on the screen. Unfortunately, we are essentially limited to a single button mouse, and many molecular display programs need more buttons and/or keyboard shortcuts. Daniel's Board Helper changes the floating cursor to a floating menu that can be used to select the relevant keyboard keys, which lets us use molecular display programs such as Mage.

- **Andrew Leaver-Fay,** in collaboration with Brian Kuhlman of UNC biochemistry, has been applying dynamic programming to get exact solutions to the problem of placing sidechains on a protein fragment to minimize interaction energy.

- **Yuanxin (Leo) Liu** and Jack Snoeyink have completed development and testing of a Delaunay tessellation program specifically tailored for protein databank (PDB) files. By assuming evenly spaced input points, their program can use Hilbert curves both to speed up point location and to reduce the arithmetic precision needed for correctness. Because it does not constantly check correctness, it is three times faster than its closet competitor. Yet it is correct for all but one of the 20,393 files in the PDB, and that one file contains only the nucleic acids for a protein stored in another PDB file, and therefore violates the basic assumption.

People

Jeffrey Roach was promoted to

Research Assistant Professor in the UNC Department of Biochemistry and Biophysics, effective May 1, 2003. His role in the ITR was elevated to that of Senior Research Associate. Jeff was trained as a mathematician and computer scientist. He has contributed steadily to fundamental advances on the X-ray crystallographic phase problem in a series of papers published in Acta Crystallographica dealing with the extension of Sayre's squaring law to macromolecular crystallography. Using Fourier transformation, Jeff was able to implement the squaring law locally in real space, whereas it had previously been possible only to apply it globally in its reciprocal space formulation. More recently, Jeff has focused on accelerating and generalizing the problem of three-dimensional macromolecular structure alignment and comparison. His algorithm for reducing 3D structures to 1D strings capturing both secondary and tertiary neighbor contacts has been submitted for publication. This work has applications across a broad range of timely problems, including probing deeper phylogenetic relations within and between protein families and characterization of clustered microstates visited during molecular dynamics trajectories.

- **Andrea Mantler** has been working on rigidity analysis as a model for the allosteric binding of glycogen phosphorylase. She and Jack Snoeyink participated in the workshop on Modeling Protein Flexibility and Motions at the Banff Research Center.

- **Deepak Bandyopadhyay** has been working with Luke Huan and his supervisor, Wei Wang, to use his almost-Delaunay tetrahedra for mining spatial motifs from proteins. He presented a poster/demo at ISMB in Glasgow.

Ph.D. Graduates

The future of Biogeometry belongs to the students who go through this program getting educated in computer science and structural biology. Perhaps even more important than the dual education is the unique experience of applying one set of knowledge to the problems in another. We are proud to report a number of students across the institutions graduating with a PhD after being deeply immersed in this new field.

Hamish Carr completed his PhD at the University of British Columbia in April 2004, with a thesis entitled *Topological Manipulation of Isosurfaces* under the joint supervision of Jack Snoeyink and Michael van de Panne. He had started his PhD at UNC Chapel Hill when the Biogeometry project began with work on interpreting electron density maps by summarizing the topology of isosurfaces. His work on simplification and flexible exploration applies to volume data from many sources, and his thesis concentrates on medical imagery such as CT and MRI scans. He has taken a position on the Computer Science faculty of the University of Dublin.

Rachel Kolodny is completing her PhD dissertation, entitled *Comparing and Modeling Protein Structure* (expected September 2004). The first part of the dissertation focuses on protein structural alignment, namely, the comparison of two structures. She formalizes this problem as the optimization of a geometric similarity score over the space of rigid body transformations. This lead to an approximate polynomial time alignment algorithm. She also presents a large-scale comparison of six publicly available structural alignment heuristics and evaluates the quality of their solutions using several geometric measures. The second part of her thesis presents and uses an efficient model of protein structure. The model concatenates elements from libraries of commonly observed protein backbone fragments into structures that approximate protein well. By varying the size of the library and the length of its fragments, structure sets of different resolution are generated. She will be spending the next year at Howard Hughes Medical Institute and Columbia University, working as a postdoctoral fellow with Prof. Barry Honig, where she will work on protein-protein interactions and protein-DNA interactions.

Bala Krishnamoorthy just graduated from UNC with a PhD in Operations Research. Bala's thesis research was somewhat unusual since he worked concurrently with two advisors on two completely different topics. With Gabor Pataki from the Operations Research department he studied various basis reduction algorithms and their applications to integer programming; most of this work was done prior to his work under the Biogeometry project with co-PI Tropsha on the applications of computational geometry for protein structure analysis. The title of his thesis was *Preconditioning Integer Programs Using Column Basis Reduction* AND *Geometry and Topology of Protein Structure*.

Bala's initial contributions in the area of Biogeometry dealt with the improvement of the four-body statistical scoring function for protein fold recognition. In addition, Bala was interested in geometrical shape analyses of the data structure provided by the application of computational geometry to proteins. In a collaborative project with Charlie Carter and Jeff Roach, Bala analyzed the distributions of several shape parameters (tetrahedrality, volume, surface area, edge length, etc.) of frequently occurring subclasses of five Delaunay simplices calculated from united residue representations of protein structure. In a related project inspired by Herbert Edelsbrunner, Bala worked with Alex Tropsha as well as with Scott Provan (Department of Operations Research, UNC) on the analysis of protein geometry using its alpha complex. He has just started as an Assistant Professor in the Department of Mathematics, Washington State University, Pullman WA.

Nabil Mustafa completed his PhD dissertation this summer under the supervision of Pankaj Agarwal. His thesis, entitled *Simplification, Estimation and Classification of Geometric Shapes*, focuses on shape analysis problems that arise in molecular biology. The first part of his thesis demonstrates that the hardware available on graphics cards for PCs can be exploited to develop simple, fast algorithms for a wide range of geometric problems, including map simplification and statistical analysis of point sets. The second part focuses on shape matching and classification algorithms. He will join Kurt Mehlhorn's group as a postdoctoral fellow at MPI, Saarbrucken.

Vijay Natarajan completed his PhD dissertation, entitled *Topological Analysis of Scalar Functions for Scientific Data Visualization*, this summer under the supervision of Herbert Edelsbrunner. It deals with real-valued functions defined over meshes of 2- and 3-dimensional domains and studies algorithms for geometric and topological simplification, for extracting topological features and for measuring how similar the topological features of two functions over the same domain are. He joined UC Davis as a post-doctoral fellow where he will continue his work on Morse complexes.

Yusu Wang completed her dissertation, entitled *Geometric and Topological Methods in Protein Structure Analysis*, this year under the supervision of Pankaj Agarwal and Herbert Edelsbrunner. The first part of her thesis describes efficient algorithms for representing and characterizing shapes, including a subquadratic algorithm for computing the writhing number of a polygonal curve in 3-space. The second part of her thesis focuses on shape-matching algorithms. Using ideas from Morse theory she describes the notion of elevation for a molecular surface and uses the maxima of the elevation function to identify cavities and protrusions on molecular surfaces. These features are used to dock two proteins. She will be spending the next year at Stanford, working as a postdoctoral fellow with Leonidas Guibas. The following year she will join the Department of Computer Science of Ohio State University as an Assistant Professor.

The BioGeometry project is funded by the National Science Foundation under grant CCR-00-86013. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.