

Welcome! This is the first issue of BioGeometry News, a monthly newsletter intended to facilitate communication among project participants at Duke, UNC-CH, Stanford, and NC A&T Universities.

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

## People

**DAVID HSU** has moved up from his postdoctoral position at UNC to become the Sung Kah Kay Assistant Professor in the School of Computing at the National University of Singapore. David continues his interest in geometric computation with emphasis on applications in computational biology and robotics. His research focuses on representations

and algorithms for synthesizing and analyzing motion at both macroscopic (e.g., robots and animated characters) and microscopic (e.g., molecules) levels. He is finishing a paper on identifying hinges in molecules from several crystallographic structures, research begun with Jack Snoeyink and Charlie Carter. His web site is [www.comp.nus.edu.sg/~dyhsu](http://www.comp.nus.edu.sg/~dyhsu).

## Projects

### ALMOST-DELAUNAY TETRAHEDRA

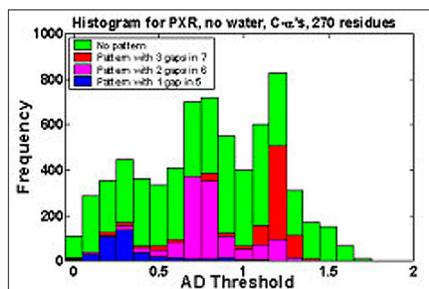
Deepak Bandyopadhyay, a PhD candidate in Computer Science at UNC, is exploring almost-Delaunay tetrahedra to demonstrate the stability of Delaunay analysis of proteins and to detect structural motifs. The idea is the following: For a given set of points, or sites, in 3D, the Delaunay tessellation captures which sites are neighbors, based on an "empty sphere property": if some circumsphere of a subset of 2 to 4 sites is empty of all other sites, then the line segment, triangle, or tetrahedron defined by that subset is included in the Delaunay tessellation. If the sites are allowed to move at most  $d \geq 0$ , then some additional segments, triangles, and tetrahedra may be able to satisfy the empty circle property. We say that a tetrahedron is almost Delaunay with threshold  $d$ , if  $d$  is the minimum movement.

Several researchers have applied Delaunay tessellation to proteins by choosing one point for each amino acid, and then analyzed the statistical frequencies of what amino acids appear in the Delaunay tetrahedra. Since atom coordinates are given to finite precision, and since there are choices of which point to select to represent an amino acid, Deepak and his colleagues wanted to see whether the additional tetrahedra that could arise from choosing different coordinates would show

the same statistics as the Delaunay - they seem to, although somewhat weaker.

For the C-alpha carbons of synthetic alpha helices, they noticed that the almost-Delaunay tetrahedra came in at certain threshold values that could be associated with certain patterns of vertices along the protein backbone. They can identify protein structures that have alpha helices based on whether peaks show up at these values in the threshold histogram, and also find individual helices and quantify helix content to high accuracy. They have also developed patterns for beta sheets and are working on alpha/beta structures, beta turns and more complex motifs.

For more information, please go to <http://biogeometry.cs.duke.edu/research/year3/almost>.



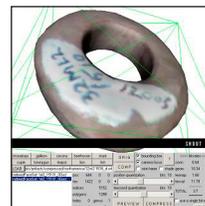
### SURFACE COMPRESSION

Martin Isenburg, a Computer Science PhD candidate at UNC, has been exploring many aspects of geometric surface compression [www.cs.unc.edu/~isenburg/research](http://www.cs.unc.edu/~isenburg/research). He has added compression to the commercial Shout3d Java applet, and is considering whether to use this to put some 3D content on the web. [biogeometry.duke.edu/software/skin/morph/](http://biogeometry.duke.edu/software/skin/morph/) shows a skin surface morph as a series of 10 GIF images, each 11K-14K. If you are running Internet Explorer with a current Java plug-in, you can look at [www.cs.unc.edu/~isenburg/pmc/morph\\_w\\_applet.html](http://www.cs.unc.edu/~isenburg/pmc/morph_w_applet.html) to see a series of 10 3D

models, each 7K-10K. (The original binary STL files were 400K each.) You can rotate with the mouse, pan with Shift-mouse, and zoom with Ctrl-mouse. This is a rough draft. It takes time to start because it must download the 100K Shout3d program, then download each model file. He hopes to change the mouse interface from the default to virtual trackball. Let him know what you think, or if you have models that you would like to put online.

Martin has also put up a benchmark coder for compressing polygonal meshes, [www.cs.unc.edu/~isenburg/pmc](http://www.cs.unc.edu/~isenburg/pmc), which offers an easily accessible online Web implementation and a downloadable

standalone version of his Polygon Mesh Compressor in pure Java. This software is meant to provide benchmark bit-rates for future research in the area of mesh compression. It compresses not



only the connectivity (conn) and geometry (geom) of a polygon mesh, but also one optional layer of texture coordinates.

This is accomplished by compressing the texture coordinate mapping (texmap) and the texture coordinate values (texval).

## Student Profile: Rachel Kolodny

---



Rachel Kolodny is a 3rd year PhD student at Stanford's Computer Science Department, working with PIs Guibas (Computer Science) and Levitt (Structural Biology). She came to Stanford from the Hebrew University in Jerusalem, after a two-year stay with Checkpoint Software (a leading Internet security company) where she played a key role in R&D.

The initial project that Rachel was involved in with the PIs aimed at defining a library of representative protein fragments, and then using this library to model protein structures. Rachel developed a novel clustering method for fragment shapes and used it to construct a family of shape libraries that differ in the fragment length (four to seven residues) and the number of representative fragments they contain. She then proceeded to encode a protein backbone by a sequence of library fragments, concatenated without any degrees of freedom. This gives a one-dimensional representation of native protein three-dimensional structures, whose quality depends on the nature and size of the library. Proteins of known structure can be approximated using a greedy construction method. Each library is characterized by the quality of fit (accuracy) and the number of allowed states per residue (complexity). Her goal was to find representations that are both accurate and economical (low complexity). The models defined using these libraries are substan-

tially better in this regard: with ten states per residue, it is possible to approximate native protein structures to  $1 \text{ \AA}$ , compared to over 20 states per residue needed in previously studied models. This work was presented in the 2001 ITR project meeting and has since appeared as a paper in the *Journal of Molecular Biology* [1].

Libraries of fragments can be used to explore protein-like structures. She implemented a protein decoy generation algorithm based on these libraries, where the idea is to sample geometrically valid structures biased by the secondary structure prediction for the protein. Unlike other methods, secondary structure prediction is the only protein-specific information used for generating the decoys. Nevertheless, these decoys are qualitatively similar to those found by others. She showed that the method works well for all-alpha proteins, and shows promising results for alpha and beta proteins.

While working on the fragments project Rachel repeatedly ran into the structural alignment problem. Structural alignment is the geometric analog of sequence alignment: given two chain structures (proteins), superimpose them so that they have a large common substructure. This is trickier than its sequence counterpart, because each of the structures is arbitrarily positioned and oriented in space. Structural alignment is a fundamental problem in biology as it can detect evolutionary relationships that are unseen when studying protein sequences. In its most general form it is considered to be very hard. Surprisingly, together with Nati Linial of the Hebrew University, Rachel observed that an approximate solution can be found in polynomial time. A fair amount of brute-force is still used in the solution, however, as they consider approximations to all rotations and translations, and

for each transformation compute the best alignment. In contrast, they showed that comparing two structures by considering only their internal distances does not allow a polynomial-time solution. This work was presented in the 2002 ITR project meeting.

In her current project, Rachel is collaborating with Ileana Streinu of Smith College to study the process of protein folding using a simplified model: namely a two-dimensional off-lattice HP model. The protein is modeled as a chain of atoms that are either H (hydrophobic) or P (polar or hydrophilic). She uses the construction of pseudo-triangulations, previously investigated by Streinu, to define a discrete set of continuous folding (or unfolding) motions. The folding process is simulated using Monte-Carlo, picking a random folding step and accepting or rejecting it based on an evaluation of an energy function. The novelty of this project is using an off-lattice model, i.e. the atoms of the chain are not constrained to lie only on grid points. Indeed, in a given step, many atoms move simultaneously but in ways that are provably non-interfering. The result is that much larger steps can be taken than is possible with a lattice model. The hope is that this approach will offer more rapid convergence to the folding state.

Besides her obvious productivity, what is most impressive about Rachel is how much she is at home in both Computer Science and Biology. Her background in theoretical computer science is an ideal match for the tough computational problems on protein structure prediction that form the focus of her research.

[1] R. Kolodny, P. Koehl, L. Guibas and M. Levitt. Small libraries of protein fragments model native protein structures accurately, *J. Mol. Biol.*, 323, 297-307 (2002).

- Profile by Leo Guibas