# BioGeometry NEWS

May 2005

biogeometry.cs.duke.edu


## People

### PI Awards

**Fred Books** of UNC has received the Eckert-Mauchly Award given jointly by the ACM and IEEE Computer Society for outstanding contributions to the field of computer and digital systems architecture.

**Herbert Edelsbrunner** of Duke University has been elected to the American Academy of Arts and Sciences for his work in computational geometry.

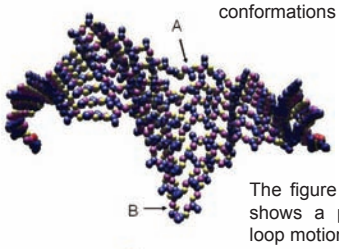
**Homme Hellinga** of Duke University has been named the James B. Duke Professor of Biochemistry.

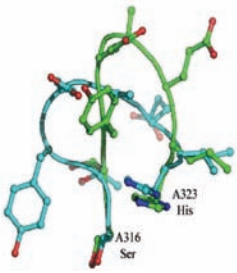
### New Students

Two new students – Nathan Marz and Ankur Dhanik – recently joined Jean-Claude Latombe's group to study the self-motion manifold of protein fragments (usually, loops) attached to the rest of a protein at two fixed "anchors" (amino-acids). This work is part of a newly funded NSF project with the Joint Center of Structural Genomics at the Stanford Linear Accelerator Center (Dr. Henry van den Bedem) and Prof. James Milgram in the Mathematics Department at Stanford.

**Nathan Marz** is a sophomore at Stanford University majoring in Computer Science. He had limited exposure to research during his freshman year and decided to devote a larger portion of his time to it this year. Nathan is developing software for the modeling, manipulation, visualization, and sampling of protein-loop conformation spaces. He is from San Diego, and on his free time enjoys basketball, music, and reading.

**Ankur Dhanik** is interested in unraveling the mysteries of protein structure determination and protein folding. His current goal is to study the conformation space of flexible protein loops, corresponding to blurred regions of electron density maps obtained with X-ray crystallography. He is a visiting researcher from the National University of Singapore, where he is pursuing a Master's degree in Mechanical Engineering. His thesis work at NUS is on modeling the dynamic behavior of a nylon thread for simulating suturing operations in a virtual-reality surgical training system. Simulating snap-through jumps requires handling energy bifurcations which forms part of his thesis. He has a Bachelor of Technology degree in Mechanical Engineering from Indian Institute of Technology (IIT), Kanpur, India.




Motion sequence between two conformations A and B

The figure (upper) shows a possible loop motion generated using the software created by Nathan and Ankur. The loop consists of residues A316 to A323 of TM0755 between two preferred states (shown green and blue at left) extracted using X-ray crystallography.

## Student Profile: Madhuwanti Vaidya

Madhuwanti Vaidya is in the second year of her Ph.D. program at Duke University. She is originally from Pune, India, and has obtained a Bachelor's degree in Computer Science from the University of Pune. She completed her final year project under the supervision of Narendra Karmarkar at the Tata Institute of Fundamental Research, devising a new method of data storage using an effective combination of hashing and sorting. While at TIFR, she used non-linear optimization methods to factor numbers and multivariate polynomials. She also extended the optimization package using multi-precision [3].

As a first year Ph.D. student at Duke, she applied semi-supervised and unsupervised learning methods to gene classification and multiple protein structure alignment problems. In her second year, she shifted her research focus towards understanding and analyzing protein structure. Her Second Year Project with Herbert Edelsbrunner started with the conjecture that the interatomic attractions and repulsions move each atom at or near the Voronoi vertices defined by its neighbors. This conjecture was easily refuted by testing it on a few high resolution protein structures. One of the reasons for the failure is that each atom is covalently bonded to some but very few atoms, and even if we ignored these neighbors, the fact that the next layer of neighbors are partially covalently bonded to each

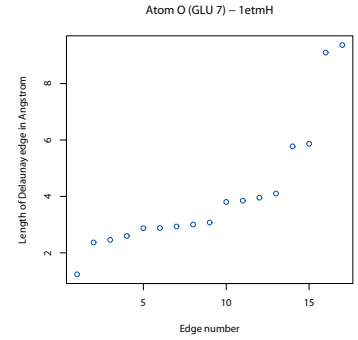



Figure 1. Delaunay edges incident to a single atom. Besides the one covalent bond we see four groups of edge lengths suggesting the neighbors occupy five shells around the central atom.

other biases the distances. She then decided to re-orient her project towards a general statistical study of basic metric and combinatorial properties of protein structures. Looking at the edges in the Delaunay triangulation, we observe that their lengths vary from 1.5Å to around 10Å. Most of the covalent bonds are approximately of length 1.5Å, while the covalent C-S bonds (in methionine and cysteine) are of length 1.8 Å. The hydrogen bonds are of length 2Å and all these are seen in the Delaunay triangulation. More generally, we observe that neighboring atoms tend to lie on shells of different radii, as illustrated in Figure 1, which is another reason why the earlier conjecture failed.

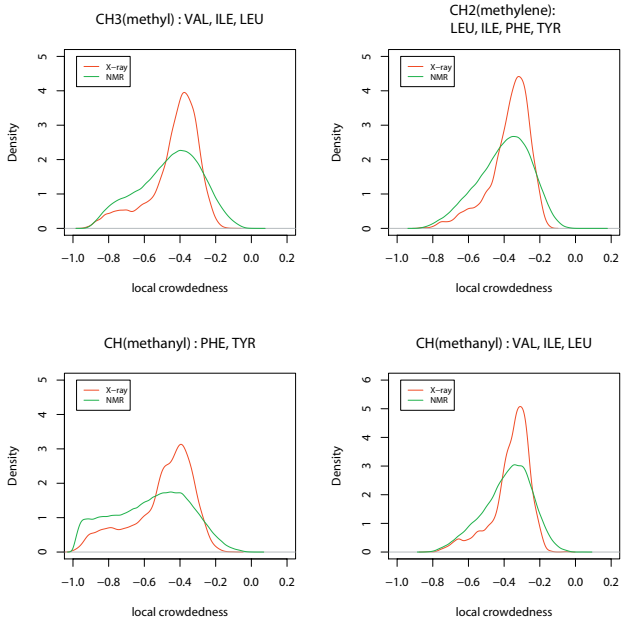



Figure 2. Local crowdedness of a set of 53 high resolution X-ray and 305 NMR protein models.

The more productive direction turned out to be the study of the local packing of atoms in a protein molecule. The Voronoi volume method [2] and the small-probe contact dot method [1] are two popular methods used to measure packing quality of proteins. In the *Voronoi method*, packing density is defined as the ratio of van der Waals volume to Voronoi volume. In an overpacked situation, the ratio usually goes up but it is also possible that the atom migrates out of its Voronoi cell and occupies a progressively smaller fraction. For this reason, the measurements are useful for detecting underpacked situations but ambiguous in over-packed arrangements of atoms. In the *small-probe contact dot method*, a dot is placed on the surface of the atom whenever the surface is within 2.5Å of another van der Waals sphere contributed by an atom that is not covalently bonded. The packing quality is estimated as a function of the gap distance between each dot and the neighboring atom's surface. It is clear that this measure is insensitive to very loose packing of atoms and therefore does not capture under-packed configurations. Madhuwanti's current research involves testing and using a new measure of packing quality that captures over-packing and under-packing and provides a seamless transition between the two opposite situations.

In this research, she follows closely the in-depth statistical study of Ban et al. [4], which is based on local density of an atom defined as the volume of the corresponding van der Waals sphere within its Voronoi cell divided by the volume of the Voronoi cell. The new measure, which we refer to as *local crowdedness*, is defined in terms of van der Waals spheres and orthospheres, as defined in [5]. For a Delaunay tetrahedron *t*, let *V* be the sum of volumes of the four van der Waals spheres that span *t* and let *W* be the volume of the orthosphere of *t*. *Local crowdedness*, denoted by $\delta_t$, is defined as:

$$\delta_t = \frac{V}{V + 2W} - 1.$$

Recall that the orthosphere intersects each one of the four van der Waals spheres in a right angle. If the four spheres contain the center of the orthosphere in their intersection then its square radius negative. We interpret such a ball to have negative volume so we get $\delta_t > 0$. The most negative value *W* can reach is -*V*/4. In the positive direction there is no limit for *W*, hence $-1 \le \delta_t \le 1$. The *local crowdedness* of an atom *a* is then the average of the $\delta_t$ over all tetrahedra that have the center of *a* as a vertex. Values close to 1 indicate highly over-packed (and physically impossible) arrangements, while values close to -1 indicate very loose and therefore under-packed arrangements.

Local crowdedness is calculated by extending the Ciel program, originally designed by Andrew Ban to construct protein-protein interface surfaces [6]. For direct comparison with [4] the high resolution X-ray data and the NMR data was provided by the first author. In Figure 2, we compare the local crowdedness averaged over the X-ray structures with that averaged over the NMR structures. Following [4], we have limited the measurements to hydrogen atoms in the CH, $CH_2$ and $CH_3$ groups of VAL, LEU, ILE, PHE and TYR residues. Similar to the results in [4], we see a dramatic difference between the distributions we get from the X-ray and the NMR structures.

References:


[1] J.M. Word, S.C. Lovell, T.H. LaBean, H.C. Taylor, M.E. Zalis, B.K. Presley, J.S. Richardson and D.C. Richardson. Visualizing and quantifying molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. *J. Mol. Biol.* **285** (1999), 1711-33.

[2] F.M. Richards. The interpretation of protein structures: total volume, group volume distributions and packing density. *J. Mol. Biol.* **82** (1974), 1-14.

[3] http://crd.lbl.gov/~dhbailey/mpdist/

[4] Y.-E. A. Ban, H. Edelsbrunner, J. Rudolph and P. Zhou. Local density: measuring the quality of NMR structures. Manuscript, 2005.

[5] H. Edelsbrunner. Deformable smooth surface design. *Discrete Comput. Geom.* **21**, (1999): 87-115.

[6] Y.-E. A. Ban, H. Edelsbrunner, and J. Rudolph. Interface surfaces for protein-protein complexes. Proc. 8th Ann. Int'l Conf. Res. Comput. Mol. Biol., 2004, 205-212.


*- Profile by Herbert Edelsbrunner*


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.