BioGeometry NEWS June 2005
biogeometry.cs.duke.edu

Software Almost-Delaunay Tetrahedra by Deepak Bandyopadhyay

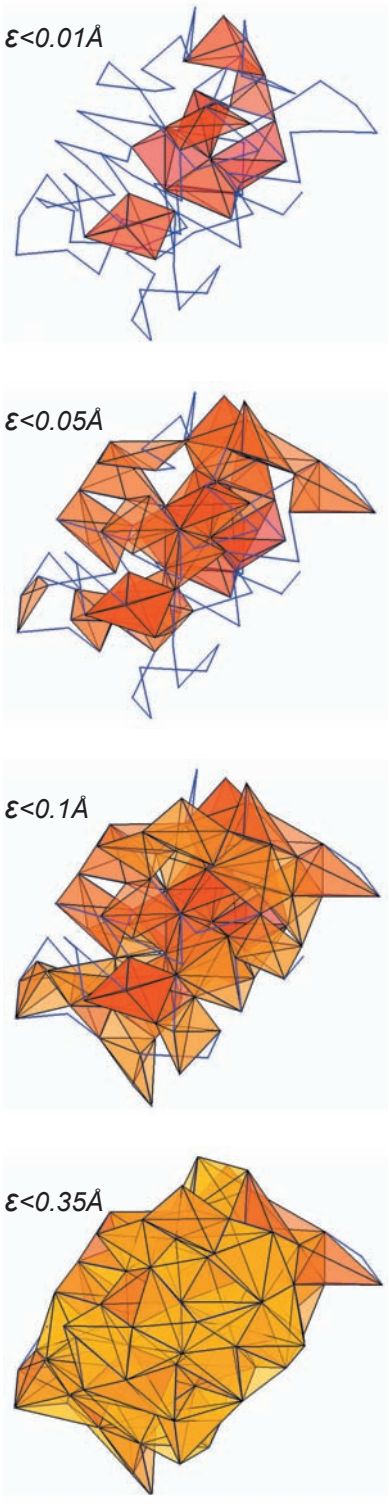

Figure 1. Visualization of Almost-Delaunay Tetrahedra with threshold ε in the range 0 (red) < ε ≤ 0.35 (yellow) in acylphosphatase (PDB 2acy). Delaunay tetrahedra (ε=0) are not shown for clarity.

To analyze a protein structure, one must often determine the nearest neighbors of each residue or atom within it. The Delaunay tessellation (DT) is a mathematically precise representation of the nearest neighbors, and many tools for protein structure analysis are based on the DT, including SNAPP, pvSOAR, ProShape and Ciel developed in the labs of Biogeometry PIs.

Because the data is often imprecise, we developed the Almost-Delaunay simplices [1], which capture potential changes in the DT under perturbation.

AlmDel is a software package for computing almost-Delaunay edges, triangles and tetrahedra from 3D point sets, for a given bound on perturbation of each point and maximum distance between neighboring points. Each almost-Delaunay edge, triangle or tetrahedron is output with an associated *threshold* signifying the minimum perturbation needed to make it part of the DT.

The AlmDel software is written in C++ and implements the algorithm of [1] using CGAL, the Computational Geometry Algorithms Library [2]. It has been released as source code and binaries for Windows and Linux. Almost-Delaunay tetrahedra may be computed directly from PDB files, visualized as MATLAB graphs (Figure 1), and exported to Kinemage format (Figures 2-3) using the ADMATLAB package.

Applications: We have explored several applications of the almost-Delaunay tetrahedra in analysis of protein structure:

1. Confirming the robustness of SNAPP (Simplicial Neighborhood Analysis of Protein Packing), a four-body statistical potential for protein packing derived from the DT.
2. Distinguishing well-packed proteins from decoys by the number of almost-Delaunay tetrahedra.
3. Assigning secondary structure (particularly alpha-helices) from packing signatures (Figure 2).
4. Detecting hinges from differences in thresholds of almost-Delaunay tetrahedra between snapshots of a conformational change (Figure 3).
5. Providing a sparse and robust graph representation for subgraph mining to identify fingerprints, local spatial motifs that characterize protein families and are useful for family classification and functional annotation.

We hope the availability of the AlmDel package will lead the computational biology/chemistry community to evaluate almost-Delaunay tetrahedra for their favorite nearest neighbor analyses, and thus come up with new applications.

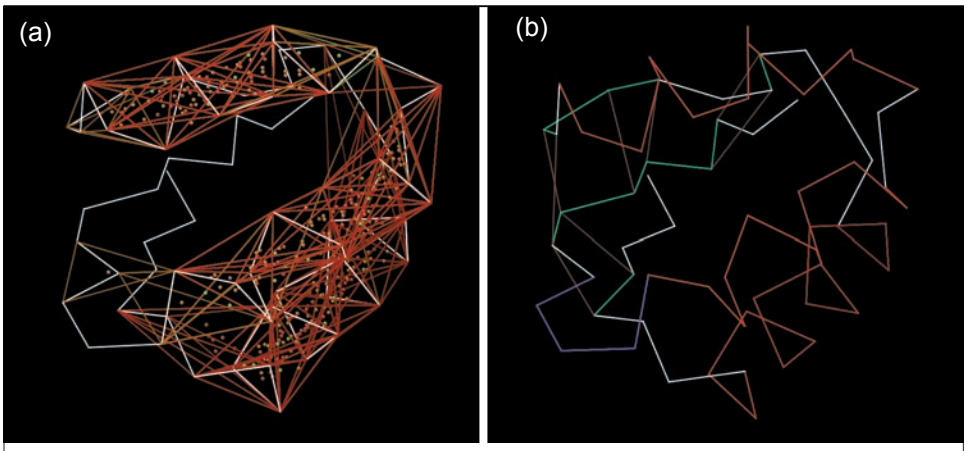

Figure 2. (a) AD tetrahedra in alpha-helices have characteristic ranges of thresholds that correspond to patterns of gaps in sequence numbers. They are shown here for a small ribosomal protein (PDB 1ctf) as tetrahedra and also as spheres at their centers, in green (threshold 0.25-0.35 Å), orange (0.6-0.8 Å) and red (1.05-1.25 Å). Such patterns were used to assign secondary structure, shown in (b) for 1ctf. Color code — red: alpha-helix, green: beta-sheet, brown: neighbors on adjacent beta strands, violet: beta turn, white: coil.

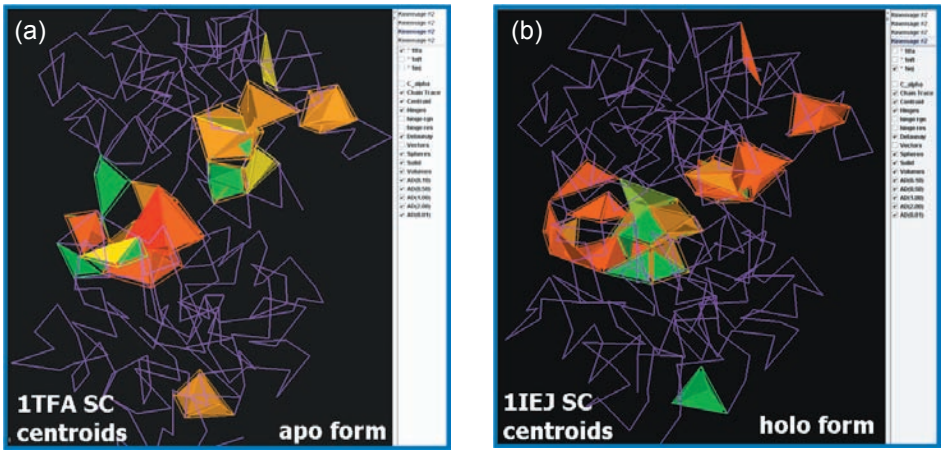

Figure 3. AD tetrahedra that differ significantly in threshold among multiple conformations of a protein help find hinges of the conformational change. The (a) open and (b) closed forms of ovotransferrin are shown. Color code — green: Delaunay, yellow: almost-Delaunay with threshold 0.01-0.1 Å, ochre: 0.1-0.5 Å, orange 0.5-1.0 Å, red 1.0-2.0 Å.

References

[1] D. Bandyopadhyay and J. Snoeyink. Almost Delaunay simplices: nearest neighbor relations for imprecise points. In Proc. 15th Ann. ACM-SIAM Sympos. Discrete Algorithms (SODA), 2004, 403-412.

[2] D. Bandyopadhyay and J. Snoeyink. Almost Delaunay simplices: robust neighbor relations for imprecise 3D points in CGAL. 2nd CGAL Workshop (2004).

[3] J. Huan, W. Wang, D. Bandyopadhyay, J. Snoeyink, J. Prins and A. Tropsha. Mining protein family-specific residue packing patterns from protein structure graphs. In Proc 8th Ann. Int. Conf. Res. Comp. Mol. Biol. (RECOMB), 2004, 308-315.

Almost-Delaunay Tetrahedra (AlmDel) software is available at http://biogeometry.cs.duke.edu/software/almdel/

Research News

Members of the BioGeometry group will continue work begun on this project with funding from various sources.

● Jean-Claude Latombe (Stanford CS) was awarded a new NSF grant entitled "Applications of Probability Measures on the Self-Motion Manifold of Deformable Fragments in Proteins" within the Joint DMS-NIGMS Initiative to Support Research in the Area of Mathematical Biology led by Program Director Dr. Junping Wang. This new project will be a joint effort 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. It will develop mathematical and computational models for a redundant, closed, protein-like kinematic chain, and study the topology and geometry of its conformational manifold (the self-motion manifold). The goal will be to calculate a probability measure on the self-motion manifold from electron density maps obtained by X-ray crystallography, in order to enable crystallographers to retrieve and study dynamic properties of protein loops. In the multimodal disordered case, the objective will be to identify all sub-states, along with their probabilities, and to reconstruct energetically plausible conformational pathways. This research is in the continuation of the recent work done by Itay Lotan in the BioGeometry project.

● Jack Snoeyink (UNC CS) and PhD candidate Xueyi Wang (UNC CS) will continue applying robotics techniques in molecular structure through a recently NIH-funded project with David and Jane Richardson of Duke Biochemistry: "Inverse Kinematics, Sterics, and Data to Fit RNA Backbone"

The goal is to produce tools to check and improve the structures of RNA that are determined by crystallography. Structures that are fitted with combined atom models (in which hydrogens just enlarge the radii of heavy atoms) show unrealizable clashes when the hydrogen atom geometry is made explicit.

RNA has more conformational freedom than protein, which actually makes it harder for people to consider the alternatives for structure threading or improvement. By combining observations from known structures, geometric analysis, and search, this project will help crystallographers find more accurate representations of RNA structures.

● Homme Hellinga (Duke Biochem) and Jack Snoeyink (UNC Biochem) are involved in a protein design project led by David Baker (U of Washington Biochem) that has been approved for funding by the DARPA directorate.

The goal of this effort is to create a set of design and synthesis processes that will enable the specification of a desired function, and be able to rapidly synthesize a protein that performs the function. To achieve this goal, significant advances must be made in the understanding of several problems including the relationship of sequence to physical structure to biological function and the definition of reusable components of proteins.

DARPA aims to begin funding in August 2005, and has scheduled a kickoff meeting in early September 2005.

Announcement

BioGeometry Meeting

Please mark your calendars and plan to attend the next BioGeometry meeting which is scheduled for August 22-23 at North Carolina A&T State University in Greensboro. Organizers are Solomon Bililign and Robert Gdanitz. The nearest airport is Piedmont Triad International Airport (GSO) in Greensboro. See details at http://biogeometry.cs.duke.edu/meetings/ITR/05aug/

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.