BioGeometry NEWS
July 2003
biogeometry.cs.duke.edu

Research

Four-Body Statistical Potentials

Bala Krishnamoorthy, a graduate student in Operations Research at UNC has been working with Alex Tropsha, a faculty from the School of Pharmacy at UNC, on four-body (4B) statistical pseudo-potentials for scoring protein conformations.

Motivation: Most scoring functions used in protein fold recognition employ two-body (pseudo) potential energies. The use of higher-order terms may improve the performance of current algorithms.

Methods: Proteins are represented by points in 3D space located at the residue sidechain centers. The Delaunay tessellation of this representation provides non-overlapping aggregates of four-body nearest neighbor clusters. The frequencies of occurrence of these residue clusters in a training data set of proteins with known structures are used to formulate the four-body statistical potentials. In addition, the chain separation of the residues in a Delaunay tetrahedron is an important factor in this formulation. Delaunay tetrahedra occurring in protein structures are classified into five classes (Fig 1). Finally the total contributions from each class of tetrahedra are weighted by numbers such that the deviations from the average fractions as observed in the training set are penalized. Please refer to [1] for the detailed formulation of the 4B potentials.

Results and Discussion: The 4B potentials were applied to different sets of native and corresponding decoy structures. These included random sequence decoys generated by permuting the primary sequence while keeping the structure of the native protein unchanged and single and multiple decoys sets from the Decoys `R Us database. The 4B potentials were able to discriminate the native structure from the corresponding non-native conformations with a high degree of accuracy in most of the cases.

A very interesting result was obtained when the 4B potentials were applied to the conformational samples from the molecular dynamics (MD) simulations of protein folding. The folding trajectory indicated the presence of a single transition state. In the original simulations both pre- and post-transition state conformations had comparable total energies and could not be discriminated from each other by the native structure specific Gō potentials used in the MD simulations. Yet, starting from a pre-transition structure resulted in an unfolded final conformation while starting from a post-transition structure almost always gave a folded conformation. The 4B potentials could clearly distinguish pre-transition and post-transition conformations and the native structure from the MD folding simulation trajectories of CI2 and SH3.

People

Alex Tropsha, Associate Professor in the UNC School of Pharmacy, division of Medicinal Chemistry, has accepted our invitation to join the project as a co-PI, effective August 15, 2003. Alex is a chemist who studies the application of geometric structures such as the Delaunay tessellation to protein structure classification and scoring. He and his group have been actively working with several members of the Biogeometry project since its inception, and we are glad for the opportunity to make this interaction more formal.

More importantly, analyzing the residue-wise log-likelihood profiles of these conformations revealed that certain residues have markedly higher scores in the post-transition conformations and the native structure as compared to the pre-transition conformations. The residues form certain core tetrahedra (see Fig 2). In both the cases of CI2 and SH3, the core tetrahedra forming residues included those residues that were identified as nucleus-forming from the original MD simulation studies.

Reference:

[1] B. Krishnamoorthy and A. Tropsha. Development of a four-body statistical pseudo-potential to discriminate native from non-native protein conformations, *Bioinformatics*, August 2003, In Press. (Authors' version is available at http://www.unc.edu/~kbala/DT/)

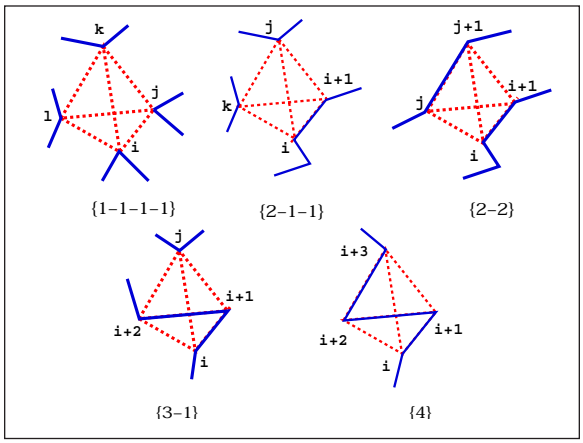


Figure 1: Five classes of Delaunay tetrahedra based on the primary sequence separation of the four residues - from Class {1,1,1,1}, with all the four residues non-consecutive in the primary sequence to class {4} in which all four residues in the tetrahedron are consecutive in the protein primary sequence.

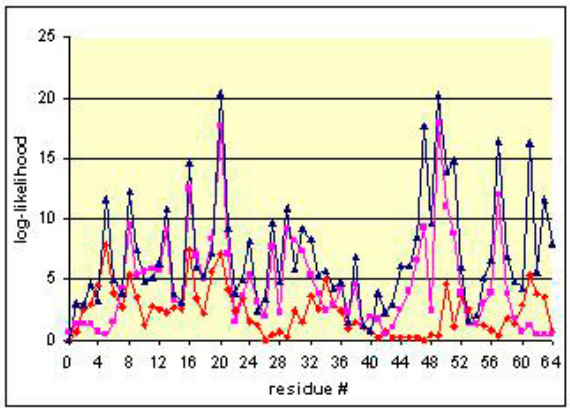


Figure 2: Structure profile of a typical pre-transition (red), post-transition (violet) and native structure (blue) of CI2. The residues 8,13,16,20,29,47,49 and 57 are observed to have markedly high profile scores in the post-transition structure and in the native structure as compared to the pre-transition structure. These residues form the *core tetrahedra*. Residues 16,49 and 57 were identified as *nucleus-forming* from original MD simulation studies.

Student Profile: Andrea Mantler

Andrea Mantler works under Prof. Jack Snoeyink at the University of North Carolina at Chapel Hill (UNC). Fresh out of her bachelor of Computer Science program at the University of Manitoba, Andrea knew that she wanted to study computational geometry, and decided to go to the University of British Columbia to work with Dr. Snoeyink. She was surprised when half a year later Dr. Snoeyink announced that he was changing from UBC to UNC. Andrea was one of three students who followed him.

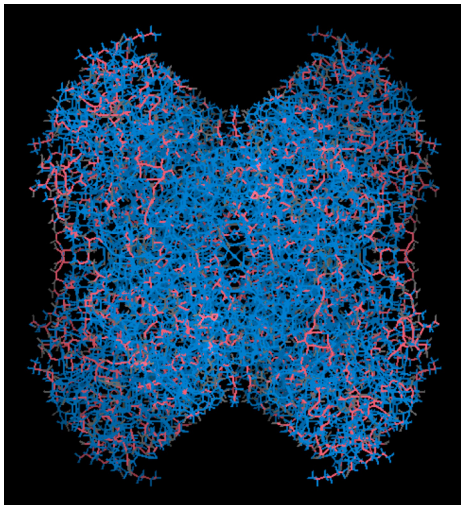
Although the transition period was challenging, Andrea finished off her UBC master's degree while starting the PhD program at UNC. For her master's thesis, she designed an optimal algorithm for the problem of intersecting line segments in the plane using the smallest possible arithmetic precision. This is one of several projects she has done that were motivated by Geographic Information Systems (GIS).

Andrea has also worked on folding and unfolding polyhedra, which led her into problems of folding and conformational analysis in protein molecules. In fall 2002, she assisted Dr. Snoeyink in leading a first-year seminar course, "Folding: from Paper to Proteins." In January 2003, Andrea participated in the Barbados workshop on the geometry of protein folding that was organized by Prof. Illeana Streinu. At the workshop, she became interested in combinatorial rigidity and its use in analyzing and representing the flexibility of structures such as proteins.

Combinatorial rigidity is the study of first-order rigidity in a generic framework --- for example, in a graph with fixed length bars for edges and universal joints at vertices, identify the sets of vertices that must move together. Although combinatorial rigidity is well understood in two dimensions, the extension to three dimensions is incomplete. Laman's condition in 2D completely determines whether or not a framework is rigid; a similar condition in 3D is conjectured, but not yet proven, to work for the special case of graphs of bonds in molecular models (the Molecular Conjecture).

Andrea recently implemented a Java version of a 3D pebble game by Jacobs et al. that takes a protein in kinemage format, and outputs overconstrained, rigid, and flexible bonds, also in kinemage format. Andrea started this project in the Richardson's biochemistry class at Duke, as a way of exploring the Molecular Conjecture, and to examine the allosteric effect in proteins.

Allostery is the process in which ligand binding at one site on a protein affects the ligand binding at another site via motions transmitted through the protein structure. One possible explanation of the allosteric effect is that adding and



removing bonds affects the flexibility of the protein. Prof. Walter Whiteley and the others at the Barbados workshop constructed combinatorial frameworks whose first-order rigidity shows allosteric behavior.

Duke professor Homme Hellinga suggested that phosphofructokinase (PFK) was the classic model for an allosteric mechanism in proteins, so Andrea has been applying her analysis to PFK. The figure shows initial output, in which bonds are drawn using three colors to indicate flexible, rigid, and overconstrained regions.

Andrea is currently developing better ways to visualize the results of the pebble game analysis --- to make the rigid regions easily visible and distinguishable, and to show how different rigid regions can move with respect to each other. Andrea is investigating many other questions about the limitations of first-order analysis and how best to apply it: which bonds are strong enough to consider in analyzing flexibility, how to allow limited motion, how to impose one-sided constraints such as steric (collision) constraints. And, of course, the Molecular Conjecture.

- Profile by Jack Snoeyink

BioGeometry News is the monthly newsletter of the BioGeometry project. For more information, please visit http://biogeometry.cs.duke.edu/
The project is funded by the National Science Foundation under grant CCR-00-86013. Any opinions, findings, and conclusions or recommendations expressed in this newsletter are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.