BioGeometry NEWS
September 2004
biogeometry.cs.duke.edu

Research

Symmetry in Protein Interfaces

by Tammy Bailey and Herbert Edelsbrunner

Tammy Bailey is a PhD student in the Department of Computer Science at Duke University working with Herbert Edelsbrunner. Tammy earned BA and MS degrees in Mathematics from Syracuse University and an MS in Computer and Information Science from the State University of New York Institute of Technology. Tammy became involved with the Biogeometry project two years ago, initially working on geometric alignment algorithms for protein structures. At present, her research focuses on the characterization and classification of protein-protein interfaces. The interaction between proteins forms the basis for most cellular processes, yet there is little understanding of how one protein recognizes and binds to another. In the field of protein structure, the characterization of structural motifs made possible the classification of proteins into families, which in turn has provided valuable insights into their function, mechanisms of folding, and evolutionary origins. Adopting a similar approach toward the understanding of protein interactions, Tammy, in conjunction with Jeff Phillips, another student in the computer science department at Duke, is investigating several methods of characterizing interfaces at the structural level. As the majority of multimeric proteins found in living cells are symmetrical complexes, she initiated research into the symmetry of protein-protein interfaces. Symmetry is an intrinsic property of a geometric object that renders it invariant under a specified group of transformations. A symmetry element of the object is a transformation which, when applied to the object, leaves the object invariant. The collection of all such transformations is the symmetry group of the object. All protein structures may be characterized and categorized by their symmetry.

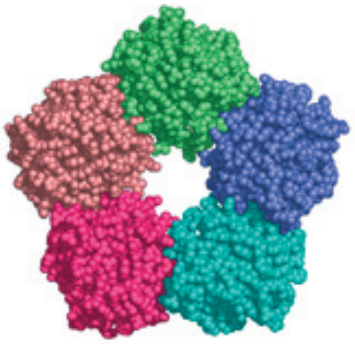

Figure 1: Protein-protein interactions include events intimately linked to human disease such as cell division and growth. The central feature of diseases such as Alzheimer's is the deposition of amyloid plaques that kill surrounding cells. Amyloid fibers are stabilized and bound by serum amyloid P component, a naturally occurring protein present in the bloodstream.

Symmetry is abundant in protein structures. The evolution of multimeric proteins to adopt symmetrical configurations often relates to their stability and function. Protein interaction sites are highly specific and directional, favoring the formation of symmetric complexes. As symmetry in protein structure provides insights into protein function and misfunction, it is natural to ask if the interface of interacting proteins also form symmetric structures. Utilizing the interface surface, a geometric representation of the region of interaction in protein complexes, we found that symmetry is evident at the interface level as well.

Figure 2: The interface surface between neighboring side chains of serum amyloid P component.

Symmetry is a binary property; an object either has a particular symmetry or it does not. Ideally, we would like to classify protein interfaces by their symmetry, but the construction of interface surfaces makes it highly unlikely that a structure with true symmetry will result. However, if we relax symmetry to a continuous property, then we can evaluate the symmetry of a particular object with respect to any symmetry element or group.

Consider for example a collection of points U in R^3. U has rotational symmetry of order *n* if it is invariant under a rotation of angle $2\pi/n$ about an axis through its center of mass. Let $\rho(U)$ be a rotation of U and pair each point in U to the closest point in $\rho(U)$. The root-mean-square (rms) distance between corresponding points in U and $\rho(U)$ is zero if ρ represents a rotational symmetry of U. If ρ is instead an approximate symmetry of U, then the rms distance between U and $\rho(U)$ provides a geometric measure of that approximation. We are currently working toward a rigorous definition for approximate symmetry and approximate symmetry groups.

We can use the rms distance to visualize the approximate symmetries of the interface surface over a collection of unit spheres. Recall that a rotation can be represented by a unit vector and an angle of revolution about that vector. Let each individual sphere correspond to a fixed

Figure 4: The interface surface for vipoxin.

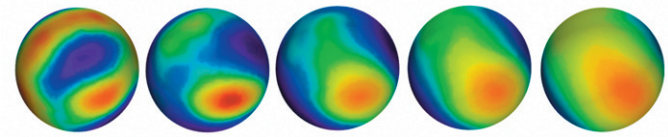

Figure 3: Approximate symmetry measures for the interface of serum amyloid P component side chains are displayed in pseudocolor over a set of spheres. From left to right, the spheres correspond to the angles $\pi/2$, $\pi/3$, $\pi/4$, $\pi/5$, and $\pi/6$. An approximate axis of symmetry is suggested by the red spot on the second sphere, which visualizes the rms distance for a rotation of $\pi/3$.

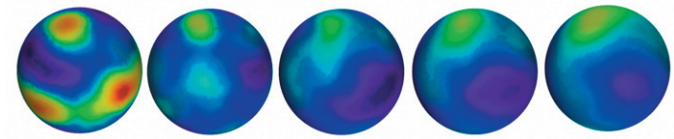

Figure 5: Approximate symmetry measures for the interface of vipoxin. Three perpendicular approximate axes of symmetry are suggested by the coloration of the first sphere, which visualizes the rms distance for a rotation of $\pi/2$.

rotation angle; any point on that sphere then corresponds to a unique axis of rotation. Figure 3 illustrates the approximate symmetries of the interface (Figure 2) between neighboring chains of serum amyloid P component. Root-mean-square distances are represented in pseudo-color with the smallest values colored by red.

The interface surface of vipoxin, a neurotoxic complex between a toxic phospholipase A2 and non-toxic protein inhibitor, is shown in Figure 4. The approximate symmetries are illustrated in Figure 5.

References

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

[2] T.L. Blundell and N. Srinivasan. Symmetry, stability and dynamics of multidomain and multicomponent protein systems. Proc. Natl. Acad. Sci. 93 (1996), 14243-14248.

[3] D.S. Goodsell and A.J. Olson. Structural symmetry and protein function. Ann. Rev. Biophys. Biomol. Struct. 29 (2000), 105-153.

BioGeometry Meetings at PSB '05

Session The upcoming Pacific Symposium on Biocomputing 2005, to be held on January 4-8, 2005 on Big Island of Hawaii, will feature the session on BioGeometry: Applications of Computational Geometry to Molecular Structure. Organized and chaired by Alex Tropsha and Herbert Edelsbrunner, this session will be the first instance of including a specialized Biogeometry session in a major computational biology conference. Biogeometry is becoming recognized by the computational biology community as an emerging scientific discipline at the interface between computational geometry, biochemistry and biophysics, statistics, and chemistry that brings together specialists in the above disciplines to develop new computational techniques and paradigms for representing, storing, searching, simulating, analyzing, and visualizing biological structures.

Only 35 papers were selected by the organizers for the presentation in Hawaii out of more than 130 submissions, and five of them will be included in the Biogeometry session. Three of these papers will be presented by the coinvestigators on the Biogeometry grant. Wang, Agarwal, Brown, Edelsbrunner, and Rudolph (Duke) will give a paper entitled Fast Geometric Algorithm for Rigid Protein Docking. A paper by Russel and Guibas (Stanford) is titled Exploring Protein Folding Trajectories Using Geometric Spanners, and Leaver-Fay, Kuhlman, and Snoeyink (UNC) will talk on An Adaptive Dynamic Programming Algorithm for the Side Chain Placement Problem. Two additional presentations are given by Li and Liang (University of Illinois at Chicago) on Computational Design of Combinatorial Peptide Library for Modulating Protein-Protein Interactions and by Karklin (Carnegie Melon) and Meraz and Holbrook (Lawrence Berkeley National Laboratory) on Classification of Non-coding RNA using Graph Representations of Secondary Structure.

PI Meeting The PSB meeting provides a natural opportunity for us to get together in Hawaii, which is why the next PI meeting has been scheduled for Saturday, January 8 immediately following the Biogeometry session. Although the oral presentations at the PSB'05 meeting have been finalized, opportunities still exist for poster submissions. All poster abstracts should be submitted online at http://psb.stanford.edu/abstractopen2005.html (choose "poster only" for Abstract Type).

We hope to see you in Hawaii in January. Aloha!

Ph.D. Graduates

Serkan Apaydin completed his PhD dissertation this summer under the joint supervision of Jean-Claude Latombe and Doug Brutlag. His thesis, entitled Stochastic Roadmap Simulation: An Efficient Representation and Algorithm for Analyzing Molecular Motion, describes a new computational framework – Stochastic Roadmap Simulation – for exploring the kinetics of molecular motion by simultaneously examining many pathways. These pathways are compactly encoded in a graph, which is constructed by sampling a molecular conformation space at random. Each arc in the graph represents a potential transition of the molecule and is associated with a probability indicating the likelihood of this transition. By viewing the graph as a Markov chain, ensemble kinetic properties can be efficiently computed. SRS was applied to the computation of the Pfold parameter in protein folding and to study ligand-protein binding. He will soon join Professor Bruce Donald's group at Dartmouth College as a postdoctoral researcher.

Itay Lotan completed his PhD dissertation this summer under the supervision of Jean-Claude Latombe. His thesis, entitled Algorithms Exploiting the Chain Structure of Proteins, studies three important problems in computational structural biology. The first part of the thesis describes an automatic method (developed in collaboration with the Joint Center for Structural Genomics at SSRL) for completing partial models of protein structures resolved using X-ray crystallography. The second part deals with the computation of structural similarity and proposes an approximation technique that speeds up this computation at the expense of introducing a small error in the similarity measure. The third part introduces a new data structure and algorithm – the ChainTree – to efficiently identify pairs of interacting atoms in a protein and update energy values during Monte Carlo simulation. Itay has recently joined the group of Dr. Head-Gordon as a postdoctoral researcher at U.C. Berkeley.

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.