BioGeometry NEWS
October 2003
biogeometry.cs.duke.edu

Research

Backbone Motion by Inverse Kinematics for Protein Design

by Andrew Leaver-Fay and Jack Snoeyink

Motivation: The problem of protein design is to construct an amino acid sequence that folds to a desired structure. PI Homme Hellinga (Duke Biochem) has had significant success in designing receptor proteins by starting with a given protein, fixing the backbone as a *scaffold*, and using Dead End Elimination (DEE) to search for changes in the sequence that will create desired structure in the binding pocket. These include changing ribose binding protein to recognize and bind to TNT [1].

UNC computer scientists (PI Snoeyink, Leaver-Fay, O'Brien, Noonan) have been working with Hellinga and David & Jane Richardson (Duke Biochem) to support design with a movable protein backbone. We have developed tools that can be applied generally, to expand the space of rotamers that fit at a given position and allow DEE to find a sequence of amino acids that better accommodates our molecule – filling unoccupied space or allowing stabilizing hydrogen bonds. They can also be applied for specific motions, such as resolving structural defects in the TNT binding protein.

Motion By Loop Closure: We apply techniques from robotics for exact analysis of motions of small fragments of protein backbone. For a robot arm with 6 torsional degrees of freedom, the inverse kinematics (IK) problem is to find the angles that will position the end of the arm in a desired position and orientations. Raghavan and Roth reduced this to finding the real roots of a degree 16 polynomial. Manocha [2] turned this into a robust algorithm by casting it as an eigenvalue problem. We apply this to fragments of protein backbones that contain six dihedral angles, which we treat as free variables, while keeping bond angle, bond lengths, and peptide dihedral angles fixed. The IK solver returns all real valued solutions (at most 16) that preserve the chain endpoints and orientations. This includes the original solution, plus alternate solutions, which usually must be discarded because of collisions with the surrounding structure. To generate motions near the original solution, we vary each of the other parameters, one by one, before calling the IK solver.

Motion Selection: With the large number of parameters we are able to manipulate, the number of solutions that are computed can be overwhelming. There are several ways we can handle this. First, since each parameter has an ideal value observed in proteins, we prefer motions within the standard deviations for ideal bond geometries that move the backbone atoms conservatively (< 0.2 Å). Even such bond angle changes can produce sizeable motions (~0.4 Å) of the off-backbone Cβ carbon or carbonyl oxygens. Second, we compute atomic motion derivatives with respect to the changing parameter and display derivative vectors in Mage. With a specific motion in mind, it is possible to look at several sets of vectors and quickly determine if the indicated motions are worthwhile. Third, we can leave the computer to sample over parameters, and filter the resulting configurations by collisions or energy criteria.

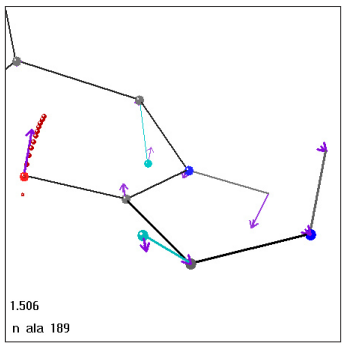

Figure 1. Motion derivative vectors for a solution branch around residue 189 in ribose binding protein. The vectors point in the direction of atom motion as the Cα bond angle is goes towards its ideal value.

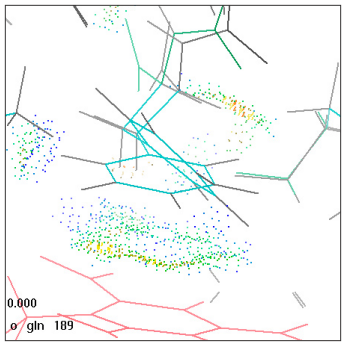

Figure 2. Phenylalanine 190 of the R3 design (green backbone) does not participate in π stacking with TNT (pink) in the original structure. Moving the gray backbone removes the 189 carbonyl oxygen, giving the phenylalanine ring room to stack nicely. Hellinga's student, Yaji Tian, is currently running experiment with this and other adjusted backbones.

Announcement

BioGeometry Meeting

The second BioGeometry meeting of 2003 will take place at Duke University on Monday-Tuesday, November 17-18. PIs and students from the four participating institutions will attend, as will a representative from the NSF.

More details are available at http://biogeometry.cs.duke.edu/meetings/ITR/03_nov.

Extensions: Backbone modification also has a role in x-ray crystallography. Model creation from electron density maps is a tricky process. A poor fitting of the backbone can lead to poor, non-rotameric choices for side chains to fit the density. A tool for backbone motion can support better choices in model creation and refinement.

References

[1] Looger, L. L., Dwyer, M. A., Smith, J. J. & Hellinga, H. W. Computational design of receptor and sensor proteins with novel functions. Nature 423, 185-190, 2003.

[2] Raghavan, M and Roth, B. Inverse kinematics of the general 6r manipulator and related linkages. Transactions of ASME, Journal of Mechanical Design, 1992.

[3] Manocha, D. and Canny, J.F. Efficient inverse kinematics for general 6R manipulators. IEEE Journal on Robotics and Automation, pages 648-657, 1994.

Student Profile: Bala Krishnamoorthy

Bala Krishnamoorthy, a fifth-year PhD student in the Department of Operations Research at UNC-Chapel Hill, is working with coPI Tropsha (who recently formally joined the project) on the applications of computational geometry for protein structure analysis. After obtaining his B.Tech. in Mechanical Engineering from the Indian Institute of Technology (where he was placed among 0.2% best students selected from 100,000 applicants), Bala entered UNC as a graduate student in 1999. Initially, he concentrated on theoretical aspects of operations research studying various basis reduction algorithms and their applications to integer programming. However, over the past two years, Bala's interests expanded towards computational geometry of proteins.

Bala's initial contributions dealt with the improvement of the four body statistical scoring function for protein fold recognition that was developed earlier in the Tropsha laboratory. Delaunay tessellation (DT) as applied to simplified, united residue representation of the protein structure produces an aggregate of space-filling, irregular tetrahedra with amino acid residues at their vertices. The original scoring function was established as a log likelihood (i.e., the ratio of observed vs. predicted frequencies) of finding any specific set of four amino acid residues forming a four-body simplex of nearest neighbors. Based on the topology of vertex-residue distribution in the primary sequence, five different classes of quadruplets have been established, ranging from a class of four consecutive residues to a class of four disconnected residues (i.e., no two residues occur next to each other in the primary sequence). Log likelihood scores were estimated for each class independently; however, in the original formulation we have used identical values of expected frequency of occurrence for all five topological classes. Bala quickly realized that each expected probability should be corrected taking into account the

relative sizes of the five topological classes in experimental protein structures. This simple yet important correction afforded a more accurate fold recognition scoring function as demonstrated by a greater separation between native and structural decoys structures (the latter were generated in Michael Levitt's group at Stanford). Another interesting feature of the modified scoring function developed by Bala was its ability to discriminate between pre-transition, post-transition, and native structures of the CI2 protein, generated in the course of folding simulations in Eugene Shakhnovich's group at Harvard. Interestingly, the Go potential used in folding simulations could not discriminate between pre- and post-transitional structures emphasizing the unique power of the DT based potential for accurate fold recognition. The results of this study were published recently [1] and were profiled in the July 2003 issue of Biogeometry News.

As a follow up to this study, Bala has been involved in a collaborative project with Prof. Dani Fischer at Ben-Gurion University, Beer Sheva, Israel, on the analysis of predicted vs. actual protein structures considered in recent CASP4 and CASP5 (Critical Assessment of Structure Prediction) competitions. The preliminary results demonstrate that Bala's scoring function on average affords the most accurate selection of native-like protein conformations among alternative predictions as compared to scoring functions or pseudo-potentials used by other groups.

Beyond the further development of the DT-based scoring functions for fold recognition, Bala is interested in geometrical shape analyses of the data structure provided by the application of computational geometry to proteins. In a collaborative project with Charlie Carter and Jeff Roach (UNC Biochemistry & Biophysics) Bala is analyzing the distributions of several shape parameters (tetrahedrality, volume, surface area, edge length, etc.) of frequently occurring subclasses of the five topological classes of Delaunay simplices discussed above. The ultimate goal of this project is to use these frequent motifs for the initial mapping of amino acid residue quadruplets onto electron density provided by x-ray diffraction patterns of protein structures. Finally, in yet another project inspired by Herbert Edelsbrunner (Duke), Bala works with Scott Provan (Department of Operations Research, UNC) on the analysis of protein geometry using its alpha complex [2]. For simplicity, each residue is represented by its alpha carbon atom. From the alpha complex, a "tube" can be defined around strands of consecutive residues. To capture adjacencies effectively, the barycenrtic subdivision of the alpha complex is used. Different structural motifs can be characterized by studying how the Betti numbers of the tubes vary when the alpha complex grows. Results are promising for the standard structural motifs such as alpha helix and beta sheets. Bala is currently trying to characterize non-standard structural motifs as well. Hopefully, these related projects will culminate next year in a successful defense of Bala's PhD thesis.

Publications

[1] Krishnamoorthy, B. and Tropsha, A. Development of a Four-Body Statistical Pseudo-Potential to Discriminate Native from Non-Native Protein Conformations. Bioinformatics 2003, 19, 1540-8.

[2] Krishnamoorthy, B. A topology-based characterization of protein structure, INFORMS Annual meeting, Atlanta, Oct 2003.

- Profile by Alex Tropsha

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.