BioGeometry NEWS
November 2004
biogeometry.cs.duke.edu

BioGeometry Events at PSB '05

An Introduction by organizers Alex Tropsha and Herbert Edelsbrunner

The analysis of structure-function relationships has traditionally been an area of interest of the Pacific Symposium on Biocomputing. Because (macro)molecular shape frequently defines the function, it seems evident that geometric methods should be an essential component of any attempt to understand and simulate biological systems. Existing techniques in computational structural biology and bioinformatics, however, rely primarily on sequence and, in some cases, structure information (in the context of 3D contacts or patterns of contacts) and use statistical and/or energy based methods to analyze the relationship between biological structure and function. They have been developed over three decades and have their roots in methods first applied by computational chemists to much smaller molecular systems. Although there have been significant advancements in the field, a systematic solution of many of the most important biological problems is still elusive, including ab initio protein structure prediction, the protein folding process, and ligand to protein docking.

Biogeometry is 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. Biogeometry embraces ideas from a wide range of areas of computer science and mathematics, including algorithms, geometry, topology, graphics, robotics, and databases to address some of the most fundamental biological problems such as structure-function relationships for biological molecules.

Although a new discipline, Biogeometry has been a subject of intensive research in several groups for a number of years. The "Computational Geometry for Structural Biology and Bioinformatics" project has been funded by NSF since 2001. It has involved researchers and students from Duke University, Stanford University, University of North Carolina Chapel Hill, and North Carolina A&T University (see http://biogeometry.duke.edu/ for additional information). Collectively, the collaborating researchers have published many dozens of papers and made numerous presentations at various national and international meetings. The Biogeometry session as part of PSB'05 is the first specialized session with such title ever included in a major computational biology conference, and the papers included in these proceedings present novel methods and developing ideas in a broad range of topics covered by Biogeometry.

The first two papers of the session apply computational geometry approaches to the problem of protein-protein recognition. The paper by Wang et al describes a coarse alignment algorithm for efficient protein-protein docking. This algorithm detects protrusions and cavities as local maxima of the novel elevation function, aligns them, and employs a simple scoring function to produce a reliable set of potential docking positions. Using a test set of 25 protein complexes, the authors demonstrate that their algorithm is able to generate near native conformations in all but one case. The paper by Li and Liang presents a novel method for designing peptide libraries to modulate protein-protein interactions. Based on the alpha shapes of antibody-antigen complexes, they develop an empirical pair potential for antigen-antibody interactions that depends on local packing. They demonstrate that this potential successfully discriminates the native interface peptides from a simulated library of 10,000 random peptides for 34 antigen-antibody complexes.

The next three papers explore various aspects of protein folding and design problems. For many practical tasks associated with the protein folding problem such as energy functions for folding simulations or fold recognition approaches to structure prediction, it is important to have a set of structure decoys. To this end, Singh and Berger describe their CHAINTWEAK algorithm for rapid generation of near native decoys starting from the native protein conformation (their paper will be published, but not presented). Russel and Guibas present the first application of so-called geometric spanners (geometric graphs with a sparse set of edges which approximate the n(n-1)/2 interatom distances with paths) to the segmentation of folding trajectories. They show that

Schedule for Saturday, January 8, 2005

Morning: Biogeometry Session at PSB

8:40A	Introduction to Biogeometry: Applications of computational geometry to molecular structure (5 papers)
8:50A	Coarse and reliable geometric alignment for protein docking by Y. Wang, P.K.Agarwal, P.Brown, H. Edelsbrunner, and J. Rudolph
9:10A	Computational design of combinatorial peptide library for modulating protein-protein interactions by X. Li and J. Liang
9:30A	Exploring protein folding trajectories using geometric spanners by D. Russel and L. Guibas
10:20A	An adaptive dynamic programming algorithm for the side chain placement problem by A. Leaver-Fay, B. Kuhlman, and J. Snoeyink
10:40A	Classification of non-coding RNA using graph representations of secondary structure by Y. Karklin, R.F. Meraz, and S.R. Holbrook
11:00A	Biogeometry discussion session: What could experimental biologists learn from computational geometry based models of biomolecules?

Afternoon: Biogeometry Project Meeting

Presentations by Students/Postdocs

- Forward and inverse kinematics in RNA backbone conformations by Xueyi Wang
- Motifs in protein-protein interfaces by Jeff Phillips
- A non-pairwise decomposable energy function to describe hydrophilic desolvation by Andrew Leaver-Fay
- Persistence diagrams as stable signatures of proteins? by Dmitriy Morozov
- Proton packing as a quality measure for NMR structures by Andrew Ban
- Real space protein model completion: an inverse kinematics approach by Itay Lotan
- Structure-based function inference using family-specific subgraph fingerprints mined from protein families by Deepak Bandyopadhyay

this representation affords easy visualization of the protein conformations over the entire folding trajectory of a protein and easy detection of the formation of secondary and tertiary structures as the protein folds. Leaver-Fay et al describe the novel application of a dynamic programming algorithm to a side chain placing problem, which facilitates the task of rational protein design.

Although most of the studies in the area of macromolecular structure and biocomputing have been done on proteins, there is a growing interest among computational biologists to study nucleic acids. The contribution from Karklin et al applies graph representation of noncoding RNA secondary structure to develop a structure classification method. They show that the combination of labeled dual graph representations and kernel machine learning methods (such as support vector machines) has potential for use in automated classification of uncharacterized RNA molecules or efficient genome-wide screens for RNA molecules from existing families.

As the Biogeometry session chairs, we are convinced that such interdisciplinary topic will continue to attract attention of leading specialists in computational, statistical, and biochemical/biophysical sciences who are interested in the role of shape in such fundamental computational problems as ligand-to-protein docking, ab initio and knowledge-based structure prediction, and visualization. Because of their fundamental role in structural biology, methods and applications to be discussed in this session will be of a great value for all participants of the PSB'05 conference.

The Biogeometry PI meeting will take place in the afternoon of January 8, following the morning Biogeometry Session. This is a very important meeting for all of us since the NSF grant is in its last year and we will need to discuss and decide on the practical steps to secure continuing funding for our joint project. PIs are encouraged to present a brief overview of their research programs with the emphasis on exciting developing and future research in the context of the entire Biogeometry project. The student presentations will reflect most recent accomplishments in individual research groups. The tentative participants and the agenda can be found on the other side of this newsletter.

Student Profile: Jeff Headd

Jeff Headd, a second year student in the Duke Program for Bioinformatics and Genome Technology (BGT), has recently joined the collaboratory of Herbert Edelsbrunner (Computer Science) and Johannes Rudolph (Biochemistry). His general research interests lie at the heart of BioGeometry. Specifically he hopes to contribute substantially towards computational advances in the study of protein-protein interactions.

Jeff was an undergraduate in Computer Science at Brown University in Providence, RI. Already as an undergraduate he delved into the world of chemistry and biology by taking courses such as organic chemistry and cancer biology. His undergraduate independent research project was with Y. Eugene Chin studying the structure-function relationships of the JAK and STAT families of protein tyrosine kinases using both computation and bench work in the laboratory.

After joining the BGT program at Duke in the Autumn of 2003, Jeff continued to acquire a broad and diverse background in computational biochemistry by taking courses in Computer Science, Statistics, Biochemistry, and Computational Biology. He supplemented this course work with three different rotation projects. In the Fall of 2003 he worked with Herbert

and Johannes on a model system for introducing protein flexibility into protein docking by molecular dynamics. In the Spring of 2004 he worked with Homme Hellinga (Biochemistry) on re-engineering protein-protein interaction surfaces using both his computational and laboratory skills. In the Summer of 2004 he worked with Merlise Clyde (Statistics) on developing algorithms for modeling peak data from mass spectrometry.

In the Fall of 2004, Jeff started working in earnest on protein docking, the computational prediction of protein binding that uses only the knowledge of their individual crystal structures. Through previous efforts by Sergei Bespamyatnikh, Vicky Choi, and Yusu Wang, our collaboratory is presently able to take the structure of any known protein-protein complex and re-dock it with high accuracy following a randomization of the original orientation. While this is a significant achievement, the real problem lies in docking two separately determined structures and taking into account the side-chain and/or backbone rearrangements that take place upon docking. Thus, Jeff's initial goal will be to implement side-chain flexibility into our computational approaches to protein-protein docking.

He faces many daunting problems in this undertaking. He needs to choose a suitable rotomer library that contains sufficient diversity to adequately dock proteins without being too computationally expensive. He needs to efficiently define the residues he is interested in reorganizing, both by their location at the potential interface and by their probable flexibility. And he needs to do all this with a delicate touch, as dramatic surface re-orientations will certainly allow almost any two proteins to be smeared together without specificity. Jeff has a benchmark database of undocked and docked protein structures on which to base his decisions and with which to evaluate his methods. Most importantly, his background, training, and research interests situate him ideally to tackle this major problem in computational biology.

In addition to his research work, Jeff does have an outside life. He is currently lead guitarist in the Durham-based band "Pleather", a Hair Metal tribute band. Additionally he is rumored to have a girlfriend.

- Profile by Johannes Rudolph

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.