Biogeometry: Applications of Computational Geometry to Molecular Structure: Session Introduction

A. Tropsha and H. Edelsbrunner

Pacific Symposium on Biocomputing 10:1-3(2005)

BIOGEOMETRY: APPLICATIONS OF COMPUTATIONAL GEOMETRY TO MOLECULAR STRUCTURE*

ALEXANDER TROPSHA
School of Pharmacy, University of North Carolina, Chapel Hill, NC, USA (alex_tropsha@unc.edu)

HERBERT EDELSBRUNNER
Duke University, Computer Science Department, Box 90129, Durham, NC 27708, USA (edels@cs.duke.edu)

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

*This work is supported by the National Science Foundation grant CCR-0086013.

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.

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. Russell 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 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 non-coding 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.