Computational Geometry in Molecular Biology

Department of Computer Science University of North Carolina at Chapel Hill February 2004

The Challenge

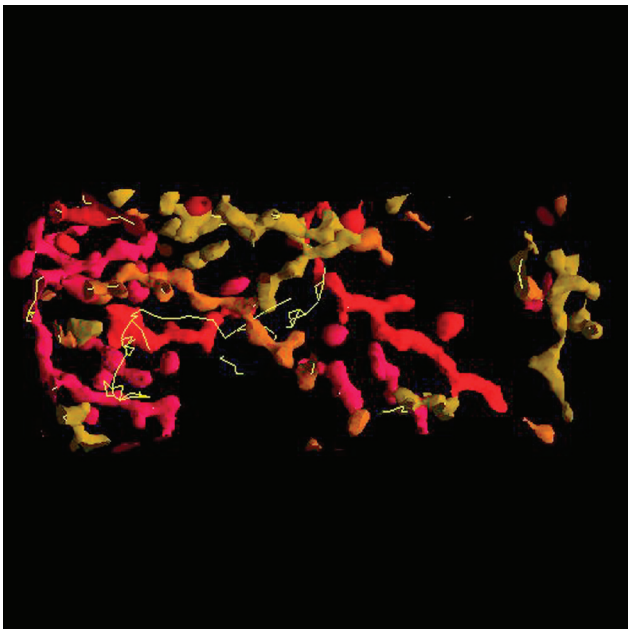
The function of all life forms depends on organization in space and time, and the effect of one part of a biological system on another is generally much greater when the two parts are in close proximity in space and/or time. In themselves, these two observations would seem to indicate that geometric methods should be an essential component of any attempt to understand and simulate biological systems. Existing techniques in computational structural biology and bio-informatics, however, rely primarily on sequence information and use statistical and energy-based methods to analyze 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.

The Approach

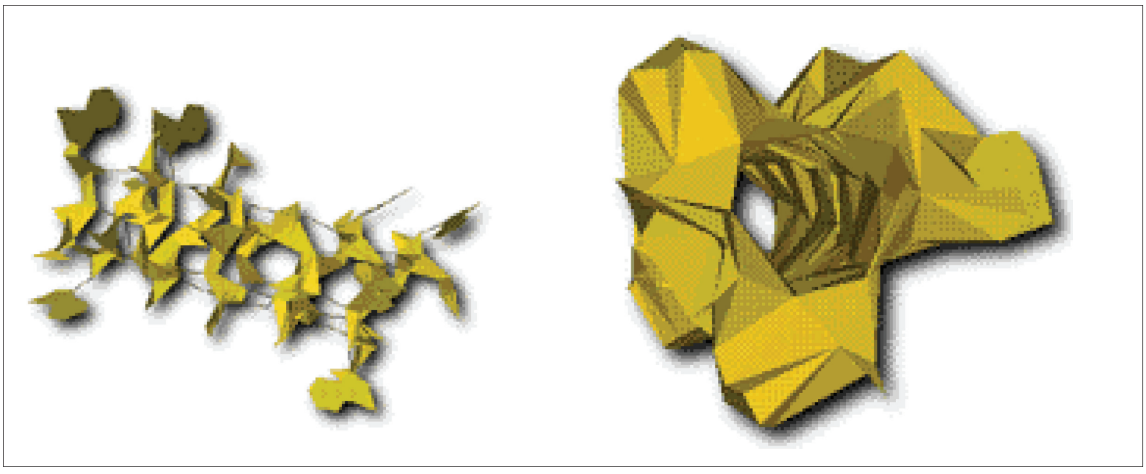
It is widely believed that the geometry of molecules plays a crucial role in molecular processes, yet geometrical methods are relatively uncommon in computational biology because of several unresolved representational and algorithmic issues. We work to develop new computational techniques and paradigms for representing, storing, searching, simulating, analyzing, and visualizing biological structures. We rely on geometry, but combine it with statistics and

Highlights

- **Developing geometry-based tools to analyze the relationship of 3D protein structure to biochemical function.**
- **Interdisciplinary project involves computer scientists, biochemists, and biophysicists from Duke University, N.C. Agricultural and Technical State University, Stanford University, and UNC–Chapel Hill.**

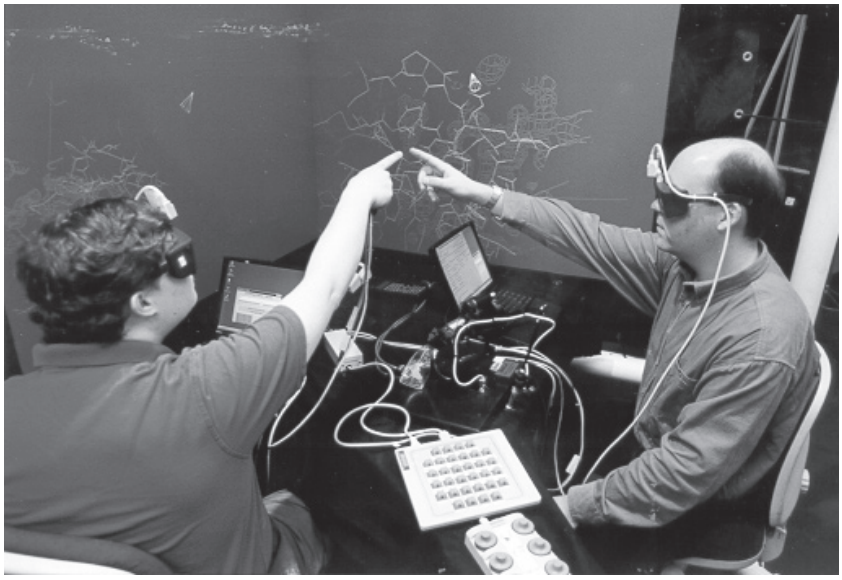


Preliminary work on automatic bond detection using Morse theory.



Tunnel inside Gramicidin as revealed by alpha shapes.

physics. We aim for methods that have practical, predictive power and validate them by comparison with the best existing techniques. In order to transfer the technology in an effective way and to have real impact on research in biology, our researchers try to create software that will help structural biologists with their research and that can be integrated with their current tools.



Using the Protein Interactive Theater (PIT) to collaboratively explore protein molecules.

Ideas from a wide range of areas of computer science and mathematics—including algorithms, geometry, topology, graphics, robotics, and databases—are needed to accomplish our goals. Some of the problem areas addressed represent great challenges for computer science itself. These include building and querying large libraries of three-dimensional and possibly flexible shapes, exploring hierarchical representations of deformable geometry, integrating geometry and physics in modeling, and properly sampling systems with many degrees of freedom.

At UNC–Chapel Hill, we work on three areas in particular:

Structure Determination. We develop tools for modification of electron density in real space and phase refinement in reciprocal space that will support automated collection of high-quality data. These include new methods based on Sayre's equations, and computation of iso-surfaces in electron density maps and their properties.

Modeling and Algorithms. We work on developing shape representations that incorporate topological persistence, on geometric and topological simplification hierarchies, on Morse theory applied to electron density maps, and on computation and evaluation of shape descriptors.

Visualization. We integrate the representations, algorithms, and software developed within the entire project for display on the unique visualization tools at UNC–Chapel Hill, such as the PIT (Protein Interactive Theater).

Project Leaders

Jack Snoeyink, professor (Computer Science)
Frederick P. Brooks Jr., Kenan professor (Computer Science)
Charlie W. Carter, professor (Biochemistry and Biophysics)

Other Investigators

David Hsu, postdoctoral researcher (Computer Science)
Jeffrey Roach, postdoctoral researcher (Biochemistry)

Graduate Research Assistants

Hamish Carr (Computer Science, University of British Columbia)
Deepak Bandyopadhyay
Andrew Leaver-Fay
Ajith Mascarenhas
David O'Brien
(All Computer Science, UNC-Chapel Hill)

Research Sponsors

National Science Foundation (Information Technology Research Grant 0076984)

Key Words

Protein folding; molecular modeling; X-ray crystallography; alpha-shapes

For More Information

Dr. Jack Snoeyink
Department of Computer Science
University of North Carolina at Chapel Hill
CB#3175, Sitterson Hall
Chapel Hill, NC 27599-3175
Phone: (919) 962-1969
Fax: (919) 962-1799
E-mail: snoeyink@cs.unc.edu
Web: www.cs.unc.edu/Research/compgeom/bioGeom/