



Computational Geometry of Structural Biology

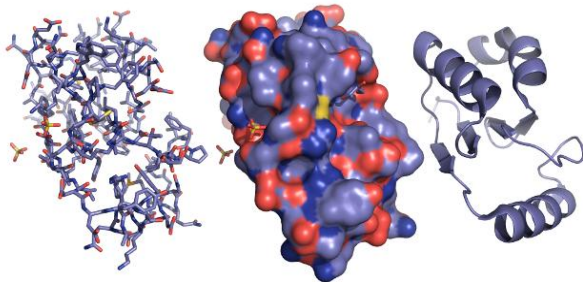
Department of Computer Science

University of North Carolina at Chapel Hill

March 2010

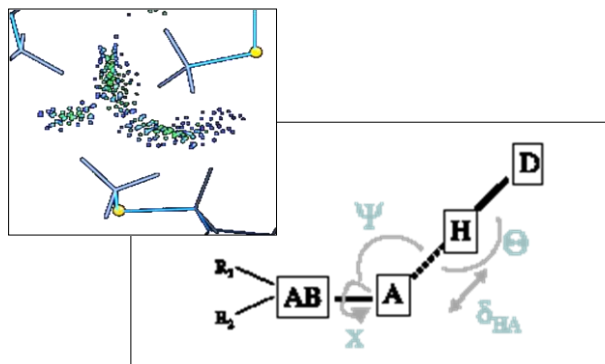
The Challenge

Diverse biological function is created by encoding the 3D atomic structure of macro-molecules such as Proteins and RNA in the 1D genetic sequence. Understanding the sequence to structure to function relationships allows biochemists to predict the activity of genes and rationally design genes with novel biological function.



Three representations of the atomic structure of a protein.

The confluence of the genomics and computer revolutions has led to new technologies that produce exponentially growing amounts of molecular structure data. Whole genomes are being sequenced, the Protein Databank has over 60,000 experimentally solved structures at atomic level accuracy and structure prediction software are generating large numbers of candidate structures on powerful compute clusters. The challenge of turning this data into knowledge is to bring together computational and biochemical thinking to develop models that are able to predict protein structure and solve important problems in biology.



Two models of hydrogen bonding from our collaborators:
(Left) MolProbity uses contact dots to evaluate H-bonds
(Right) Rosetta uses four geometric parameters to evaluate H-bonds

The Approach

Computational models in structural biology are often surprisingly geometric: Atoms are represented as points in space and chemical bonds are represented as constraints

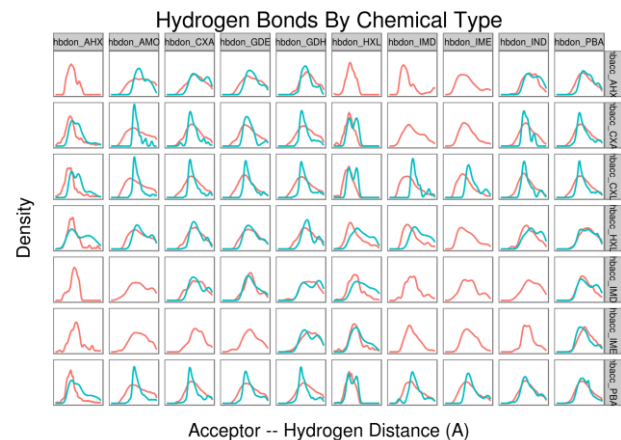
Highlights

- Close collaborations with leading structural modeling systems including Rosetta (UNC, UW), Molprobity (Duke), and Phenix (Duke, Cambridge, Berkeley)
- Large scale refactoring of hydrogen bond modeling in Rosetta to accommodate higher accuracy parametrizations
- Exploration of large experimental structural biology datasets for geometric feature statistics

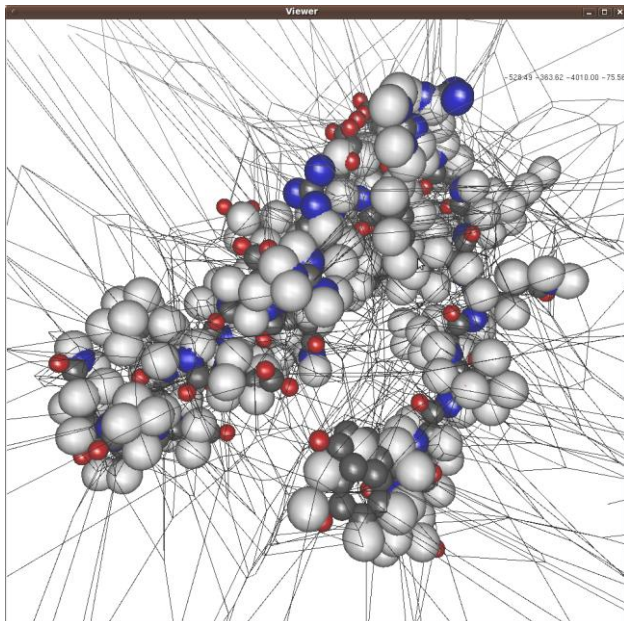
on distances and angles between atoms. With this interpretation the structure of a macromolecule is an articulated body in a high dimensional configuration space. We can apply computational geometry techniques drawn from robotic motion planning, graphical information systems, and combinatorial optimization to build robust and scalable predictive models of molecular structure.

Hydrogen Bonding

A focus of our group is building and evaluating models of hydrogen bonding with our collaborators. How do the geometric configurations of hydrogen bonds in predicted molecular structures differ from naturally occurring structures? To answer this question we are developing tools to explore and analyze ensembles of geometries. We then incorporate our findings to improve leading structure prediction and evaluation tools such as Rosetta and Molprobity.



Hydrogen bonding geometries split by the chemical type of the donor and acceptors. The Red densities are from experimental crystal structures. The blue densities are from Rosetta predictions.



Accessibility Radius

We leverage the tools and methodology we develop in improving the modeling of hydrogen bond geometries to identify new computational geometry flavored observables of biological molecules such as the accessibility radius.

The strength of a hydrogen bond has been shown to depend upon its proximity to polar solvents such as water. We developed an approach to rapidly identify which bonds are accessible to water molecules by using a generalization of the Voronoi Diagram called the Power Diagram. If used in structure prediction software, the quality of a bond can be modulated to more accurately evaluate subtle properties of the hydrogen bonding interaction.

Current Project Members

Jack Snoeyink (Principal Investigator)
Matthew O'Meara, Graduate Research Assistant
Vishal Verma, Graduate Research Assistant

Rosetta Commons Collaboration

Kuhlman Lab, UNC Department of Biochemistry
Baker Lab, UW-Seattle Department of Biochemistry

MolProbity Collaboration

Richardson Lab, Duke Department of Biochemistry

Phenix Crystallography Collaboration

Richardson Lab, Duke Department of Biochemistry

Selected Publications

Verma and Snoeyink. *Defining and computing accessibility radius* In Proceedings of the 21st Canadian Conference on Computational Geometry (CCCG2009), pages 123-126, 2009.

Wang, Kapral, Murray, Richardson, Richardson, and Snoeyink. *RNABC: forward kinematics to reduce all-atom steric clashes in RNA backbone*. J Math Biology, 56:253–278, 2008.

Wang, Snoeyink. *Defining and Computing Optimum RMSD for Gapped and Weighted Multiple-Structure Alignment*. IEEE/ACM Trans. Comput. Biology Bioinform. 5(4): 525-533 (2008)

Leaver-Fay Snoeyink, and Kuhlman, *On-the-fly rotamer pair energy evaluation in protein design*. In Proc. 4th Int'l. Symp. Bioinf. Research & App. (ISBRA), volume 4983 of Lecture Notes in Computer Science, pages 343–354. Springer-Verlag, 2008.

Leaver-Fay, Butterfoss, Snoeyink and Kuhlman. *Maintaining solvent accessible surface area under rotamer substitution for protein design*. J. Comp. Chem., 28, 1336-41.

Davis, Leaver-Fay, Chen, Block, Kapral, Wang, Murray, Arendall III, Snoeyink, Richardson, and Richardson, *MolProbity. All-atom contacts and structure validation for proteins and nucleic acids*. Nucleic Acids Research, 35:W375–W383, 2007. Web server issue.

Bandyopadhyay, Huan, Liu, Prins, Snoeyink, Wang, and Tropsha, *Structure-based Function Inference Using Protein Family-specific Fingerprints*, Protein Science, 15:1-7, 2006.

Research Sponsors

NIH / DARPA

Keywords

Computational geometry; structural biology; exploratory data analysis, model selection, geometric algorithms, parameter fitting, proteomics, geometric optimization, statistical validation, quantum modeling

For More Information

Matthew O'Meara
 E-mail: momeara@cs.unc.edu