BioGeometry NEWS January 2004
biogeometry.cs.duke.edu

Research News

Predicting Molecular Crystal Structures by Robert Gdanitz


Robert Gdanitz at a seminar talk

Since about 10 years ago, computer algorithms are available that can generate probable structures of molecular crystals [1]. Such predictions are important for virtually any scientific research and industrial application that deals with molecular crystals. The modified Monte-Carlo Simulated Annealing method [2] of PI Robert Gdanitz (NCAT, Physics) is, in fact, the first method that allows for automatic predictions of this kind and it is still considered as being among the best available [3]. The method owes its success to a multi-scale algorithm that confines the random search to regions that correspond to fairly dense crystal packings.

Unfortunately, at the time being, the method is only available in a commercial software package [4]. In collaboration with PIs Pankaj Agarwal (Duke), Herbert Edelsbrunner (Duke), Michael Levitt (Stanford), and Jack Snoeyink (Chapel Hill), the popular molecular modeling code CHARMM of Prof. Martin Karplus (Harvard) will be used to generate an improved academic version as well. Since major aspects of molecular crystal packing are purely geometrical in nature, significant improvements due to the expertise of the BioGeometry group are to be expected. Indeed, under the supervision of Pankaj Agarwal and Herbert Edelsbrunner, Ibrahima 'Ibou' Mbaye (former Physics, NCAT) did some initial work last summer on using normalized quaternions (i.e. a four-component generalization of complex numbers) to specify the orientation of a (rigid) molecule in space.

To facilitate this work, together with PI Solomon Bililign (Physics, NCAT) and the computer expert Prof. Floyd James (NCAT), a new state-of-the-art computer system consisting of 8 dual processor Apple Macintosh G5 ("Big-Mac") is currently being installed at the Physics Dept. of NCAT. Moreover, we are in the process of hiring a graduate student from Pakistan who will support this project beginning in spring.

[1] R.J. Gdanitz. Ab initio prediction of molecular crystal structures. Current Opinion in Solid State and Material Science 3 (1998), 414.

[2] R.J. Gdanitz. Prediction of molecular crystal structures by Monte Carlo simulated annealing without reference to diffraction data. Chem. Phys. Lett. 190 (1992), 391.

[3] W.D.S. Motherwell, H.L. Ammon, J.D. Dunitz, A. Dzyabchenko, P. Erk, A. Gavezzotti, D.W.M. Hofmann, F.J.J. Leusen, J.P.M. Lommerse, W.T.M. Mooji and S.L. Price, H. Scheraga, B. Schweizer, M.U. Schmidt, B.P. van Eijck, P. Verwer, and D.E. Williams: Crystal structure prediction of small organic molecules: a second blind test. Act. Cryst. B 58 (2002), 647.

[4] Polymorph Predictor in Cerius² of Accelrys.

Education

An important component of education in computational structural biology is the development and delivery of topical courses. We present short descriptions of one new and one mature course.

CPS 296.2: Shape Analysis (Spring 2004; Agarwal; Duke)

The course is intended to provide a systematic introduction to the modeling and algorithmic techniques behind the geometric and statistical analysis of 3D shapes that arise in many applications including molecular biology, computer graphics, and computer aided design. The primary emphasis will be on recent algorithms developed for representing, analyzing, comparing, classifying, and indexing 3D shapes.

The topics covered will include:
I. Shape representation: Basic representation methods, shape simplification, hierarchical methods, deformable shapes; II. Shape descriptors: Histograms, harmonic maps, distance distribution, medial axis, topology based methods; III. Statistical shape analysis: Shape space, coordinate systems, procrustes distances and their generalizations, deformations; IV. Shape matching and registration: Combinatorial methods, geometric hashing, ICP and its variants, graph matching, entropy based methods; V. Shape classification and clustering: Geometric clustering, graph based methods, spectral methods, decision trees, support vector machines; VI. Shape indexing: Indexing multidimensional data, proximity search, search engines.

http://www.cs.duke.edu/education/courses/spring04/cps296.2

MEDC 276 (BIOC 276): Macromolecular Modeling (Annually, Spring; Tropsha; UNC)

This course introduces major modeling concepts as applied to the analysis, prediction, and use of protein structure. The course is divided into three sections. The first part addresses general aspects of protein structure organization (secondary and tertiary structure) and classification of proteins into structural and functional classes. It explores available methods for protein structure prediction from sequence using knowledge based homology modeling and fold recog-

nition approaches. The second part concentrates on methodologies for molecular simulations including molecular dynamics and free energy perturbation techniques. Finally, the third part considers methods for structure-based drug discovery using molecular docking and scoring. Elements of computational geometry, i.e., Voronoi and Delaunay tessellations, are taught in the context of the first and the third sections of the class.

Student Profile: Vijay Natarajan

Vijay Natarajan is a Ph.D. student in the Computer Science Department at Duke University and works with Herbert Edelsbrunner and John Harer. He did his undergraduate studies in India at the Birla Institute of Technology and Sciences, Pilani, majoring in mathematics and computer science.

Vijay's first project at Duke was on computing coarse representations of three-dimensional density maps that arise in various application areas, including x-ray crystallography, medical imaging, geometric modeling and computer graphics. These datasets are ever increasing in size. In order to study and analyze this data efficiently, it is useful to work with a simplified and small representation. The data is typically specified over a region of three-dimensional space which is triangulated. Vijay extended a well known quadric error metric to combine the goals of determining a good approximation of the map and generating good quality elements in the resulting simplified mesh. He presented this work in the 2001 ITR meeting at Stanford and the results are to appear in the IEEE Transactions on Visualization and Computer Graphics [1].

While performing the density map simplification, it is beneficial to retain the important features of the map. The approximation error measure used in the above project was geometric in nature. Vijay also studied the behavior of topological features, described by the critical points of the map, during the simplification. He got interested in exploring other topological structures that could be used to represent features and form the basis of an analysis tool. He started working on Morse-Smale complexes, which partition the domain of a given real-valued function into regions with uniform gradient flow behavior. Together with Edelsbrunner and Harer he developed an algorithm for computing the Morse-Smale complex of a real-valued function defined over a 3-manifold. He then got an offer to work as an intern at the Lawrence Livermore National Laboratory in California. In total, he has spent almost a year working with Valerio Pascucci on the construction of Morse-Smale complexes and effective methods for visualizing them.

The Morse-Smale complex has a rich structure and displaying the entire complex results in visual clutter. In many cases, displaying just the one-dimensional substructures seems to be useful. Vijay is currently working with some electron microscopy data of DNA helicase, trying to understand how its structure is represented by the Morse-Smale complex. He presented this work as a poster during the 2002 ITR meeting at UNC, Chapel Hill and gave a talk in the 2003 ITR meeting at Duke. This work was also published in the proceedings of the Symposium on Computational Geometry [2].

In addition to the research described above, Vijay is working on the development of a topology-based similarity measure that can be used to compare the features of two scalar functions defined over a common two-dimensional domain. This measure, referred to as the topological correlation between the two functions, computes the variation of one function over the isocontours of the other. Three different formulations of the similarity measure are given, each leading to a different algorithm. Two of these algorithms run

in time linear in the size of the mesh representing the domain and the functions. A visualization tool helps the user perform a comparative study of the functions. The tool computes the similarity measure also as a scalar function over the domain which is then optionally mapped to colors or to a terrain for visualization.

Besides research work, Vijay is also active in various departmental activities. He is the graduate student liaison and serves regularly on the graduate student admissions committee and on faculty search committees.

[1] V. Natarajan and H. Edelsbrunner. Simplification of three-dimensional density maps. IEEE Transactions on Visualization and Computer Graphics, to appear.

[2] H. Edelsbrunner, J. Harer, V. Natarajan and V. Pascucci. Morse-Smale complexes for piecewise linear 3-manifolds. In Proc. Symposium on Computational Geometry 2003, 361--370.

- Profile by Herbert Edelsbrunner

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.

BioGeometry News is the project's monthly newsletter. For more information, please visit http://biogeometry.cs.duke.edu/newsletter