DOE Grant DE-FG02-95ER25253 Final Report

Development of Simulation Tools for Virus Shell Assembly

Bonnie Berger

1 Introduction

Prof. Berger's major areas of research have been in applying computational and mathematical techniques to problems in biology, and more specifically to problems in protein folding and genomics. Significant progress has been made in the following areas relating to virus shell assembly: development has been progressing on a second-generation self-assembly simulator which provides a more versatile and physically realistic model of assembly; simulations are being developed and applied to a variety of problems in virus assembly; and collaborative efforts have continued with experimental biologists to verify and inspire the local rules theory and the simulator. The group has also worked on applications of the techniques developed here to other self-assembling structures in the material and biological sciences. Some of this work has been conducted in conjunction with Dr. Sorin Istrail when he was at Sandia National Labs.

2 Notable Events

The following notable events have occurred:

- Berger was promoted to Associate Professor with tenure in the MIT Dept. of Mathematics.
- Berger gave an Invited Address at the American Math Society annual meeting.
- Russell Schwartz completed the PhD program in the MIT Dept. of Electrical Engineering and Computer Science, then worked with Prof.s Berger and King as a postdoc, and began work at Celera in Nov. 2000.

DOE Patent Clearance Granted

MP Dvorscak

10/30/01

Mark P Dvorscak

Date

(630) 252-2393

E-mail: mark.dvorscak@ch.doe.gov

Office of Intellectual Property Law

DOE Chicago Operations Office

3 Summary of Research Progress

3.1 Simulating Virus Shell Assembly

The assembly pathway of viruses represents an interesting example of a self-assembling structure and an attractive yet under-explored antiviral target. This problem has previously proven experimentally difficult due to the lack of a general framework for approaching the problem. Berger et al. [4, 3] have developed a local rules-based theory of virus shell assembly, making it possible to develop

computational tools to motivate and interpret biochemical experiments. Suitable computational tools should expedite and direct general understanding of protein folding and antiviral efforts.

Work on this project has been conducted by Berger and EECS graduate student Russell Schwartz in collaboration with Dr. Peter Shor at AT&T Labs, Dr. Peter Prevelige of the University of Alabama at Birmingham Dept. of Microbiology, Dr. Robert Garcea of the University of Colorado Dept. of Pediatric Hematology/Oncology, and Dr. Jonathan King of the MIT Dept. of Biology. Progress on simulating virus assembly has been concentrated in three areas: improving the assembly simulator, developing and running simulations, and devising local rules for assembly. In addition, the project has been the subject of several publications and presentations.

Schwartz and Berger have continued work on the development and improvement of the "laws of physics" virus assembly simulator. This has involved further exploration of reasonable mathematical models for binding interactions. These explorations have particularly focused on developing more realistic simplified models of the tolerance of binding to inexact positioning of subunits and the development of reasonable models of the process of forming and breaking binding interactions. In addition, extensive testing is currently being conducted on the dependence of simulation behaviors, such as the average time to form or break bonds, on a variety of simulation parameters. These experiments are meant to validate the simulation model, determine limits on its applicability, and provide predictive mathematical models of its behavior for use in simulation design.

Schwartz and Berger have also worked on reducing the computational cost of running simulations. Several optimizations have focused on the problem of efficiently integrating rotational particle motion; mathematical techniques developed for handling similar problems in computer graphics have been adapted to accelerate these calculations and improve their accuracy. In addition, some extensions have been added to the simulation control language to allow for more time- and memory-efficient simulations.

One area that has received considerable attention is the creation of an alternate implementation of the simulator's numerical methods using a waveform-relaxation method. Waveform-relaxation methods have been developed for circuit simulations in order to take advantage of shared-memory parallel architectures. They treat individual variables in a system of many variables as waveforms that can be evaluated in parallel, using methods for single-variable systems, and refined iteratively. A waveform-relaxation method has been adapted to work with particle simulations, treating individual particles rather than individual variables as isolated systems of equations, in order to improve parallel performance of the assembly simulator. This has involved developing an appropriate numerical method for single-particle time integration and creating the necessary data structures and auxiliary algorithms to implement it. Furthermore, a suite of tests was conducted to assess the serial and parallel performance of both the old and new numerical methods under a variety of simulation conditions and problem sizes.

In addition, Pulim has begun work on an alternate method for implementing local rules simulations of particle dynamics without the use of continuous numerical methods. A new simulation engine is under development that will evaluate particle positions and implement collisions and binding events using a discrete, event-driven method. This experimental method is intended to avoid many of the issues of numerical instability that affect traditional continuous numerical methods, thus requiring less knowledge on the part of the user to create efficient simulations. In addition,

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

the method is being explored as a possible means of accelerating simulations overall.

Work has continued on creating and running local rules models for different virus systems. Further work has been conducted, in collaboration with Garcea, on developing non-kinetic models of papovaviruses, a class of animal viruses including some believed to cause cancer in humans. This work has focused on creating models capable of explaining observed malformations of papovavirus capsids. In addition, simple kinetic models for the "laws of physics" simulator have been developed for papovaviruses; these models are currently being refined. Other work has been conducted on developing "laws of physics" models of T=3 virus capsids. Another project has focused on developing a more realistic treatment of earlier work, done in collaboration with Prevelige, exploring the relative merits of different models for nucleation-limited assembly. Additional work was also conducted on a project to develop local rules explanations for the link between T=4 and T=7 geometries.

Berger, Schwartz, Prevelige, and Shor have presented their work to the biology, computer science, and mathematics communities. A paper by Schwartz, Shor, Prevelige, and Berger introducing the kinetic simulation model has been published [10]. Berger, King, Schwartz, and Shor have also published a paper on the work relating T=4 and T=7 geometries [3]. Schwartz, Prevelige, and Berger have prepared a technical memo on earlier work applying local rules simulations to compare different assembly models [8]. Schwartz, Garcea, and Berger have published a paper describing their papovavirus work [9]. In addition, Schwartz spoke on local rules modeling at the DIMACS Reconnect conference in July 1998, presented a poster on the kinetic simulation work at the January 1999 PMMB meeting, and delivered a talk on experiments conducted with local rules simulations at the Phage and Virus Assembly Meeting in June 1999.

3.2 Materials Science

We have found a computationally efficient way to model radiation damage of crystal lattices. Usually this is done with *ab initio* molecular dynamics simulations, which for the time being are prohibitively complex for structures with more than a few hundred atoms. In such small arrangements, the boundary is very close to all the atoms involved, and this has a strong influence on the structures generated. Thus while molecular dynamics is very realistic physically, it is hard to draw conclusions from it about radiation damage to more reasonably-sized structures.

We have been studying disordering of silicates. It is believed that following radiation bombardment of silicates, short range SiO_4 tetrahedra form very rapidly, and the final structure is built from these. Thus we believe that efficient packings of tetrahedra may provide a good model for disordered crystal structures.

3.2.1 Methods for Constructing Packings

First a regular crystalline arrangement is constructed. Then the tetrahedra within the region that we wish to disorder are randomly moved and rotated slightly. A graph is constructed whose vertices correspond to all the tetrahedral vertices that were freed during the random motions. An edge between two such vertices is included if a bond between the corresponding tetrahedral vertices is energetically feasible, and a weight is assigned to this edge according to just how feasible it is -

the lower the energy required by the bond, the higher the edge's weight. Next a *maximal match* is constructed from this graph. (A *match* is a subgraph in which every vertex has degree at most one, and a maximal match is a match for which the sum of the weights on the edges it includes is as large as possible.) This match specifies a set of bonds between the tetrahedra, and energetically speaking, this collection of bonds is quite efficient. Finally, this structure is optimized according to a simple potential.

We have not yet attempted it, but we can see no theoretical impediment to applying this construction to the disordering of other crystal structures.

3.2.2 Results

Very highly connected and energetically efficient disordered arrangements have been constructed in this way. No other method for producing highly connected disordered structures is known to us, so these arrangements provide a unique window into the types of packings that are possible in disordered structures.

They have already proven to be an extremely useful tool in testing conjectures regarding such structures. For instance, we started to explore a possible topological invariant of crystal networks that would have been a two-dimensional analogue of the local cluster of a node. We were looking for subgraphs of the network in which every node had degree three. However, the large suite of examples of crystal networks now available to us allowed us to quickly find a counter-example in which such subgraphs did not exist.

We have also worked on presenting these results to the materials science community. Articles published include: [5, 6]. In addition, Hobbs has recently spoken about this work.

3.3 Simulating Protein Aggregation

Graduate student Russell Schwartz has also worked with Dr. Sorin Istrail of Sandia National Labs and Prof. Jonathan King of the MIT Biology Dept. on a project involving simulating protein aggregation. Protein aggregation occurs when proteins, in the process of folding from their linear amino acid sequences to their three-dimensional functional forms, become stuck to each other and cannot reach their correct final forms. Protein aggregation is implicated in several human diseases and has also become a significant problem in the commercial production of large volumes of proteins through recombinant DNA. It is hoped that gaining a better understanding of the problem and its causes may provide means for reducing its effects.

Schwartz and Istrail have been approaching this problem through the use of simplified Monte Carlo lattice models. Such models have been used extensively to study protein folding dynamics and were therefore chosen to model the dynamics of the aggregation process. Using these models, Schwartz and Istrail have conducted folding simulation experiments and have used these to gather statistics on how peptide chain aggregation relates to other chain properties. It is hoped that such statistics can form a basis for studying how to predict or control the likelihood of aggregation of a peptide.

Work on this topic has continued in the past year, focusing on ways to adapt the methods to

actual diseases related to aggregation and on computational methods for modeling multi-protein aggregates. Attempts at application have particularly focused on applying the method to data on Alzheimer's disease, which is known to be linked to the formation of protein aggregates in the brain. One idea being explored to extend the model is to combine it with a differential equation model of the progress of aggregate formation. Another project involved using local rules techniques to predict likely structures for the beta amyloid fibrils found to produce plaques in the brains of Alzheimer's patients. It is hoped that such models will suggest mechanisms of assembly for the fibrils and possible avenues by which their growth might be affected. In addition, preliminary work has been conducted on merging some aspects of the protein aggregation project with the virus assembly work described above; in particular, it is hoped that the simulation techniques developed for analyzing virus capsid assembly can be applied to study protein aggregate assembly as well.

Schwartz began this project while working as a summer intern under the direction of Dr. Sorin Istrail of Sandia National Laboratories. Istrail, Schwartz, and King have published a paper describing the work [7]. In addition, the work has been described in *Science News* and in *UniSci*, an online journal of scientific news. Brian Walenz of Sandia National Laboratories has worked on adapting simulation code developed for exploring aggregation for use with Sandia's Tortilla software package.

3.4 Bipole Assembly

Graduate student Serafim Batzoglou in conjunction with Dr. Istrail and Dr. Bill Hart of Sandia Labs, and Brian Walenz of UNM and Sandia Labs, have been working on lattice protein folding approximation algorithms, and on the mathematics of materials self-assembly used for rigorous estimation of the lattice native conformations.

The self-assembly of bipoles can be modeled on the cubic and face-centered cubic lattices by representing a bipole as two adjacent points on the lattice, one of them labeled with an *H* (Hydrophobic) and the other with a *P* (Polar). A conformation of bipoles has an energy that depends on the number of *H-H* contacts on the lattice. Batzoglou, Hart, Istrail, and Walenz have been working on proving rigorous lower bounds on the native conformation energies for this model. Moreover, they demonstrate what implications these bounds have on protein folding on the face-centered cubic lattice, under the side-chain *H/P* folding model.

Serafim has also been working with Dr. Istrail on Physical Mapping, in the presence of repeated probes. The algorithmic aspects of Physical Mapping involve finding the "minimal explanation" arrangement of clones, given hybridization data that could contain errors such as repeated probes. Batzoglou and Istrail have new results on the complexity of Physical Mapping with repeated probes, as well as efficient approximation algorithms for biologically relevant simple versions of the problem. In particular the sparse version of the problem is proven MAXSNP-complete and efficient algorithms with good approximation ratios are demonstrated. The algorithms are implemented and tested on simulated data. This work was presented in the CPM '99 conference, and appears in the conference proceedings [1, 2].

References

- [1] S. Batzoglou and S. Istrail. Physical mapping with repeated probes: The hypergraph superstring problem. *Lecture Notes in Computer Science, Special Issue on CPM '99*, 1645, 1999.
- [2] S. Batzoglou and S. Istrail. Physical mapping with repeated probes: The hypergraph superstring problem. *Journal of Discrete Algorithms: Special Issue on Matching Patterns*, 2001. In press.
- [3] B. Berger, J. King, R. Schwartz, and P. W. Shor. Local rule mechanism for selecting icosahedral shell geometry. *Discrete Applied Mathematics*, 105:55–69, 2000.
- [4] B. Berger, P. W. Shor, L. Tucker-Kellogg, and J. King. Local rule-based theory of virus shell assembly. *Proc. of the Natl. Academy of Sci., USA*, 91(16), aug 1994.
- [5] L. Hobbs, C. Jesurum, V. Pulim, and B. Berger. Local topology of silica networks. *Philosophical Magazine*, A 78:679–711, 1998.
- [6] L. Hobbs, C. Jesurum, V. Pulim, and B. Berger. Topological modeling of cascade amorphization in network structures using local rules. *Materials Science and Engineering*, A 253(1-2):16–29, 1998.
- [7] S. Istrail, R. Schwartz, and J. King. Lattice simulations of aggregation funnels for protein folding. *J. Comp. Biol.*, 6:143–162, 1999.
- [8] R. Schwartz, , P. E. Prevelige, and B. Berger. Local rules modeling of nucleation-limited virus capsid assembly. MIT LCS Technical Memo 584, 1998.
- [9] R. Schwartz, R. L. Garcea, and B. Berger. “local rules” theory applied to polyomavirus polymorphic capsid assemblies. *Virology*, 268:461–70, 2000.
- [10] R. Schwartz, P. W. Shor, P. E. Prevelige, and B. Berger. Local rules simulation of the kinetics of virus capsid self-assembly. *Biophys. J.*, 75:2626–2636, 1998.

2011 MAR 6 - 6:11 PM