DOE/ER/25253--T1

DOE Grant Progress Report '95–'96 for The Mathematics of 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. Significant progress has been made in the following areas relating to virus shell assembly: the local rules theory has been further developed; development has begun on a second-generation simulator which provides a more physically realistic model of assembly, collaborative efforts have continued with an experimental biologist to verify and inspire the local rules theory; an investigation has been initiated into the mechanics of virus shell assembly; laboratory experiments have been conducted on bacteriophage T4 which verify that the previously believed structure for the core may be incorrect.

2 Notable Events

The following notable events have occurred.

- The virus shell assembly work was featured in a *Science News* article (3/25/95).
- The paper entitled "Structural freedom, topological disorder, and the irradiation-induced amorphization of ceramic structures" by Esther Jesurum and Berger, in conjunction with Prof. Linn Hobbs and A.N. Sreeram of MIT's Material Science and Engineering, Dept., received a Best Paper award at the 8th International Conference on Radiation Effects in Insulators, Sicily, Italy.
- Russell Schwartz's M.Eng. thesis entitled "A multi-threaded simulator for the kinetics of virus shell assembly" received the MIT-EECS William A. Martin Memorial Prize for Best M.Eng. Thesis in Computer Science.
- Berger was selected as one of three new members to the board of the Program in Mathematics and Molecular Biology.
- Berger is the guest editor of the *Algorithmica* Special Issue on Computational Biology.
- Berger is on the program committee for the Eighth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA) and the 1st Annual International Conference on Computational Molecular Biology (RECOMB 97). She is publications chair of RECOMB 97 as well.

- Berger is a Plenary Speaker at the 1996 SIAM Conference in Discrete Math, where she will be speaking on "Mathematical Challenges in Protein Folding".
- Mona Singh received her PhD in EECS in September of 1995 and is now a DIMACS postdoc at Princeton.
- Postdoc Scott Decatur was awarded a DIMACS postdoc in computational molecular biology for September, 1996.
- The MIT ACM Programming team, which Berger, Leighton, and Kaashoek coach, took first place in the Northeast Regional Competition and fifth place in the International Competition, which awarded the team a $1500 prize.

3 Summary of Research Progress

The assembly pathway of viruses represents an interesting example of a self-assembling structure and an attractive yet underexplored 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] 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.

3.1 Further Development of Local Rules

Berger and Dr. Peter Shor of AT&T Bell Labs [3] have continued to work on the development of local rule theories for virus shell assembly and misassembly. They have theorized that the local rules may provide a possible switching mechanism between one size shell or another, suggesting an evolutionary pathway between different size virus shells [3]. In particular, they have given alternate local rules that assemble a $T = 7$ shell such that a small change in these rules produces a $T = 4$ shell. The theory relies on particular interactions of the shell and scaffolding proteins, which aid in the formation of virus structure. The presumed placement of the scaffolding proteins has since been shown experimentally to be correct by biologists for both the $T = 4$ and $T = 7$ shells. These local rules also predict that hexamers in the assembled procapsid would have approximate two-fold rotational symmetry. This symmetry is exemplified by the elongation of hexamers observed in many $T = 7$ viruses. This work was presented at the 14th Biennial Conference on Phage/Virus Assembly in the Summer of 1995.

Berger and Shor are working out the details of a similar switching mechanism for hepatitis virus, which is known to produce $T=4$ and $T=3$ shells.

They are also able to explain for the first time several anomalies that have been observed in nature within the context of the local rules. These include spiral and tubular variants of the icosahedral shell as well as polyomaviruses such as SV40, which is believed to cause cancer and whose structure has puzzled researchers since it is asymmetric and violates quasi-equivalence theory. Moreover, they used the simple simulator discussed below to implement local rules that interrupt

the formation process of an icosahedral shell, thereby resulting in these anomalies. Indeed, by adding just one "poisoned protein" subunit to a growing shell, assembly could be disrupted.

3.2 Simulating Virus Shell Assembly

To verify various local rule theories of assembly and misassembly, a simple simulator of the assembly process was first developed by Berger and undergraduate student Doug Muir. In fact, Muir received an MIT undergraduate thesis award for this work.

This earlier simulator worked by representing coat proteins abstractly as spherical masses connected by springs. With this simulator, a user could define a set of rules which could generate a shell. The simulator would begin with a single protein and add new proteins one at a time, relaxing the stresses in the shell after each addition until the energy of the inter-protein bonds dropped below a fixed tolerance. This simulator was very successful in producing easily understood output for a variety of shells, including some that cannot be modeled by the quasi-equivalence theory. In addition, it made it possible to model some kinds of malformations and to test the tolerance of shells to random perturbations of bond angles.

Working with undergraduate student Russell Schwartz over the summer of 1994, Berger adapted the code for the simple simulator to run on a parallel CM-5 supercomputer. This substantially sped up the algorithm, thereby allowing them to test many more variations within a reasonable amount of time. A virus shell picture they produced appeared in the Supercomputing '94 calendar.

Despite the simple simulators successes, the basic model it employed still made it unsuitable for gathering various kinds of information. In particular, the simple simulator could model local rules but could not provide information on how they might be physically implemented, such as how important the shape of a subunit is, how subunits bind to each other, or how strong inter-subunit binding interactions must be. The simulator could not provide some potentially important quantitative data, such as that concerning reaction rates and pathways. Such information could be significant in finding likely avenues for attacking shell growth. Moreover, while the simulator could suggest strategies for disrupting shell growth, it could not provide specifics; it could indicate that poisoned subunits might be successful in attacking a shell, but could not say what physical properties such a subunit would need, or what concentration would be required in a cell in order to reliably inhibit shell growth. These are important questions in evaluating the biological feasibility of such techniques. The simulator also could not adequately model the ability of proteins to detach after attaching to a shell, or the effect of different models of binding kinetics on the order of assembly. Finally, it could not realistically model subunit conformational switching, which appears to be crucial to understanding and simulating shell growth; this is particularly problematic in modeling alternate rule sets, for which conformational switching is proposed to occur after a protein has attached to the shell.

This year Berger and Schwartz (now a graduate student), in conjunction with Prof. Prevelige of U. of Alabama-Birmingham Medical School and Dr. Shor, continued work on simulating the growth of virus shells, and have begun development of a second-generation simulator, which is tailored to examining the specific limitations of the existing simulator. The new simulator will provide a more realistic physical model of viral coat protein interactions. They have started to incorporate several features that make it both powerful and easy to use, including a versatile model of the underlying

physical systems, a high-level control language, a graphical user interface, and a multi-threaded design that allows the use of parallel architectures. Preliminary work with the simulator suggests that it will be a valuable tool for understanding icosahedral virus shell assembly [5].

This work was presented (by Schwartz) at the 1996 annual meeting of the *Biophysical Society* [2]. Schwartz's M.Eng. thesis on this topic [5], entitled "A Multi-Threaded Simulator for the Kinetics of Virus Shell Assembly," received the MIT-EECS William A. Martin Memorial Prize for best M.Eng. thesis in computer science. Schwartz also presented the work at the EECS Masterworks Colloquim. Schwartz will spend the summer working at the virology lab of Prevelige on experimental work related to evaluating potential inhibitors to viral shell assembly.

3.3 Collaborative Work

Berger and Prevelige have collaborated over the past two years in the area of virus assembly. This collaboration, while limited in scope, has proven to be extremely productive.

Prevelige's lab has determined that bacteriophage P22 scaffolding protein, a protein required to assemble shells, consistently formed tetramers in solution, but this fact was highly puzzling because the final capsid structure has 5-3-2 symmetry. Using the local rules theory, Berger was able to demonstrate that exactly this configuration of the scaffolding protein would be necessary to support one of the more complicated local rules theories [3]. The local rules model developed by Berger and Shor predicted the location of the scaffolding protein in the growing shell. Prevelige and collaborators used cryo-electron microscopy and image analysis to determine the location of the scaffolding protein in the shell and found it was located in the precise positions predicted by the local rule theory. The structural data obtained provided strong evidence for a) the predictive value of the local rules theory, and b) the use of a particular set of rules in the case of bacteriophage P22.

The set of local rules that are consistent with the existence of scaffolding protein tetramers predicts that assembly should initiate from a pentamer of coat protein. Prevelige et al. have experimentally determined that initiation of procapsid assembly does involve a pentamer of coat protein.

A frequent misassembly structure is the spiral form. Spiral forms are seen in viruses ranging from bacteriophage to HIV. Local rules theory predicts that in at least some cases these spiral forms arise from nucleation via hexamers, rather than pentamers. Based upon these results, Prevelige et al. are in the process of experimentally determining the size of the nucleus that leads to spiral formation.

Berger and Prevelige will continue their collaboration on the second-generation simulator.

3.4 Mechanics of Virus Shell Assembly

This past year, graduate student Homsy and Berger have been working to develop a theory of the underlying mechanics of local rule-based virus shell assembly. They started by using rotationally asymmetric hamiltonians to calculate the bond energies, and found that they could build some of the smaller icosahedral shells with only a single type of subunit. They can also *almost* assemble

some of the medium size shells with this method; however, in this context the assembly is not nearly tolerant enough to parameter variation and assembly order.

Their current work therefore attempts to introduce stochastic variation into the simulation in such a way as to accurately model the statistical mechanics of the underlying system, while retaining all or most of the computational efficiency gained by using the simple "ball and stick" model. In this way they hope to be able to build much larger shells reliably and robustly, with only a single subunit type.

3.5 On The Structure of Bacteriophage T4

Berger and Shor have previously postulated that the currently accepted structure for the scaffolding core of bacteriophage T4, one of the earliest-discovered and most-studied viruses, may be incorrect. In the currently accepted structure for the scaffolding core in bacteriophage T4, there is a symmetry mismatch between the protein shell, which has five-fold symmetry, and the scaffolding core, which is believed to consist of six helical chains. Berger and Shor alternatively postulated that a ten-helix core model is particularly attractive because it suggests a Vernier mechanism which is able to explain the process of length determination in giant head mutants of T4. If this Vernier mechanism could be confirmed experimentally, this would be the first biological evidence of a Vernier mechanism in morphogenesis.

Berger and undergraduate student Hoest have conducted experiments on Paulson's original T4 data at Brandeis' Rosenstiel Basic Medical Research Center. The analysis of T4 giant head data that was used to determine the six helical chains made an implicit assumption about the manner in which giant heads flatten during the preparation for electron microscopy, but reexamination of the experimental data shows that this assumption may be incorrect. Reanalysis of the data shows that it could be consistent with six, eight, or ten helical chains. All these results appear in [1]

References

[1] B. Berger, G. W. Hoest, J. R. Paulson, and P. W. Shor. On the structure of the scaffolding core of bacteriophage T4 and its role in head length determination. Technical Report 686, MIT Lab. for Computer Science, January 1996. Submitted for publication.

[2] B. Berger, R. Schwartz, P. W. Shor, and P. E. Prevelige. Kinetic modeling of virus capsid assembly. In *Biophysical J.*, volume 70, Baltimore, MD, February 1996.

[3] B. Berger and P. W. Shor. Local rule switching mechanism for viral shell geometry. In *Proc. 14th Biennial Conference on Phage/Virus Assembly*, June 1995. Full version appears in MIT-LCS TM #527.

[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):7732–7736, aug 1994.

[5] R. Schwartz. A multi-threaded simulator for the kinetics of virus shell assembly. Master of Engineering Thesis. Massachusetts Institute of Technology, 1996. Received the MIT-EECS William A. Martin Memorial Prize for best thesis in computer science.

NSF Career Award Progress Report '95–'96

Bonnie Berger

1 Introduction

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. She has also worked on applications of these techniques to the material sciences.

2 Notable Events

The following notable events have occurred.

- The virus shell assembly work was featured in a *Science News* article (3/25/95).
- The paper entitled "Structural freedom, topological disorder, and the irradiation-induced amorphization of ceramic structures" by Esther Jesurum and Berger, in conjunction with Prof. Linn Hobbs and A.N. Sreeram of MIT's Material Science and Engineering, Dept., received a Best Paper award at the 8th International Conference on Radiation Effects in Insulators, Sicily, Italy.
- Russell Schwartz's M.Eng. thesis entitled "A multi-threaded simulator for the kinetics of virus shell assembly" received the MIT-EECS William A. Martin Memorial Prize for Best M.Eng. Thesis in Computer Science.
- Berger was selected as one of three new members to the board of the Program in Mathematics and Molecular Biology.
- Berger is the guest editor of the *Algorithmica* Special Issue on Computational Biology.
- Berger is on the program committee for the Eighth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA) and the 1st Annual International Conference on Computational Molecular Biology (RECOMB 97). She is publications chair of RECOMB 97 as well.
- Berger is a Plenary Speaker at the 1996 SIAM Conference in Discrete Math, where she will be speaking on "Mathematical Challenges in Protein Folding".
- Mona Singh received her PhD in EECS in September of 1995 and is now a DIMACS postdoc at Princeton.

- Scott Decatur was awarded a DIMACS postdoc in computational molecular biology for September, 1996.
- The MIT ACM Programming team, which Berger, Leighton, and Kaashoek coach, took first place in the Northeast Regional Competition and fifth place in the International Competition, which awarded the team a $1500 prize.

3 Summary of Research Progress

Berger and her students are working on the following problems.

3.1 Motif Recognition

The motif recognition problem is to predict how a protein will fold in three dimensions when we only have access to its one-dimensional amino acid sequence. Biologists are interested in this problem since the structure or fold of a protein provides the key to understanding its biological function. Unfortunately, determining the three-dimensional fold of a protein is very difficult. Experimental approaches such as NMR and X-ray crystallography are expensive and time-consuming (they can take more than a year), and often do not work at all. Therefore, computational techniques that predict protein structure based on one-dimensional sequence data can help speed up the understanding of proteins. Berger had previously developed a statistical-based pairwise correlation method for protein motif recognition and successfully applied it to the coiled-coil motif [1].

3.1.1 Recognizing 2-stranded Coiled Coils

Berger, graduate students David Wilson and Ethan Wolf, and undergraduate student Tonchev have continued work in conjunction with Prof. Peter S. Kim of the MIT Biology Dept., Whitehead and Howard Hughes Medical Institutes on predicting the (2-stranded) coiled-coil motif in sequence data [8]. The `PairCoil` program derived from this work is now available on the Web (http://theory.lcs.mit.edu/ bab/paircoil) through a Web interface and ftp. Already since January, the website has had over 1000 hits for running or ftp'ing the program.

3.1.2 Recognizing 3-stranded Coiled Coils

Berger and Wolf, working in conjunction with Prof. Kim, have extended the `PairCoil` program from the domain of 2-stranded coiled coils to that of 3-stranded coiled coils in the program `MultiCoil` [11]. This program uses a new multi-dimensional scoring approach to identify and distinguish 3-stranded and 2-stranded coiled coils. Creating the `MultiCoil` program also involved the construction of a new 3-stranded coiled-coil database. The methods used to construct and test the program also gave insight into a number of interesting biological features of coiled coils, including a statistically justifiable estimate for the fraction of all protein residues that form 3-stranded coiled coils and the fraction that form 2-stranded coiled coils. The `MultiCoil` program is currently being readied for release, and we expect it to be widely used by biologists who are currently using `Paircoil`.

3.1.3 Recognizing the Beta-helix Motif

Berger, postdoc Dr. Scott Decatur, and undergraduate student Brian Dean, working in conjunction with Prof. Jonathan King of the MIT Biology Dept., are developing a method and program for recognizing beta-helices from sequence data. Currently, beta-helix motifs are little understood; to date there are few proteins whose structure has been confirmed to contain a beta-helix, and biologists do not yet know what causes this motif to arise.

They have formed a database of beta-helix structures and annotated it with topological features which they have determined to be relevant to the motif formation. These features are computed from the 3D solved structures in the Brookhaven X-ray crystallography Protein Database (PDB). They have extended the pairwise correlation framework from the simple domain of coiled coils, whose local spatial interactions are also close together in sequence, to accommodate the spatial interactions of distant residues in sequence that are found in the more complicated beta-helix. They also developed new dynamic programming algorithms which allow for efficient scoring of candidate proteins. Through this analysis, they have identified structural features which will aid both their program as well as biologists in their understanding of these structures. They will be working to further improve the recognition capability of their programs and hope to make available a Web interface analogous to the `PairCoil` interface.

3.1.4 Learning Protein Motifs

A key feature of recognizing protein motifs is the prior knowledge of relevant structures. This prior knowledge has come from biologists who have extensively studied many examples of the motifs. However, there are many cases where previously known methods fail because there exists only limited data on the motif.

Berger and Dr. Mona Singh (who is now a DIMACS postdoc at Princeton) have devised a learning algorithm that improves existing methods for recognizing protein structural motifs [6], especially in the case where large numbers of examples of the motif are not known. The algorithm is an iterative method that exploits randomness and statistical techniques to obtain good performance. They have implemented the algorithm and demonstrate its performance on the coiled-coil motif. They have tested the program `Learn-Coil` on the domain of 3-stranded coiled coils and subclasses of 2-stranded coiled coils. They show empirically that for these motifs, their method overcomes the problem of limited data. The program has been used to identify potential 3-stranded coiled-coils in mouse hepatitis virus, human rotavirus, human T-cell lymphotropic virus, and 14-3-3 proteins. The program has also identified a new coiled-coil region in HIV and SIV which no previous program has found. The Kim lab has independently shown through experimental studies that this new coiled-coil region interacts with a previously presumed coiled-coil-like region to form what may be a new motif. The program indicates a similar motif may exist in other retroviruses. This coiled-coil-like region is thought to be the mechanism by which the virus binds to the cell membrane during infection.

Berger and Singh are currently working with with the Kim lab on investigating predictions relevant to the lab's research. With Kim's postdoc Dr. Andrea Cochran, they are exploring applications of `LearnCoil` to a large class of bacterial receptors, and expect to have a biology paper in

release in the near future.

3.1.5 Hidden Markov Models

The hidden Markov model (HMM) approach is considered important in protein motif recognition, but it has not been generalized successfully to statistically take into account spatial interactions between non-consecutive pairs of residues. Moreover, until now pairwise correlations were considered to be ineffective when used in HMMs (Dr. Sean Eddy, personal communication). Even for threading methods, pairwise correlations have not been shown to work better than taking into account only single residues (Temple Smith, personal communication). According to Dr. Eddy it would be interesting to show that pairwise correlations can be used effectively in a hidden Markov models framework.

Berger and Wilson [7] have extended the pairwise correlation framework of Berger to HMMs. They have this year been joined by graduate student Serafim Batzoglou in adapting and implementing this algorithm for the coiled-coil and beta-helix motifs. A program `HMMCoil` has been developed which has no false positives or false negatives when tested on a database of 2-stranded coiled coils and non-coiled-coils from the PDB database of solved structures. Extensive testing of the algorithm and improvements in performance by using standard learning techniques for HMMs are in progress. In preliminary testing, the program appears to perform similarly to `PairCoil`, which alternatively utilizes a window-based algorithm. The program and the effects of learning will also be tested on coiled coils with skip regions, particularly 3-stranded coiled coils that contain many skip regions. Finally, the program will be generalized to apply to the more complicated pairwise correlations in the beta-helix motif.

3.2 Virus Shell Assembly

The assembly pathway of viruses represents an interesting example of a self-assembling structure and an attractive yet underexplored antiviral target. This problem has previously proven experimentally difficult due to the lack of a general framework for approaching the problem. Berger et al. [5] 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.

3.2.1 Further Development of Local Rules

Berger and Dr. Peter Shor of AT&T Bell Labs [4] have continued to work on the development of local rule theories for virus shell assembly and misassembly. They have theorized that the local rules may provide a possible switching mechanism between one size shell or another, suggesting an evolutionary pathway between different size virus shells [4]. In particular, they have given alternate local rules that assemble a $T = 7$ shell such that a small change in these rules produces a $T = 4$ shell. The theory relies on particular interactions of the shell and scaffolding proteins, which aid in the formation of virus structure. The presumed placement of the scaffolding proteins has since been shown experimentally to be correct by biologists for both the $T = 4$ and $T = 7$ shells. These

local rules also predict that hexamers in the assembled procapsid would have approximate two-fold rotational symmetry. This symmetry is exemplified by the elongation of hexamers observed in many $T = 7$ viruses. This work was presented at the 14th Biennial Conference on Phage/Virus Assembly in the Summer of 1995.

Berger and Shor are working out the details of a similar switching mechanism for hepatitis virus, which is known to produce $T=4$ and $T=3$ shells.

They are also able to explain for the first time several anomalies that have been observed in nature within the context of the local rules. These include spiral and tubular variants of the icosahedral shell as well as polyomaviruses such as SV40, which is believed to cause cancer and whose structure has puzzled researchers since it is asymmetric and violates quasi-equivalence theory. Moreover, they used the simple simulator discussed below to implement local rules that interrupt the formation process of an icosahedral shell, thereby resulting in these anomalies. Indeed, by adding just one "poisoned protein" subunit to a growing shell, assembly could be disrupted.

3.2.2 Simulating Virus Shell Assembly

To verify various local rule theories of assembly and misassembly, a simple simulator of the assembly process was first developed by Berger and undergraduate student Doug Muir. In fact, Muir received an MIT undergraduate thesis award for this work.

This earlier simulator worked by representing coat proteins abstractly as spherical masses connected by springs. With this simulator, a user could define a set of rules which could generate a shell. The simulator would begin with a single protein and add new proteins one at a time, relaxing the stresses in the shell after each addition until the energy of the inter-protein bonds dropped below a fixed tolerance. This simulator was very successful in producing easily understood output for a variety of shells, including some that cannot be modeled by the quasi-equivalence theory. In addition, it made it possible to model some kinds of malformations and to test the tolerance of shells to random perturbations of bond angles.

Working with undergraduate student Russell Schwartz over the summer of 1994, Berger adapted the code for the simple simulator to run on a parallel CM-5 supercomputer. This substantially sped up the algorithm, thereby allowing them to test many more variations within a reasonable amount of time. A virus shell picture they produced appeared in the Supercomputing '94 calendar.

Despite the simple simulators successes, the basic model it employed still made it unsuitable for gathering various kinds of information. In particular, the simple simulator could model local rules but could not provide information on how they might be physically implemented, such as how important the shape of a subunit is, how subunits bind to each other, or how strong inter-subunit binding interactions must be. The simulator could not provide some potentially important quantitative data, such as that concerning reaction rates and pathways. Such information could be significant in finding likely avenues for attacking shell growth. Moreover, while the simulator could suggest strategies for disrupting shell growth, it could not provide specifics; it could indicate that poisoned subunits might be successful in attacking a shell, but could not say what physical properties such a subunit would need, or what concentration would be required in a cell in order to reliably inhibit shell growth. These are important questions in evaluating the biological feasibility

of such techniques. The simulator also could not adequately model the ability of proteins to detach after attaching to a shell, or the effect of different models of binding kinetics on the order of assembly. Finally, it could not realistically model subunit conformational switching, which appears to be crucial to understanding and simulating shell growth; this is particularly problematic in modeling alternate rule sets, for which conformational switching is proposed to occur after a protein has attached to the shell.

This year Berger and Schwartz (now a graduate student), in conjunction with Prof. Prevelige of U. of Alabama-Birmingham Medical School and Dr. Shor, continued work on simulating the growth of virus shells, and have begun development of a second-generation simulator, which is tailored to examining the specific limitations of the existing simulator. The new simulator will provide a more realistic physical model of viral coat protein interactions. They have started to incorporate several features that make it both powerful and easy to use, including a versatile model of the underlying physical systems, a high-level control language, a graphical user interface, and a multi-threaded design that allows the use of parallel architectures. Preliminary work with the simulator suggests that it will be a valuable tool for understanding icosahedral virus shell assembly [10].

This work was presented (by Schwartz) at the 1996 annual meeting of the *Biophysical Society* [3]. Schwartz's M.Eng. thesis on this topic [10], entitled "A Multi-Threaded Simulator for the Kinetics of Virus Shell Assembly," received the MIT-EECS William A. Martin Memorial Prize for best M.Eng. thesis in computer science. Schwartz also presented the work at the EECS Masterworks Colloquim. Schwartz will spend the summer working at the virology lab of Prevelige on experimental work related to evaluating potential inhibitors to viral shell assembly.

3.2.3 Collaborative Work

Berger and Prevelige have collaborated over the past two years in the area of virus assembly. This collaboration, while limited in scope, has proven to be extremely productive.

Prevelige's lab has determined that bacteriophage P22 scaffolding protein, a protein required to assemble shells, consistently formed tetramers in solution, but this fact was highly puzzling because the final capsid structure has 5-3-2 symmetry. Using the local rules theory, Berger was able to demonstrate that exactly this configuration of the scaffolding protein would be necessary to support one of the more complicated local rules theories [4]. The local rules model developed by Berger and Shor predicted the location of the scaffolding protein in the growing shell. Prevelige and collaborators used cryo-electron microscopy and image analysis to determine the location of the scaffolding protein in the shell and found it was located in the precise positions predicted by the local rule theory. The structural data obtained provided strong evidence for a) the predictive value of the local rules theory, and b) the use of a particular set of rules in the case of bacteriophage P22.

The set of local rules that are consistent with the existence of scaffolding protein tetramers predicts that assembly should initiate from a pentamer of coat protein. Prevelige et al. have experimentally determined that initiation of procapsid assembly does involve a pentamer of coat protein.

A frequent misassembly structure is the spiral form. Spiral forms are seen in viruses ranging

from bacteriophage to HIV. Local rules theory predicts that in at least some cases these spiral forms arise from nucleation via hexamers, rather than pentamers. Based upon these results, Prevelige et al. are in the process of experimentally determining the size of the nucleus that leads to spiral formation.

Berger and Prevelige will continue their collaboration on the second-generation simulator.

3.2.4 Mechanics of Virus Shell Assembly

This past year, graduate student Homsy and Berger have been working to develop a theory of the underlying mechanics of local rule-based virus shell assembly. They started by using rotationally asymmetric hamiltonians to calculate the bond energies, and found that they could build some of the smaller icosahedral shells with only a single type of subunit. They can also *almost* assemble some of the medium size shells with this method; however, in this context the assembly is not nearly tolerant enough to parameter variation and assembly order.

Their current work therefore attempts to introduce stochastic variation into the simulation in such a way as to accurately model the statistical mechanics of the underlying system, while retaining all or most of the computational efficiency gained by using the simple "ball and stick" model. In this way they hope to be able to build much larger shells reliably and robustly, with only a single subunit type.

3.2.5 On The Structure of Bacteriophage T4

Berger and Shor have previously postulated that the currently accepted structure for the scaffolding core of bacteriophage T4, one of the earliest-discovered and most-studied viruses, may be incorrect. In the currently accepted structure for the scaffolding core in bacteriophage T4, there is a symmetry mismatch between the protein shell, which has five-fold symmetry, and the scaffolding core, which is believed to consist of six helical chains. Berger and Shor alternatively postulated that a ten-helix core model is particularly attractive because it suggests a Vernier mechanism which is able to explain the process of length determination in giant head mutants of T4. If this Vernier mechanism could be confirmed experimentally, this would be the first biological evidence of a Vernier mechanism in morphogenesis.

Berger and undergraduate student Hoest have conducted experiments on Paulson's original T4 data at Brandeis' Rosenstiel Basic Medical Research Center. The analysis of T4 giant head data that was used to determine the six helical chains made an implicit assumption about the manner in which giant heads flatten during the preparation for electron microscopy, but reexamination of the experimental data shows that this assumption may be incorrect. Reanalysis of the data shows that it could be consistent with six, eight, or ten helical chains. All these results appear in [2]

3.3 Modeling the Structure of Materials

Material science researchers are interested in knowing more about the structure of various materials, in particular their amorphizability, i.e., how easily they could change shape. They are interested in inexpensive materials that they would like to use in the design of nuclear reactors and waste

containment, but if the material changes shape too easily in the presence of radiation, then it is not a safe or reasonable option.

3.3.1 Local Rules for Silica Glasses

Silica glass is an inexpensive material that provides a good starting point for analyzing the amorphizability of cystalline structures. Previous research had determined empirically that silicon dioxide is fairly amorphizable. Recently researchers asked if silicon carbide might be a better option, due to its fewer degrees of freedom, but found that it resembles silicon dioxide in amorphizability. Alternatively, silicon nitride is (inexplicably) over 7000 times more stable in practice. Previously, insight into these structures was based on global repeating patterns of polytopes.

Working in conjunction with Prof. Linn Hobbs of the Mat. Sci. and Eng. Dept. at MIT, Berger, graduate student Jesurum, and UROP Pulim hope to gain a better and more formal understanding of the amorphizability of silica glasses by developing local rule-based models of its growth and regrowth. They have focused initially on the modes of tetrahedron connectivity found in silica (two tetrahedra to a vertex), silicon nitride (three tetrahedra to a vertex), and silicon carbide (four tetrahedra to a vertex, but with the possibility of anti-site disorder). They model a silicon atom and its four neighbors as a tetrahedron. The set of local rules governing growth accounts for the number of connections each corner of a tetrahedron seeks (in the case of silicon nitride, for example, each corner seeks to connect to two other tetrahedra for a total of eight neighbors per tetrahedron) and the relative rotation of those neighbors with respect to the orientation of that tetrahedron. With very simple rules sets they are able to successfully grow a variety of crystal forms for these compounds. They now turn to studying the effects of faulty rules sets – that is a set of rules which does not yield a crystal. These come in two basic forms: rules that are close to a rules set for a crystal but with the rotation angles off and rules that have random tolerance ranges for rotations.

The model gives them an easy tool for studying these structures. They can simulate hand modeling but get much more information. Visually, they can rotate and inspect a structure for any kind of repeated patterns through the graphics package they are using (Silicon Gaphics Indigo workstation) or, as in crystallography, they can look for tunnels. Also, they have the entire adjacency matrix at their disposal to analyse the graph properties of these structures including, but not limited to, the ring structure. UROP Frankel has been working on creating a Web interface to the silica glass simulator.

Initial accounts of this approach have appeared at the conference on Experimental and Simulation Challenges in Nanostructural Materials, the American Physical Society March 1996 meeting, and the 1996 American 98th Ceramic Society meeting.

3.3.2 Topological Freedom of Ceramics

Berger and Jesurum, in conjunction with Hobbs [9], have applied ideas from network fault tolerance to explain why ceramic structures such as olivine and spinel amorphize at greatly varied rates when their bonds are randomly broken. Their alternate way of viewing the structures spinel and olivine seems to account for the major discrepancy in their amorphizability. Using the previous Cooper-Gupta measure of total polytope edge, face, and vertex sharing does not, in itself, point to such a

large variance in the amorphizability of these structures. Consider instead the underlying structure of the polyhedra in a different way. If we imagine that each octahedron and each tetrahedron is a vertex and we have an edge between two vertices if their corresponding polyhedra share an edge, then the resulting graph gives us a clue to underlying structure. It so happens that this representation of spinel, while leaving all the tetrahedral atoms isolated, is a strongly connected graph of the octahedral molecules. On the other hand, this representation yields a disconnected graph for olivine. What this implies is that by breaking only corner shared atoms olivine can be disconnected, and not by simply isolating a single atom. In fact, there appears to be entire planes of weakness (corner sharing only) in olivine.

An account of these approaches given at the eighth international conference on Radiation Effects in Insulators was awarded the best paper award in September 1995.

References

[1] B. Berger. Algorithms for protein structural motif recognition. *J. of Computational Biology*, 2:125–138, 1995.

[2] B. Berger, G. W. Hoest, J. R. Paulson, and P. W. Shor. On the structure of the scaffolding core of bacteriophage T4 and its role in head length determination. Technical Report 686, MIT Lab. for Computer Science, January 1996. Submitted for publication.

[3] B. Berger, R. Schwartz, P. W. Shor, and P. E. Prevelige. Kinetic modeling of virus capsid assembly. In *Biophysical J.*, volume 70, Baltimore, MD, February 1996.

[4] B. Berger and P. W. Shor. Local rule switching mechanism for viral shell geometry. In *Proc. 14th Biennial Conference on Phage/Virus Assembly*, June 1995. Full version appears in MIT-LCS TM #527.

[5] 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):7732–7736, aug 1994.

[6] B. Berger and M. Singh. Learning-based algorithms for protein motif recognition. Technical Report TM-539, MIT Lab. for Computer Science, November 1995. Submitted for publication.

[7] B. Berger and D. B. Wilson. Improved algorithms for protein motif recognition. In *Proc. of the Sixth Annual ACM–SIAM Symposium on Discrete Algorithms*, pages 58–67, Philadelphia, January 1995. SIAM.

[8] B. Berger, D. B. Wilson, E. Wolf, T. Tonchev, M. Milla, and P. S. Kim. Predicting coiled coils by use of pairwise residue correlations. *Proc. of the Natl. Academy of Sci., USA*, 92:8259–8263, Aug. 1995.

[9] L. Hobbs, A. Sreeram, C. Jesurum, and B. Berger. Structural freedom, topological disorder, and the irradiation-induced amorphization of ceramic structures. In *8th International Conference on Radiation Effects in Insulators*, Acriale, Sicily, Italy, September 1995. Full version to appear in *Nucl. Instrum. Meth. Phys. Res. B.* Best Paper award.

[10] R. Schwartz. A multi-threaded simulator for the kinetics of virus shell assembly. Master of Engineering Thesis. Massachusetts Institute of Technology, 1996. Received the MIT-EECS William A. Martin Memorial Prize for best thesis in computer science.

[11] E. Wolf, P. S. Kim, and B. Berger. `MultiCoil`: a program for predicting two and three stranded coiled coils, 1996. Unpublished manuscript.