

# Multiscale Computations of Molecular Assemblies

A.D.J Haymet  
S.S. Akhtar  
T. Basak  
B.W. Beck  
J.M. Cabrera  
K. Dyer  
S.C. Gay  
R. Glowinski  
S.L. Johnsson

S. Kim  
G.C. Lynch  
S. Mejia-Rosales  
J.S. Perkyns  
B.M. Pettitt  
T.C. Rogala  
E.R. Smith  
A. Vainrub  
K. Wong

May 2004



**Molecular Science  
Computing Facility**



*This research was performed in part using the Molecular Science Computing Facility (MSCF) in the William R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the U.S. Department of Energy's Office of Biological and Environmental Research and located at the Pacific Northwest National Laboratory. Pacific Northwest is operated for the Department of Energy by Battelle.*

## 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 Battelle Memorial Institute, 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, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY  
*operated by*  
BATTELLE  
*for the*  
UNITED STATES DEPARTMENT OF ENERGY  
*under Contract DE-AC05-76RL01830*

## **Multiscale Computations of Molecular Assemblies**

A.D.J Haymet	S. Kim
S.S. Akhtar	G.C. Lynch
T. Basak	S. Mejia-Rosales
B.W. Beck	J.S. Perkyms
J.M. Cabrera	B.M. Pettitt
K. Dyer	T.C. Rogala
S.C. Gay	E.R. Smith
R. Glowinski	A. Vainrub
S.L. Johnsson	K. Wong

May 2004

Published by Pacific Northwest National Laboratory for the  
Environmental Molecular Sciences Laboratory

## **Abstract**

The goal of this project is to provide a comprehensive computational model, spanning many length-and time-scales, of the characteristics of biological molecular assemblies (such as associating proteins, lipid vesicles, and viruses) when placed in a modern analytical ultracentrifuge. Sedimentation equilibria and velocities for polydisperse and self-assembling systems will be studied. This problem has an explicit coupling of length-scales from the microscopic correlations between molecules in the assembly to the shape of the resulting assembly to the hydrodynamic flow around the irregularly shaped bodies during sedimentation. The results will be tested against experiments and checked for consistency with known thermodynamic and structural data.

## **Research Team Leader and Members**

### Team Leader

A.D.J. Haymet, University of Houston

### Team Member

### Institution

B.M. Pettitt	University of Houston
R. Glowinski	University of Houston
S.L. Johnsson	University of Houston
S.S. Akhtar	University of Houston
K. Dyer	University of Houston
S. Kim	University of Houston
T.C. Rogala	University of Houston
E.R. Smith	University of Houston
T. Basak	University of Houston
B.W. Beck	University of Houston
J.M. Cabrera	University of Houston
S.C. Gay	University of Houston
G.C. Lynch	University of Houston
S. Mejía-Rosales	University of Houston
J.S. Perkyns	University of Houston
A. Vainrub	University of Houston
K. Wong	University of Houston

## **Number of Hours Allocated in the Past 3 Years**

Year 1 – 500,000 hours allocated

Year 2 – 500,000 hours allocated

Year 3 – 500,000 hours allocated.

## **Number of Hours Used in the Past 3 Years**

Year 1 – 200,000 hours used

Year 2 – 500,000 hours used

Year 3 – 500,000 hours used

## Overview of the Past 3 Year's Accomplishments and Activities

### Characteristics of Biological Molecular Assemblies

The major effort of this project was directed at the study and understanding of the characteristics of molecular assemblies. To perform these studies a two-pronged approach was utilized in which computer experiments were conducted on two different types of macromolecular systems. The importance and function of the interfacial environment of aggregates, in particular those of biomacromolecular systems, is still unanswered. The role of the solvent interface in the stability and function of macromolecules is also not fully understood. The studies performed on these systems will provide a large amount of statistical data that will be used to try to comprehend some of these outstanding issues.

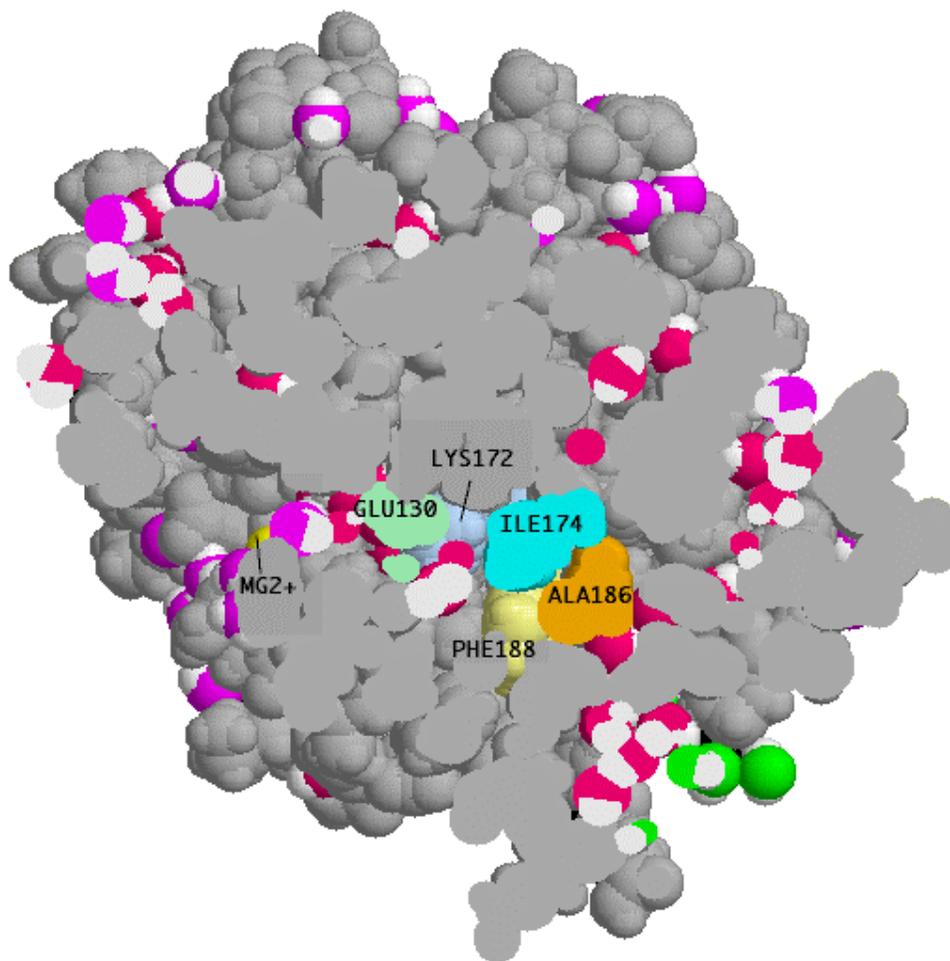
The first system is an enzyme, *Serratia Marcescens* endonuclease, which is found in its naturally occurring state only as a dimer. A mutation of one amino acid can force this system entirely to a stable monomeric state. This system contains a "wet" active site and there is a highly ordered solvent layer in the interfacial region between the two subunits of the dimer. There are very high resolution (0.9 Å) x-ray crystallographic structures for the dimer that include well resolved data for the surrounding solvent and unusual buried solvent molecules associated with the proteins; this also includes the distinct solvent layer between the two subunits of the dimer. A third system in this series is the complex formed when the deoxyribonucleic acid (DNA) is bound to the enzyme. Although there is no x-ray structure for the DNA-bound species, other studies indicate that the complex composed of the protein dimer and the DNA also includes well-ordered water molecules. The hypothesis is that the enzyme's structure and activity are dependent on the internal and interfacial water molecules. But, another unanswered question is why does nature prefer the dimer instead of the monomer?

A systematic study of the monomer, dimer, and complex will be performed to try to resolve the role of the interfacial and buried water molecules. Three molecular dynamics simulations with explicit water molecules have been performed: 1) contains the monomer, 2) contains the dimer, and 3) contains the complex composed of dimer with a double-stranded DNA molecule bonded to one of the active sites. The initial structures of the monomer and dimer species were taken directly from the x-ray crystallography data. Because there is no complex structure, this was built using the A-form DNA duplex d(ATAGGGGCGCCCTAT)<sub>2</sub>. The simulations were all performed with the following: ESP program, NVE ensemble, CHARMM27 force field parameters with TIP3 water, and the temperature set to 300K.

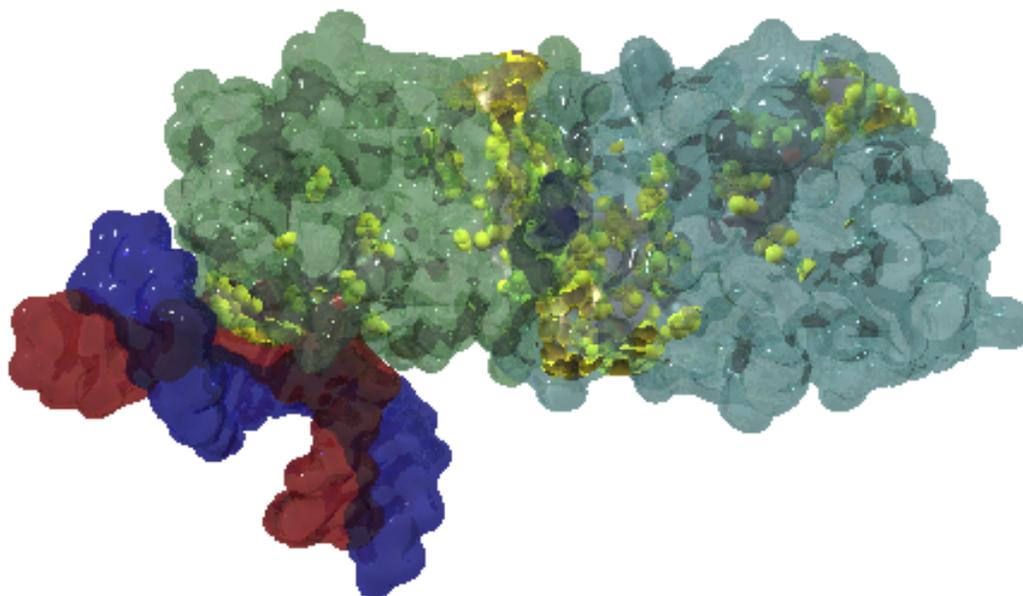
Preliminary results from the monomer simulation indicates that the structure is stable; the root-mean-square-deviation (RMSD) of the monomer versus the x-ray crystallography structure is less than 2.1 Å and the radius of gyration from the molecular dynamics simulation is 16.99 Å, compared to 16.69 Å from the x-ray data. These results provided confidence in both the force field parameters and the chosen technique. This monomer simulation reveals the presence of a solvent filled channel leading from the active site of the enzyme to three distinct openings near the dimer interface. The channel appears to be mediated by a barrier, located about midway between the active site and the final dimer interface near residues Ile174, Ala186, and Phe188, that behaves like a "gate" controlling the diffusion of water between the active site and the dimer

interface. The presence of this “gate” appears to have implications in both stability and regulation of the enzymatic activity; this may also be part of the explanation of the evolution of the dimeric state of this enzyme. In Figure 1, a snapshot at 454 picoseconds of this “gated” water channel for the monomer is illustrated.

The dimer and complex simulations are in the preliminary stages. The simulations are continuing to produce more statistical data. Initial results from both simulations confirm the solvent channel discovered in the monomer simulation. In Figure 2, the average solvent channel for the complex is displayed. This average structure was determined from 285 picoseconds of data and reproduces the large solvated cavity leading to three openings near the interface as initially seen in the monomer simulations.



**Figure 1.** Snapshot of “gated” water channel from monomer simulation at 454 picoseconds.



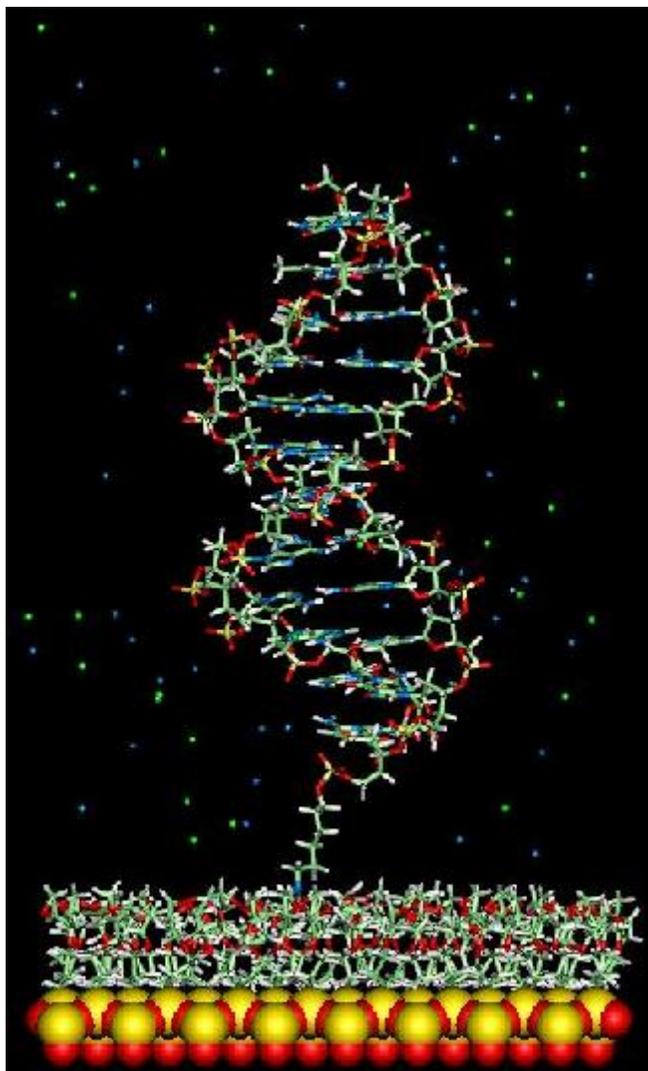
**Figure 2.** Average solvent density over 285 picoseconds in the channel region of the protein-DNA complex.

These simulations continue to try to elucidate the role of the solvent channel in both the stability and activity of this system. The current set of calculations is being performed in the semi-grand canonical ensemble. The semi-grand method allows the control of the solvent density through the chemical potential. This technique, which has been successfully applied to bovine pancreatic trypsin inhibitor (BPTI) simulations, is being used to try to understand the necessity of the water in the channel region. Those experiments are utilizing a biasing algorithm to study the increase and decrease of the water density in the channel region.

The second system in this study involves DNA microarrays; these are powerful devices used in areas like genetics, drug discovery, etc. Although these devices have exploded onto the biotechnology arena, little is understood about their underlying physics and chemistry thereby making the design and optimization processes hit or miss.

To begin the investigations of the complicated interactions of DNA microarrays, a computer experiment that mimics the DNA on the surface had to be designed. The first obstacle in this design was a physical representation of the microarray in a molecular dynamics simulation but within a computationally efficient manner; i.e., the number of particles in the system could not be so large that the calculation was no longer feasible. The solution to this problem was the first success of this project and resulted in the design of a new molecular dynamics boundary condition called the glide-plane boundary condition (GBC). This new boundary condition uses the lead space group and reduces the number of particles necessary to represent the system by one-half compared to the traditionally used periodic boundary condition (PBC). The test case for this new boundary condition was a system of liquid water interacting with a hydroxylated silica surfaces.

The first all-atom DNA microarray molecular dynamics simulation utilized the glide-plane boundary conditions and consisted of a 12-base-pair DNA duplex tethered to a silica surface coated with an epoxide monolayer. This configuration mimics a microarray with a surface density of 0.07 DNA/nm<sup>2</sup>. The simulation was run for 7 nanoseconds so as to collect a large quantity of data for statistical analysis of the complicated interactions between the oligonucleotide, the solvent particles and the surface. The results from this simulation showed that the DNA relaxed from the initial canonical B conformation (see Figure 3) and fluctuated between the A and B forms. However, despite these fluctuations the overall average DNA structure resembled the B form.

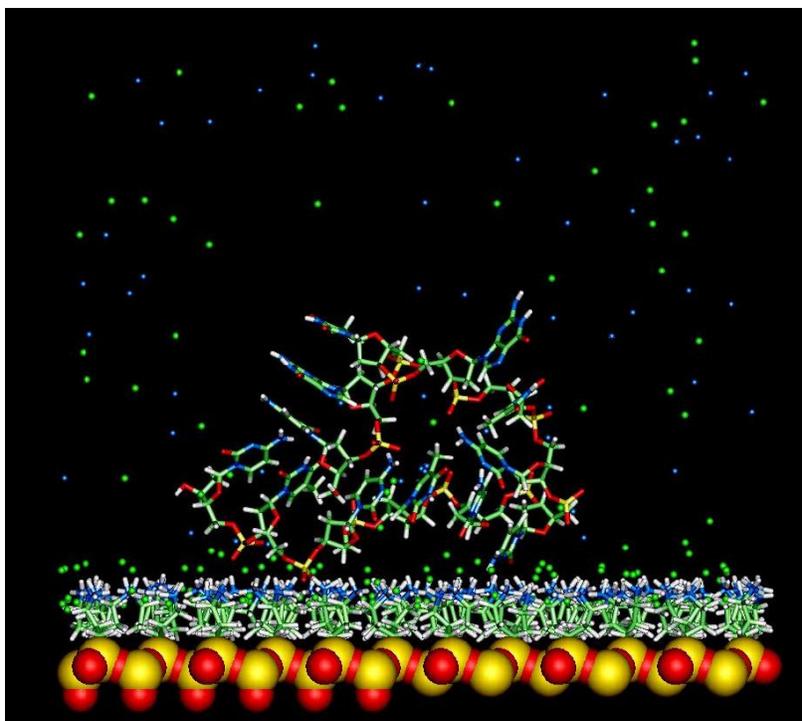


**Figure 3.** Snapshot of the simulation at 0.5 ns. The DNA is in **B**-form, standing nearly perpendicular to the surface. The amine linker is in an extended conformation, pointing upwards. The silica layer is represented as a van der Waals spheres model, the epoxides, amine linker and the DNA as a liquorice model, and the sodium and chloride ions as blue and green spheres, respectively.

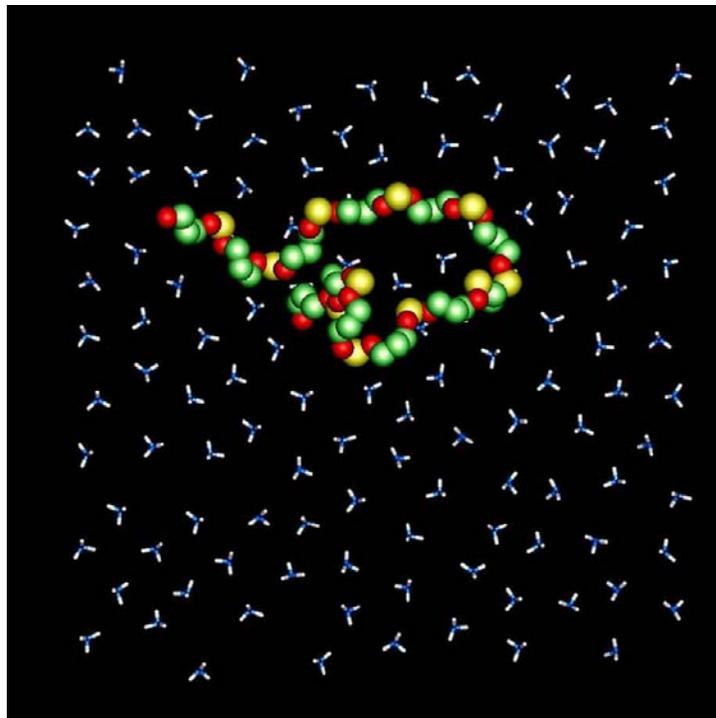
The second phase of these studies involved the specificity and affinity of DNA associated or hybridized to a target oligonucleotide molecule near a surface. These investigations are a collaborative effort using experimental data, an analytical model, and molecular dynamics simulations. The recent experimental studies indicate that there is a kinetically rapid hybridization between a target DNA fragment, usually large, and oligonucleotides electrostatically immobilized to a surface. An analytic solution to the linear Poisson-Boltzmann theory of the electric double layer interaction between DNA and a hard surface predicts tight binding in this system. Molecular simulations were performed for a modified silicon dioxide surface with positively charged groups at neutral pH and an unattached oligonucleotide molecule. The oligonucleotide associated with the surface in salt water in such a way that some of the bases remained stacked and the bases closest to the surface were pointed preferentially toward the solution, away from the surface.

The computer simulation experiments consisted of two independent all-atom molecular dynamics simulations, called **I** and **II**. In both experiments the model consists of a glass surface coated with an ammonium monolayer, a 12-base B-form DNA single strand, and a surrounding solution of 0.8 M NaCl with explicit water molecules. The CHARMM27 all-atom force field parameters are used to describe the interactions. This configuration is equivalent to a DNA coated interface with a surface density of 0.04 DNA nm<sup>-2</sup>, if two-dimensional periodicity parallel to the surface is considered, and approximates half the density of an adsorbed monolayer. The composition of the DNA strands is CGTGTCCCTCTC, which is the same composition used in the earlier study for the tethered duplex experiment. The force field parameters of the silica layer were adopted from the modified CVFF force field, and the ammoniums were from the all-atom CHARMM22 proteins parameters. CHARMM27 was also used for the nucleic acid, salt, and water interactions. The DNA single strands were started with the helical axes parallel to and about 12 and 22 Å above the surface in **I** and **II**, respectively. The empty space of the simulation boxes was filled with water molecules of which 215 were randomly chosen and replaced by 53 sodium and 162 chloride ions. Final numbers of water molecules in the simulations were 3461 and 3452, respectively.

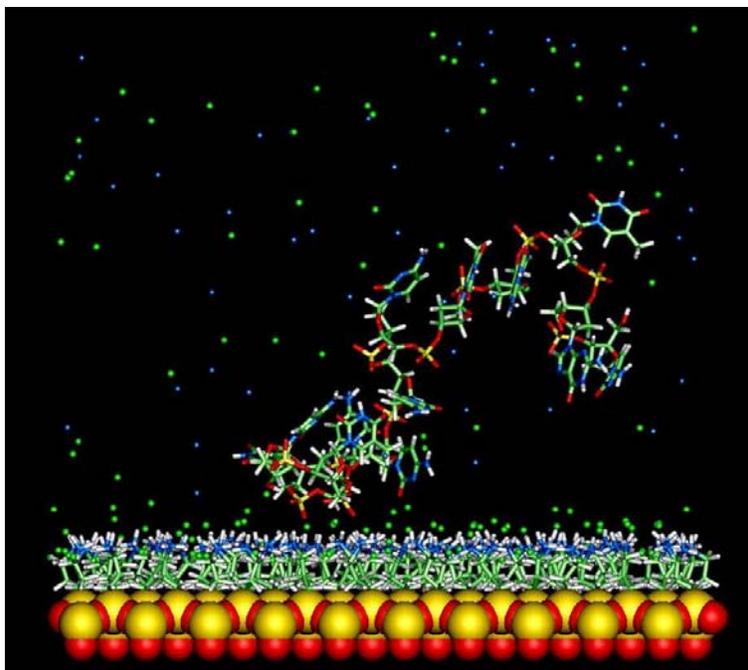
Because only two simulations with different initial conditions were performed, there is not enough data for a complete kinetic description. However, both simulations yield consistent structures and, as such, a plausible and testable hypothesis for the structure of DNA formation that is parallel to the observed binding characteristics of the theoretical model. In **I**, when the DNA approached the surface, adsorption layers of the chloride ions above the ammoniums were only partially formed. The phosphate of T(11), which was closest to the surface, bound tightly to the surface ammoniums. Only chloride ions and almost no water molecules were found between them. One portion of the oligonucleotide, from C(7) to C(12), formed a curve segment and stayed close to the surface. The remaining part of the oligonucleotide bent upwards. Figures 4 and 5 are different views of the snapshot at 9.0 ns for this simulation. Primarily, the DNA was in a triangular shape. In simulation **II**, a substantial surface double layer was formed involving the surface ammonium ions and the free chlorides. The strand of DNA quickly formed a salt bridge with phosphates of C(10) and T(11), which were closest to surface ammoniums. This created a complicated layered structure of ions and water molecules between the phosphates and surface ammoniums. The remaining of the oligonucleotide curved upwards into the solution. Overall, the DNA formed an S shape conformation. Different views of a snapshot at 9.0 ns are shown in Figures 6 and 7. This feature remained for the rest of the simulation.



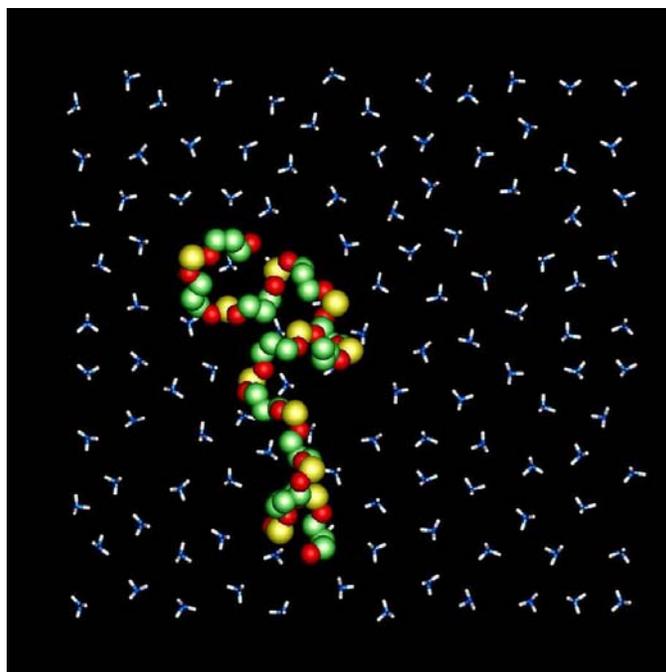
**Figure 4.** A side view of a snapshot of simulation **I** at 9 ns. Note that the phosphate of T(11) is tightly bound to a surface ammonium and the bases are predominantly stacked.



**Figure 5.** A top view, backbone atoms only, of the same snapshot of simulation **I** at 9 ns as illustrated in Figure 4.



**Figure 6.** A side view of a snapshot of simulation **II** at 9 ns. Note that there is a bigger gap between surface ammoniums and the nearest phosphates compared to **I** (Figure 4). The other end of the DNA reached further into the solution.



**Figure 7.** A top view, backbone atoms only, of the snapshot (Figure 6) of simulation **II** at 9 ns.

## Sedimentation and Hydrodynamic Flow

Another aspect of this project involves flow phenomena, in particular sedimentation and hydrodynamic flow. Previous calculations involved systems described by two and three dimensions and produced numerical results for both fluid flow and sedimentation. In all of the earlier calculations, the rigid body was composed of one or two spheres. These calculations have now been extended to a simulation of a two-dimensional system composed of three spheres, called a tripole-like body, in an incompressible viscous fluid.

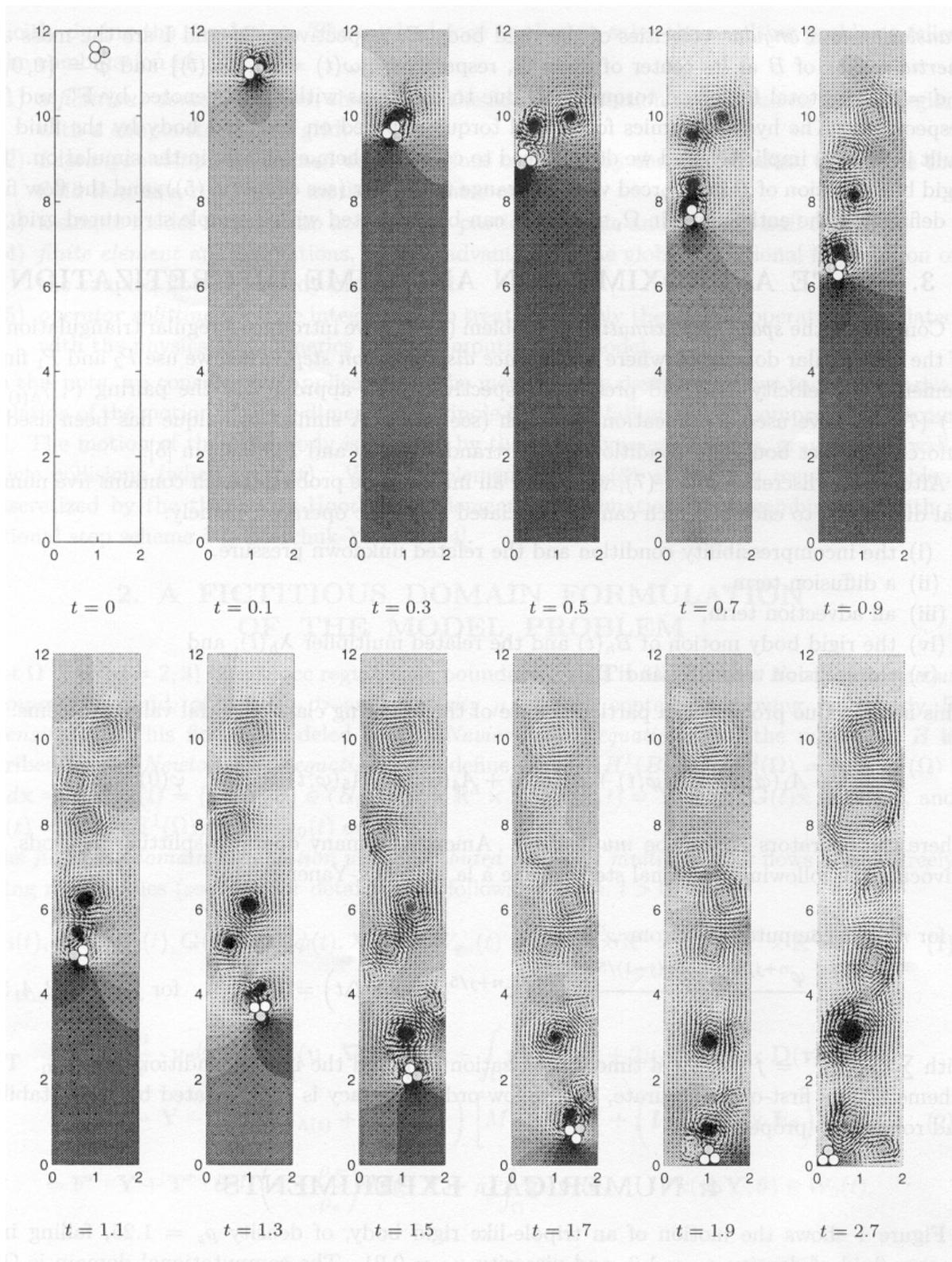
The methodology developed to perform this simulation is composed of several parts that will be briefly described below; full details are described by Juárez et al. (2002). The first part consists of a *fictitious domain formulation with distributed Lagrange multipliers*. This combination allows the flow computation to be limited to a fixed region that can be described by a simple structured grid with the hydrodynamic forces and torque on the rigid body built in. The rigid body motion is then enforced via the Lagrange multiplier. The particle-particle and particle-wall collisions are taken into account via a simple model. A *finite element* approximation for the global variational formulation of the coupled motion of the flow and the rigid body and *operator splitting* for the time integration, have also been utilized.

The results from this calculation are presented in Figure 8. This shows the motion of the tripole-like body, density of 1.25, falling in a viscous fluid of density 1.0 and viscosity 0.01, with a time step of 0.001. The black arrows describe the velocity flow field. The rigid body rotates in a clockwise direction soon after being released; this is believed to be a result of initial orientation. The body also drifts to the left until it “touches” the wall but with an orientation that is similar to the initial orientation. This produces another collision with the same wall but at a different orientation that results in a rotation in the opposite direction. Before colliding with the right wall, the rigid body sediments at the bottom.

## SimDB – Database of Simulation Data

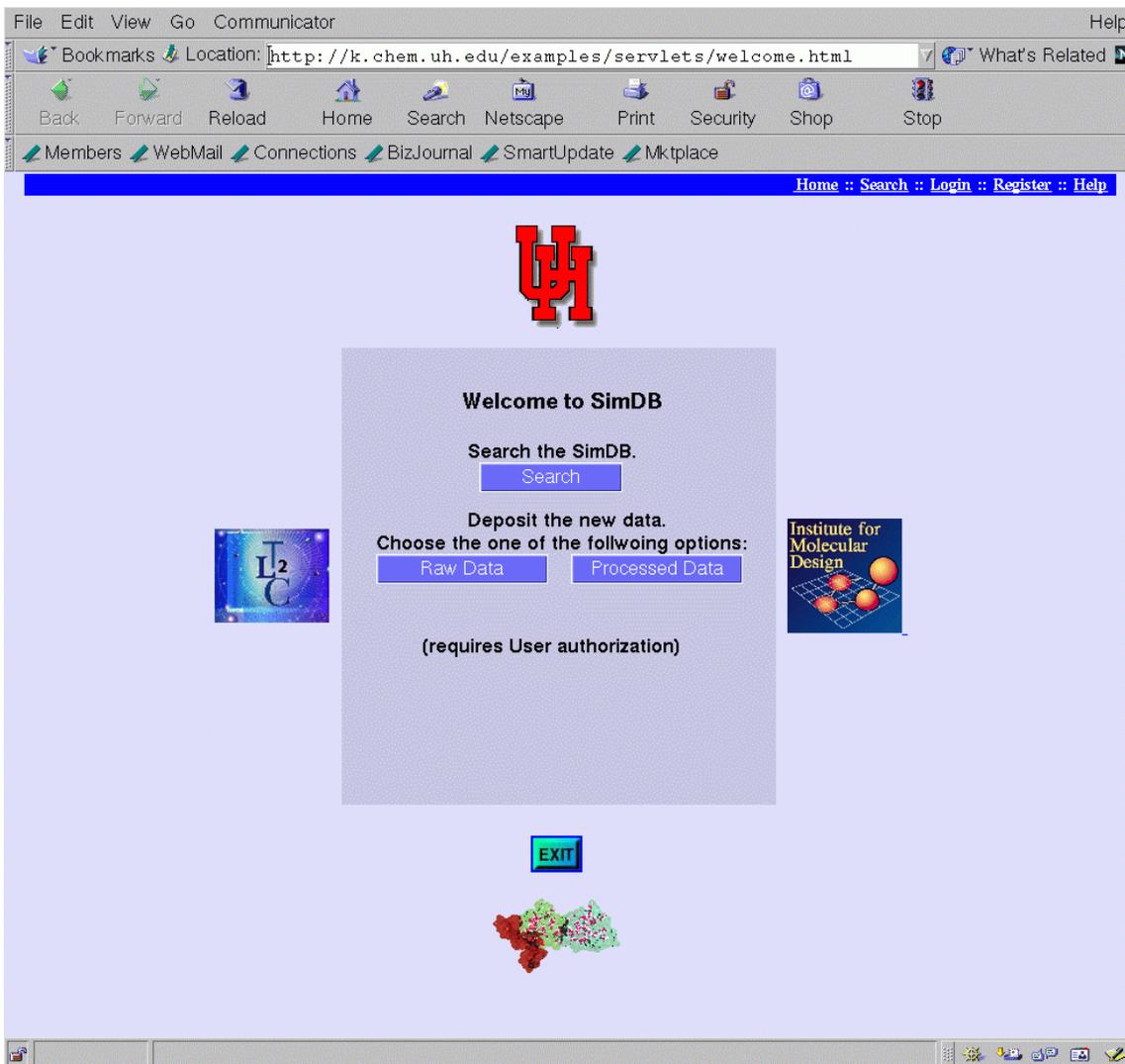
Molecular dynamics simulations generate large quantities of data. This data is usually used by the research group performing the calculations but is unavailable to the general scientific community. A repository of this data would provide a means of sharing this scientific resource that usually requires thousands of computer hours to generate.

*SimDB* has been designed to provide a repository of simulation data for the scientific community. This design is composed of *raw* and *processed* data where *raw* data is the data generated by the modeling programs (i.e., positions, velocities, energies, etc.) and *processed* data is any data not directly generated by the simulation programs. The design includes the storage of data at distributed locations and a web-based graphical user interface for depositing and searching the system. A database, in this case Oracle, has been implemented to handle the details for each entry. Details include location, system type (protein, DNA, ribonucleic acid [RNA], membrane, etc.), force field, authors, journal references, etc.

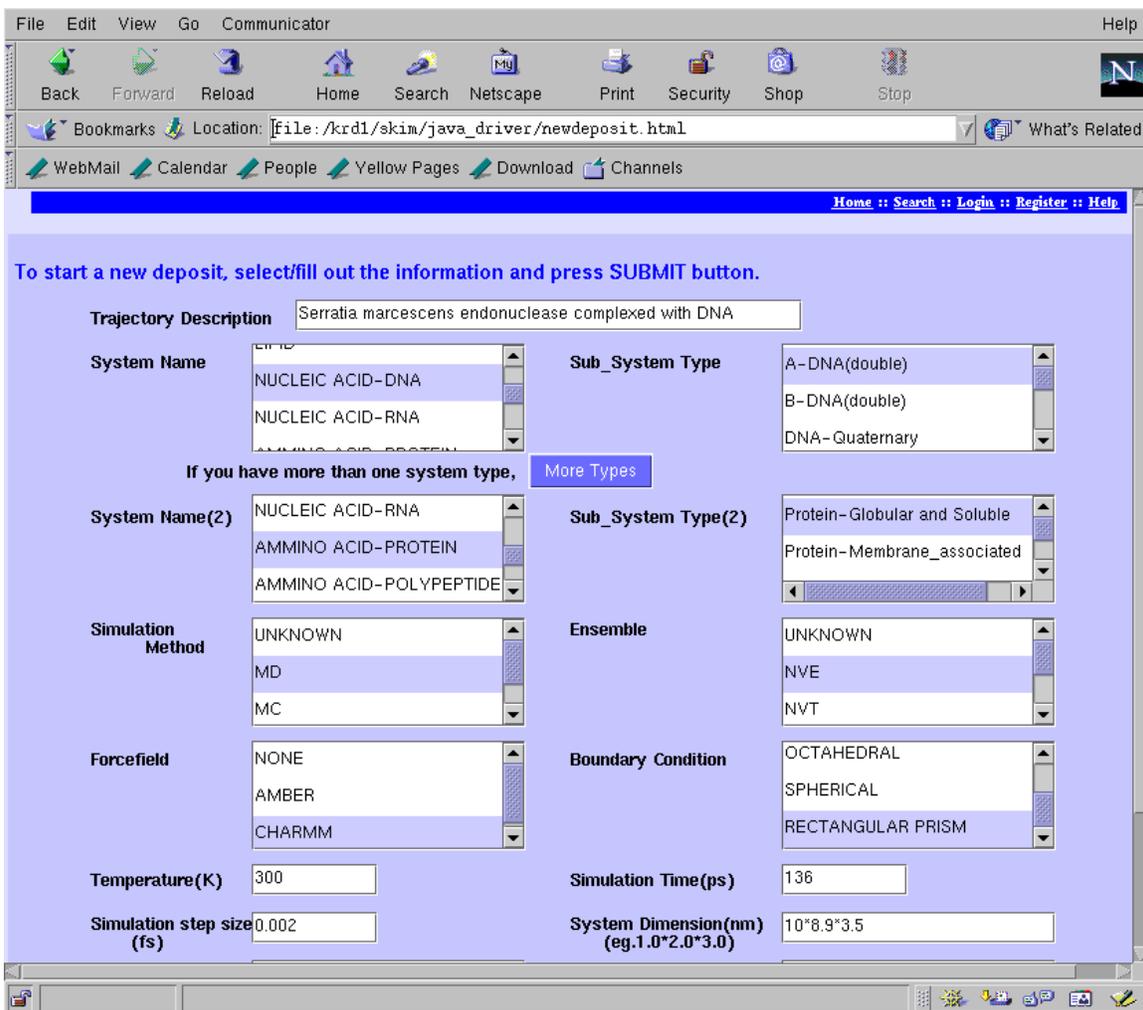


**Figure 8.** A numerical simulation of a tripole-like body falling through a viscous fluid.

The user-friendly web browser interface has been developed and implemented. The database tables for storing the necessary information for each entry have also been designed and configured. This implementation includes the flexibility to easily add new families of data. The Storage Resource Broker (SRB) has been chosen to handle the distributed storage of data. The SRB will also handle the movement of data from one site to another as requested by an authorized user. This is a necessary feature because users may need to perform different types of calculation on data in the system and therefore, would need to access the data directly. The link between the SRB and the *SimDB* database server is in its final stages. Examples of the web-interface are illustrated in Figures 9 and 10.



**Figure 9.** Snapshot of the welcome window for the SimDB project.



**Figure 10.** Snapshot of one of the web-base windows for depositing data into the repository.

## **Publications**

Dyer, KM, JS Perkyins, and BM Pettitt. 2002. "Computationally Useful Bridge Diagram Series: III. Lennard-Jones Mixtures." *J. Chem. Phys.*, 116:9413-9421.

Juárez, LH, R Glowinski, and BM Pettitt. 2002. "Numerical Simulation of the Sedimentation of a Tripole-Like Body in an Incompressible Viscous Fluid." *App. Math. Lett.*, 15:743-747.

Perkyins, JS, K Dyer, and BM Pettitt. 2002. "Computationally Useful Bridge Diagram Series. II. Diagrams in h-Bonds." *J. of Chem. Phys.*, 116:9404-9412.

Vainrub, A and BM Pettitt. 2000. "Thermodynamics of Association to a Molecule Immobilized in an Electric Double Layer." *Chem. Phys. Lett.*, 323:160-166.

Vainrub, A and BM Pettitt. 2003. "Surface Electrostatic Effects in Oligonucleotide Microarrays: Control and Optimization of Binding Thermodynamics." *Biopolymers*, 68:265-270.

Wong K and BM Pettitt. 2001. "A Study of DNA Tethered to a Surface by an All-atom Molecular Dynamics Simulation." *Theor. Chem. Acc.*, 106:233-235.

Wong, K, and BM Pettitt. 2000. "A New Boundary Condition for Computer Simulations of Interfacial Systems." *Chem. Phys. Lett.*, 326:193-198.

Wong, KY, AVainrub, T Powdrill, M Hogan, and BM Pettitt. 2004. "A Non-Watson-Crick Motif of Base-Pairing on Surfaces for Untethered Oligonucleotides." *Molecular Simulation*, 30:121-129.

## **Presentations**

Beck, BW, and BM Pettitt. 2002. "Biomacromolecular Solvation: 'Gated' Solvent-filled Channels in *Serratia marcescens* Endonuclease Mediate Active-site and Dimer-interface Interaction." Gulf Coast Consortium for Bioinformatics and the Keck Center for Computational and Structural Biology Bioinformatics Symposium, October 14-15, 2002, Houston, Texas.

Beck, BW, and BM Pettitt. 2001. "Buried Solvent Channels: Cross-Protein Interactions Between the Active Site and Dimer Interface in *Serratia marcescens* Endonuclease." WM Keck Center for Computational Biology, April 20, 2001, Houston, Texas.

Beck, BW and BM Pettitt. 2000. "Biomacromolecular Solvation: Solvent Mediated Contacts Between the Dimer Interface and the Active Site in *Serratia marcescens* Endonuclease." WM Keck Center for Computational Biology 2000 Symposium, October 16-17, 2000. Houston, Texas.

Kim, S, L Prabu, M Abdullah, BW Beck, M Feig, GC Lynch, L Johnsson, and BM Pettitt. 2002. "Database of Simulation Data: SimDB Design and Implementation." Gulf Coast Consortium for Bioinformatics and the Keck Center for Computational and Structural Biology Bioinformatics Symposium, October 14-15, 2002, Houston, Texas.

Mejía-Rosales, S, GC Lynch, and BM Pettitt. 2002. "Semi-Grand Canonical Molecular Dynamics for Ionic Systems." Gulf Coast Consortium for Bioinformatics and the Keck Center for Computational and Structural Biology Bioinformatics Symposium, October 14-15, 2002, Houston, Texas.

Wong, K and BM Pettitt. 2002. "Molecular Dynamics Simulation of DNA on a Surface." Gulf Coast Consortium for Bioinformatics and the Keck Center for Computational and Structural Biology Bioinformatics Symposium, October 14-15, 2002, Houston, Texas.

### **List of Significant Methods/Routines or Codes Developed**

GBC – glide boundary conditions (GBC) that uses the Pb space group and reduces the number of particles necessary to represent an interfacial system by a half compared to the traditionally used periodic boundary condition (PBC)

ESP – molecular dynamics program that is designed to perform simulations using traditional and extended-system ensembles; it includes the grand canonical ensemble. The glide boundary conditions are also incorporated into this program and the code is portable, parallelized, and optimized for several different platforms including the MCSF's IBM SP<sup>®</sup>.

Viscous Fluid Flow – a method combining distributed Lagrange multiplier based fictitious domain techniques, finite element approximations and operator splitting to the numerical simulation of a rigid body in a Newtonian incompressible viscous fluid.

## **Appendix A - Full Report of First Year Activities and Accomplishments**

### **Multiscale Computations of Molecular Assemblies**

#### **Team Members**

A.D.J. Haymet, Distinguished University Professor, Department of Chemistry and Institute for Molecular Design

B.M. Pettitt, Cullen Distinguished Professor, Department of Chemistry and Institute for Molecular Design

S.S. Akhtar, Department of Chemistry and Institute for Molecular Design

T. Basak, Ph.D., Department of Chemistry and Institute for Molecular Design

B.W. Beck, Ph.D., Department of Chemistry and Institute for Molecular Design

J.M. Cabrera, Ph. D., Department of Chemistry and Institute for Molecular Design

K. Dyer, Department of Chemistry and Institute for Molecular Design

S.C. Gay, Ph. D., Department of Chemistry and Institute for Molecular Design

G.C. Lynch, Ph.D., Department of Chemistry and Institute for Molecular Design

T.C. Rogala, Department of Chemistry and Institute for Molecular Design

E.R. Smith, Department of Chemistry and Institute for Molecular Design

K. Wong, Ph.D., Department of Chemistry and Institute for Molecular D

#### **Number of Hours Allocated, Used, and Requested**

Hours allocated: 500,000

Hours used: 200,000

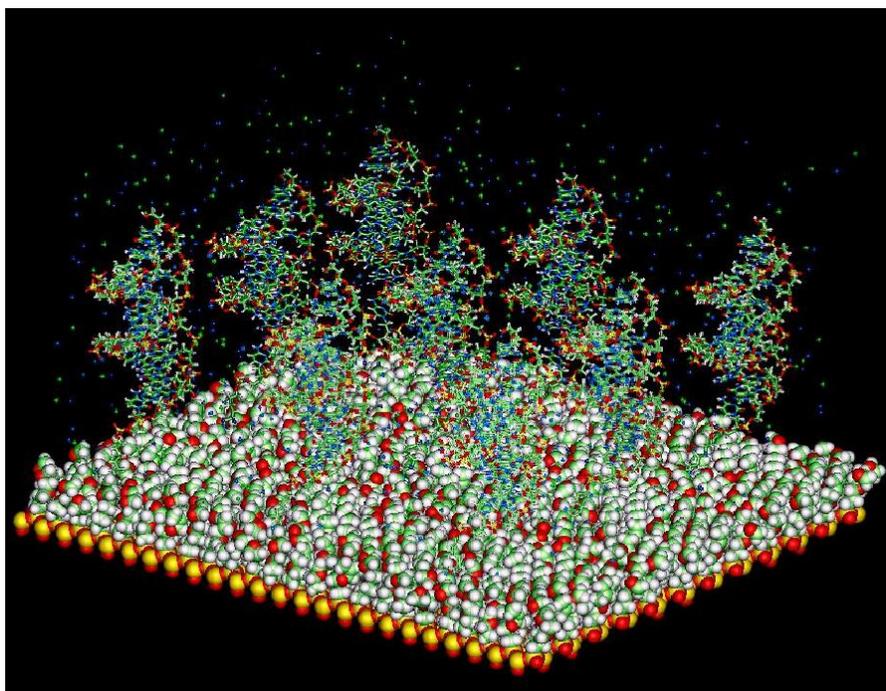
Hours requested: 500,000

**Note:** The small number of hours used was due to security delays in obtaining computer accounts for research members who were not United States nationals.

## Overview of Past Year's Accomplishments

One of the Molecular Science Computing Facility (MSCF) Grand Challenge Teams is developing an extensive multiscale computational model to study the characteristics of biological and nonbiological assemblies. Before this final goal can be achieved, it will be necessary to develop many computational building blocks that can be integrated into the general computational schemes.

One of the first tools to be developed is a new boundary condition for molecular dynamics simulations. When the modeled system is composed of an interface, the generally used PBC doubles the particles necessary to represent the system by replicating the interface. A new boundary condition called the GBC uses the lead space group and reduces the representation to a single interface. This new boundary condition has been successfully applied to a system of liquid water interacting with hydroxylated silica surfaces (Wong and Pettitt, submitted for publication June 2000). To further test the capabilities of the GBC a new all-atom molecular dynamics simulation of a 12-base-pair DNA duplex tethered on a silica surface coated with an epoxide monolayer is being performed (Figure A.1). The Environmental Molecular Sciences Laboratory's (EMSL) MSCF has enabled this simulation, composed of 10 thousand explicit atoms, to be performed. The electrostatic interactions in this simulation are determined via the Ewald sum technique, which is a central processing unit (CPU) intensive, but more accurate, methodology. The molecular dynamics program used for these computations (ESP, University of Houston) has been parallelized and runs efficiently on the MSCF's IBM SP<sup>®</sup> (NWmpp1).



**Figure A.1.** DNA strands on a silica surface coated with an epoxide monolayer. Created by Dr. Ka-Yiu Wong using the program QUANTA through a cooperative agreement between the Institute for Molecular Design and Molecular Simulations, Inc.

Another calculation being performed at the MSCF is one composed of a polymer melt with carbon aggregates. This is another multiscale project that will look at the flow around regular and irregular particles and the motions of non-Newtonian fluids. This project will require the combination of molecular dynamics simulations to describe the motions of the particles and Navier-Stokes equations to model the fluid dynamics. The new GBC boundary conditions will be incorporated into these calculations to manage the aggregate interface.

The computational resources at PNNL allow the simultaneous study of these systems for both short and long-time scales. The DNA simulations are based on the microarray experiments currently being conducted in the labs of their collaborators at Baylor College of Medicine and Genometrix Inc. The carbon simulations are helping to analyze experiments done at Continental Carbon in Houston, Texas. The data from these trajectories will be used to formulate an understanding of the basic chemistry and physics of these aggregate systems and cooperative with the experimentalist will allow the design and testing of new experiments to add to the description. The structural information obtained from these studies will also be used in the following step of this project - a continuum mechanics model.

### **Publications**

Dyer, KM, JS Perkyms, and BM Pettitt. 2002. "Computationally Useful Bridge Diagram Series: III. Lennard-Jones Mixtures." *J. Chem. Phys.*, 116:9413-9421.

Wong K and BM Pettitt. 2001. "A Study of DNA Tethered to a Surface by an All-atom Molecular Dynamics Simulation." *Theor. Chem. Acc.*, 106:233-235.

### **Significant Methods/Routines or Codes Developed**

ESP – molecular dynamics program – parallelized and optimized to run on the MCSF's IBM SP®.

GBC – Uses the lead space group and reduces the number of particles necessary to represent an interfacial system by a half compared to the traditionally used PBC – incorporated into the ESP program

## **Appendix B - Full Report of Second Year Activities and Accomplishments**

### **Multiscale Computations of Molecular Assemblies**

#### **Team Members**

A.D.J. Haymet, Distinguished University Professor, Department of Chemistry and Institute for Molecular Design

B.M. Pettitt, Cullen Distinguished Professor, Department of Chemistry and Institute for Molecular Design

S.S. Akhtar, Department of Chemistry and Institute for Molecular Design

T. Basak, Ph.D., Department of Chemistry and Institute for Molecular Design

B.W. Beck, Ph.D., Department of Chemistry and Institute for Molecular Design

J.M. Cabrera, Ph. D., Department of Chemistry and Institute for Molecular Design

K. Dyer, Department of Chemistry and Institute for Molecular Design

S.C. Gay, Ph. D., Department of Chemistry and Institute for Molecular Design

G.C. Lynch, Ph.D., Department of Chemistry and Institute for Molecular Design

T.C. Rogala, Department of Chemistry and Institute for Molecular Design

E.R. Smith, Department of Chemistry and Institute for Molecular Design

K. Wong, Ph.D., Department of Chemistry and Institute for Molecular D

#### **Sources of Financial Support**

“Theory of Saline Solutions” P.I. Pettitt ‘00-’03 5%

Agency: R.A. Welch Fndn.

Major goals of this proposal are derivations of new theories of the liquid state structure of saline solutions.

“Peptides in Saline Solution” P.I. Pettitt ‘01-’05 15%

Agency: National Institutes of Health (NIH) (this grant)

The major goal of this is to determine the conformations of proteins, peptides and eptidomimetics in saline solution.

“Multiparameter analysis of DNA on surfaces in solution” P.I. Hogan ‘97-’02 15%  
Agency: NIH/Nuclear Control Institute (NCI)

The goals of this project are to model and simulate the properties of single-strand DNA tethered on surfaces for the purpose of hereditarily detection.

“Computational fluid dynamics on particulate flows” P.I. Glowinski ‘98-’02 5%  
Agency: National Science Foundation (NSF)

The main goals of the work are to advance computational algorithms and methods of parallel processing to achieve the ability to routinely perform accurate very-large scale computations on meso scale biomolecular systems.

“Multiscale simulation of Rubber” P.I. Pettitt ‘99-’01 5%  
Agency: Texas ARP

The main goals of the work are to advance multiscale computational algorithms and methods in computations on systems composed of carbon black and isoprene rubbers.

### **Number of Hours Allocated, Used, and Requested**

Hours allocated: 500,000

Hours used: 500,000

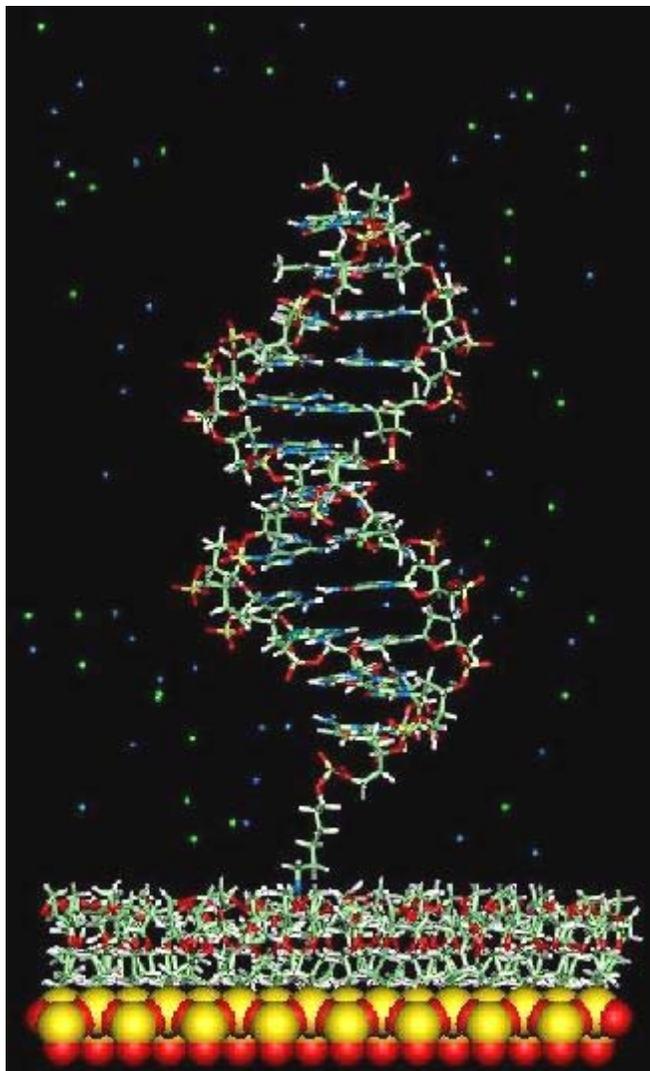
Hours requested: 500,000

### **Overview of Past Year’s Accomplishments**

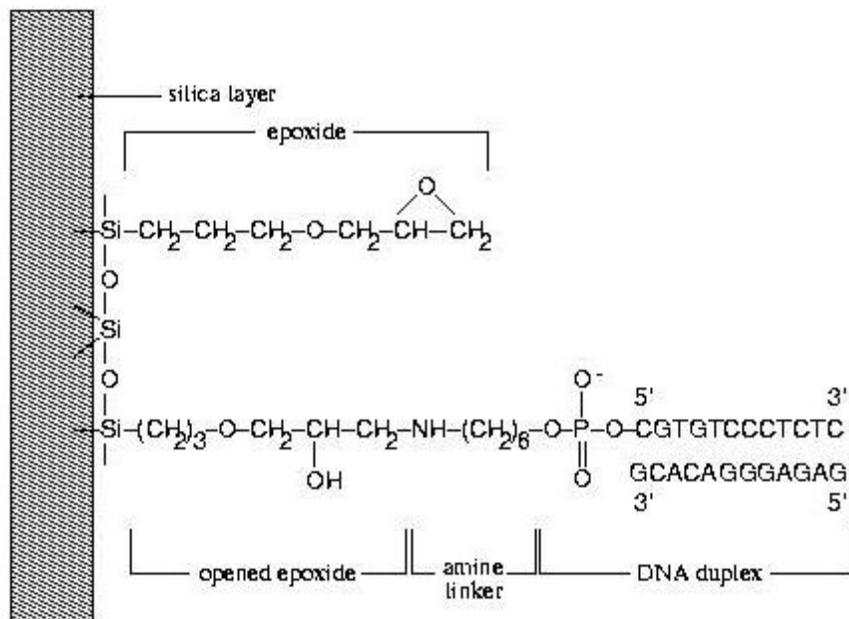
The goal of this project is to design computational tools that would span many length and time scales for both biological and nonbiological assembles. This involves developing techniques that would tie modeling methods up and down the time and system size scales, i.e., methods that would reduce the system size by decreasing the number of explicit atoms and increase the time scale to meso-scale lengths. Beginning at the smallest scale, molecular fluid calculations of several test macromolecular assembles have been performed. These systems contain  $10^4$  atoms and the calculations are performed for several nanoseconds. In addition to conformational fluctuations, thermodynamic and kinetic information can also be collected.

Microarrays are devices composed of oligonucleotides immobilized on surfaces. On one hand, they caught an explosion of interest and became powerful tools in areas like genetics, drug discovery, etc. On the other hand, little is understood about the underlying physics and chemistry in microarrays and that information is crucial to the design and optimization of the devices. To further investigate the complicated interactions in DNA microarrays, particularly at atomic length scales, an all-atom molecular dynamics simulation of a DNA microarray was performed. The composition of this simulation consisted of a 12-base-pair DNA duplex tethered to a silica surface coated with an epoxide monolayer; this mimics a real microarray with a surface density of  $0.07 \text{ DNA/nm}^2$ . The inclusion of a fixed surface into the dynamics simulations was handled using the glide-plane boundary conditions [1] developed last year as part of this project. The simulation was run for 7 nanoseconds, thereby providing a large quantity of data for statistical analysis of the complicated interactions between the oligonucleotide, the solvent particles and the surface at atomic length scales.

The results of the simulation indicated that the DNA relaxed from the initial canonical B conformation, Figure B.1, and fluctuated between the A and B forms. However, its overall average conformation resembled the B form. Even though the amine linker connecting the DNA to the surface epoxide is hydrophobic, it remained mostly in an extended conformation, Figure B.2.



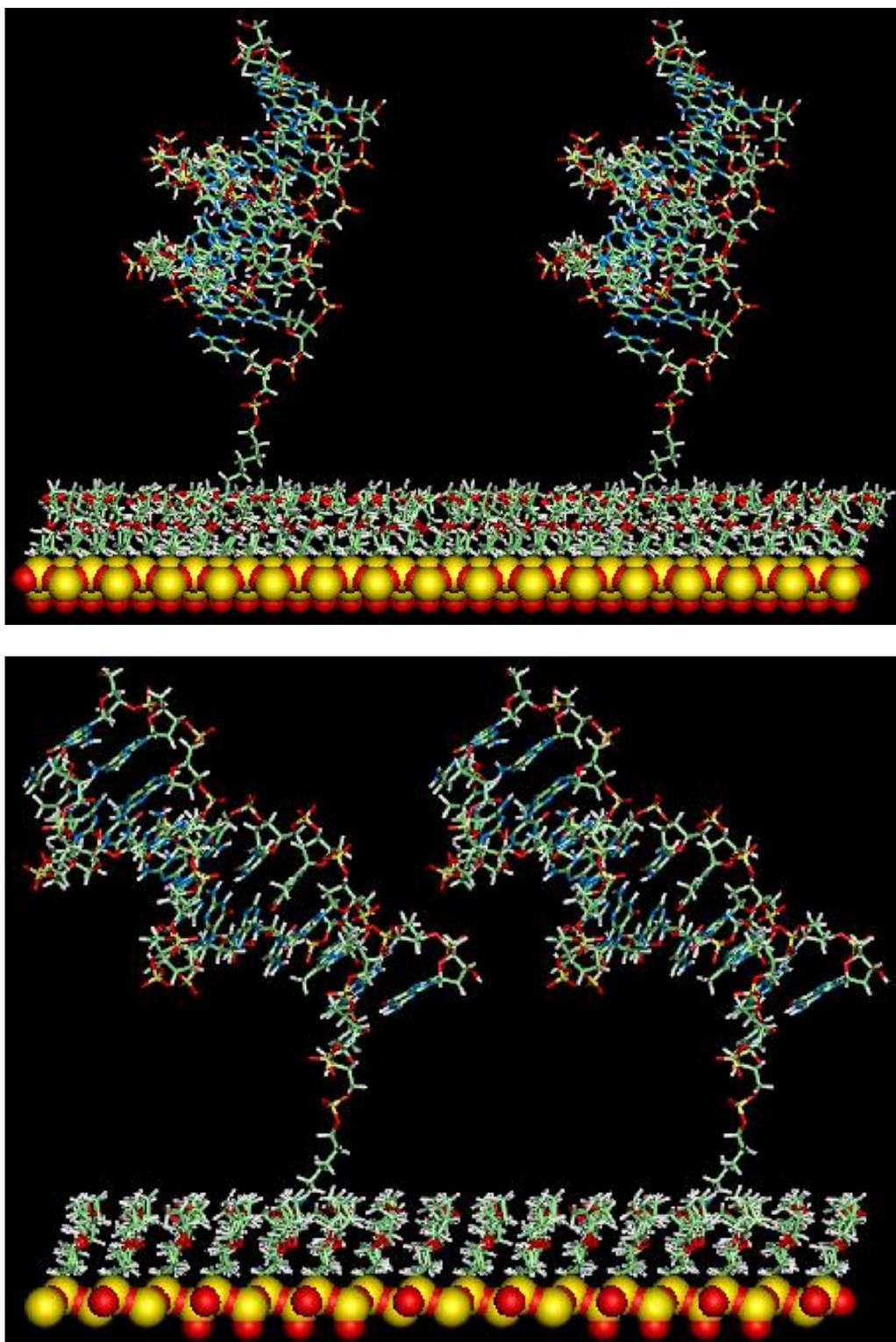
**Figure B.1.** Snapshot of the simulation at 0.5 ns. The DNA is in **B**-form, standing nearly perpendicular to the surface. The amine linker is in an extended conformation, pointing upwards. The silica layer is represented as a van der Waals spheres model, the epoxides, amine linker and the DNA as a liquorice model, and the sodium and chloride ions as blue and green spheres, respectively.



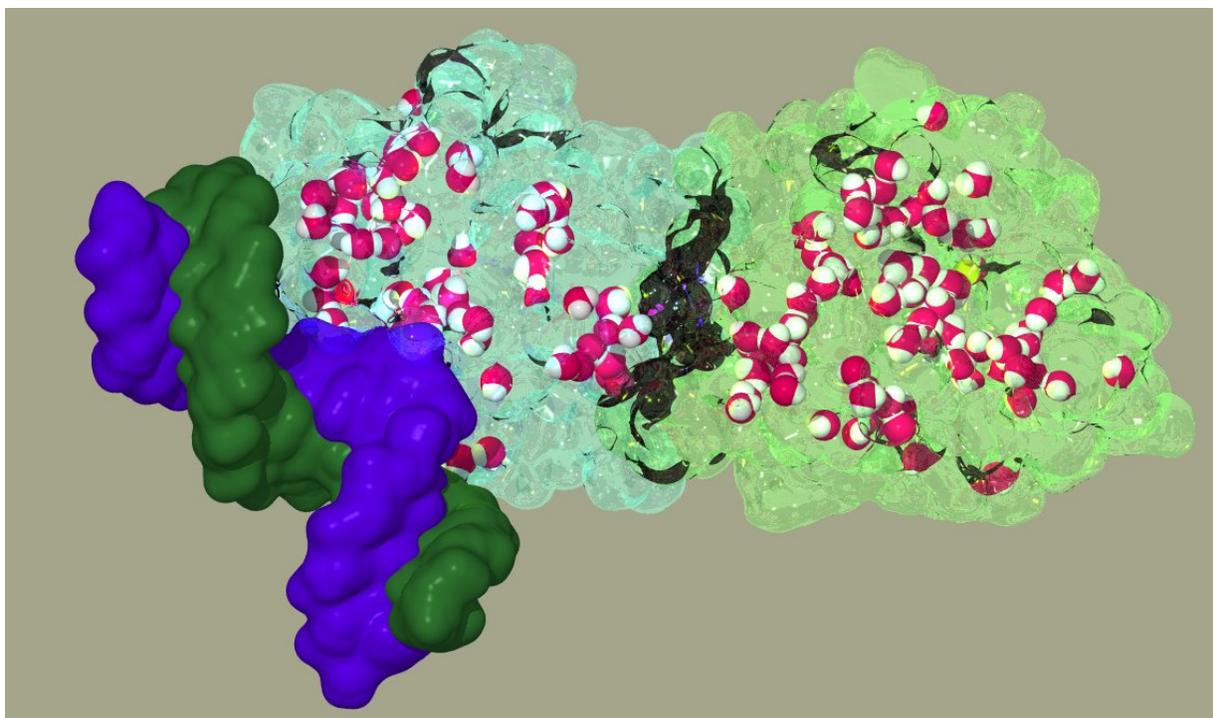
**Figure B.2.** This figure shows the chemical structure of the unreacted epoxides on the silica surface, the opened epoxide that bonds with the amine linker, and the DNA duplex.

Attractive interactions between the DNA duplexes tethered on the surface were also present and the attraction caused the DNA duplexes to settle into a tilted position, with an angle of 55 degrees between the helical axis and the surface normal (Figure B.3). An interaxial distance of about 2.3 nm also suggested that the DNA strands are in a closed-packed phase. This conformational information of the tethered oligonucleotides is important in the design of new DNA microarrays. The results of this calculation have been published by Lynch and Pettitt (2000).

Another aspect of this project is the investigation and understanding of the role of specific molecular species in the formation and stability of macromolecular interfaces. An example of this type of phenomena is the case of solvent particles in the interface region of DNA binding proteins—are they uninvolved bystanders or are they crucial to the process? The infectious bacteria *Serratia marcescens* is an excellent system to use in these investigations; there are crystal structures for both the monomer and the dimer species and there is a definite solvent layer at the dimer interface. Studies have also been carried out on the protein: DNA complex and imply a three species complex of protein, DNA, and solvent. To elucidate the importance of the presence and role of the solvent particles buried in interior channels in this protein and in the interfacial regions of the macromolecules molecular dynamics simulations were performed on a system consisting of both the dimer and the DNA duplex (Figure B.4). Preliminary results indicate that there are distinct solvent channels connecting the dimeric interface to the nucleotide binding site.



**Figure B.3.** Snapshot of the simulation at 6.3 ns **a** and **b** are the views from two perpendicular directions. Each figure displays two DNA duplexes. **(a)** In this view, the DNAs were pointing out of the page. **(b)** This view shows the DNA tilted towards its periodic image. Note the narrow space between the DNAs.



**Figure B.4.** The DNA duplex is rendered as blue and green CPK-sized smoothed “blobs.” The *Serratia marcescens* dimer is also rendered as transparent light green and light blue CPK-sized smoothed blobs. Inside the protein are buried waters found in the channels. Active site MG2+ are rendered as yellow CPK spheres.

Solvent pathways were also discovered between the binding site and different regions of the surface of the protein. Investigation into the function of these pathways is continuing.

One of the goals of this project is to study systems on multiple time scales using information from one time scale to feed the next calculation at a longer time scale. In addition to the utilization of different modeling methods, techniques to improve the speed of the calculations with little or no cost to accuracy are also necessary. One such technique would reduce the number of explicit atoms, usually  $10^6$  to  $10^8$ , by implementing a potential surface environment that would mimic the presence of the particles. It is proposed to develop bridge functions to generate radial distribution functions and thermodynamic quantities which it turn can be used to develop such potential surfaces. Ideally these will be for the solvent mediums of interest such as water and salt water. The first step in this portion of the project was accomplished when calculations necessary to generate the radial distribution function and the corresponding thermodynamic quantities for a binary mixture of Lennard-Jones fluids were completed (Smith and Haymet 1992).

The EMSL’s MCSF Grand Challenge grant provided access to the massively parallel IBM SP2<sup>®</sup> at the Pacific Northwest National Laboratory without which these calculations would not have been feasible in a reasonable amount of time.

## **Proposed Research for the Next Year**

### Multiscale Computations of Biological and Nonbiological Assemblies

Recently, Vainrub and Pettitt (2000) developed a thermodynamic theory to study the hybridization of oligonucleotides near a surface. Using linear Poisson Boltzmann theory of the electric double-layer interaction between an ion-penetrable sphere and a hard plate, they calculated the free energies of hybridization and studied the electrostatic effects on microarrays. This theory will be extended to utilize the data collected from the molecular dynamics calculations instead of the linear Poisson Boltzmann data. The molecular dynamics calculations for the DNA microarray will be extended to provide more statistical data to the thermodynamic theory.

An alternative path to thermodynamic information will be taken with the protein: DNA calculation; semi-grand canonical molecular dynamics (Lynch and Pettitt 2000). The semi-grand canonical molecular dynamics simulation technique uses the excess chemical potential to modify the density of a particular type of particle in a simulation system. This method will be used to remove the water molecules at the interface region. These calculations will provide a direct route to study the system's response to the localized modification of water density.

The modeling of the nonbiological aggregates will continue with the two test systems, N<sub>2</sub> on graphite and C<sub>60</sub> in isoprene. Understanding the deposition onto and/or removal of particles from carbon aggregates is fundamental to the design of new tires. This project, in collaboration with Continental Carbon in Houston, Texas, will provide computational data that will be used to aid in the analysis and interpretation of experiments currently underway.

To begin, molecular dynamics simulations of N<sub>2</sub> adsorbed on graphite at 77 K were performed. The graphite adsorbent area was held fixed at 15.088 nm<sup>2</sup> and the number of adsorbate molecules were varied to span different coverage regimes. The calculations were 300 picoseconds in length. An improvement over the earlier results of Vernov and Steele (1986) was found at the monolayer regime showing a sharper dip in energy with slight change in coverage (Figure B.1). The molecular area calculated for monolayer coverage hovered at 16.2 Å<sup>2</sup>/molecule; this is the same as the value used for the standard B.E.T. (Brunauer, Emmett, and Teller) adsorption experiment of Piper et al. (1983). These calculations will continue for longer time periods and the fluid dynamics will be modeled using Navier-Stokes equations.

To extend the time scale of the computational techniques, Brownian and Stochastic dynamics methods will also be used. This work will allow the reduction in the number of explicit atoms used in the calculations because the solvent will be represented as a medium the model parameters like the frictional force. The system that will be used for this development will be C<sub>60</sub> and the fluid will be isoprene.

Extending the bridge diagram calculations to more complex systems so that the number of explicit atoms can be reduced will continue the quest for better implicit force fields.

This will begin with the next logical chemical system in the increasing series of complexity and that is the nitrogen diatomic.

## Molecular Dynamics Simulations of Simple Solutes in Water, Ice, and at the Ice-Water Interface

A fundamental understanding of aqueous systems is key to understanding life processes on a molecular level. Molecular dynamics simulations are an important tool for gaining insight about these systems, and the constantly increasing computational power accessible to scientists through programs such as EMSL's MSCF mean that the team can look at simple and more complex systems in greater and more thorough detail than ever before.

The research team is currently installing its molecular dynamics (MD) (Smith and Haymet 1992 and 1993; Smith and Dang 1994) and program at PNNL which models systems of up to four atomic/ionic components. Specifically, the team is implementing the Nose-Hoover thermostat (Nose 1986; Hoover 1985) and Rahman-Parinello box-shape change (Parinello and Rahman 1980) to calculate properties of systems in both the fluid and solid phase in the constant NVT and constant NPT ensemble. The improved program will accept any box shape which is a deformation of a cube (of which the box with lowest symmetry is a triclinic unit cell). This flexibility in the box shape is important to ensure the system exhibits zero internal stress when in the solid phase, as small internal stresses in the system may lead to significant changes in the free energy calculated via molecular simulation.

The team's initial focus is on systems of simple solutes (both charged and uncharged) in water, ice and at the ice-water interface. The team will use the rigid SPC/E model of water (Berendsen et al. 1987) to avoid having to resolve vibrational motions of the water molecules. Team members will calculate the free energy difference between solute particles in liquid water and solid ice in the limit of infinite dilution. Structural and transport properties will also be calculated. Future simulations will focus on systems involving more complicated solute molecules such as trehalose and their effects on the ice-water interfaces. The MSCF computational resources will be a large factor in the success of this project. To improve the overall performance of the simulation code, parallel processing techniques may be implemented in the Ewald sum algorithm and in the loop for the calculation of the interatomic forces.

### Electrolytes Near a Charged Surface: Integral Equation Approximations

The team is currently installing at PNNL the code required to study the equilibrium structure of various electrolytes. Current team projects include dissolved salts, molten salts, and autoionizable water-in the presence of an electrode. Along with electrolyte structure, the team will also calculate charge profiles, the mean electrostatic potential, and the double-layer capacitance. The team's initial focus is on simple 2-2 electrolytes or molten KCl near an infinite planar electrode, but the same methodology applies to the much more exotic cases, such as that of water structure (and pH) near an arbitrarily shaped electrode such as a strand of DNA or a protein.

In principle, the team will solve the following inhomogeneous Ornstein-Zernike equation:

$$[g_{ss'}(\vec{r}_1, \vec{r}_2) - 1] = c_{ss'}(\vec{r}_1, \vec{r}_2) + \sum_{s''} \int d\vec{r}_3 \rho_{s''}(\vec{r}_3) c_{ss''}(\vec{r}_1, \vec{r}_3) [g_{s''s'}(\vec{r}_3, \vec{r}_2) - 1] \quad (1)$$

where  $g_{ss'}(\vec{r}_1, \vec{r}_2)$  is the species  $s$ -species  $s'$  radial distribution function and  $c_{ss'}(\vec{r}_1, \vec{r}_2)$  is the associated direct correlation function, along with associated density

$$\rho_{s'}(\vec{r}_3) = \rho_{s'}[\{g_{ss'}\}, \{c_{ss'}\}, \vec{r}_3] \quad (2)$$

and “closure”

$$c_{ss'}(\vec{r}_1, \vec{r}_2) = c_{ss'}[\{\rho_s\}, \{g_{ss'}\}, \vec{r}_1, \vec{r}_2] \quad (3)$$

relations.

There are several remaining complications. Although exact density relations (Equation [2]) exist, there are only approximate closure relations (Equation [3]) available; the team will try several. Long-range forces prevent a naïve implementation of Equation (1) from converging, but analytical techniques exist to solve the long-ranged portions of Equation (1), thereby allowing the team to replace Equation (1) with a short-ranged version. The most significant complication, however, is the fact that Equation (1), at its most general, comprises six degrees of freedom  $r_1, r_2$  (in three dimensions). Numerically solving Equation (1) involves mapping the degrees of freedom onto a grid, and then performing a high-dimensional root search. Using a grid of 100 points for each degree of freedom, six degrees of freedom corresponds to a  $10^{12}$ -dimensional root search. The team will use EMSL’s IBM SP<sup>®</sup> along with an off-the-shelf parallelized nonlinear equation solver (such as PETSc, <http://www.mcs.anl.gov/petsc>, by S. Balay, W. D. Gropp, L. C. McInnes, and B. F. Smith) to satisfy the large computational requirements.

## **References**

- Berendsen, HJC, JR Grigera, and TP Straatsma. 1987. “The Missing Term in Effective Pair Potentials.” *J. Phys. Chem.*, 91:6269-6271.
- Hoover, WG. 1985. “Canonical Dynamics: Equilibrium Phase-Space Distributions.” *Phys. Rev. A*, 31:1695-1967.
- Lynch, GC, and BM Pettitt. 2000. “Semi-Grand Canonical Molecular Dynamics Simulation of Bovine Pancreatic Trypsin Inhibitor.” *J. Chem. Phys.*, 258:405-413.
- Nose, S. 1986. “An Extension of the Canonical Ensemble Molecular-Dynamics Method.” *Mol. Phys.*, 57:187-191.
- Parinello, M and A Rahman. 1981. “Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method.” *J. Appl. Phys.*, 52:7182-7190.

Parinello, M and A Rahman. 1980. "Crystal Structure and Pair Potentials: A Molecular-Dynamics Study." *Phys. Rev. Lett.*, 45:1196-1199

Piper, J, JA Morrison, C Peters, and Y Ozki. 1983. "Heat and Entropies of Adsorption of N-2 on Grafoil at 79.3-K." *J Chem. Soc. Faraday Trans.*, I 79:2863-2874.

Smith, DE and LX Dang. 1994. "Computer Simulations of NaCl Association in Polarizable Water." *J. Chem. Phys.*, 100:3757-3766.

Smith, DE and ADJ Haymet. 1993. "Free Energy, Entropy, and Internal Energy of Hydrophobic Interactions: Computer Simulations." *J. Chem. Phys.*, 98:6445-6454.

Smith, DE and ADJ Haymet. 1992. "Structure and Dynamics of Water and Aqueous Solutions: The Role of Flexibility." *J. Chem. Phys.*, 96:8450-8459.

Vainrub, A and BM Pettitt. 2000. "Thermodynamics of Association to a Molecule Immobilized in an Electric Double Layer." *Chem. Phys. Lett.*, 323:160-166.

### **Publications**

Dyer, KM, JS Perkyms, and BM Pettitt. 2002. "Computationally Useful Bridge Diagram Series: III. Lennard-Jones Mixtures." *J. Chem. Phys.*, 116:9413-9421.

Wong, K, and BM Pettitt. 2001. "A Study of DNA Tethered to a Surface by an All-Atom Molecular Dynamics Simulation." *Theor. Chem. Acc.*, 106:233-235.

Wong, K, and BM Pettitt. 2000. "A New Boundary Condition for Computer Simulations of Interfacial Systems." *Chem. Phys. Lett.*, 326:193-198.

### **Presentations**

Beck, BW and BM Pettitt. 2001. "Buried Solvent Channels: Cross-protein Interactions Between the Active Site and Dimer Interface in *Serratia marcescens* Endonuclease." WM Keck Center for Computational Biology, April 20, 2001, Houston, Texas.

Beck, BW and BM Pettitt. 2000. "Biomacromolecular Solvation: Solvent Mediated Contacts Between the Dimer Interface and the Active Site in *Serratia marcescens* Endonuclease." WM Keck Center for Computational Biology 2000 Symposium, October 16-17, 2000, Houston, Texas.

### **Significant Methods/Routines or Codes Developed**

ESP – molecular dynamics program – parallelized and optimized to run on the MCSF's IBM SP®.

GBC – Uses the lead space group and reduces the number of particles necessary to represent an interfacial system by a half compared to the traditionally used PBC - incorporated into the ESP program.

## **Appendix C - Full Report of Third Year Activities and Accomplishments**

### **Multiscale Computations of Molecular Assemblies**

#### **Team Members**

A.D.J. Haymet, Distinguished University Professor, Department of Chemistry and Institute for Molecular Design

B.M. Montgomery Pettitt, Cullen Distinguished Professor, Departments of Chemistry, Physics, Computer Science, Biology and Biochemistry, and Institute for Molecular Design

R. Glowinski, Cullen Professor, Departments of Mathematics and Mechanical Engineering

S.L. Johnson, Cullen Distinguished Professor, Departments of Computer Science, Mathematics, and Electrical Engineering

S.S. Akhtar, Department of Chemistry and Institute for Molecular Design

K. Dyer, Department of Chemistry and Institute for Molecular Design

S. Kim, Department of Computer Science and Institute for Molecular Design

T.C. Rogala, Department of Chemistry and Institute for Molecular Design

E.R. Smith, Department of Chemistry and Institute for Molecular Design

T. Basak, Ph.D., Department of Chemistry and Institute for Molecular Design

B.W. Beck, Ph.D., Department of Chemistry and Institute for Molecular Design

J.M. Cabrera, Ph. D., Department of Chemistry and Institute for Molecular Design

S.C. Gay, Ph. D., Department of Chemistry and Institute for Molecular Design

G.C. Lynch, Ph.D., Department of Chemistry and Institute for Molecular Design

S. Mejía-Rosales, Ph. D., Department of Chemistry and Institute for Molecular Design

J.S. Perkyns, Ph. D., Department of Chemistry and Institute for Molecular Design

A. Vainrub, Ph. D., Department of Chemistry and Institute for Molecular Design

K. Wong, Ph.D., Department of Chemistry and Institute for Molecular Design

### **Sources of Financial Support**

RA Welch Foundation, “Theory of Saline Solutions”

(Pettitt)

NIH, “Peptides in Saline Solution”

(Pettitt)

NIH/NCI, “Multiparameter analysis of DNA on surfaces in solution”

(Hogan and Pettitt)

NSF, “Computational fluid dynamics on particulate flows”

(Glowinski and Pettitt)

Texas ARP, “Multiscale simulation of Rubber”

(Pettitt and Glowinski)

### **Number of Hours Allocated and Used in the Past Year**

Year 3 – 500,000 hours allocated and 500,000 hours used

### **Publications**

Dyer, K, J Perkyns, and BM Pettitt. 2002. “Computationally Useful Bridge Diagram Series. III. Lennard-Jones mixtures.” *J. of Chem. Phys.*, 116:9413-9421.

Juárez, LH, R Glowinski, and BM Pettitt. 2002. “Numerical Simulation of the Sedimentation of a Tripole-Like Body in an Incompressible Viscous Fluid.” *App. Math. Lett.*, 15:743-747.

Perkyns, JS, K Dyer, and BM Pettitt. 2002. “Computationally Useful Bridge Diagram Series. II. Diagrams in h-Bonds.” *J. Chem. Phys.*, 116:9404-9412.

Vainrub, A and BM Pettitt. 2003. “Surface Electrostatic Effects in Oligonucleotide Microarrays: Control and Optimization of Binding Thermodynamics.” *Biopolymers*, 68:265-270.

Vernov, AV and WA Steele. 1986. “Dynamics of Nitrogen Molecules Adsorbed on Graphite by Computer-simulation.” *Langmuir*, 2:606-612.

Wong, KY, AVainrub, T Powdrill, M Hogan, and BM Pettitt. 2004. “A Non-Watson-Crick Motif of Base-Pairing on Surfaces for Untethered Oligonucleotides.” *Molecular Simulation*, 30:121-129.

## **Presentations**

Beck, BW and BM Pettitt. 2002. “Biomacromolecular Solvation: “Gated” Solvent-filled Channels in *Serratia marcescens* Endonuclease Mediate Active-site and Dime-interface Interaction.” Gulf Coast Consortium for Bioinformatics and the Keck Center for Computational and Structural Biology Bioinformatics Symposium, October 14-15, 2002, Houston, Texas.

Kim, S, L Prabu, M Abdullah, BW Beck, M Feig, GC Lynch, L Johnsson, and BM Pettitt. 2002. “Database of Simulation Data: SimDB Design and Implementation.” Gulf Coast Consortium for Bioinformatics and the Keck Center for Computational and Structural Biology Bioinformatics Symposium, October 14-15, 2002, Houston, Texas.

Mejía-Rosales, S, GC Lynch, and BM Pettitt. 2002. “Semi-Grand Canonical Molecular Dynamics for Ionic Systems.” Gulf Coast Consortium for Bioinformatics and the Keck Center for Computational and Structural Biology Bioinformatics Symposium, October 14-15, 2002, Houston, Texas.

Wong, K and BM Pettitt. 2002. “Molecular Dynamics Simulation of DNA on a Surface.” Gulf Coast Consortium for Bioinformatics and the Keck Center for Computational and Structural Biology Bioinformatics Symposium, October 14-15, 2002, Houston, Texas.

## **Significant Methods/Routines or Codes Developed**

GBC – Uses the lead space group and reduces the number of particles necessary to represent an interfacial system by a half compared to the traditionally used PBC.

ESP – molecular dynamics program designed to perform simulations using traditional and extended-system ensembles; it includes the grand canonical ensemble. The glide boundary conditions are also incorporated into this program and the code is portable, parallelized, and optimized for several different platforms including the MCSF’s IBM SP<sup>®</sup>.

Viscous Fluid Flow – a method combining distributed Lagrange multiplier-based fictitious domain techniques, finite element approximations, and operator splitting to the numerical simulation of a rigid body in a Newtonian incompressible viscous fluid.

## **Overview of the Past Year’s Accomplishments and Activities**

### **Characteristics of Biological Molecular Assemblies**

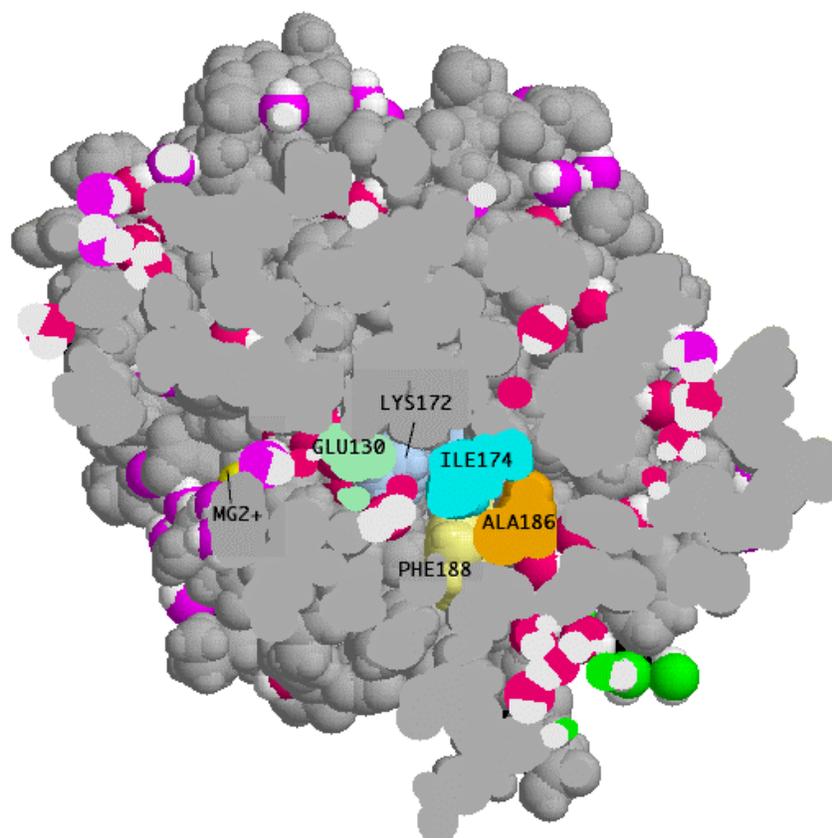
The major effort of this project was directed at the study and understanding of the characteristics of molecular assemblies. To perform these studies, a two-pronged approach was utilized in which computer experiments were conducted on two different types of macromolecular systems. The importance and function of the interfacial environment of aggregates, in particular those of biomacromolecular systems, is still unanswered. The role of the solvent interface in the stability and function of macromolecules is also not fully understood. The studies performed on these systems will provide a large amount of statistical data that will be used to try to comprehend some of these outstanding issues.

The first system is an enzyme, *Serratia Marcescens* endonuclease, which is found in its naturally occurring state only as a dimer. A mutation of one amino acid can force this system entirely to a stable monomeric state. This system contains a “wet” active site and there is a highly ordered solvent layer in the interfacial region between the two subunits of the dimer. There is very high resolution (0.9 Å) x-ray crystallographic structures for the dimer and they include well resolved data for the surrounding solvent and unusual buried solvent molecules associated with the proteins; this also includes the distinct solvent layer between the two subunits of the dimer. A third system in this series is the complex formed when the DNA is bound to the enzyme. Although there is no x-ray structure for the DNA-bound species, other studies indicate that the complex composed of the protein dimer and the DNA also includes well-ordered water molecules. The hypothesis is that the enzyme’s structure and activity are dependent on the internal and interfacial water molecules. But, another unanswered question is why does nature prefer the dimer instead of the monomer?

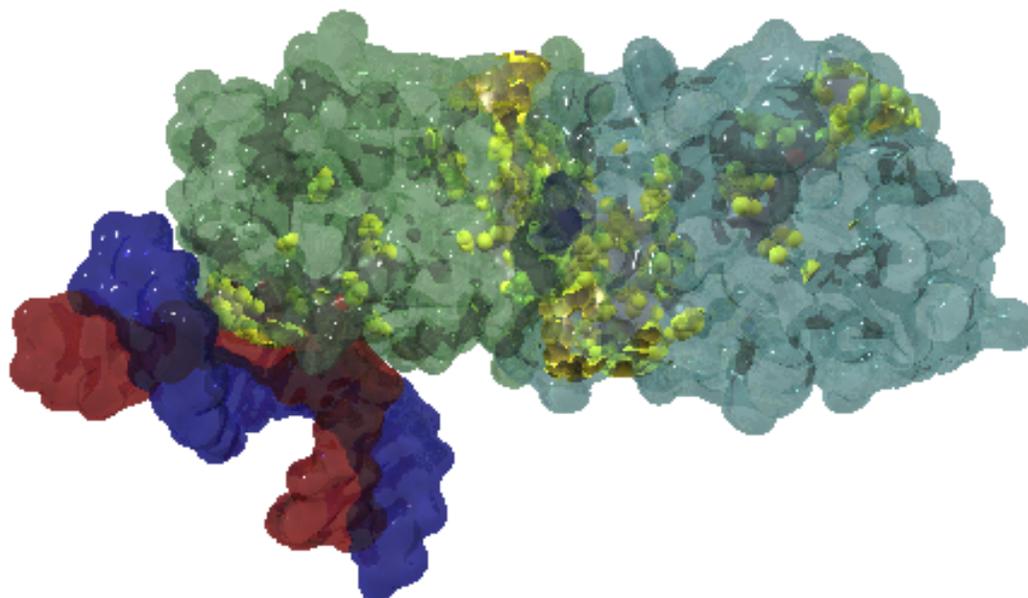
A systematic study of the monomer, dimer, and complex will be performed to try to resolve the role of the interfacial and buried water molecules. Three molecular dynamics simulations with explicit water molecules have been performed: (1) contains the monomer, (2) contains the dimer, and (3) contains the complex composed of dimer with a double-stranded DNA molecule bonded to one of the active sites. The initial structures of the monomer and dimer species were taken directly from the x-ray crystallography data. Because there is no complex structure this was built using the A-form DNA duplex d(ATAGGGGCGCCCCTAT)<sub>2</sub>. The simulations were all performed with the following: ESP program, NVE ensemble, CHARMM27 force field parameters with TIP3 water, and the temperature set to 300K.

Preliminary results from the monomer simulation indicate that the structure is stable; the RMSD of the monomer versus the x-ray crystallography structure is less than 2.1 Å and the radius of gyration from the molecular dynamics simulation is 16.99 Å compared to 16.69 Å from the x-ray data. These results provided confidence in both the force field parameters and the technique chosen. This monomer simulation reveals the presence of a solvent-filled channel leading from the active site of the enzyme to three distinct openings near the dimer interface. The channel appears to be mediated by a barrier, located about midway between the active site and the final dimer interface near residues Ile174, Ala186, and Phe188, that behaves like a “gate” controlling the diffusion of water between the active site and the dimer interface. The presence of this “gate” appears to have implications in both stability and regulation of the enzymatic activity; this may also be part of the explanation of the evolution of the dimeric state of this enzyme. In Figure C.1, a snapshot at 454 picoseconds of this “gated” water channel for the monomer is illustrated.

The dimer and complex simulations are in the preliminary stages. The simulations are continuing to produce more statistical data. Initial results from both simulations confirm the solvent channel discovered in the monomer simulation. In Figure C.2, the average solvent channel for the complex is displayed. This average structure was determined from 285 picoseconds of data and reproduces the large solvated cavity leading to three openings near the interface as initially seen in the monomer simulations.



**Figure C.1.** Snapshot of “gated” water channel from monomer simulation at 454 picoseconds.



**Figure C.2.** Average solvent density over 285 picoseconds in the channel region of the protein-DNA complex.

These simulations continue to try to elucidate the role of the solvent channel in both the stability and activity of this system. The current set of calculations is being performed in the semi-grand canonical ensemble. The semi-grand method allows the control of the solvent density through the chemical potential. This technique, which has been successfully applied to BPTI simulations, is being used to try to understand the necessity of the water in the channel region. Those experiments are utilizing a biasing algorithm to study the increase and decrease of the water density in the channel region.

The second system in this study involves DNA microarrays; these are powerful devices used in areas like genetics, drug discovery, etc. Although these devices have exploded onto the biotechnology arena, little is understood about their underlying physics and chemistry thereby making the design and optimization processes hit or miss.

The second phase of these studies involved the specificity and affinity of DNA associated or hybridized to a target oligonucleotide molecule near a surface. These investigations are a collaborative effort using experimental data, an analytical model, and molecular dynamics simulations. The recent experimental studies indicate that there is a kinetically rapid hybridization between a target DNA fragment, usually large, and oligonucleotides electrostatically immobilized to a surface. An analytic solution to the linear Poisson-Boltzmann theory of the electric double layer interaction between DNA and a hard surface predicts tight binding in this system. Molecular simulations were performed for a modified silicon dioxide surface with positively charged groups at neutral pH and an unattached oligonucleotide molecule. The oligonucleotide associated with the surface in salt water in such a way that some of the bases remained stacked and the bases closest to the surface were pointed preferentially toward the solution, away from the surface.

The computer simulation experiments consisted of two independent all-atom molecular dynamics simulations, called **I** and **II**. In both experiments the model consists of a glass surface coated with an ammonium monolayer, a 12-base B-form DNA single strand, and a surrounding solution of 0.8 M NaCl with explicit water molecules. The CHARMM27 all-atom force field parameters are used to describe the interactions. This configuration is equivalent to a DNA coated interface with a surface density of 0.04 DNA nm<sup>-2</sup>, if two-dimensional periodicity parallel to the surface is considered, and approximates half the density of an adsorbed monolayer. The composition of the DNA strands is CGTGTCCCTCTC which is the same composition used in the earlier for the tethered duplex experiment. The force field parameters of the silica layer were adopted from the modified CVFF force field, and the ammoniums were from the all-atom CHARMM22 proteins parameters and CHARMM27 was used for the nucleic acid, salt, and water interactions. The DNA single strands were started with the helical axes parallel to and about 12 and 22 Å above the surface in **I** and **II**, respectively. The empty space of the simulation boxes was filled with water molecules of which 215 were randomly chosen and replaced by 53 sodium and 162 chloride ions. Final numbers of water molecules in the simulations were 3461 and 3452, respectively.

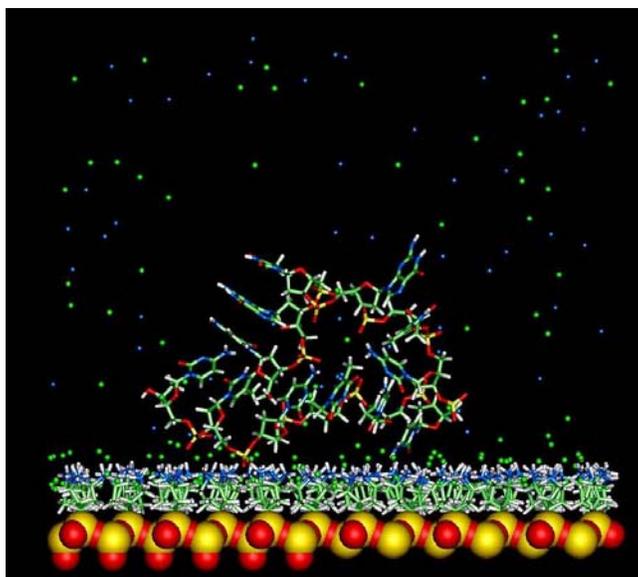
Because only two simulations with different initial conditions were performed, there is not enough data for a complete kinetic description. However, both simulations yield consistent structures and, as such, a plausible and testable hypothesis for the structure of DNA formation

that is parallel to the observed binding characteristics of the theoretical model. In **I**, when the DNA approached the surface, adsorption layers of the chloride ions above the ammoniums were only partially formed. The phosphate of T(11), which was closest to the surface, bound tightly to the surface ammoniums. Only chloride ions and almost no water molecules were found between them. One portion of the oligonucleotide, from C(7) to C(12), formed a curve segment and stayed close to the surface. The remaining part of the oligonucleotide bent upwards. Figures C.3 and C.4 are different views of the snapshot at 9.0 ns for this simulation. Primarily, the DNA was in a triangular shape. In simulation **II**, a substantial surface double layer was formed involving the surface ammonium ions and the free chlorides. The strand of DNA quickly formed a salt bridge with phosphates of C(10) and T(11), which were closest to surface ammoniums. This created a complicated layered structure of ions and water molecules between the phosphates and surface ammoniums. The remaining of the oligonucleotide curved upwards into the solution. Overall, the DNA formed an S shape conformation. Different views of a snapshot at 9.0 ns are shown in Figures C.5 and C.6. This feature remained for the rest of the simulation.

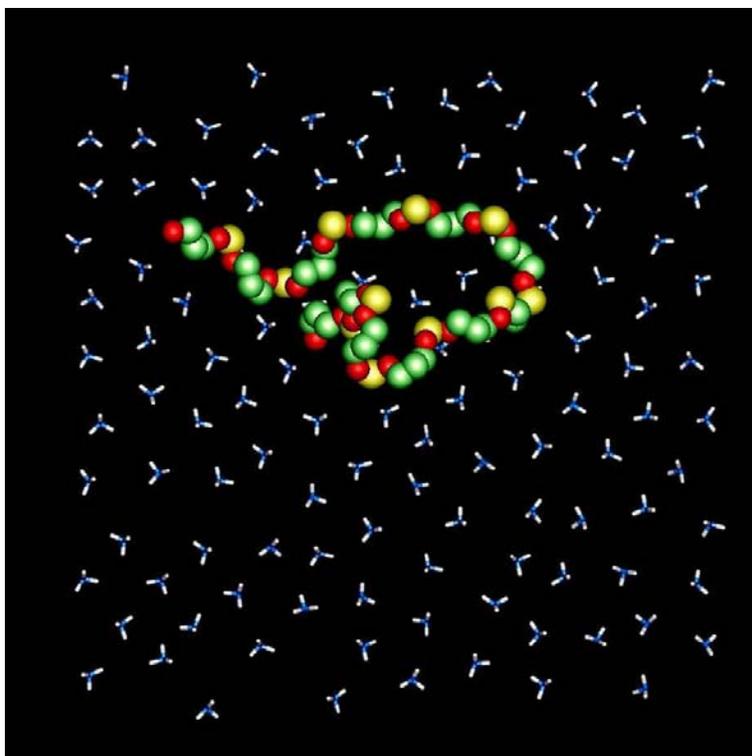
### Sedimentation and Hydrodynamic Flow

Another aspect of this project involves flow phenomena, in particular sedimentation and hydrodynamic flow. Previous calculations involved systems described by two and three dimensions and produced numerical results for both fluid flow and sedimentation. In all of the earlier calculations the rigid body was composed of one or two spheres. These calculations have now been extended to a simulation of a two-dimensional system composed of three spheres, called a tripole-like body, in an incompressible viscous fluid.

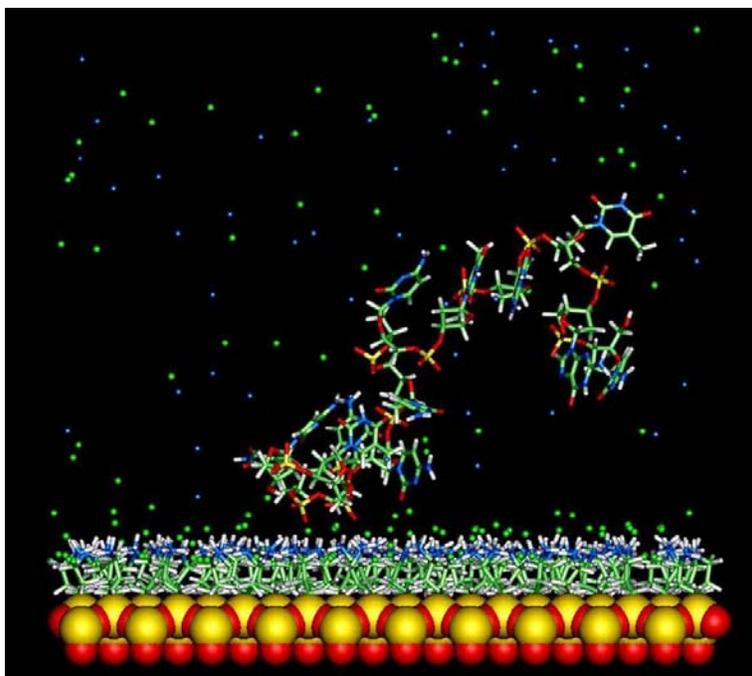
The methodology developed to perform this simulation is composed of several parts that is briefly described below. Full details are provided in Juárez et al. 2002. The first part consists of a *fictitious domain formulation with distributed Lagrange multipliers*. This combination allows



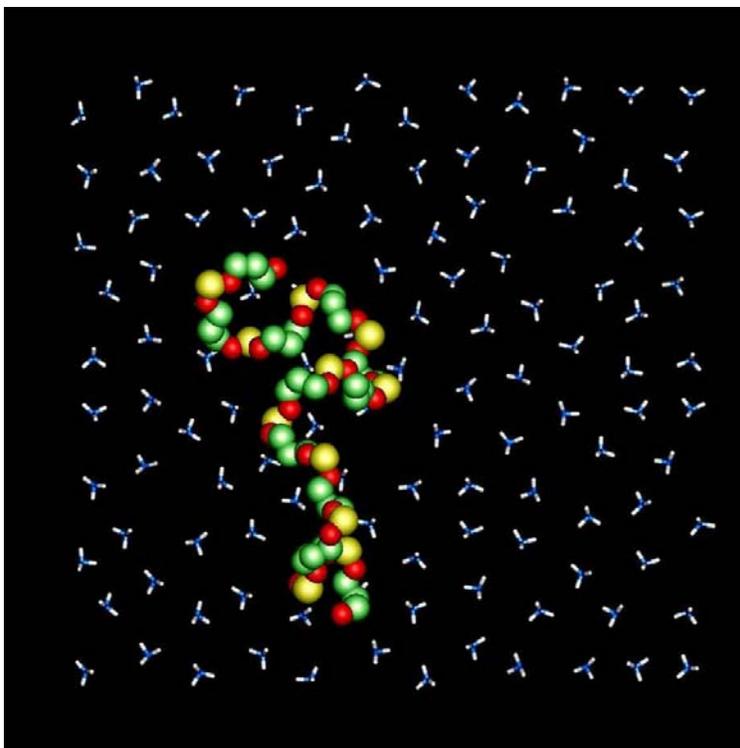
**Figure C.3.** A side view of a snapshot of simulation **I** at 9 ns. Note that the phosphate of T(11) is tightly bound to a surface ammonium and the bases are predominantly stacked.



**Figure C.4.** A top view, backbone atoms only, of the same snapshot of simulation **I** at 9 ns as illustrated in Figure C.3.



**Figure C.5.** A side view of a snapshot of simulation **II** at 9 ns. Note that there is a bigger gap between surface ammoniums and the nearest phosphates compared to **I** (Figure C.3). The other end of the DNA reached further into the solution.



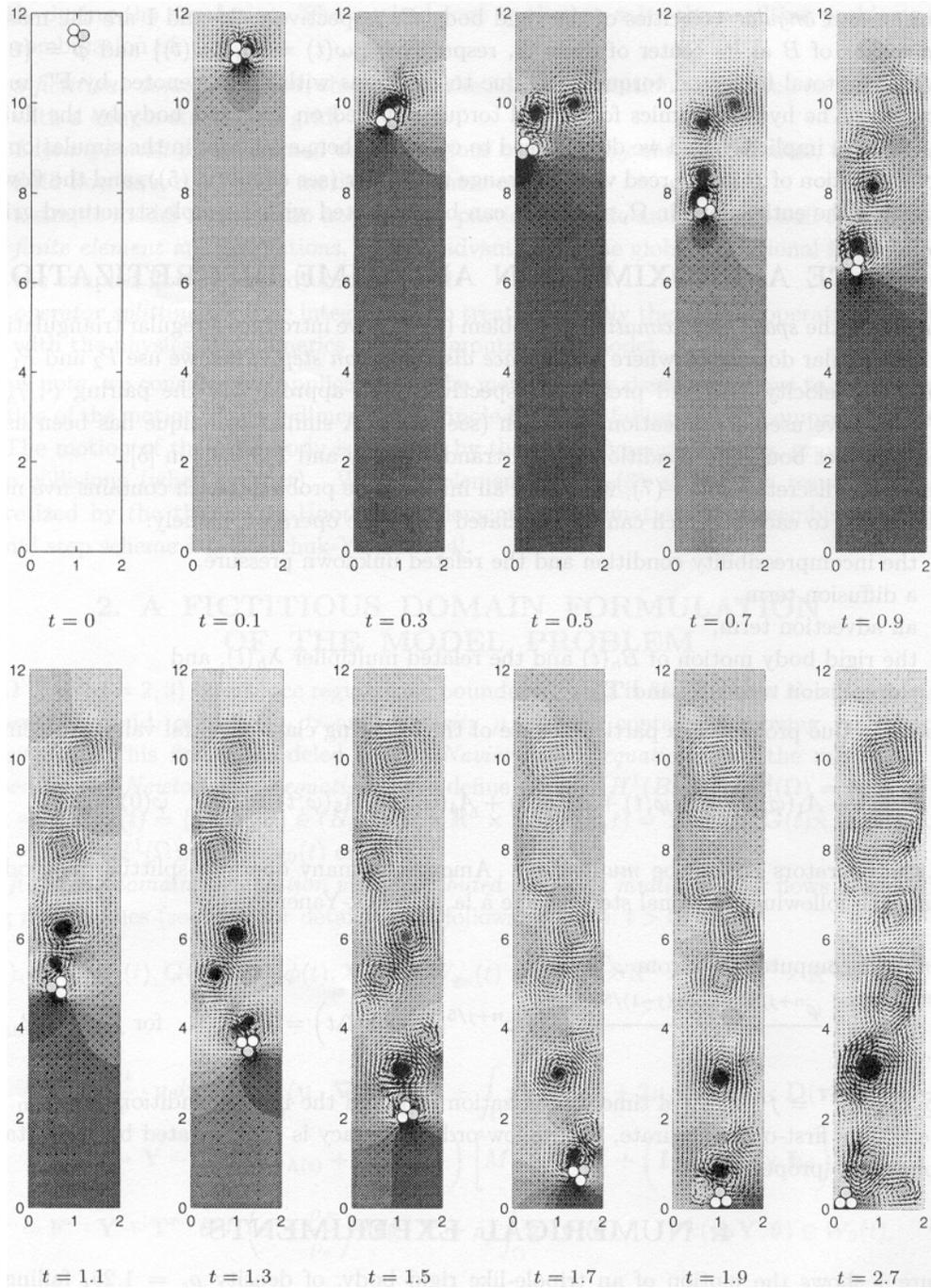
**Figure C.6.** A top view, backbone atoms only, of the snapshot (Figure C.5) of simulation II at 9 ns.

the flow computation to be limited to a fixed region that can be described by a simple structured grid with the hydrodynamic forces and torque on the rigid body built in. The rigid body motion is then enforced via the Lagrange multiplier. The particle-particle and particle-wall collisions are taken into account via a simple model. A *finite element* approximation for the global variational formulation of the coupled motion of the flow and the rigid body and *operator splitting* for the time integration, have also been utilized.

The results from this calculation are presented in Figure C.7. This shows the motion of the tripole-like body, density of 1.25, falling in a viscous fluid of density 1.0 and viscosity 0.01, with a time step of 0.001. The black arrows are used to describe the velocity flow field. The rigid body rotates in a clockwise direction soon after being released; this is believed to be a result of initial orientation. The body also drifts to the left until it “touches” the wall but with an orientation that is similar to the initial orientation. This produces another collision with the same wall but at a different orientation that results in a rotation in the opposite direction. Before colliding with the right wall, the rigid body sediments at the bottom.

#### SimDB – Database of Simulation Data

Molecular dynamics simulations generate large quantities of data. This data is usually used by the research group performing the calculations but is unavailable to the general scientific community. A data repository would provide a means of sharing this scientific resource that usually requires thousands of computer hours to generate.

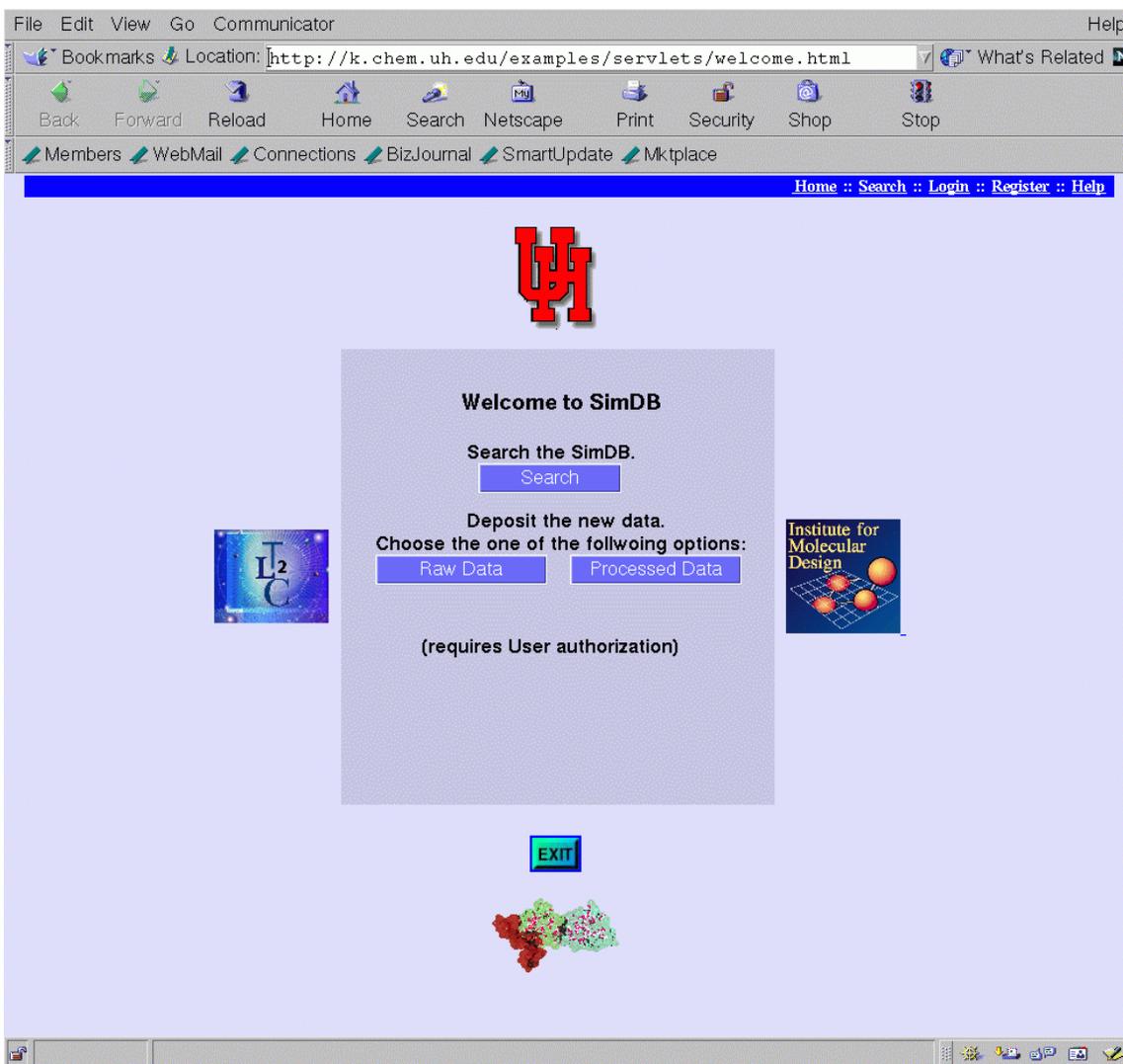


**Figure C.7.** A numerical simulation of a tripole-like body falling through a viscous fluid.

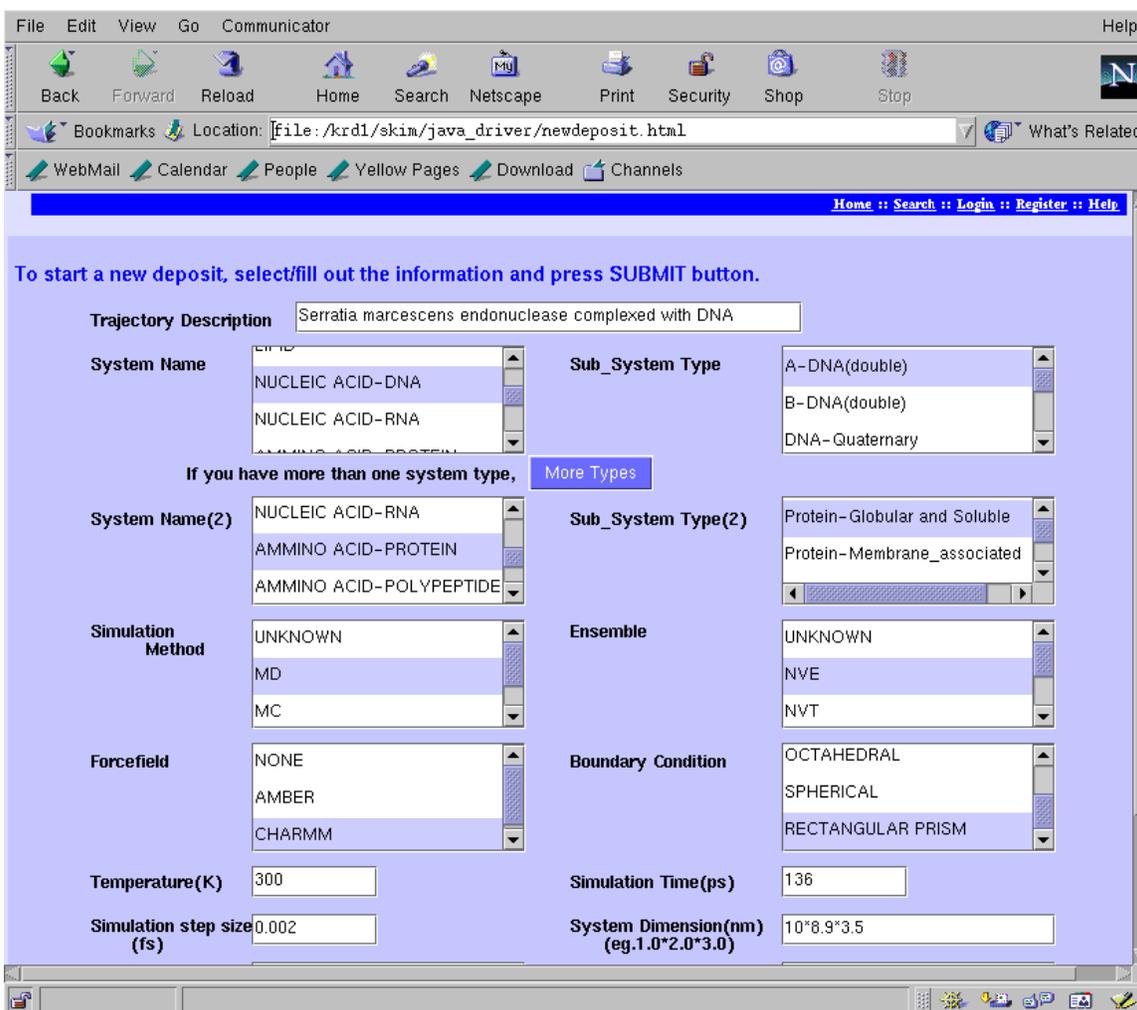
*SimDB* has been designed to provide a repository of simulation data for the scientific community. This design is composed of *raw* and *processed* data where *raw* data is the data generated by the modeling programs (i.e., positions, velocities, energies, etc.) and *processed* data is any data not directly generated by the simulation programs. The design includes the storage of data at distributed locations and a web-based graphical user interface for depositing and searching the

system. A database, in this case Oracle<sup>®</sup>, has been implemented to handle the details for each entry. Details include location, system type (protein, DNA, RNA, membrane, etc.), force field, authors, journal references, etc.

The user-friendly web browser interface has been developed and implemented. The database tables for storing the necessary information for each entry have also been designed and configured. This implementation includes the flexibility to easily add new families of data. The SRB has been chosen to handle the distributed data storage. The SRB will also handle the data movement from one site to another as requested by an authorized user. This is a necessary feature because users may need to perform different types of calculations on data in the system, and therefore would need to access the data directly. The link between the SRB and the *SimDB* database server is in its final stages. Examples of the web-interface are illustrated in Figures C.8 and C.9.



**Figure C.8.** Snapshot of the welcome window for the SimDB project.



**Figure C.9.** Snapshot of one of the web-base windows for depositing data into the repository.