

First-Principles Methods for Predicting the Chemistry of Environmentally Relevant Systems

R.A. Friesner	J.H. Johnson
E.R. Batista	G.A. Kaminski
T. Cagin	R. Martin
M.S. Diallo	R.B. Murphy
J.L. Faulon	H.A. Stern
B.F. Gherman	G. Wang
W.A. Goddard III	M.E. Wirstam
V. Guallar	

August 2005



**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

First-Principles Methods for Predicting the Chemistry of Environmentally Relevant Systems

R.A. Friesner	J.H. Johnson
E.R. Batista	G.A. Kaminski
T. Cagin	R. Martin
M.S. Diallo	R.B. Murphy
J.L. Faulon	H.A. Stern
B.F. Gherman	G. Wang
W.A. Goddard III	M.E. Wirstam
V. Guallar	

August 2005

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

Abstract

The ability to predict the transport and transformations of contaminants within the subsurface environment is critical for decisions on virtually every waste disposal option facing the U.S. Department of Energy (DOE), from remediation technologies such as *in situ* bioremediation to evaluations of the safety of nuclear waste repositories. With this fact in mind, DOE recently sponsored a series of workshops that were held to develop Strategic Simulation Plan (SSP) on applications of high-performance computing to national problems of significance to the DOE. Subsurface Transport and Environmental Chemistry was one of the areas chosen for this plan. Within this general area, several areas were identified where applications of high-performance computing could potentially advance our knowledge of contaminant fate and transport in a significant way. Within each of these areas, molecular-level simulations were specifically identified as a key capability necessary for the development of a fundamental mechanistic understanding of complex biogeochemical processes.

Research Team Leader and Members

Team Leader

Name: R.A. Friesner

Institution: Columbia University

<u>Team Member</u>	<u>Institution</u>
E.R. Batista	Columbia University
B.F. Gherman	Columbia University
V. Guallar	Columbia University
H.A. Stern	Columbia University
M.E. Wirstam	Columbia University
M.S. Diallo	Howard University
J.H. Johnson	Howard University
J.-L. Faulon	Los Alamos National Laboratory
G.A. Kaminski	Columbia University
R.B. Murphy	Schrodinger, Inc.
T. Cagin	California Institute of Technology
W.A. Goddard III	California Institute of Technology
R. Martin	California Institute of Technology
G. Wang	California Institute of Technology

Number of Hours Allocated for the Past Three Years

Year 1: 620,000

Year 2: 505,000

Year 3: 675,000

Total: 1,800,000

Number of Hours Used During the Past Three Years

Year 1: 604,147

Year 2: 502,505

Year 3: 660,983

Total: 1,767,635

Overview of the Past Three Year's Accomplishments and Activities

During the past three years, a wide variety of projects have been carried out using the Molecular Science Computing Facility (MSCF). These projects including modeling of enzymatic reactions in iron-containing proteins, modeling the chemistry of iron oxide surfaces, force field development, modeling of organic soil contaminants, modeling of hydrodesulfurization catalysts, investigation of dendrimers as chelating agents for metal ions, molecular-level predictions for electronic noses, and first principles prediction of vapor-liquid equilibrium phase diagrams. A significant amount of code development was also carried out to be able to perform the calculations required for these projects. A brief summary of the progress made on each project follows.

Modeling of Enzymatic Reactions in Iron-Containing Proteins

Iron-containing proteins catalyze a wide variety of reactions in nature, for example the conversion of methane to methanol, at room temperature. From an environmental point of view, such reactions are of interest for deployment as possible bioremediation agents. The objective of this project has been to produce a detailed atomic-level picture of the catalytic pathways in a number of iron containing proteins, including methane monooxygenase (MMO), hemerythrin (Hr), and cytochrome P450 (cP450). MMO and cP450 activate dioxygen to oxidize an aliphatic C-H bond to an alcohol. Hr is an oxygen transport protein that binds and releases dioxygen in certain marine invertebrates.

These three enzymes have been studied using *ab initio* quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) techniques. Density functional (DFT) methods were used to carry out the QM aspects of the calculation. The Jaguar program, developed over the past several decades, provides the ability to routinely optimize QM models in the 100–200 atom range, which is typically sufficient to include the key reactive portions of the enzyme active site. QM/MM methods then allow the remainder of the protein environment to be incorporated into the calculations in a rigorous fashion. The studies, which are discussed below, represent a leading-edge effort to build realistic models of enzymatic reactions and investigate them with high quality calculations. The combination of Jaguar's high efficiency (running in parallel as well as serially) and the computation time provided by PNNL enabled a substantial increase in the size, accuracy, and complexity of the calculations performed as compared to prior results reported in the literature for these and related systems.

For both MMO, a complete description of the catalytic cycle has been produced, characterizing all of the intermediates and transition states at an atomic level of detail. Activation energies have been computed in good agreement (2–3 kcal/mol) with experimental results for a number of MMO reactions, a highly challenging task for a metal-containing enzyme. The rate determining

hydrogen atom abstraction step has been modeled for a number of substrates, and the detailed kinetics of the abstraction pathway have been explored via semiclassical quantum dynamics simulations, using a potential energy surface generated via DFT calculations. This simulation quantitatively predicted the per cent racemization obtained in chiral ethane experiments carried out by the Lippard group. Variations in substrate kinetics have recently been explained by carrying out QM/MM calculations of the abstraction barrier, which exhibit important substrate-dependent environmental effects. For cytochrome P450, the great majority of intermediates and transition states in the catalytic cycle were similarly characterized. In this case, an interesting and apparently general mechanism for stabilization of the transition state of hydrogen abstraction by the porphyrin chromophore, via salt bridges between peripheral carboxylate groups on the porphyrin and positive charged protein residues, was discovered. Finally, for Hr, the free energy of binding of dioxygen was computed, obtaining agreement to within 2 kcal/mol of experimental data.

Modeling of Iron Oxide Surfaces

Iron oxide is an important constituent of the soil, and is thought to play a major role in catalyzing the decomposition of small organic pollutants. A large-scale experimental effort to study iron oxide surfaces has been ongoing at Columbia University over the past several years in conjunction with the Environmental Molecular Sciences Institute (EMSI). Calculations for this project have been performed in close collaboration with these experimental groups, with the objective of elucidating basic structural and mechanistic questions that have arisen upon examination of the experimental data.

The initial goal was to develop a method capable of modeling iron oxide surfaces efficiently using *ab initio* DFT methods so systems with net charge and to use hybrid functionals could be treated. Therefore, an embedded cluster method in which classical ions surround a quantum cluster of atoms was developed. Ewald methods are used to represent the far field of the ions; this field is then fit to a finite set of point charges that can be used to compute the interaction of the long-range field with quantum cluster. The charges on the ions are calculated self-consistently by extracting values from the cluster using electrostatic potential fitting.

Initial application of this method to computation of the work function of the (001) surface of hematite yielded a value of 5.6eV. No direct experimental measurements have been carried out for this hematite, but the hematite work function has been measured relative to that of platinum, and other iron oxide surface work functions have been measured directly. This data yields an

estimate of approximately 5.3-5.9eV for the experimental value, which is in good agreement with the value calculated in this project and in poor agreement with alternative calculations of Scheffler and coworkers who obtained 4.3eV for the same system.

More recently, the methodology has been used to calculate the interaction of CCl₄ with magnetite. Flynn and coworkers (personal communication) have observed that phosgene is ultimately released from the surface, presumably via a decomposition of CCl₄ into CCl₂ and chlorine atoms, and a subsequent binding of the carbon to an oxygen atom followed by release of the phosgene moiety. In this project, models for the binding of CCl₂ to an oxygen molecule on the surface have been constructed, and at present, barrier heights for formation of phosgene, which is believed to be the rate-limiting step, are being computed.

Finally, the project team has started to apply the same methodology to titanium dioxide, which is another important metal oxide material, for example, in the development of solar energy devices such as the Gratzel cell as alternatives to silicon based devices. The work function of the anatase form of TiO₂ has been completed, and a paper describing this work will be submitted for publication shortly.

Polarizable Force Field Development

A methodology for constructing polarizable force fields for an arbitrary molecule directly from *ab initio* quantum chemistry was developed. Parameters for both small molecules and proteins were developed, and tests for the small molecules in the condensed phase (the protein force field has been tested in the gas phase; condensed phase testing is just starting) were performed. A key aspect of the project has been obtaining basis set limit MP2 level quantum chemical pair binding energies. The Columbia and PNNL group have collaborated to validate such calculations. A paper describing the MP2 basis set limit methodology in Jaguar is in preparation, and another paper presenting results obtained for the liquid state simulations has been submitted for publication.

Molecular Modeling of Hydrophobic Organic Contaminants Uptake and Sequestration by Soil Organic Matter (California Institute of Technology and Howard University)

The interactions of hydrophobic organic contaminants (HOCs) with soil organic matter (SOM) determine to a large extent their fate and transport in subsurface systems. The primary objective of this research has been to develop a molecular-level understanding of the uptake of HOCs (*e.g.*, phenanthrene and PCE) by two model SOMs: Chelsea humic acid (CHA) and Lachine kerogen (LK). A hierarchical approach for predicting the structure and properties of complex petroleum geomacromolecules, such as asphaltenes, has been developed. This novel approach combines experimental characterization data, computer assisted structure elucidation (CASE), and atomistic simulations to generate a sample of three-dimensional models that statistically represent the entire population of models that can be built from a given analytical data set. The CASE program was then used to generate a sample of three-dimensional structural models for CHA that is representative of the entire population of structural models that can be built from the available experimental data. These computer-generated structural models for CHA were then

used as starting three-dimensional structures for the molecular dynamics (MD) simulations of the uptake and sequestration of HOCs by CHA.

The first objective of this project was to determine the extent to which the enthalpic interaction parameter χ_{HOC} could be correlated with experimentally measured parameters of HOC uptake and sequestration such as the Freundlich capacity parameter (K_f) and site energy heterogeneity factor (2). Using phenanthrene as model HOC, MD simulations performed to estimate c_{ij} as a function of temperature yielded a linear relationship between $\log c_{ij}$ and measured $\log K_f$ for phenanthrene sorption onto CHA. Not surprisingly, the site energy heterogeneity factor (n), a measure of the extent of “non linearity and hysteresis” of HOC sorption onto SOM, was poorly correlated with c_{ij} . Additional MD simulations were then performed to test the validity of this correlation for several HOCs (*e.g.*, naphthalene, dichlorobenzene, and trichlorobenzene, etc.).

Atomistic Models of Hydrodesulfurization Catalysts over Co/MoS₂ (California Institute of Technology)

The need to meet more stringent standards limiting the sulfur content of gas oils urges a deeper understanding of the mechanism by which sulfur-containing compounds are destroyed over hydrodesulfurization (HDS) catalysts. Thiols and sulfides are readily treated by the Co- and Ni-promoted MoS₂ catalysts currently employed, but satisfying the imminent restrictions will require the removal of the most refractory species, mainly alkyl-substituted polyaromatic thiophenes. Unfortunately, the nature of the catalyst’s active sites is not well known. In this project, density functional calculations were performed on models of pure and Co-promoted MoS₂, along with relevant organic substrates, to determine 1) the structural and electronic factors indispensable to the “CoMoS” active phase, 2) the mechanism by which thiophene and its derivatives react, and 3) how alkyl substituents promote and inhibit different branches of the reaction network.

Since the detailed surface reactions are modeled intermediate-by-intermediate, *ab initio* DFT electronic structure calculations employing the B3LYP functional were used. Small clusters (three to six metal atoms) were used to imitate highly dispersed MoS₂ crystallites. The geometries of organic adsorbates and their binding sites were relaxed to find stable intermediates and transition states are located when relevant. To establish whether the specific pattern of reconstruction along the crystallite edges witnessed in scanning tunneling microscopy (STM) studies remains in the finely dispersed MoS₂ particles believed to provide most of the active sites, three- and six-metal atom clusters were generated, and their behavior in reconstruction and sulfidation were studied. Exposing the reconstructed (10 $\bar{1}$ 0) face is preferred over the ($\bar{1}$ 010) face by the small clusters, as in the larger STM subjects, but more disordered states will be prevalent. Since vacancies at the edges MoS₂ crystallites left by abstracted sulfur atoms are believed to be the active sites, the number of vacancies present and the energy required to remove sulfur atoms from the edges are important quantities. For these small models, it is less appropriate to discuss a “vacancy formation energy” than an “equilibrium degree of sulfidation,” as the binding energies of sulfur at edges depends strongly on the stoichiometry of the entire crystallite.

Dendrimers: A New Class of High Capacity Chelating Agents for Metals Ions

Dendrimers are highly branched polymers with controlled composition and architecture consisting of three structural components: a core, interior branch cells, and terminal branch cells. In this project, MD simulations were performed to investigate the performance of dendrimers and metal chelating agents. The first studies focused on the binding of Cu (II) ions to poly (amidoamine) (PAMAM) dendrimers in aqueous solutions. Compared to traditional chelating agents (*e.g.*, triethylene tetramine) and macrocycles (*e.g.*, cyclams) with nitrogen donors, which can typically bind only one Cu (II) ion per molecule, a generation eight (G8) PAMAM dendrimer can bind up 153 ± 20 Cu (II) ions per molecule. This clearly illustrates a distinct advantage of dendrimers over traditional chelating agents and macrocycles; that is, the covalent attachment of nitrogen ligands to conformationally flexible PAMAM chains enclosed within a nanoscopic structure results in substantial increase in binding capacity. Finally, in more general work on dendrimer simulation technology, the three-dimensional molecular structure of various dendrimers with continuous configurational Boltzmann biased direct Monte Carlo method was determined, and their energetic and structural properties were studied using MD techniques after annealing these molecular representations.

Electronic Noses

Electronic noses are finding applications in the food industry as well as in military missions. Monitoring of air quality on the international space station and the space shuttle also makes use of electronic noses fitted to monitor the concentrations of a dozen or so air pollutants. In spite of the significant progress made on such applications, there is substantial progress to be made in predicting and understanding the functioning of such devices from first principles. We report on the development and use of MD algorithms for first-principles predictions of vapor/polymer equilibrium.

A common electronic nose device consists of chemical sensors that are constructed with thin polymer films loaded with electrically conducting carbon black particles. Carbon black percolates in the polymer film, making the film conductive if an appropriate amount is added. When the analyte is absorbed into the sensor, the volume of the polymer film swells, causing some of the percolated carbon black pathways to break. This results in an increase in the electric resistivity of the polymer film which is the signal detected for the analyte. It has been found experimentally that the sensor response is only a function of the swelling volume of the film. This experimental evidence indicates that one could predict the response of such electronic nose if the solubility of the analyte in the polymer film is known. We have developed MD simulation methods for the prediction of the solubility of chemical molecules (analytes) in polymers, needed to predict the response of electronic noses.

In this project, a combination consisting of an MD simulation and a Flory-Huggins model was used to predict solubility. An excellent agreement between the calculated and experimental solubility parameter was found and the predicted solubilities are in good agreement with values obtained from experimental approaches.

First-Principles Prediction of Vapor-Liquid Equilibrium (VLE) Phase Diagrams

The ability to predict the vapor-liquid equilibrium (VLE) phase behavior of binary and multi-component systems is essential in order to refine chemical design processes sufficiently that the focus of environmental protection can be shifted from pollution clean-up to pollution prevention. Reliable understanding of the VLE for mixtures of organic compounds in both aqueous and non-aqueous solutions is also of great importance in understanding the processes in atmospheric chemistry. The primary challenge to predicting VLE is estimating the activity coefficients, γ , for various components. Most commonly γ is determined by fitting to experimental data on equations of state (EOS) or excess Gibbs free energy. These models have been useful in using fits to binary experimental data to predict the VLE of ternary and higher systems. However, there has been no reliable way to predict the binary data.

The UNIQUAC (UNIversal QUAsi-Chemical) model and the Wilson model provide a method for predicting the excess Gibbs free energy without using experimental data. These models calculate γ directly from molecular interaction energies sampled over an appropriate ensemble. Experiments do not readily yield these parameters. The known interaction energies are most frequently derived from fitting experimental phase behavior to the UNIQUAC equation for binary systems. This expands nicely to describe multi-component systems, affording prediction of ternary and higher systems based on binary experiments.

Since the equations use molecular interaction energies as fundamental input parameters, predictions of phase behavior should follow from fundamental analysis of interaction energies. This could provide a method of property prediction using only theoretical methods and free from experimental bias. Previous to this work, first-principles methods developed to predict activity coefficients were limited by the type of molecular interactions and the computational expense. Research by this project team has produced an efficient method for the theoretical determination of the fitting parameters for polar/polar, polar/non-polar, and non-polar/non-polar systems. In this method, the liquid phase is simulated with force-field dynamics calculations, and the interaction energies of molecular pairs are determined with high-level (LMP2/cc-pvtz(-f)) quantum calculations. The Cohesive Energy Density model in Cerius2 produces the ensemble of pairs used in the quantum calculations. These calculations determine the interaction parameters for both the UNIQUAC and the Wilson models.

With a modified UNIQUAC equation, the experimental phase diagrams were modeled successfully for a range of systems. All systems were correctly predicted except for phase diagrams for binary systems involving a polar molecule and a molecule containing a carbonyl group. Comparison to experimental data indicates that our modeling of these pairs gives binding energies that are too strong. The consistency of this failure suggests the source of the error arrives from the molecular dynamics simulations of these systems. This error may be due to the hydrogen-bonding term in the force field.

Publications

Baik, M-H, M Newcomb, RA Friesner, and SJ Lippard. 2003. "Mechanistic Studies on the Hydroxylation of Methane by Methane Monooxygenase." *Chemical Reviews*, 103:2385-2419.

Baik, M-H, BF Gherman, RA Friesner, and SJ Lippard. 2002. "Hydroxylation of Methane by Non-Heme Diiron Enzymes: Molecular Orbital Analysis of the C-H Bond Activation by Reactive Intermediate Q." *J. Am. Chem. Soc.*, 124:14608-14615.

Baik, M-H, D Lee, RA Friesner, and SJ Lippard. 2001. "Theoretical Studies of Diiron (II) Complexes that Model Features of the Dioxygen-Activating Centers in Non-Heme Diiron Enzymes." *Israel Journal of Chemistry*, 41:173-186.

Batista, E and RA Friesner. 2002. "A Self-Consistent Charge-Embedding Methodology for *Ab Initio* Quantum Chemical Cluster Modeling of Ionic Solids and Surfaces: Application to the (001) Surface of Hematite (α -Fe₂O₃)." *J. Phys. Chem. B*, 106(33):8136-8141.

Cagin, T, G Wang, R Martin, G Zamanakos, N Vaidehi, DT Mainz, and WA Goddard III. 2001. "Multiscale Modeling and Simulation Methods with Applications to Dendritic Polymers." *Comput. Theor. Polym. S.*, 11(5):345-356.

Cagin, T, G Wang, R Martin, N Breen, and WA Goddard III. 2000. "Molecular Modeling of Dendrimers for Nanoscale Applications." *Nanotech*, 11:77-84.

Cagin, T, PJ Miklis, G Wang, G Zamanakos, R Martin, H Li, DT Mainz, V Nagarajan, and WA Goddard III. 1999. "Recent Advances in Simulation of Dendritic Polymers." *Mat. Res. Soc. Symp. Proc.*, 543:299-310.

Diallo, MS, T Cagin, JL Faulon, and WA Goddard III. 2000. "Thermodynamic Properties of Asphaltenes: A Predictive Approach Based on Computer Assisted Structure Elucidation and Atomistic Simulations." In *Asphaltenes and Asphalts*, 2. Developments in Petroleum Science, 40B, TF Yen and GV Chilingarian, Editors (Elsevier Science 2000) Chapter 5, pp. 103-127.

Dunietz, BD, MD Beachy, Y Cao, DA Whittington, SJ Lippard, and RA Friesner. 2000. "Large Scale *Ab Initio* Quantum Chemical Calculation of the Intermediates in the Soluble Methane Monooxygenase Catalytic Cycle." *J. Am. Chem. Soc.*, 122(12):2828-2839.

Friesner, RA, M-H Baik, BF Gherman, V Guallar, M Wirstam, RB Murphy, and SJ Lippard. 2003. "How Iron-Containing Proteins Control Dioxygen Chemistry: A Detailed Atomic Level Description via Accurate Quantum Chemical and Mixed Quantum Mechanics/Molecular Mechanics Calculations." *Coordination Chemistry Reviews*, 238-239:267-290.

Gherman, BF, BD Dunietz, DA Whittington, SJ Lippard, and RA Friesner. 2001. "Activation of the C-H Bond of Methane by Intermediate Q of Methane Monooxygenase: A Theoretical Study." *J. Am. Chem. Soc.*, 123:3836-3837.

Goddard, WA, T Cagin, M Blanco, N Vaidehi, S Dasgupta, W Floriano, M Belmares, J Kua, G Zamanakos, S Kashihara, M Iotov, and G Gao. 2001. "Strategies for Multi-Scale Modeling and Simulation of Organic Materials: Polymers and Biopolymers." *Comput. Theor. Polym. S.*, 11(5):329-343.

Guallar, V, M-H Baik, SJ Lippard, and RA Friesner. 2003. "Peripheral Heme Substituents Control the Hydrogen-Atom Abstraction Chemistry in Cytochromes P450." *P. Natl. Acad. Sci. USA*, 100:6998-7002.

Guallar, V, BF Gherman, SJ Lippard, and RA Friesner. 2002. "Quantum Chemical Studies of Methane Monooxygenase: Comparison with P450." *Curr. Opin. Chem. Biol.*, 6:236-242.

Guallar, V, BF Gherman, WH Miller, SJ Lippard, and RA Friesner. 2001. "Dynamics of Alkane Hydroxylation at the Non-Heme Diiron Center in Methane Monooxygenase." *J. Am. Chem. Soc.* 124:3377-3384.

Kaminski, GA, HA Stern, BJ Berne, RA Friesner, YX Cao, RB Murphy, R Zhou, and TA Halgren. 2002. "Development of a Polarizable Force Field for Proteins via *Ab Initio* Quantum Chemistry: First Generation Model and Gas Phase Tests." *J. Comput. Chem.*, 23(16):1515-1531.

Murphy, RB, DM Philipp, and RA Friesner. 2000. "A Mixed Quantum Mechanics/Molecular Mechanics (QM/MM) Method for Large-Scale Modeling of Chemistry in Protein Environments." *J. Comp. Chem.*, 21(16):1442-1457.

Vargas, R, J Garza, RA Friesner, H Stern, BP Hay, and DA Dixon. 2001. "Strength of the N-H \cdots O=C and C-H \cdots O=C Bonds in Formamide and N-Methylacetamide Dimers." *J. Phys. Chem. A*, 105(20):4963-4968.

Wirstam, M, SJ Lippard, and RA Friesner. 2003. "Reversible Dioxygen Binding to Hemerythrin." *J. Am. Chem. Soc.*, 125:3980-3987.

Presentations (Oral and Poster)

Speaker	Title	Conference	Location	Date
R. Friesner	Polarizable models for water from <i>ab initio</i> quantum chemistry	National ACS Meeting	New Orleans, LA	9/99
R. Friesner	<i>Ab initio</i> quantum chemistry: From small molecules to proteins	Pitzer Symposium	Berkeley, CA	1/00
R. Friesner	Quantum chemical modeling of metalloprotein active sites	Metals in Biology Gordon Conference	Ventura, CA	1/00
R. Friesner	Molecular modeling in the 21 st Century	National ACS Meeting	San Francisco, CA	3/00
W. Goddard	Strategies for multiscale modeling and simulation of organic materials: polymers and biopolymers	ACS National Meeting	San Francisco, CA	3/00

W. Goddard	<i>De novo</i> simulations of catalysts, materials, and biochemistry	Dow Chemical Lecture Lecture	Dow Chemical Corp.	6/00
W. Goddard	Modeling and simulation of materials for industrial	National ACS Meeting	Washington, D.C.	8/00
W. Goddard	Investigation of heterogeneous and homogeneous transition metal catalysis using density functional theory	National ACS Meeting	Washington, D.C.	8/00
M. Diallo	Thermodynamics and structural properties of asphaltenes from molecular dynamics simulations	ACS National Meeting	Washington, D.C.	8/00
R. Friesner	DFT Methods for Inorganic Chemistry	National ACS Meeting	Washington, D.C.	8/00
R. Friesner	Development of a polarizable force field from <i>ab initio</i> quantum chemistry	National ACS Meeting	Washington, D.C.	8/00
R. Friesner	Modeling of metalloprotein active site chemistry using DFT methods	Sanibel meeting	Sanibel, FL	3/01
R. Friesner	QM/MM methods for modeling metalloprotein active site chemistry	Emerging methods workshop (ARO)	Aberdeen, MD	5/01
R. Friesner	Development of a polarizable force field from <i>ab initio</i> quantum chemistry	NIST Workshop on Thermophysical Properties of Fluids	Gaithersburg, MD	6/01
R. Friesner	QM/MM methods for modeling metalloprotein active site chemistry	ACS National Meeting	Chicago, IL	9/01
R. Friesner	Computational modeling of proteins	Nations Symposium	Georgia Tech Atlanta GA	10/01
R. Friesner	Development of a polarizable force field from <i>ab initio</i> quantum chemistry	Symposium on Modeling of Biological Systems	Florida State U. Tallahassee, FL	1/02
R. Friesner	Quantum chemical modeling of iron oxide surfaces	ACS National Meeting	Orlando, FL	3/02
R. Friesner	QM/MM methods for modeling metalloprotein active site chemistry	QBIC satellite meeting on quantum bioinorganic chemistry	Malmo Sweden	7/02
R. Friesner	Development of a polarizable force field from <i>ab initio</i> quantum chemistry	National ACS meeting	Boston, MA	8/02
R. Friesner	<i>Ab initio</i> electronic structure calculations of nanoscale materials	EPENS 02 Meeting	Montauk, NY	9/02
R. Friesner	QM/MM methods for modeling metalloprotein active site chemistry	Welch Meeting	Houston, TX	10/02

Significant Methods/Routines or Codes Developed

Incorporation of Ewald Coulomb Field in Jaguar

A model was developed to represent the Coulomb field due to ions in the solid state in *ab initio* electronic structure programs in Jaguar. This was the initial step in building a full QM/MM representation of ionic solids such as iron oxide. The method has been tested in iron oxide cluster calculations and performs satisfactorily.

Development of QM/MM Methods

Development of QM/MM methods for metalloprotein active sites continued. A particular focus is carrying out calculations in aqueous solution with a central goal being determination of pKa's of ionizable groups.

Development of a Polarizable Force Field Simulation Code

The polarizable force field simulation code has been modified to carry out NPT (as opposed to NVT) simulations. Various improvements in computational efficiency have been incorporated. A path integral code was also developed to perform polarizable simulations using a quantum representation of nuclear degrees of freedom.

Appendix A - First Year Activities and Accomplishments

Modeling of the Catalytic Cycle of Methane Monooxygenase

In collaboration with Prof. Steve Lippard's group at MIT, the project team has been studying the functioning of the enzyme MMO. This enzyme catalyzes the reaction of dioxygen and methane to form methanol, via a diiron site that undergoes a series of transitions through several different intermediate states, which have been elucidated by a variety of experimental techniques. For the first two states in the cycle (oxidized and reduced enzyme), the Lippard group has determined high-resolution structures via x-ray crystallography (Rosenzweig *et al.* 1993, 1997). For the remaining states (designated P and Q), there is less information; the spin and charge states of the iron atoms have been deduced via Mossbauer spectroscopy, and an EXAFS measurement on Q (Shu *et al.* 1997) has yielded an estimate for the Fe-Fe distance in that species of approximately 2.5Å, a remarkably short distance for a diiron system of this type.

Dunietz (2000) reported a large-basis-set DFT quantum chemical calculations using a approximately 100 atom model for all four catalytic states of the enzyme. This model included atoms in the second coordination shell, which the researchers found necessary to recover an accurate representation of the crystal structures upon geometry optimization. The reduced form of the enzyme is actually unstable without the surrounding network of protein hydrogen bonds to hold the diiron core and associated ligands in place. Previous modeling of MMO using DFT methods has employed smaller models that were unable to capture these effects. The model used in this work reproduces the reduced and oxidized enzyme structures to approximately 0.5Å RMSD.

Generation of a structure for the P intermediate from the reduced enzyme structure can be accomplished via addition of dioxygen and removal of the one of the structural waters that is weakly bound to the complex. In this project, both superoxo and peroxo structures, for which the products are lower in energy than the reactants by a few kcal/mol, were obtained. Studies of barrier heights, currently ongoing, will be needed to determine how these structures participate in the catalytic cycle; their relative energies with respect to each other are smaller than the accuracy of the DFT methodology. Finally, a number of possibilities for the Q intermediate, which has been the subject of intense speculation by many workers, were examined. In typical models for Q in the literature (Siegbahn 1999), a structural water ligated to one of the Fe atoms is displaced by a mobile carboxylate. The calculations from this project demonstrate that such a displacement is energetically highly unfavorable, because the water molecule forms very strong hydrogen bonds to charged oxygens in the core. An alternative model is presented in which the water is retained, but switches its hydrogen bonding pattern to bind with the mobile carboxylate, thereby inducing a short Fe-Fe distance. This model is thermodynamically consistent with Q having a lower energy than P and, hence, with the experimentally inferred catalytic cycle.

All of this work is summarized in Dunietz (2000), which provides a complete description of the four intermediates in the catalytic cycle at an atomic level of detail. In more recent unpublished work undertaken in this project, the step in which methane is converted into methanol was investigated. An activation energy of 18 kcal/mol was obtained, which is in good agreement with an estimate of 15–17 kcal/mol from experimental kinetic data using simple transition state

theory. Furthermore, although the transition state appears to be consistent with a hydrogen abstraction reaction, the project team has been able to develop an explanation for the retention of configuration seen in chiral ethane experiments by the Lippard group (Valentine *et al.* 1997), which is inconsistent with formation of a radical intermediate. As the hydrogen atom moves towards the oxygen, the barrier to rotation out of the C---H---O linear configuration becomes much lower, whereupon the system can proceed directly to products with no barrier.

References

Dunietz, BD, MD Beachy, Y Cao, DA Whittington, SJ Lippard, and RA Friesner. 2000. "Large Scale *Ab Initio* Quantum Chemical Calculation of the Intermediates in the Soluble Methane Monooxygenase Catalytic Cycle." *J. Am. Chem. Soc.*, 122(12):2828-2839.

Rosenzweig, AC, P Nordlund, PM Takahara, C Frederick, and SJ Lippard. 1995. Geometry of the Soluble Methane Monooxygenase Catalytic Diiron Center in Two Oxidation States. *Chem. & Biol.* 2, 6:409-418.

Rosenzweig, AC, CA Frederick, SJ Lippard, and P Nordlund. 1993. Crystal Structure of a Bacterial Non-Heme Iron Hydroxylase that Catalyses the Biological Oxidation of Methane. *Nature*, 366:537-543.

Shu, L, JC Nesheim, K Kauffman, E Münck, JD Lipscomb, and L Que, Jr. 1997. An Fe₂^{IV}O₂ Diamond Core Structure for the Key Intermediate Q of Methane Monooxygenase, *Sci.*, 275:515-518.

Siegbahn PEM. 1999. Theoretical Model Studies of the Iron Dimer Complex of MMO and RNR, *Inorg. Chem.*, 38:(12)2880-2889.

Valentine, AM, B Wilkinson, KE Liu, S Komar-Panicucci, ND Priestley, PG Williams, H Morimoto, HG Floss, and SJ Lippard. 1997. "Tritiated Chiral Alkanes as Substrates for Soluble Methane Monooxygenase from *Methylococcus capsulatus* (Bath): Probes for the Mechanism of Hydroxylation." *J. Am. Chem. Soc.*, 119:1818-1827.

Polarizable Force Field Development From *Ab Initio* Quantum Chemistry

Parameter Development from Quantum Chemical Calculations

The objective of constructing a polarizable force field development protocol has been to derive as much of the parametrization from *ab initio* quantum chemistry as is feasible while still retaining high accuracy and stability of the model. For small molecules where there is a vast amount of experimental data and relatively few parameters to adjust (water being the paradigmatic example), it is quite possible to build a very good force field with little or no input of quantum chemical data at all (although even in this case, some input of this type will undoubtedly improve the model). However, as one desires to treat increasingly complex structures, the amount of experimental data rapidly diminishes; in the extreme case, there is no experimental data at all. Furthermore, the number of parameters required increases substantially. The standard approach to parametrization of such molecules is to transfer parameters from small

molecule fragments, which are themselves optimized to fit experimental data. To achieve truly next generation accuracy, quantum chemical treatment of larger atom groups is necessary. The protocol developed in this project (Banks *et al.* 1999; Stern *et al.* 1999) integrates this type of computation in a seamless fashion, while transferring only the data from small molecule condensed phase simulations that cannot be reliably obtained from current levels of quantum chemistry.

The first component of the protocol is construction of an electrostatic model entirely from quantum chemical calculations, with no input of experimental data. The electrostatic model consists of a fixed charge distribution (which can be constructed from atom center charges, dipoles, and lone pairs) supplemented with a linear response polarization model that incorporates fluctuating charges and inducible dipoles. The project team realized that dipoles are absolutely necessary to treat certain important phenomena, such as bifurcated hydrogen bonds, and out-of-plane responses of aromatic rings; hence, a model based only on fluctuating charges, as was originally investigated, is not sufficiently robust to serve as a basis for a next generation polarizable force field for biomolecular simulation. Flexible parametrization software has been developed that, in response to user specifications, will place these components on specified atoms and functional groups. The fixed charge distribution is fit to the electrostatic field of the molecule as computed via high level quantum chemistry (DFT with large basis sets gives accurate charge distributions), while the linear response model is fit to quantum chemical response data generated by placing probes around the target molecule at strategic locations (*e.g.*, hydrogen bonding positions). Placement of the probes, collection of the linear response data, and fitting using data from an ensemble of conformations have been automated so that treatment of a new molecule requires minimal effort.

Given an electrostatic model, the second component of the force field that must be specified is the remainder of the nonbonded pair interactions. The great majority of widely used fixed-charge force fields employ an atom-atom Lennard-Jones, 6-12 functional form. A Lennard-Jones pair term is an acceptable functional form, but improved accuracy can be achieved by replacing the Lennard-Jones function with a three-parameter form combining a $1/r^6$ term with a short-range exponential term. The reasoning is as follows. The $1/r^6$ term represents the dispersive tail of the atom-atom interaction with the correct asymptotic form; it is the one feature of the potential that the project team believes cannot be reliably and conveniently fit to *ab initio* quantum chemistry for larger organic molecules with currently available quantum chemical methods. However, this term has a very strong effect on liquid state densities (or pressures, if one is carrying out simulations in the NVT ensemble as is typically done in an MD simulation). Therefore, the coefficient of this term was fit to the pressure for small molecule liquid state simulations.

The exponential parameters, in contrast, are fit to quantum chemical structures and binding energies for gas phase molecular dimers. In a hydrogen bonding interaction, there are two key parameters: the hydrogen bond distance and the binding energy. Accurate calculation of the hydrogen bond distance requires geometry optimization of the complex at the MP2 level; accurate calculation of the binding energy necessitates a single point MP2 calculation at the cc-pVQZ (-g) level, coupled to an extrapolation procedure. In this project, a standardized

protocol was developed using the LMP2 functionality in Jaguar, which reproduces values in the literature (Tsuzuki *et al.* 1999) computed at even higher levels of theory, and incorporating CCSD (T) calculations as well, which have a minimal effect on the binding energy) to within 0.1–0.2 kcal/mol, but require orders of magnitude less CPU time than alternative approaches because of the outstanding performance of pseudospectral local MP2 methods for large basis set calculations. Table 1 compares binding energies obtained from this protocol for five hydrogen bonded complexes with the results from Tsuzuki *et al.* 1999. It can be seen that the Jaguar results agree with these results within approximately 0.2 kcal/mol for all of the complexes, which is adequate for the target accuracy for this generation of the force field; furthermore, the calculations are fast enough the project team to easily generate results for the large number of molecular dimer interactions required to construct force fields for a wide range complex organic molecules. This protocol is used in all of the work that follows, including the polarizable protein force field for which a large number of hydrogen bonding energies were required.

Table 1. Comparison of binding energies of hydrogen bonded dimers, kcal/mol. The first column shows results obtained with the new LMP2 extrapolation method described above; the second and third columns are from Tsuzuki *et al.* (1999) and are obtained from ultralarge basis set calculations plus basis set extrapolation procedures using canonical MP2 methods with counterpoise corrections. The last column includes a CCSD (T) correction. See Tsuzuki *et al.* (1999) for details.

Dimer	Jaguar LMP2/extrap.	MP2/extrap.	CCSD(T)/extrap
water-methanol	4.80	4.99	4.90
water-dimethyl ether	5.74	5.70	5.51
water-formaldehyde	5.15	5.21	5.17
methanol-methanol	5.54	5.58	5.45
formic acid-formic acid (cyclic)	13.92	13.79	13.93

As a further test of this methodology, team members at PNNL computed dimer binding energies for four structures of the formamide dimer using canonical MP2 methods and extrapolation. These results agree with those obtained via the pseudospectral LMP2 protocol discussed above to within 0.2 kcal/mol, thus providing independent confirmation of the validity of that methodology. Additional benchmark studies along these lines will be performed in the next granting period.

With the non-bonded interactions defined, stretches, bends, and torsions can be fit to quantum chemical data using standard fitting methodologies.

Development of Polarizable Models for Pure Liquid Simulations

The initial objective of this work was to demonstrate that the protocol described above produces reasonably accurate potential functions that give a good description of condensed phase as well as gas phase properties. This philosophy is similar to that employed by Jorgensen and coworkers in developing the Optimized Potentials for Liquid Simulations (OPLS) force fields with the exception that a protocol is being validated via small molecule liquid state simulations rather

than directly transferring large numbers of parameters. The key differences between this approach and much of the work that has been published to date in this area are twofold: 1) a very small number of adjustable parameters are being used—only the dispersive tails of the atom-atom pair function are fit to reproduce condensed phase data, and 2) the protocol used in this project can readily be scaled up to treat molecules on the order of 20 to 50 atoms, and could be applied to even larger systems if desired.

The first step in the protocol is to choose the defining features of the model, such as the basis sets for the quantum chemical calculations and the specifics of the electrostatic description (*i.e.*, which atoms have dipoles, lone pairs, etc.). The intention is then to use the same protocol for all molecules to be studied, thus removing this as an adjustable aspect of the methodology. Once this is done, the electrostatic parameters are automatically generated as described above. The case of water was used to illustrate how the remainder of the nonbonded potential is constructed. The binding energy and geometry of the water dimer is first computed quantum chemically. Then, a series of models using different B ($1/r^6$) coefficients (the range of values to search becomes obvious after a small number of computational experiments) is produced, each of which reproduces the binding energy and hydrogen bonding distance of the dimer by adjustment of the two exponential parameters discussed above. Finally, all of these models are run in MD simulations and the B value that yields the pressure in best agreement with experimentally obtained values (new values can be tried if refinement is necessary) is selected.

Because only the pressure is fit, predictions for the remaining condensed phase properties constitute a severe test of the protocol. Table 2 summarizes results for three liquids. It can be seen that excellent agreement with experiment is obtained for a wide range of properties. The radial distribution function for water (not shown) also displays excellent agreement with experimental results. There are errors in predicting dynamical properties, such as the diffusion constant and dielectric constant. These quantities are extremely sensitive to the details of the model, and would be altered by more than the theory/experiment deviation by changing parameters by few tenths of a kcal/mol; this energy scale is approaching the accuracy of the input data. In contrast, previously developed force fields routinely have 1 to 2 kcal/mol errors for intermolecular hydrogen bonds, which probably can be systematically eliminated.

Table 2. Liquid state properties of three organic liquids as determined from polarizable force field simulations.

Water			Methanol			Formamide		
Energy (kcal/mol)	-9.89	-9.9	Pressure (kbar)	0.001	-0.026	Pressure (kbar)	0.001	0.200
Pressure (kbar)	0.17	0.00	Heat of vap (kcal/mol)	8.95	8.92	Heat of vap (kcal/mol)	15.55	15.41
Dielectric constant	80	78	Diffusion const (10^9 m ² /sec)	2.2	2.2	Diffusion const (10^9 m ² /sec)	0.52	0.42
Diffusion constant (10^9 m ² /s)	1.74	2.30	Dielectric const	33	23	O-N RDF peak dist (Å)	1.9	2.0
NMR relaxation time	2.5	2.1	O-O RDF peak height	3.1	3.4	O-HN RDF peak dist (Å)	2.9	2.9
			O-O RDF peak dist (Å)	2.8	2.8			

References

Banks, JL, GA Kaminski, R Zhou, DT Mainz, BJ Berne, and RA Friesner. 1999. "Parametrizing a Polarizable Force Field from *Ab Initio* Data. I: The Fluctuating Point Charge Model." *J. Chem. Phys.*, 110(2):741-754.

Stern, HA, GA Kaminski, JL Banks, R Zhou, BJ Berne, and RA Friesner. 1999. "Fluctuating Charge, Polarizable Dipole, and Combined Models: Parameterization from *Ab Initio* Quantum Chemistry." *J. Phys. Chem. B.*, 103(22):4730-4737.

Tsuzuki S, T Uchimaru, K Matsumura, M Mikami, and K Tanabe. 1999. "Effects of Basis Set and Electron Correlation on the Calculated Interaction Energies of Hydrogen Bonding Complexes: MP2/cc-pV5Z Calculations of H₂O-MeOH, H₂O-Me₂O, H₂O-H₂CO, MeOH-MeOH, and HCOOH-HCOOH Complexes." *J. Chem. Phys.*, 110(24):11906-11910.

Studies of Iron Oxide Surfaces (Columbia University, PNNL)

Development of a quantum chemical approach to the study of iron oxide surfaces has begun. Previous work in this area has primarily employed plane wave DFT methods using gradient corrected functionals. Because the focus of this work is on catalytic chemistry at the surface in solution, methods are being developed that can use more accurate hybrid functionals and also that can incorporate solvation effects via a self-consistent reaction field continuum solvation model. At the same time, the project team intends to move beyond a cluster model and incorporate a representation of the entire solid. An embedded cluster approach is being taken. As a first step, a modified version of the methodology of Derenzo and coworkers (Derenzo *et al.* 2000) for embedding a cluster in an ionic solid by using a classical representation of the ions has been developed. The idea is to determine the classical electrostatic potential from the solid using Ewald methods and then fit charges for a finite number of ions that reproduce that field. This methodology has been generalized to the semi-infinite boundary conditions required for a surface. The field of the ions is then incorporated into DFT cluster calculations, and the charges on the classical ions are iterated self-consistently and new results for the cluster are obtained. As a test of the method, ionization potentials of the cluster are being examined and compared with results obtained by Scheffler and coworkers using plane wave methods (Wang *et al.* 1998). Convergence of the ionization potential and other properties with cluster size and shape is being investigated. Future work will involve collaborations with PNNL researchers to apply this methodology to systems of environmental interest.

References

Derenzo, SE, MK Klintonberg, and MJ Weber. 2000. "Determining Point Charge Arrays That Produce Accurate Ionic Crystal Fields for Atomic Cluster Calculations." *J. Chem. Phys.*, 112(5):2074-2081.

Wang, XG, W Weiss, SK Shaikhutdinov, M Ritter, M Petersen, F Wagner, R Schlogl, and M Scheffler. 1998. "The Hematite (α -Fe₂O₃) (0001) Surface: Evidence for Domains of Distinct Chemistry." *Phys. Rev. Lett.* 81(5):1038-1041.

Molecular Modeling of Hydrophobic Organic Contaminants Uptake and Sequestration by Soil Organic Matter (California Institute of Technology and Howard University)

This project, which is being carried out by Dr. Mamadou S. Diallo (California Institute of Technology and Howard University) in collaboration with Dr. William A. Goddard III (California Institute of Technology), Dr. Jean-Loup Faulon (Sandia National Laboratories), and Dr. James H. Johnson, Jr. (Howard University), focuses on the uptake and sequestration of HOCs). The interactions of HOCs with SOM determine to a large extent their fate and transport in subsurface systems. These interactions can result in strong binding and slow release of HOCs from contaminated soil and sediments (Luthy *et al.* 1997). This apparent sequestration of HOCs has a significant impact on their mobility, reactivity and bioavailability in natural systems and engineered environmental systems (Luthy *et al.* 1997). Despite decades of investigations, the physical chemistry of HOC uptake and sequestration by SOM is not understood. The primary objective of this research is to develop a molecular-level understanding of the uptake of hydrophobic organic compounds (*e.g.*, phenanthrene and PCE) by two model SOM: Chelsea humic acid (CHA) and Lachine kerogen (LK). They are representative of geosorbents having significantly different geological ages and composition (Huang and Weber 1997). The results of the simulations will be used to address some key questions regarding the uptake and sequestration of HOCs by SOM:

1. What are the molecular-scale locations of HOCs sorbed to SOM?
2. How strongly bound are HOCs sorbed to SOM?
3. How mobile are HOCs sorbed to SOM?
4. What are the effects of HOC and SOM molecular properties on the binding strength and mobility of HOCs sorbed to SOM?

Computational Methods and Procedures

Because environmental geomacromolecules such as CHA are operationally defined as “solubility” classes of compounds, the development structural models for these compounds has been a major challenge to environmental chemists. Diallo *et al.* (2000a) have recently described a hierarchical approach for predicting the structure and properties of complex petroleum geomacromolecules such as asphaltenes. This novel approach combines experimental characterization data, computer assisted structure elucidation (CASE) and atomistic simulations to generate a sample of three-dimensional models that statistically represent the entire population of models that can be built from a given analytical data set (Diallo *et al.* 2000a). Diallo *et al.* (2000b) have recently used the CASE program developed by to generate a sample of three-dimensional structural models for CHA representative of the entire population of structural models that can be build from the available experimental data. These computer generated structural models for CHA are currently being used as starting three-dimensional structures for the MD simulations of the uptake and sequestration of HOCs by CHA.

Progress to Date

The first objective was to determine the extent to which the enthalpic interaction parameter could be correlated with experimentally measured parameters of HOC uptake and sequestration such as the Freundlich capacity parameter (K_f) and site energy heterogeneity factor (Huang and Weber 1997). The enthalpic interaction parameter (c_{ij}) between compounds i and j is given by (Kashihara *et al.* 1997):

$$\chi_{ij} = \frac{\Delta E_{ij}}{RT} \quad (1)$$

$$\Delta E_{ij} = \frac{1}{2}(E_{ij} + E_{ji}) - \frac{1}{2}(E_{ii} + E_{jj}) \quad (2)$$

where ΔE_{ij} is the differential energy of interaction of an unlike pair, E_{ij} is the condensed phase energy of the pair of compounds ij , R is the ideal gas temperature, and T is the system temperature. Using phenanthrene as model HOC, MD simulations were performed to estimate $\chi_{i\phi}$ as a function of temperature. A linear relationship was found between $\log \chi_{i\phi}$ and measured $\log K_f$ for phenanthrene sorption onto CHA. Not surprisingly, the site energy heterogeneity factor (n), a measure of the extent of “non linearity and hysteresis” of HOC sorption onto SOM was poorly correlated with c_{ij} . Additional MD simulations are being performed to test the validity of this correlation for several HOCs (*e.g.*, naphthalene, dichlorobenzene and trichlorobenze, etc.).

References

Diallo, MS, T Cagin, JL Faulon, and WA Goddard III. 2000a. “Thermodynamic Properties of Asphaltenes: A Predictive Approach Based on Computer Assisted Structure Elucidation and Atomistic Simulations.” *Asphaltenes and Asphalts. II. Developments in Petroleum Science Series*, 40 B, Chapter 5, pp. 103-127, Ed. Yen T F and Chilingarian, GV, Elsevier Science, Amsterdam, Netherland.

Diallo, MS, JL Faulon, WA Goddard III, JH Johnson, and WJ Weber. 2000b. “Molecular Modeling of Chelsea Humic Acid: A Hierarchical Approach Based on Experimental Characterization, Computer-Assisted Structure Elucidation, and Atomistic Simulations.” ACS, Division of Environmental Chemistry, Preprints of Extended Abstracts, 219th ACS National Meeting, March 26-30, 2000, San Francisco, California.

Huang, W and WJ Weber, Jr. 1997. “A Distributed Reactivity Model for Sorption by Soils and Sediments. 10. Relationships between Desorption, Hysteris and the Chemical Characteristics of Organic Domains.” *Env. Sci. Technol.*, 31:2562-2569.

Kashihara, S, M Iotov, S Dasgupta, G Gao, M Belmares, and WA Goddard III. 1997. “Gas Diffusion Through Polymers. Cerius2. Property Prediction.” *BIOSYM/Molecular Simulations*, in preparation.

Luthy, RG, GR Aiken, ML Brusseau, SD Cunningham, PM Gschwend, JJ Pignatello, M Reinhard, SJ Traina, WJ Weber, and JC Westall. 1997. "Sequestration of Hydrophobic Organic Contaminants by Geosorbents." *Env. Sci. Technol.*, 31:3341-3347.

Atomistic Models of Hydrodesulfurization Catalysts Over Co/MoS₂ (California Institute of Technology)

Objectives

The need to meet more stringent standards limiting the sulfur content of gas oils urges a deeper understanding of the mechanism by which sulfur-containing compounds are destroyed over hydrodesulfurization (HDS) catalysts. Thiols and sulfides are readily treated by the Co- and Ni-promoted MoS₂ catalysts currently employed, but satisfying the imminent restrictions will require the removal of the most refractory species, mainly alkyl-substituted polyaromatic thiophenes. Unfortunately, the nature of the catalyst's active sites is not well known. In this project, density functional calculations were undertaken on models of pure and Co-promoted MoS₂ along with relevant organic substrates to determine 1) the structural and electronic factors indispensable to the "CoMoS" active phase, 2) the mechanism by which thiophene and its derivatives react, and 3) how alkyl substituents promote and inhibit different branches of the reaction network.

Computational Methods and Procedures

Since the detailed surface reactions are modeled intermediate-by-intermediate, electronic structure calculations are used; DFT methods employing the B3LYP functional are utilized. Small clusters (three to six metal atoms) are used to imitate highly dispersed MoS₂ crystallites. The geometries of organic adsorbates and their binding sites are relaxed to find stable intermediates and transition states are located when relevant.

Progress to Date

To establish whether the specific pattern of reconstruction along the crystallite edges witnessed in STM studies (Helveg *et al.* 2000) remains in the finely dispersed MoS₂ particles believed to provide most of the active sites (Eijsbouts *et al.* 1993), three- and six-metal atom clusters were generated, and their behavior was studied in reconstruction and sulfidation. Exposing the reconstructed (10 $\bar{1}$ 0) face is preferred over the ($\bar{1}$ 010) face by the small clusters, as in the larger STM subjects, but more disordered states will be prevalent. Since vacancies at the edges MoS₂ crystallites left by abstracted sulfur atoms are believed to be the active sites, the number of vacancies present and the energy required to remove sulfur atoms from the edges are important quantities. We find that for these small models it is less appropriate to discuss a "vacancy formation energy" than an "equilibrium degree of sulfidation," as the binding energies of sulfur at edges depends strongly on the stoichiometry of the entire crystallite.

References

Eijsbouts, S, JIL Heinerma, and HJW Elzerman. 1993. "MoS₂ Structures in High-Activity Hydrotreating Catalysts. 1. Semiquantitative Method for Evaluation of Transmission Electron-Microscopy Results – Correlations between Hydrodesulfurization and Hydrodenitrogenation Activities and MoS₂ Dispersion." *Appl. Catal. A: General*, 105:53-68.

Eijsbouts, S, JIL Heinerma, and HJW Elzerman. 1993. "MoS₂ Structures in High-Activity Hydrotreating Catalysts. 2. Evolution of the Active Phazse During the Catalyst Life-Cycle-Deactivation Model." *Appl. Catal. A: General*, 105:53-68.

Helveg, S, JV Lauritsen, E Laegsgaard, I Stensgaard, JK Norskov, BW Clausen, H Topsoe, and F Besenbacher. 2000. "Atomic-Scale Structure of Single-Layer MoS₂ Nanoclusters." *Phys. Rev. Lett.*, 84(5):951-954.

Dendrimers: A New Class of High Capacity Chelating Agents for Metals Ions

Dendrimers are highly branched polymers with controlled composition and architecture consisting of three structural components: 1) a core, 2) interior branch cells, and 3) terminal branch cells. Dendritic polymers can provide hundreds of equivalent metal binding sites all within one molecule. These observations suggest strategies of selectively extracting toxic metals by designing the dendrimers to have the proper *external sites* to partition to the region of soil or aqueous medium containing the ions and the proper internal site to selectively complex the toxic metals. The overall size could then be chosen so that the molecule could be filtered out of an aqueous system or designed to remain in place for a buried system. Despite the attractiveness, the relationship between the binding of these systems and the number of bound metals, the pH, and the composition of the dendrimer needs to be characterized.

PAMAM dendrimers possess functional nitrogen and amide groups arranged in regular "branched-upon-branched" patterns that are displayed in geometrically progressive numbers as a function of generation level. Terminal groups of PAMAM dendrimers may be any organic substituent such as primary amines, carboxylic groups, etc. In aqueous solutions, PAMAM dendrimers can serve as high capacity nanoscale containers for toxic metal ions such as Cu (II). The first studies carried out were for the binding of Cu (II) ions to poly (amidoamine) (PAMAM) dendrimers in aqueous solutions. Compared to traditional chelating agents (*e.g.*, triethylene tetramine) and macrocycles (*e.g.*, cyclams) with nitrogen donors, which can typically bind only one Cu (II) ion per molecule, a generation eight (G8) PAMAM dendrimer can bind 153 ± 20 Cu (II) ions per molecule. This clearly illustrates a distinct advantage of dendrimers over traditional chelating agents and macrocycles; that is, the covalent attachment of nitrogen ligands to conformationally flexible PAMAM chains enclosed within a nanoscopic structure results in substantial increase in binding capacity.

In more general work on dendrimer simulation technology, the three-dimensional molecular structure of various dendrimers was determined with a continuous configurational Boltzmann biased direct Monte Carlo method. The energetic and structural properties of these dendrimers were studied using MD methods after annealing these molecular representations.

Appendix B - Full Report of Second Year Activities and Accomplishments

Modeling of the Catalytic Cycle of Methane Monooxygenase and other Iron Containing Enzymes

Methane Monooxygenase

In the first year of this grant, initial work on modeling the catalytic cycle of methane monooxygenase (MMO), using DFT quantum chemical methods and large basis sets, was completed. This enzyme catalyzes the reaction of dioxygen and methane to form methanol via a diiron site which undergoes a series of transitions through several different intermediate states, which have been elucidated by a variety of experimental techniques (Rosenzweig *et al.* 1993, 1995; Shu *et al.* 1997; Siegbahn 1999). Dunietz *et al.* (2000) describes the initial computational characterizations of these intermediates, which are in good agreement with the available experimental data. A major aspect of this work was the use of an approximately 100-atom model of the active site that was able to reliably represent structural changes in the second coordination shell, which are essential in understanding the catalytic cycle.

In the past year, the project team has worked on developing a detailed description of the dynamics of the hydroxylation reaction for methane and ethane substrates; this work is described in Gherman (2001) and Guallar (2002). All of this work used the same active site model and quantum chemical approach as in Dunietz *et al.* (2000). Following is a brief summary of the results:

1. The transition state for the reaction of methane with the intermediate Q was determined first. This transition state is a classic hydrogen atom abstraction transition state in which the hydrogen is roughly equidistant between the methane carbon and a bridging oxygen of the core of the Q intermediate. The transition state is shown in Figure 1.
2. The project team then addressed in detail the question of how methanol is formed once the transition state is produced. A reduced dimensional model potential surface was constructed in four dimensions (C-H distance, CHO angle, C-O distance, and, in the case of ethane, C-C torsional angle) using restrained quantum chemical geometry optimizations in which the projected degrees of freedom were adiabatically minimized. Quantum dynamics simulations on this surface using semiclassical trajectory methods were then carried out. An analysis of these trajectories revealed that there are two major channels for the formation of methanol. In the first, a bound substrate radical species is formed as an intermediate, along with an O-H bond on the core of the active site complex. When the O-H bond rotates upwards by roughly 90 degrees (this step is rate determining and has a barrier of approximately 4 kcal/mol), the carbon of the substrate will rapidly move forward to form a C-O bond. Once the C-O bond is formed, the resulting alcohol moiety will spontaneously leave the active site complex, completing the catalytic process. In the second channel, the -OH rotation occurs prior to the formation of the bound radical intermediate (this has a barrier of only 1 kcal/mol) and methanol formation now takes place (via rapid formation of the C-O bond, as in the bound radical channel) as a concerted reaction without passing through the intermediate.

3. These results provide an explanation of experiments with chiral ethane (one carbon is substituted with H, D, and T) carried out by the Lippard group (Valentine 1997). These experiments show that the chirality of the optically active carbon is retained approximately 70% of the time. Retention of configuration arises from two sources. First, the concerted channel, which accounts for approximately 16% of all trajectories, does not allow racemization. Secondly, there is a barrier to rotation around the C-C torsion of the ethane in the bound radical state arising from steric interactions with the protein. Because of this barrier, there is some retention of configuration in the bound radical channel despite the extremely low intrinsic barrier to rotation of the ethyl radical in the gas phase (a few tenths of a kcal/mol) as compared to the barrier to -OH rotation in the protein (approximately 4 kcal/mol). Our quantitative computation of the per cent retention of configuration is in good qualitative agreement with the experimental data (accurate quantitative prediction is not feasible at present due to uncertainties about the initial conditions for the semiclassical simulations).

Finally, modeling of the kinetics of the remaining reactions in MMO, involving addition of dioxygen to the reduced enzyme to form the intermediate P and conversion of P into the catalytically competent species, Q, has been initiated. Transition states for all of these processes have been produced, and the vibrational frequencies of these transition states as well as structures and energetics are being characterized, allowing computation of zero point energies and isotope effects. A publication describing this work is being prepared.

Modeling of Cytochrome P450 and Hemerythrin

Work on both of these enzymes is in its initial stages. In contrast to MMO, these calculations are being pursued using QM/MM methods optimized for modeling of protein active site chemistry; these methods are described in detail in Murphy *et al.* 2000. A brief report of work on each system to date follows.

Cytochrome P450

Cytochrome P450 carries out hydroxylation reactions, as does MMO, but it uses a heme group as the catalytic core rather than a nonheme iron. The P450cam structure is the focus of this study, as it is the best characterized experimentally, particularly in view of the recent work of Petsko, Sligar, and coworkers who used cryogenic crystallography to obtain structures of previously unobservable intermediates (Schlichting *et al.* 2000). In the initial stage of this project, geometry optimizations of several of the known intermediates in the catalytic cycle were performed. Good agreement was obtained with the experimental x-ray structures (approximately 0.5 Å) and ordering of spin states under different conditions. At present, the focus of the project is on the protonation of the dioxygen adduct of the heme group. The research team recently ascertained that the initial protonation comes from a proximate threonine residue rather than one of the structural waters. A transition state for this process is being computed to obtain a more quantitative description of the reaction. The second protonation step is more complicated in that it requires transport of a proton along a chain of water molecules to the location of the dioxygen.

That process is being modeled as well, and an initial paper describing this work will be completed during the coming year.

Hemerythrin

Hemerythrin is a protein that reversibly binds dioxygen. To accomplish this, the free energy of the reaction must be tuned to a small value. This must be accomplished by the protein environment. We have been performing computations of the binding enthalpy of dioxygen to hemerythrin using both a quantum chemical active site model and via QM/MM calculations. The results are considerably different in the protein environment. We are in the process of carrying out a more careful treatment of the protein environment (*e.g.*, inclusion of solvation, proper assignment of protonation states of ionizable groups) which will be necessary if accurate results are to be obtained.

References

Dunietz, BD, MD Beachy, Y Cao, DA Whittington, SJ Lippard, and RA Friesner. 2000. "Large Scale *Ab Initio* Quantum Chemical Calculation of the Intermediates in the Soluble Methane Monooxygenase Catalytic Cycle." *J. Am. Chem. Soc.*, 122(12):2828-2839.

Gherman, BF, BD Dunietz, DA Whittington, SJ Lippard, and RA Friesner. 2001. "Activation of the C-H bond of Methane by Intermediate Q of Methane Monooxygenase: A Theoretical Study." *J. Am. Chem. Soc.*, 123(16):3836-3837.

Guallar, V, BF Gherman, WH Miller, SJ Lippard, and RA Friesner. 2002. "Dynamics of Alkane Hydroxylation at the Non-heme Diiron Center in Methane Monooxygenase." *J. Am. Chem. Soc.*, 124:3377-3384.

Rosenzweig, AC, P Nordlund, PM Takahara, C Frederick, and SJ Lippard. 1995. "Geometry of the Soluble Methane Monooxygenase Catalytic Diiron Center in Two Oxidation States." *Chem & Biol*, 2(6):409-418.

Rosenzweig, AC, CA Frederick, SJ Lippard, and P Nordlund. 1993. "Crystal Structure of a Bacterial Non-Heme Iron Hydroxylase that Catalyses the Biological Oxidation of Methane." *Nature*, 366:537-543.

Shu, L, JC Nesheim, K Kauffman, E Münck, JD Lipscomb, and L Que, Jr. 1997. "An Fe^I₂V^O₂ Diamond Core Structure for the Key Intermediate Q of Methane Monooxygenase." *Sci.*, 275:515-517.

Siegbahn, PEM. 1999. Theoretical Model Studies of the Iron Dimer Complex of MMO and RNR. *Inorg. Chem.*, 38(12):2880-2889.

Valentine, AM, B Wilkinson, KE Liu, S Domar-Panicucci, ND Priestley, PG Williams, H Morimoto, HG Floss, and SJ Lippard. 1997. "Tritiated Chiral Alkanes as Substrates for Soluble Methane Monooxygenase from *Methylococcus capsulatus* (Bath): Probes for the Mechanism of Hydroxylation." *J. Am. Chem. Soc.*, 119:1818-1827.

Polarizable Force Field Development From *Ab Initio* Quantum Chemistry

Benchmark Quantum Chemical Results for Hydrogen Bonding Energies of the Formamide and Acetamide Dimers (PNNL, Columbia University)

A key aspect of the force field development strategy is the use of benchmark quantum chemical calculations of hydrogen bonding interactions to parametrize the atom-atom pair terms in the force field. Benchmark calculations at the MP2 and coupled cluster level have been performed for a series of formamide and N-methylacetamide dimers (Vargas *et al.* 2001). For these molecules, comparison with substantially less expensive local MP2 calculations with basis set extrapolation show good agreement. Testing of the LMP2 protocol for more complex molecules such as benzene will be carried out during the next granting period.

Second Generation Polarizable Force Field Model

One important objective of our polarizable force field development methodology is to be able to determine polarizable force field parameters for an arbitrary molecule without the use of experimental data. While this approach will not achieve the accuracy of direct fitting to experimentally obtained data when such data is available, it allows models to be constructed when little or no experimental data is available, and thus straightforward application to a wide range of chemistry (including ultimately inorganic systems).

The basic structure of the methodology used in this project has been described in a series of previous publications (Banks *et al.* 1999; Stern *et al.* 1999), as well as in the report from the first year of this grant. The electrostatic model, including point charges, dipoles, and dipolar polarizabilities, is fit to quantum chemical DFT calculations with no fitting to experiment. This model is then supplemented with an atom-atom pair potential. Traditional force fields employ a Lennard-Jones 6-12 functional form for the pair potential. The research team found that this functional form does not have enough flexibility to allow both the short and long-range parts of the potential to be correctly fit to computational and experimental data. In the second generation model, a potential containing the following three terms is used:

1. A $1/r^{12}$ term is used to prevent polarization collapse at short range.
2. A $1/r^6$ term is used to represent long-range dispersion.
3. A term $C\exp(-\lambda r)$ is adjusted to fit the well depth and hydrogen bonding distance of a molecular pair (for small molecules, a dimer of the molecule to be modeled).

The $1/r^{12}$ term is adjusted to become zero in the allowed intermolecular interaction region. The parameters C and λ are fit to quantum chemical pair data. The B coefficient of $1/r^6$ term is now determined for each atom (C, O, N, H, etc.) from fitting to condensed phase experiment for a

small number of molecules. The idea is to generate universal B parameters for each atom (or for a small number of atom types, such as aromatic carbon as opposed to aliphatic carbon) which are then transferable to new molecules. If this can be done, then new molecules can be modeled without recourse to any experimental data.

Initial results are presented below in Table 1. The calculation began by fitting B parameters for aliphatic carbon and hydrogen to reproduce liquid state results for small hydrocarbons. Next, the B value for oxygen was fit to liquid-phase data for methanol. The value for nitrogen was then interpolated between C and O (nitrogen having an intermediate number of lone pairs). Thus the remaining molecules in the table were all simulated without fitting experimental data. As can be seen, agreement with liquid state heat of vaporization and density is quite good, at approximately the accuracy achieved by the OPLS-AA force field, which is explicitly adjusted to fit these quantities. While many more tests remain to be carried out (*e.g.*, simulation of crystal data, prediction of solvation free energies in aqueous solution), these results are highly encouraging.

Table 1. Comparison of calculated and experimental thermodynamic properties for a number of molecular liquids using the second generation polarizable force field development protocol.

System	Dimer E, kcal/mol		Dimer R, Å		DH _{vap} , kcal/mol		Molecular Volume, Å ³		Density, g/cm ³	
	pff	QM	pff	QM	pff	Exptl	pff	exptl	pff	exptl
CH ₄ ^a	-0.44	-0.50	3.86	4.06	1.89	1.96	62.2	62.8	0.428	0.424
C ₂ H ₆ ^a					3.32	3.62	94.4	91.5	0.529	0.546
C ₃ H ₈ ^a					4.79	4.49	123.9	126.0	0.591	0.581
C ₄ H ₁₀ ^a					5.62	5.35	157.2	160.3	0.614	0.602
CH ₃ OH ^b	-5.63	-5.59	2.81	2.80	8.84	8.95	67.0	67.7	0.794	0.786
Acetone ^b	-5.83	-5.76	3.37	3.28	7.73	7.48	128.5	123.0	0.750	0.784
Acetamide ^b	-12.9	-12.8	3.02	2.94	14.1	13.4	111.7	109.3	0.878	0.897
DME ^b	-1.45	-1.46	3.46	3.09	5.73	5.14	106.0	104.1	0.722	0.735

^aExperimental data from *J. Phys. Chem.*, 98, 1994, 13077. Liquid simulations done at boiling temperatures: -161.49°C for methane, -88.63°C for ethane, -42.1°C for propane, -0.5°C for butane.
^bExperimental data from *J. Am. Chem. Soc.*, 118, 1996, 11225. Liquid simulations done at 25°C for methanol, 25°C for acetone, 221.15°C for acetamide, -24.60°C for DME. Acetamide liquid state density compared with the OPLS-AA result from the same reference

Development of a Polarizable Force Field for Water

A new polarizable force field for water has been developed directly from *ab initio* quantum chemistry, using no adjustable parameters (Stern *et al.* 2001). Water is a crucial component of the environmentally relevant systems to be modeled. Whereas the work described above concentrated only on the heat of vaporization and the density in the liquid state, this work expanded the scope to examine dynamical properties such as the dielectric constant ϵ , diffusion constant D, and NMR relaxation time τ .

The table below presents results obtained with two different basis sets (augmented triple-*zeta* [TZ] and quadruple-*zeta* [QZ], Dunning correlation consistent basis sets). For comparison, results of the most recent fixed charge force field of Jorgensen and coworkers, which was fit directly to experimental results, are also presented. The TZ results are in closer agreement with experimental results than the QZ results. There are two possible reasons for this: 1) use of classical rather than quantum simulation of nuclear motion and 2) diminishment of polarization response in the condensed phase due to Pauli exclusion. Both effects would tend to reduce the dielectric constant and increase the diffusion constant. In the work discussed above, the TZ basis was used as a heuristic approach to account for these effects. Here, the TZ basis is augmented with diffuse functions that may account for the fact that the dielectric constant is still slightly too large. Finally, in data not shown here, the TZ model does a reasonably good job of predicting the temperature dependence of water density, and arguably better than TIP5P, which was fit to experimental results to reproduce this property.

In general, one cannot expect direct fitting to *ab initio* data to yield very highly accurate results for dynamical properties (note that the internal energy and density are in very good agreement, however), for the reasons given above. Thus, a water model to be used in production runs should be further optimized to fit the available experimental data. On the other hand, the dynamical properties reported below in Table 2 for the TZ model are respectable, and accuracy at this level will suffice for many purposes and is likely the best obtainable when suitable experimental data for parametrization is not available.

Table 2. Liquid state properties of various water models.

Property	POL5/TZ	POL5/QZ	TIP5P	Expt.
U(kcal/mol)	-9.9	-10.3	-9.8	-9.9
R(gm/cm ³)	0.997	0.998	0.999	0.997
ϵ_0	98 +/- 8	105 +/-7	82	78
D (10 ⁻⁹ m ² /sec)	1.81	1.25	2.6	2.3
t (ps)	2.6	4.0	1.4	2.1

References

- Banks, JL, GA Kaminski, R Zhou, DT Mainz, BJ Berne, and RA Friesner. 1999. "Parametrizing a Polarizable Force Field from *Ab Initio* Data. I: The Fluctuating Point Charge Model." *J. Chem. Phys.*, 110(2):741-754.
- Stern, HA, GA Kaminski, JL Banks, R Zhou, BJ Berne, and RA Friesner. 1999. "Fluctuating Charge, Polarizable Dipole, and Combined Models: Parameterization from *Ab Initio* Quantum Chemistry." *J. Phys. Chem. B*, 103(22):4730-4737.
- Stern, HA, F Rittner, BJ Berne, and RA Friesner. 2001. "Combined Fluctuating Charge and Polarizable Dipole Models: Application to a Five-Site Water Potential Function." *J. Chem. Phys.*, 115:2237-2251.

Vargas, R, J Garza, RA Friesner, H Stern, BP Hay, and DA Dixon. 2001. "Strength of the N-H ·O=C and C-H O=C Bonds in Formamide and N-methylacetamide Dimers." *J. Phys. Chem. A*, 105(20):4963-4968.

Studies of Iron Oxide Surfaces (Columbia University, PNNL)

Development of a quantum chemical approach to the study of iron oxide surfaces has begun. Previous work in this area has primarily employed plane wave DFT methods using gradient corrected functionals. Because catalytic chemistry at the surface in solution is the area of interest, methods that can use more accurate hybrid functionals and also can incorporate solvation effects via a self-consistent reaction field continuum solvation model are being developed. At the same time, the project team intends to move beyond a cluster model and incorporate a representation of the entire solid. An embedded cluster approach is being taken. As a first step, a modified version of the methodology of Derenzo and coworkers (Derenzo *et al.* 2000) for embedding a cluster in an ionic solid by using a classical representation of the ions has been developed. The idea is to determine the classical electrostatic potential from the solid using Ewald methods and then fit charges for a finite number of ions which reproduce that field. This methodology has been generalized to the semi-infinite boundary conditions required for a surface. The field of the ions is then incorporated into DFT cluster calculations, and the charges on the classical ions are iterated self-consistently and new results for the cluster are obtained.

As a test of the method, the ionization potentials of the cluster are currently being examined and are being compared to the results obtained by Scheffler and coworkers using plane wave methods (Wang *et al.* 1998) and also with available experimental data. The convergence of large clusters has been a significant challenge but one that must be accomplished. The surface area and depth of the cluster has been systematically increased until a converged result for the work function has been obtained. The computed value of 5.6 eV is within a few tenths of an eV of two different experimental measurements (which are in good agreement with each other). In contrast, the project team disagrees with Scheffler's computed value of 4.3 eV, and has established that the particular DFT functional used cannot account for a difference of this size, nor can basis set effects; the cause of the discrepancy is therefore at present unknown.

References

Derenzo, SE, MK Klintonberg, and MJ Weber. 2000. "Determining Point Charge Arrays that Produce Accurate Ionic Crystal Fields for Atomic Cluster Calculations." *J. Chem. Phys.*, 112(5):2074-2081.

Wang, XG, W Weiss, SK Shaikhutdinov, M Ritter, M Petersen, F Wagner, R Schlogl, and M Scheffler. 1998. "The Hematite (α -Fe₂O₃) (0001) Surface: Evidence for Domains of Distinct Chemistry." *Phys. Rev. Lett.*, 81(5):1038-1041.

Molecular Modeling of Hydrophobic Organic Contaminants Uptake and Sequestration by Soil Organic Matter (California Institute of Technology and Howard University)

This project focuses on the uptake and sequestration of HOCs by SOM and dissolved organic matter (DOM). During the last 12 months, significant progress has been made toward development and validation of a multiscale modeling strategy for predicting HOC binding to SOM and DOM. First, the project showed that its hierarchical approach for modeling complex petroleum geomacromolecules such as asphaltenes can be used to develop three-dimensional structural models for CHA using experimental characterization data provided by Dr. Weilin Huang of Drexel University and Dr. Walter Weber of the University of Michigan. Building upon this work, a multiscale modeling approach for predicting the binding of organic solutes to DOM was developed. This novel approach combines CASE with atomistic simulations and Flory-Huggins solution theory to estimate the binding constants of organic solutes to dissolved humic acid *without using any empirically derived thermodynamic parameter*. Recently, work began on the development a second generation of three-dimensional models for CHA based on a more extensive characterization data set in collaboration with Dr. Andre Simpson and Dr. Patrick Hatcher of Ohio State University (1-D/2-D ¹H and ¹³C NMR spectroscopy and mass spectrometry [MS]) and Mr. Paul Gassman [FT-IR spectroscopy] of the EMSL at PNNL. A manuscript describing this work will be submitted to *Environmental Science and Technology*.

Dendritic Nanoscale Chelating Agents: (California Institute of Technology, Howard University, and University of Michigan)

This project explores the fundamental science of metal ion uptake by poly(amidoamine) dendrimers in aqueous solutions. Dendrimers are monodisperse and highly branched nanostructures with controlled composition and architecture. Poly(amidoamine) (PAMAM) dendrimers possess functional nitrogen and amide groups arranged in regular “branched-upon-branched” patterns. This high density of nitrogen ligands enclosed within a nanoscale container makes PAMAM dendrimers particularly attractive as high capacity chelating agents metal ions. During the last 12 month, significant progress has been made in the experimental characterization of metal ion uptake by PAMAM dendrimers. Bench scale laboratory measurements strongly suggest that, in aqueous solutions, dendrimers such as PAMAM G3, G4, and G5 can serve as nanoscale containers for the partitioning of toxic metal ions such as Cu(II) and Ag(I) and redox active metal ions such as Fe(II). Preliminary x-ray absorption spectroscopy (XAS) studies carried out at the Stanford Schynchotron Radiation Laboratory showed different coordination environments for Cu(II) ions bound to PAMAM dendrimers. These results suggest that the uptake of Cu(II), Fe(II) and Ag(I) by PAMAM dendrimers is more like a partition process rather than the traditional chelation process by which polydentate ligands, linear polymeric ligands, and macrocycle bind metal ions in aqueous solutions. To further assess this behavior, the pseudophase theory of solute binding was extended to surfactant micelles in aqueous solutions, and the uptake of metal ion by PAMAM dendrimers is expressed through its dendrimer-water partition coefficient (K_{dwi}):

$$K_{dwi} = \frac{X_{di}}{X_{ai}}$$

In all cases, the values of $\text{Log } K_{\text{dw}}$ have been found to be greater than 5.0 for Cu(II), Fe(II), and Ag(I) ions in aqueous solutions of G3, G3.5, G4, and G5 PAMAM dendrimers at $\text{pH} = 7.0$. A manuscript describing these results will be submitted to *Journal of Colloid and Interface Science*.

Application of the MS-Q Force Field to Clay Minerals, Oxides and Zeolites (California Institute of Technology)

Molecular modeling studies of mineral structures (clays, oxides, zeolites) might help stimulate research in the areas of pollution remediation, clean-up, as well as more efficient catalytic processes that avoid the production of undesirable byproducts. In addition clay/polymer nanocomposites are rapidly becoming an important commercial source of materials for low gas permeability barriers. In particular the structure of confined polymers, such as polystyrene, polyethylene, polyethyleneterephthalate, polyethylene oxide etc., have been investigated experimentally (Gianellis *et al.* 1998; Gilman *et al.* 1999) as well as modeled using Monte Carlo methods (Skipper *et al.* 1995). The molecular dynamics of some clay minerals has been recently undertaken and enthalpies of adsorption for hydrocarbons on smectite clay and trichloroethene in kaolinite and pyrophyllite humic substances in soil (oxidized lignins) calculations have been performed. Molecular modeling studies of clay, oxides, and zeolites are focused, in addition to structural properties, on the mineral water interface, hydration, and the role of ions in swelling process.

Here, the use of a Morse-Charge Equilibration Force Field (MS-Q FF) (Cagin *et al.* 1998), originally developed for the bulk oxides SiO_2 and Al_2O_3 , to model various minerals and their properties is described. The MS-Q FF (Rappé *et al.* 1991) could be employed for studies on the adsorption thermodynamics of water, hydrocarbons, and polar organic compounds (surfactants) (Kitao *et al.* 1999). Understanding the interaction between organic molecules and minerals is important, among other applications, for enhanced oil recovery processes (Hwang *et al.* 2001). Surface energies and the water and oil wettabilities of the exposed clay minerals are of great importance in heavy oil retention processes. At the same time, the MS-Q FF was successfully used to predict structural parameters for various minerals (talc, vermiculite, sepiolite, muscovite, kaolinite, mica, [Demiralp *et al.* 1999] amphibole, calcite, chlorite, gibbsite, jadeite, montmorillonite, pyrophyllite) and zeolites (ZSM-11, ZSM-12, alpha cristobalite). To further predict physical and chemical properties of the above-mentioned compounds, structural information is of fundamental importance. From the comparison of predicted and experimental (x-ray diffraction) data for muscovite (Table 3) (Fyfe *et al.* 1990) and ZSM-12 (Table 4) (Demiralp 1989) one can see that the predicted values are in good agreement with the experimental data. Future research will focus on calculating clay exfoliation energetics, clay-surfactants and polymer compatibilities, interactions of ions (NH_4^+) and precious metals (Pd, Pt) in zeolites, and catalytic processes involving metal oxides.

Table 3. Experimental [9] and predicted monoclinic (C 2/c) structural parameters of muscovite.

Parameter	Experimental	MSQ-predicted
A (Å)	5.19	5.33
B (Å)	9.01	9.19
C (Å)	20.07	20.59
α deg)	90	89.99
β deg)	95.77	95.81
γ deg)	90	90
V (Å ³)	933.43	1002.67
ρ g cm ⁻³)	2.83	2.53

Table 4. Experimental [10] and predicted monoclinic (C 2/c) structural parameters of ZSM-12 zeolite.

Parameter	Experimental	MSQ-predicted
A (Å)	24.86	24.68
B (Å)	5.01	4.93
C (Å)	24.33	24.16
α deg)	90	89.99
β deg)	107.72	107.73
γ deg)	90	90
V (Å ³)	2887.92	2802.15
ρ g cm ⁻³)	1.94	1.99

References

- Cagin, T, E Demiralp, and WA Goddard III. 1998. "Pressure Induced Phase Transformations in Silica." In *Microscopic Simulation of Interfacial Phenomena in Solids and Liquids*, Symposia Series 492, pp. 287-292, SR Phillpot. Materials Research Society, Pittsburgh, Pennsylvania.
- Demiralp, E, T Cagin, and WA Goddard III. 1999. "Morse Stretch Potential Charge Equilibrium Force Field for Ceramics: Application to the Quartz-Stishovite Phase Translation and to Silica Glass." *Phys. Rev. Lett.*, 82:1708-1711.
- Demiralp, E, T Cagin, NT Huff, and WA Goddard III. "The MS-Q Force Fields for Metal Oxides." To be submitted.
- Fyfe, CA, H Gies, GT Kokotailo, B Marler, and DE Cox. 1990. "Crystal Structure of Silica-ZSM-12 by the Combined Use of High-Resolution Solid-State MAS NMR Spectroscopy and Synchrotron X-ray Powder Diffraction." *J. Phys. Chem.*, 94:3718-3721.
- Catti, M, G Ferrarus, and G Ivaldi. 1989. "Thermal Strain Analysis in the Crystal-Structure of Muscovite at 700-Degrees-C." *European Journal Mineralogy*, 1:625-632.
- Gianellis, EP, R Krishnamoorti, and E Manias. 1998. "Polymer-Silicate Nanocomposites: Model Systems for Confined Polymers and Polymer Brushes." *Advances in Polymer Science*, 138:107-147.
- Gilman, JW, AB Morgan, R Harris, E Manias, EP Giannelis, and M Wuthenow. 1999. "Polymer Layered-Silicate Nanocomposites: Polyamide-6, Polypropylene and Polystyrene." In: *Proceedings of the New Advances in Flame Retardant Technology, Fire Retardant Chemical Association*, October 24-27, 1999, Tucson, Arizona, Fire Retardant Chemicals Assoc., Lancaster, Pennsylvania, 9-22 pp.

Hwang, S, M Blanco, E Demiralp, T Cagin, and WA Goddard III. 2001. "The MS-Q Force Field for Clay Minerals: Application to Oil Production." *J. Phys. Chem B*, 105:4122-4127.

Kitao, O, E Demiralp, T Cagin, S Dasgupta, M Mikami, K Tanabe, and WA Goddard III. 1999. "Theoretical Studies on VPI-5. 3. The MS-Q Force Field for Aluminophosphate Zeolites." *Comp. Mater. Sci.*, 14:135-137.

Rappé, AK and WA Goddard III. 1991. "Charge Equilibration for Molecular Dynamics Simulations." *J. Phys. Chem.*, 95:3358-3363.

Skipper, NT, G Sposito, and FRC Chang. 1995. "Monte-Carlo Simulation of Interlayer Molecular-Structure in Swelling Clay-Minerals. 1. Methodology." *Clays and Clay Minerals*, 43:285-293.

Skipper, NT, G Sposito, and FRC Chang. 1995. "Monte-Carlo Simulation of Interlayer Molecular-Structure in Swelling Clay-Minerals. 2. Monolayer Hydrates." *Clays and Clay Minerals*, 43:294-303.

Appendix C - Full Report of Third Year Activity (June 2001-present)

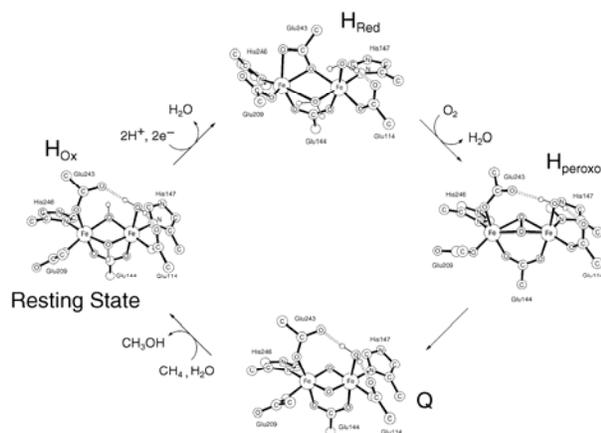


Figure 1. Catalytic cycle of the methane hydroxylation reaction promoted

level description of the structures and energies of all of the stationary points (minima and transition states) implied by Figure 1.

The great majority of the calculations for MMO in this project have been carried out using QM methods applied to a model containing approximately 100 atoms. The attack of dioxygen on the reduced enzyme and the formation of intermediate P have been investigated in detail. The first step in the reaction is the displacement of a weakly bound water molecule by dioxygen. This reaction can produce either a superoxo or peroxo intermediate, with a variety of spin couplings. The entire manifold of charge and spin states has been enumerated in detail; this report focuses only on what has been identified to be the lowest energy pathway. The main features of this pathway, along with their associated energies, are shown schematically in Figure 2.

The initial superoxo or peroxo complexes of dioxygen with the active site model are significantly higher in free energy than the initial state consisting of dioxygen in water and reduced enzyme. Note that entropy terms associated with both water and dioxygen and the solvation free energies of water and dioxygen in water must be taken into account in computing free energy differences; in this project, this in an approximate but reasonable

Modeling of Iron Containing Protein Active Sites

Methane Monooxygenase (MMO)

MMO is a bacterial enzyme that converts methane into methanol. The catalytic cycle of MMO has been elucidated via an extensive series of experiments (Rosenzweig *et al.* 1993; Rosenzweig *et al.* 1995; Merx *et al.* 2001; Liu *et al.* 1995; Shu *et al.* 1997; Liu *et al.* 1993; Valentine *et al.* 1997; Choi *et al.* 1999), and is shown in Figure 1, along with current proposed structures of the catalytic core for each intermediate in the cycle. The objective of this project has been to provide an accurate atomic-

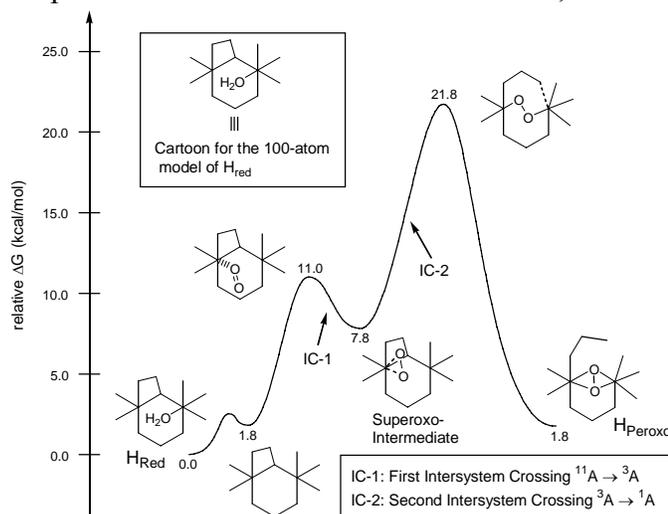


Figure 2. Computed reaction energy profile for the reductive activation of O₂ using the approximately 100-atom model.

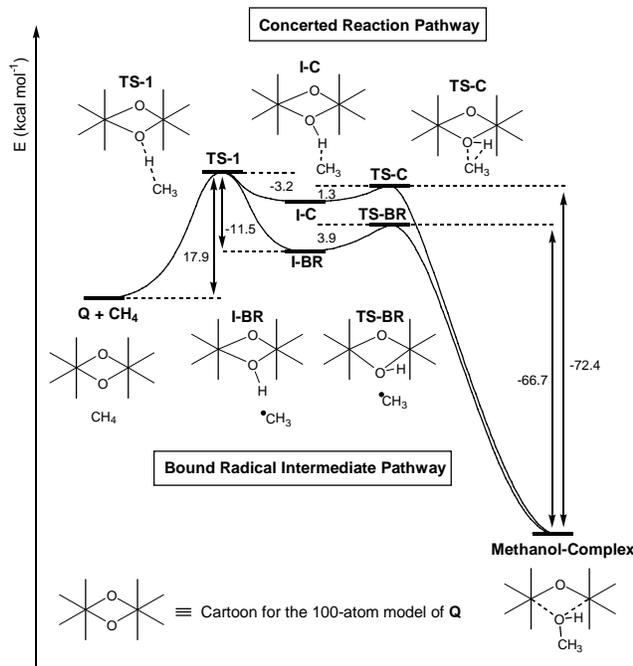


Figure 3. Proposed mechanism for methane hydroxylation reaction.

this represents a more appropriate comparison and leads to a difference of 3–5 kcal/mol (within the range of the estimated accuracy of the DFT methods used) between the calculated and experimental results.

Next, the transition from P to Q was investigated. Again, the computed free energy barrier is in reasonable agreement with that obtained from the experimental turnover rate of P and the use of transition state theory to extract the barrier (17.9 kcal/mol calculated vs. 15.7–16.6 kcal/mol experimental).

The reaction of Q with methane and ethane was studied in detail, and the results for methane are summarized in Figure 3. The calculated free energy barrier for the methane reaction, 17 kcal/mol, is within 2 to 3 kcal/mol of the experimental result. In this case, enthalpy and entropic contributions to the barrier have been determined experimentally, and calculated results from this project are in good agreement for these as well.

After the hydrogen atom abstraction, the –OH moiety that is formed must rotate upwards to allow attack of the methyl radical on the oxygen of the –OH to yield alcohol product. This aspect of the reaction was studied in detail (Gherman *et al.* 2001; Guallar *et al.* 2002), and an atomic-level picture that explains the partial racemization of chiral ethane reported by Lippard and coworkers (Valentine *et al.* 1997) emerged. There are two channels for the reaction; one is a concerted channel in which –OH rotation and alcohol formation occurs prior to formation of a bound radical intermediate, while the second involves formation of the bound radical intermediate first followed by –OH rotation. Semi-classical dynamics calculations were carried

fashion. Thus one would not predict that these states are experimentally observable. The first experimentally observable intermediate is the structure labeled Hperoxo in Figure 3, and it is therefore what the project team identified as the spectroscopically observed state known as compound P in the MMO literature. It is a symmetrically bridged structure which is in qualitative agreement with the Mössbauer data for P. The free energy barrier for the transition from reduced enzyme to P is calculated to be 22.1 kcal/mol. This is 6–8 kcal/mol larger than the barrier estimated from the experimental rate of production of P and application of simple transition state theory to extract a barrier height in the native system. However, experimental data indicates that the reaction rate for production of P is approximately 1000 times slower in the absence of protein B. Since protein B is not present in the model developed in this project,

out on a reduced dimensional potential surface generated from the quantum data, and this yielded quantitative agreement with the chiral ethane racemization results.

The next major project objective is to be able to explain and predict the thermodynamics and kinetics of substrates other than methane. Lippard and coworkers have carried out measurements of the rate constants for hydroxylation for a number of small molecule substrates, as is reported in Liu *et al.* (1993). Based on modeling efforts undertaken in this project to date, a hypothesis that explains the substantial variation in rates observed for these substrates has been developed. This hypothesis involves (among other things) differential stabilization of various substrates by the protein, an effect that requires QM/MM calculations to verify. These calculations are currently in progress.

Cytochrome P450

The family of cytochrome P450 monooxygenases are ubiquitous in biology, playing a key role in human metabolism of pharmaceutical agents and other ingested exogenous compounds. These enzymes insert an oxygen atom from O₂ into a wide variety of substrates, with substrate specificity determined by the nature of the protein cavity. In the catalytic cycle, dioxygen is bound to an iron porphyrin and then undergoes a series of transformations to produce an intermediate capable of cleaving a C-H bond of the substrate. The currently accepted steps in this cycle are depicted in Figure 4. The focus of preliminary studies undertaken for this project was on the bacterial isozyme cytochrome P450cam, for which camphor is the substrate, in order to exploit the recent, extensive experimental data measured for this system (Schlichting *et al.* 2000; Davydov *et al.* 2001). All of the steps in the catalytic cycle shown in Figure 2 were studied using both QM model systems and QM/MM calculations.

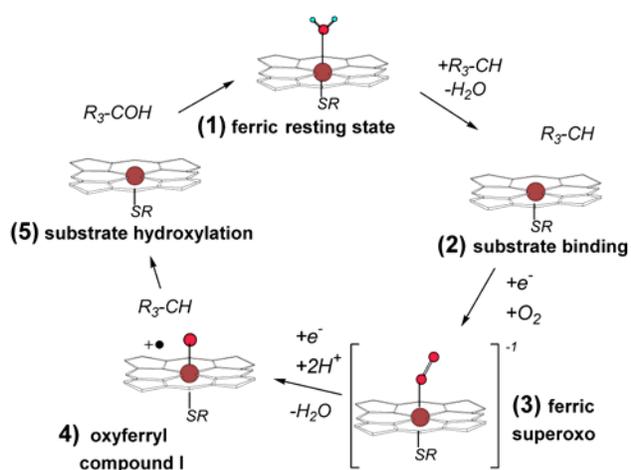


Figure 4. Catalytic cycle for alkane hydroxylation performed by P450.

Major results obtained for the various steps in Figure 4, prior to the hydroxylation step, are briefly summarized as follows. The first step in the catalytic pathway, the binding of the camphor substrate to the active site, involves a spin change in the metal core of the active site. This spin change, which has been difficult to reproduce with QM models, is accurately obtained with the extensive QM/MM description of the system developed in this project. The steps leading from the ferric superoxide to compound I, involving a reduction and addition of two protons, have eluded direct experimental characterization. The project team's calculations demonstrate that the first proton must come from Thr₂₅₂ through a 3.9 kcal/mol barrier. The proton from the threonine is replaced by migration of a proton down from a water channel that

connects the active site with the Glu₃₆₆. The results are consistent with isotopically labeled molecular oxygen studies. They indicate a disruption of the water channel followed by a direct attack of two crystallographic waters to the dioxygen when mutating Thr252 to alanine. Such non-specific attack opens the possibility of a peroxo decoupling channel.

The project team has now turned to a more detailed discussion of the hydroxylation reaction. Experimental efforts to directly observe compound I have been unsuccessful. One possible explanation for the failure to observe the oxyferryl intermediate is that the lifetime of this species is sufficiently short to preclude its detection with the experiments carried out to date. Previous theoretical studies, however, obtained using a minimal porphyrin model and methane as substrate, are inconsistent with this hypothesis. In particular, the computed enthalpy for the hydrogen abstraction, defined as the energy of the products minus that of the reactants, was highly endothermic, 24.0 kcal/mol, and the corresponding activation barrier for the model was 27.5 kcal/mol.

This problem has been addressed using more realistic models and QM and QM/MM techniques. Figure 5 depicts the transition state for hydrogen atom abstraction computed via QM/MM methods using a 126-atom QM region; the activation energy is now only 11.7 kcal/mol. If the zero-point contribution to the barrier is estimated to be comparable to the value obtained for an MMO model system (*i.e.*, on the order of 3.5 kcal/mol), the free energy of the barrier reduces to 8.2 kcal/mol. This exceptionally small barrier would permit rapid formation of the hydroxylated product from the oxyferryl intermediate and, thus, would explain the experimental results.

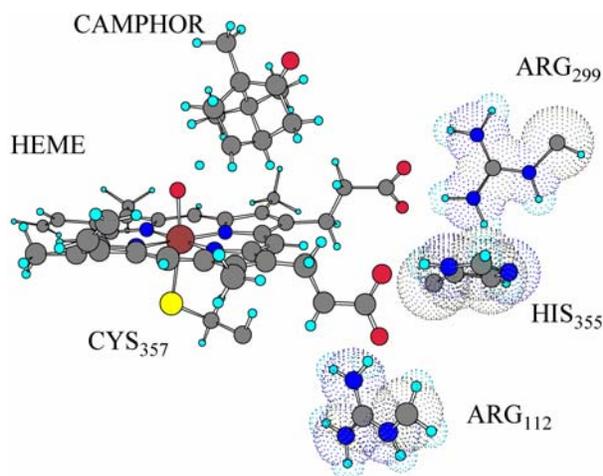


Figure 5. QM/MM model used in the P450-study. For clarity, only a few fragments of the protein environment that constitute the MM-region are shown and marked by dot spheres.

By carrying out QM model system calculations as well as the QM/MM study described above, the overall reduction of the energy barrier by 15.8 kcal/mol (compared with the minimal porphyrin model) was decomposed into components. A fraction amounting to 6.5 kcal/mol is due to the use of camphor, rather than methane, as the substrate; the energy of the transition state is lowered by delocalization of the radical created on the substrate. Strain energy contributes 1.0 kcal/mol. The protein environment contributes the remaining 8.3 kcal/mol of stabilization. The largest contribution of the protein to the stabilization of the transition state, 5.3 kcal/mol, comes from electrostatic interactions between the QM and MM regions; the remaining 3.0 kcal/mol is attributable to reorganization of the MM region of the protein. The physical origin of the

electrostatic stabilization (and quite possibly the MM reorganization as well, since this can easily

occur in response to electrostatic forces in the environment) can be ascertained in detail and is discussed below.

One simple physical picture that can be used to understand the electrostatic stabilization of the transition state is to envision the hydrogen atom abstraction reaction described above as a proton-coupled electron-transfer. The electron of the C–H σ -bond is transferred to the ferryl-oxygen atom of the heme accompanied by C–H bond cleavage with concomitant proton transfer. Calculations performed by the project team indicate that there are two main electronic effects operative in accommodating the excess charge injected into the heme. A significant fraction of the electron charge stays localized on the oxygen atom and a minor portion of the charge is transferred to the iron atom through direct coupling, introducing partial Fe(III) character. In addition, there is an indirect electronic relaxation effect present that causes a significant reorganization of the occupied molecular orbitals which do not mix directly with the redox-active orbitals, giving rise to a notable charge redistribution in the porphyrin ring and its substituents. Formal charges derived from the electrostatic potentials on both the product and reactant indicate that one of the peripheral carboxylate substituents of the heme in Figure 6 has an excess charge of approximately 0.06 electron units in the product as compared to the reactant. This additional negative charge interacts favorably with the positively charged hydrogen atoms on Arg₂₉₉ and translates into a differential electrostatic stabilization energy of 3.0 kcal/mol. There are similar, although smaller, interactions between the other propionate group and Arg₁₁₂ and His₃₃₅.

This remarkably large stabilization of the transition state by use of a carboxylate-arginine salt bridge—despite the significant distance of the carboxylate from the metal center and attached oxygen—may be representative of a more general motif in metalloproteins. Oxidation state changes at the metal ion in the catalytic cycle of a metalloenzyme are well known to be delocalized out onto the ligands. Stabilization of the substrate-donated electron by a positively charged arginine residue in the active site cavity has been observed. As is discussed in Wirstam *et al.* 2003, a similar transmittal of electronic charge from the metal center to a salt bridge is seen to stabilize dioxygen binding in hemerythrin, although in this case with the polarity reversed. Future work will be required to see whether this mechanism is manifested in other metalloprotein active sites.

Hemerythrin

Hemerythrin (Hr) is a nonheme iron protein that transports dioxygen in a number of marine invertebrates. It has been extensively studied by a variety of experimental techniques, and high resolution crystal structures are available for both the oxy and deoxy forms of the protein. The objectives of this project were to understand the mechanism by which oxygen binding to the protein is stabilized (it has proven very difficult to construct model compounds based on the Hr structure with similar properties) and to attempt quantitative calculation of binding free energy, which can be directly compared with experimental data. Both QM and QM/MM models were used in this endeavor. The major conclusions are as follows:

1. A binding free energy (referenced to aqueous solvation of dioxygen) of -5.3 kcal/mol was obtained, which compares favorably with the -7.2 kcal/mol determined experimentally via partitioning experiments.
2. Important features stabilizing dioxygen binding are the hydrogen bonding to the bridging –OH group, van der Waals interactions with the protein, and electrostatic interactions with the protein.
3. The dominant component of the electrostatic stabilization is due to the interaction of His 54 in the first coordination shell of Fe²⁺ with a glutamic acid (Glu 24) that forms a hydrogen bond with this histidine, which is protonated in the relevant hydrogen bonding position. This is a very strong hydrogen bond due to the presence of a full negative charge on the glutamic acid, and a substantial partial positive charge on the histidine. When dioxygen binds to Fe²⁺, it withdraws electrons from the iron (to form a peroxo species) and some of this electron withdrawal is transmitted to the histidine, making the proton binding to Glu more positive and, hence, increasing the hydrogen bond strength by 2.7 kcal/mol. The project team has verified that this mechanism is operative, and yields nearly identical numbers for the stabilization energy, in a fully QM model (including the Glu 24 carboxylate) as well as QM/MM calculations. As discussed above, this mechanism for stabilization of different charge states of the metal in metalloproteins via peripheral salt bridges is also operative (and critically important to function) in cytochrome P450.

Modeling of Metal Oxide Surfaces

Work on modeling metal oxide surfaces using an embedded cluster approach has continued. In this approach, a cluster of 10-100 atoms is treated at the QM level (typically with hybrid DFT methods, such as the B3LYP functional). The cluster is then embedded in an infinite solid consisting of classical ions, whose charges are determined self-consistently from the quantum cluster calculations. The electrostatic potential of the classical ions at long range is determined by Ewald methods, and the Ewald field is fit to a finite set of point charges that reproduce the field in the region of the quantum cluster. These charges can then be used to calculate the effects of the ionic environment on the quantum cluster via simple one electron matrix element calculations.

Initial calculations involved determining the work function of the (111) surface of the hematite form of iron oxide. These calculations yielded a value of 5.6 eV that is in good agreement with the experimentally estimated range of 5.3 to 5.9 eV and is considerably closer to these estimates than the plane wave DFT calculations of Scheffler and coworkers for the same surface, which yielded a value of 4.3eV. More recently, a similar calculation for the anatase form of titanium dioxide has been made. In both calculations, convergence of the self-consistent iteration process, and as a function of cluster size (in both surface area and depth) was demonstrated.

Having achieved some level of validation of the methodology, the project team has moved on to investigate reactive chemistry on the surface of the magnetite form of iron oxide. In the context of the EMSI at Columbia University, Flynn and coworkers have been studying the interactions

of CCl_4 with the magnetite surface. A central observation is that, at the appropriate temperature, phosgene (COCl_2) desorbs from the surface. The process of phosgene formation and desorption is being modeled using the embedded cluster methods discussed above. The conclusions reached thus far are as follows:

1. Phosgene can be formed by the carbon atom in CCl_2 bonding to a special oxygen site on the surface (in qualitative agreement with the experimental observations). The decomposition of CCl_4 to form two Cl atoms bonded to iron atoms on the surface, plus CCl_2 , is relatively straightforward. Once the C-O bond is formed, it is thermodynamically downhill for phosgene to leave the surface, and there is a very low barrier that is inconsistent with the experimental desorption temperature. Hence, project team concludes that the actual desorption event is not the rate-limiting step in the observation of phosgene production.
2. Formation of the C-O bond does have a nontrivial barrier, which the project team is in the process of calculating. Once a transition state is found for this process, the team will be able to determine if the magnitude of the activation free energy is compatible with the experimental desorption data. If this is the case, a more or less complete mechanism for the surface chemistry can be developed.

Force Field Development

Development of a second generation polarizable force field model has been developed. In this model, all parameters are obtained directly from quantum chemical calculations with the exception of those representing the dispersive van der Waals tail (*i.e.*, the coefficient of the $1/r^6$ term in the molecular mechanics atom-atom pair potential). Dispersive parameters are defined for every atom (H, O, C, etc.) and then used to represent all such atoms no matter what their bonding pattern in a molecule. Electrostatic parameters—point charges, point dipoles, and atomic polarizabilities—are fit to the electrostatic potentials arising from large basis set DFT calculations, with the polarizabilities determined by computing the effects of perturbing electric fields on the molecular charge distribution. The short range part of the atom-atom pair potential is adjusted to reproduce the binding energies of molecular dimers, generally calculated in the MP2 basis set limit. A highly efficient approach, based on pseudospectral methods, has been developed for estimating the MP2 basis set limit of the binding energy; in most cases, this value is accurate to a few tenths of a kcal/mol, but for a few cases (*e.g.*, aromatic ring interactions), higher order corrections are important. Sherrill has recently carried out such calculations for the benzene dimer, and the binding energies obtained are in good agreement with those determined in this project by optimizing the dimer binding energy based on polarizable liquid state simulation data.

We have to date carried out two different types of calculations with this second generation approach. Firstly, we determined the universal dispersion parameters for first row atoms and sulfur by fitting these to a few small molecule liquid state simulations. Then, we used the parameters to carry out additional small molecule liquid state simulations. The heat of vaporization and density obtained from the latter test cases are generally within 0.5 kcal/mol of the experimental values, without the use of any adjustable parameters. These test cases validate the

fact that the method can now be used to generate parameters for an arbitrary organic molecule, without any experimental data as input. This work has been submitted for publication. Secondly, we have constructed a complete second generation protein polarizable force field, which thus far has been evaluated primarily via gas phase energetics, where it performs quite well. We are in the process of developing a continuum solvation model for this force field and carrying out tests of the accuracy of the force field in aqueous solution.

Molecular-Level Predictions for Electronic Noses

A common electronic nose device consists of chemical sensors that are constructed with thin polymer films loaded with electrically conducting carbon black particles. Carbon black percolates in the polymer film, making the film conductive, if an appropriate amount is added. When the analyte is absorbed into the sensor, the volume of the polymer film swells, causing some of the percolated carbon black pathways to break, resulting in an increase in the electric resistivity of the polymer film, which is the signal detected for the analyte. It has been found experimentally that the sensor response is only a function of the swelling volume of the film. This experimental evidence indicates that one could predict the response of such electronic nose if the solubility of the analyte in the polymer film is known. In this discussion, MD simulation methods for the prediction of the solubility of chemical molecules (analytes) in polymers, which are needed to predict the response of electronic noses, are presented.

A combination of MD simulation and the Flory-Huggins model is used to predict solubility. In particular, two approaches for calculating the solvent-polymer segment interaction parameter (c) in the Flory-Huggins model are examined. In the first approach, the c -parameter is determined from the solubility parameter (square root of the cohesive energy density, CED) of the liquid analyte and the polymer. An in-house CED module is used to calculate the solubility parameters of eight analytes and two polymers, poly(ethylene oxide) and poly(caprolactone). An excellent agreement between the calculated and experimental solubility parameter is found and the predicted solubilities are in good agreement with the experimentally obtained results (Figure 6). In the second approach, the interaction c -parameter is obtained from calculation of energy of mixing. MD simulations for polymers loaded with different amount of analyte molecules are performed and the energies of mixing are calculated. Polymers filled with eight analyte molecules are initially prepared (Figure 7). Wilson's velocity autocorrelation mode analysis method is then used to determine the free energy of each analyte molecule in the polymer matrix, and high free energy analytes (poor sites) are removed. This process is repeated until all the analytes are removed. Qualitative agreement between the experimental solubility and that calculated from the energy of mixing for different analyte-polymer pairs was achieved.

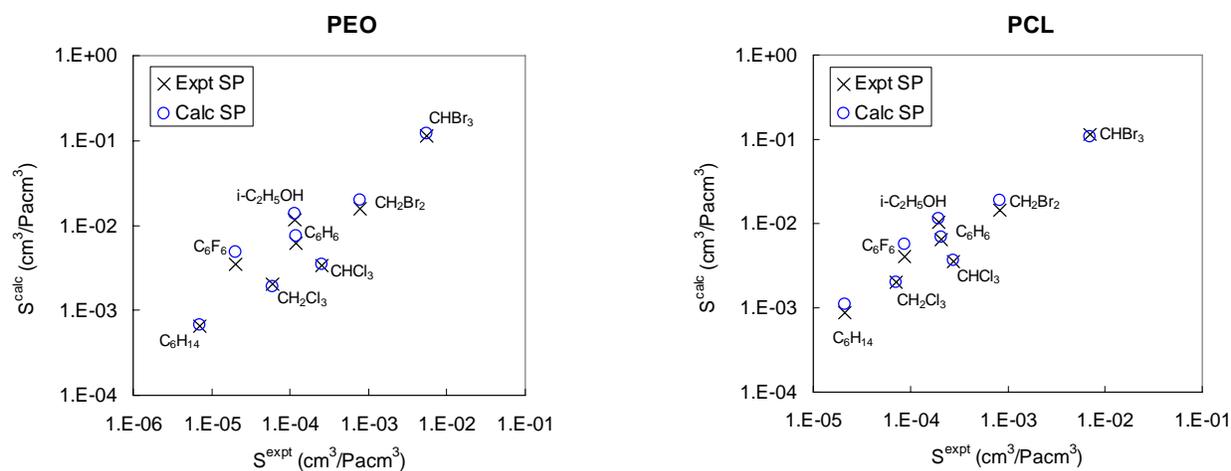


Figure 6. Comparison of predicted analyte solubility in (a) poly(ethylene oxide) and (b) poly(caprolactone) using both experimentally obtained (x) and calculated (o) solubility parameters.

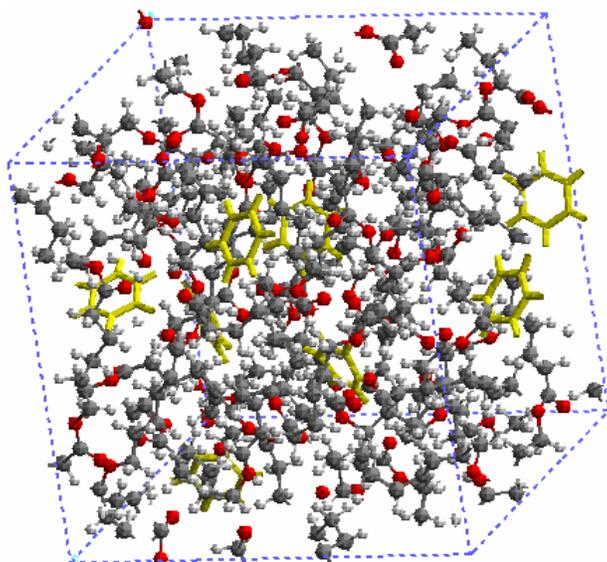


Figure 7. An atomistic model of Poly(caprolactone) filled with eight benzene molecules.

First-Principles Predictions of VLE Phase Diagrams

The ability to predict the vapor-liquid equilibrium (VLE) phase behavior of binary and multi-component systems is essential in order to refine chemical design processes sufficiently that the focus of environmental protection can be shifted from pollution clean-up to pollution prevention. Reliable understanding of the VLE for mixtures of organic compounds in both aqueous and non-aqueous solutions is also of great importance in understanding the processes in atmospheric chemistry.

The primary challenge to predicting VLE is estimating the activity coefficients, γ , for various components. Most commonly, γ is determined by fitting to experimental data on equations of state (EOS) or excess Gibbs free energy. These models have been useful in using fits to binary experimental data to predict the VLE of ternary and higher systems. However, there has been no reliable way to predict the binary data.

The UNIQUAC (UNIversal QUAsi-Chemical) model and the Wilson model provide methods for predicting the excess Gibbs free energy without using experimental data. These models calculate γ directly from molecular interaction energies sampled over an appropriate ensemble. Experiments do not readily yield these parameters. The known interaction energies are most frequently derived from fitting experimental phase behavior to the UNIQUAC equation for binary systems. This expands nicely to describe multi-component systems, affording prediction of ternary and higher systems based on binary experiments.

Since the equations use molecular interaction energies as fundamental input parameters, predictions of phase behavior should follow from fundamental analysis of interaction energies. This could provide a method of property prediction using only theoretical methods and free from experimental bias. Previous to this work, first-principles methods developed to predict activity coefficients were limited by the type of molecular interactions and the computational expense. Research in this project has produced an efficient method for the theoretical determination of the fitting parameters for polar/polar, polar/non-polar, and non-polar/non-polar systems. The liquid phase was simulated with force-field dynamics calculations, and the interaction energies of molecular pairs were determined with high-level (LMP2/cc-pvtz(-f)) quantum calculations. The Cohesive Energy Density model in Cerius2 produces the ensemble of pairs used in the quantum calculations. These calculations determine the interaction parameters for both the UNIQUAC and the Wilson models.

Results and Discussion

With a modified UNIQUAC equation, the experimental phase diagrams for a range of systems were successfully modeled. Meaningful results are obtained when the interaction parameters are divided by $z/2$, where z is a coordination number traditionally set to 10. Tables 1 and 2 list the systems studied and the predicted interaction parameters for the UNIQUAC and Wilson models, respectively. The binary pairs studied are generally soluble, making these models appropriate.

The approach developed in this project predicts all systems correctly except those phase diagrams for binary systems involving a polar molecule and a molecule containing a carbonyl group. Comparison to experimental data indicates that modeling of these pairs with this approach yields binding energies that are too strong. The consistency of this failure suggests the source of the error is due to the MD simulations of these systems. This error may be due to the hydrogen-bonding term in the force-field.

The success of this method demonstrates the appropriateness of the Cohesive Energy Density module in Cerius2 for modeling liquids with small numbers of molecules; in the simulations performed in this project, there were only 32 molecules.

Table 1. The UNIQUAC interaction parameters for the studied systems in J/mol.

Molecule 1	Molecule 2	ΔU_{12}	ΔU_{21}
Acetic Acid	Water	-1304.469961	12101.1123
Acetone	Water	-2490.997	-7441.912435
Acetone	Benzene	3809.85466	-2061.15336
Chloroform	Acetone	-4669.054	-9435.784629
Carbon Tetrachloride	Acetonitrile	11682.0063	549.4753249
Methanol	Benzene	5741.31851	9195.400921
Ethyl Acetate	Ethanol	-2692.0054	-4078.787298
Ethanol	Ethylene Glycol	2391.79048	-10330.22273
Hexane	Ethanol	13564.3217	5545.674862
Hexane	Nitroethane	7476.90735	-73.67494488
Methanol	2-Methoxy Ethanol	2042.72669	4569.149215
Methyl Acetate	Ethanol	-1713.6558	-682.4819051
Methyl Acetate	Methanol	-1770.6893	-1715.958462
Nitromethane	Benzene	1793.08293	3373.583656
Methylcyclopentane	Benzene	5888.26238	-2615.604063
Ethanol	Water	-531.19067	1482.27205
Methanol	Water	-67.335065	1060.821611
Water	Ethylene Glycol	340.87341	-1095.041936

Table 2. The calculated Wilson parameters for the studied systems in J/mol

Molecule 1	Molecule 2	ΔU_{12}	ΔU_{21}
Acetic Acid	Water	11929.771	-1133.12837
Acetone	Water	-7785.366	-2147.543432
Acetone	Benzene	1109.06707	639.6342306
Chloroform	Acetone	-10899.907	-3204.931525
Carbon Tetrachloride	Acetonitrile	946.480416	11285.0012
Methanol	Benzene	9425.24827	5511.471158
Ethyl Acetate	Ethanol	-8550.6362	1779.843526
Ethanol	Ethylene Glycol	2813.20594	-10751.63819
Hexane	Ethanol	818.680701	13529.98649
Hexane	Nitroethane	-441.75056	7844.982962
Methanol	2-Methoxy Ethanol	538.538071	6073.337834
Methyl Acetate	Ethanol	-2737.7865	341.6488003
Methyl Acetate	Methanol	-1610.2218	-1876.425999
Nitromethane	Benzene	4611.90657	554.7600182
Methylcyclopentane	Benzene	-1473.0626	4745.720928
Ethanol	Water	1743.75592	-792.6745416
Methanol	Water	1023.01329	-29.52673933
Water	Ethylene Glycol	-1512.6577	758.4891869

References

Choi, S-Y, PE Eaton, DA Kopp, SJ Lippard, M Newcomb, R Shen. 1999. "Cationic Species Can Be Produced in Soluble Methane Monooxygenase-Catalyzed Hydroxylation Reactions; Radical Intermediates are not Formed." *J. Am. Chem. Soc.*, 121:12198-12199.

- Davydov, R, TM Makris, F Kofman, DE Werst, SG Sligar, and BM Hoffman. 2001. "Hydroxylation of Camphor by Reduced Oxy-Cytochrome P450cam: Mechanistic Implications of ERP and ENDOR Studies of Catalytic Intermediates in Native and Mutant Enzymes." *J. Am. Chem. Soc.*, 123:1403-1415.
- Gherman, BF, BD Dunietz, DA Whittington, SJ Lippard, and RA Friesner. 2001. "Activation of the C-H Bond of Methane by Intermediate Q of Methane Monooxygenase: A Theoretical Study." *J. Am. Chem. Soc.*, 123:3836-3837.
- Guallar, V, BF. Gherman, WH Miller, SJ Lippard, and RA Friesner. 2002. "Dynamics of Alkane Hydroxylation at the Non-Heme Diiron Center in Methane Monooxygenase." *J. Am. Chem. Soc.*, 124:3377-3384.
- Liu, KE, AM Valentine, D Wang, BH Huynh, DE Edmondson, A Salifoglou, and SJ Lippard. 1995. "Kinetic and Spectroscopic Characterization of Intermediates and Component Interactions in Reactions of Methane Monooxygenase from *Methylococcus Capsulatus* (Bath)." *J. Am. Chem. Soc.*, 117:10174-10185.
- Liu, KE, CC Johnson, M Newcomb, and SJ Lippard. 1993. "Radical Clock Substrate Probes and Kinetic Isotope Effect Studies of the Hydroxylation of Hydrocarbons by Methane Monooxygenase." *J. Am. Chem. Soc.*, 115(3):939-947.
- Merkx, M, DA Kopp, MH Sazinsky, JL Balzyk, J Muller, and SJ Lippard. 2001. "Dioxygen Activation and Methane Hydroxylation by Soluble Methane Monooxygenase: A Tale of Two Irons and Three Proteins." *Angew Chem. Int. Edit Engl.*, 40:(15):2782-2807.
- Rosenzweig, AC, CA Frederick, SJ Lippard, and P Nordlund. 1993. "Crystal Structure of a Bacterial Non-Heme Iron That Catalyses the Biological Oxidation of Methane." *Nature*, 366:537-543.
- Rosenzweig, AC, P Nordlund, PM Takahara, C Frederick, and SJ Lippard. 1995. "Geometry of the Soluble Methane Monooxygenase Catalytic Diiron Center in Two Oxidation States." *Chemistry & Biology*, 2:6, 409-418.
- Schlichting, I, J Berendzen, K Chu, AM Stock, SA Maves, DE Benson, RM Sweet, D Ringe, GA Petsko, and SG Sligar." 2000. "The Catalytic Pathway of Cytochrome P450cam at Atomic Resolution." *Science*, 287:1615-1622.
- Shu, L, JC Nesheim, K Kauffmann, E Münck, JD Lipscomb, and L Que. 1997. "An Fe^{IV}O₂ Diamond Core Structure for the Key Intermediate Q of Methane Monooxygenase." *Science*, 24:275, 515-518.

Valentine, AM, B Wilkinson, KE Liu, S Komar-Panicucci, ND Priestly, PG Williams, H Morimoto, HG Floss, and SJ Lippard. 1997. Tritiated Chiral Alkanes as Substrates for Soluble Methane Monooxygenase from *Methylococcus capsulatus* (Bath): Probes for the Mechanism of Hydroxylation." *J. Am. Chem. Soc.*, 119:1818-1827.

Wirstam, M, SJ Lippard, and RA Friesner. 2003. "Reversible Dioxygen Binding to Hemerythrin." *J. Am. Chem. Soc.*, 125(13):3980-3987.