Molecular Science Computing Facility

The MSCF supports a wide range of computational activities in environmental molecular research, from benchmark calculations on small molecules to reliable calculations on large molecules and solids to simulations of large biomolecules, as well as reactive chemical transport modeling. The MSCF provides an integrated production computing environment, with links to external facilities within the Department of Energy, collaborating universities, and industry.

The MSCF completed the procurement of a new replacement supercomputer for the aging IBM computing resource. The new replacement system will be from Hewlett-Packard Company. The system will be delivered in phases with the expected completion of an 11.4 teraflop Linux-based supercomputer in September of 2003.

MSCF Resources

The MSCF provides a combination of production computing hardware and software resources to support the scientific research activities of the Computational Grand Challenge (GC) and EMSL pilot projects. The year brought much needed performance upgrade opportunities for the MSCF users. The MSCF completed the procurement of a new HP/Linux based supercomputer. MSCF began this transition to this new system late in the year. The year began with the previous existing IBM system as described below and as shown in Figure 1.

Molecular Science Computing Facility

- NWMPP1 – Production IBM SP, 512 P2SC Nodes, 248 Gigaflops
- NWTEST – Test IBM SP, 8 P2SC Nodes, 3.8 Gigaflops
- Scalable graphics engine with 4 monitors
- NWECS1 – Experimental IBM SP, 4 16-way NH2 Nodes, 96 Gigaflops
- JUPITER – Production PNNL IBM SP, 26 4-way WH2 Nodes, 156 Gigaflops
- Internet – T3 (45 Mbit/sec), OC3 (155 Mbit/Sec) 2000
- NWARCHIVE – ADIC AML/E Data Storage System, 20 Terabyte capacity
- SGI Challenge XL and Ciprico RAID 384 GB Disk Cache
- IBM 3494 Tape Library ADSM Backup System, 2.4 Terabyte Capacity
- 7133 SSA Disk Storage Units, 906 Gigabytes
- Graphics & Visualizations Lab – 2 SGI Onxy Systems with 4 Graphic Pipes
- COLONY – Experimental Dell – Giganet Linux Cluster, 96 Nodes, 192 Processors – 100 Gigaflops

The MSCF Previous Computational System

A 512-processor IBM SP computer system with 262 gigabytes of memory and 5 terabytes of disk storage, with an aggregate peak theoretical performance of 247 gigaflops provided high quality, reliable, production-computing cycles to support very large parallel calculations for the GC projects. To provide this level of support, system management, monitoring, allocation management, and scheduling capabilities have been developed by the MSCF.

A 48-terabyte EMASS hierarchical storage management system with Scientific Data Management (SDM) software designed to store mass amounts of data for EMSL researchers and collaborators. The archive consists of two SGI Challenge XL computer systems containing approximately 400 gigabytes of high-speed disk that acts as a cache to 20 terabytes of tape storage capacity. The system is intended to provide a location for long-term storage of valuable scientific data. The SDM software, which was developed at PNNL, provides the interface for storing, searching, and retrieving

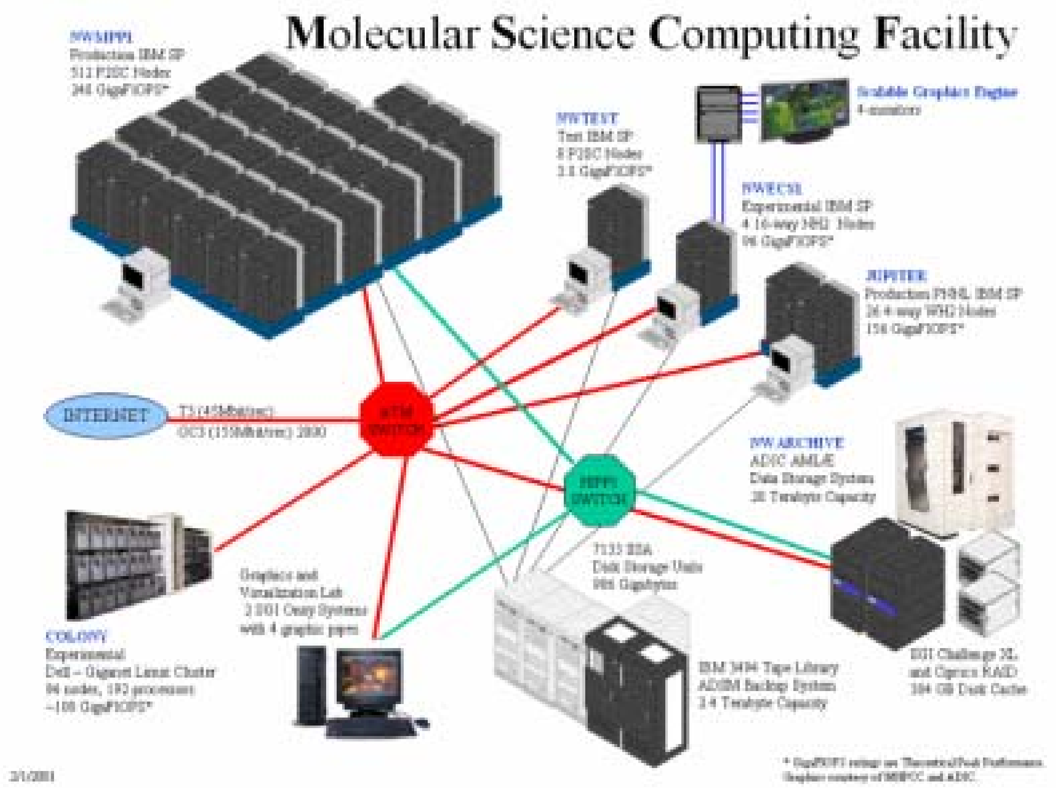


Figure 1. Previous Molecular Science Computing Facility Hardware

data, thus adding a unified view of the two separate Unix file systems on the two SGI systems, provide improved security and access control, and the ability to capture and search metadata.

MSCF New HP/Linux System

The transition to the new HP system began in the summer of 2002. The MSCF received the first of the HP/Linux prototype systems. In December of 2002 the MSCF completed the first phase of the multiphase installation process for the new supercomputer. The MSCF has also upgraded the network infrastructure and has made close ties with experimental facilities in the EMSL. Figure 2 shows the current configuration of the MSCF.

- The High Performance Computing System-2 (HPCS2) provides a balanced supercomputer that has been tailored to meet the operational needs of the users of the EMSL. The first phase of the new Hewlett-Packard Linux Supercomputer is operational now and consists of 256 Intel Itanium 2 processors representing a peak performance of 1 teraflop. The final phase of the HP supercomputer will deliver over 11.4 teraflops with 6.8 terabytes of RAM and 256 Terabytes of disk. The system is connected to the lab via multiple gigabit Ethernet connections, allowing EMSL users to perform remote visualizations and transfer data to remote storage.
- The EMSL's archive is located inside the MSCF. The archive uses a unique approach to disk storage by clustering many low cost commodity disks to provide fault tolerant high performance storage. The current EMSL archive has greater than 25 terabytes of storage and has the ability to grow to over a petabyte. The unique software developed by the MSCF to manage the large pools of clusters disks is called NWfs and is made freely available to users of the EMSL.

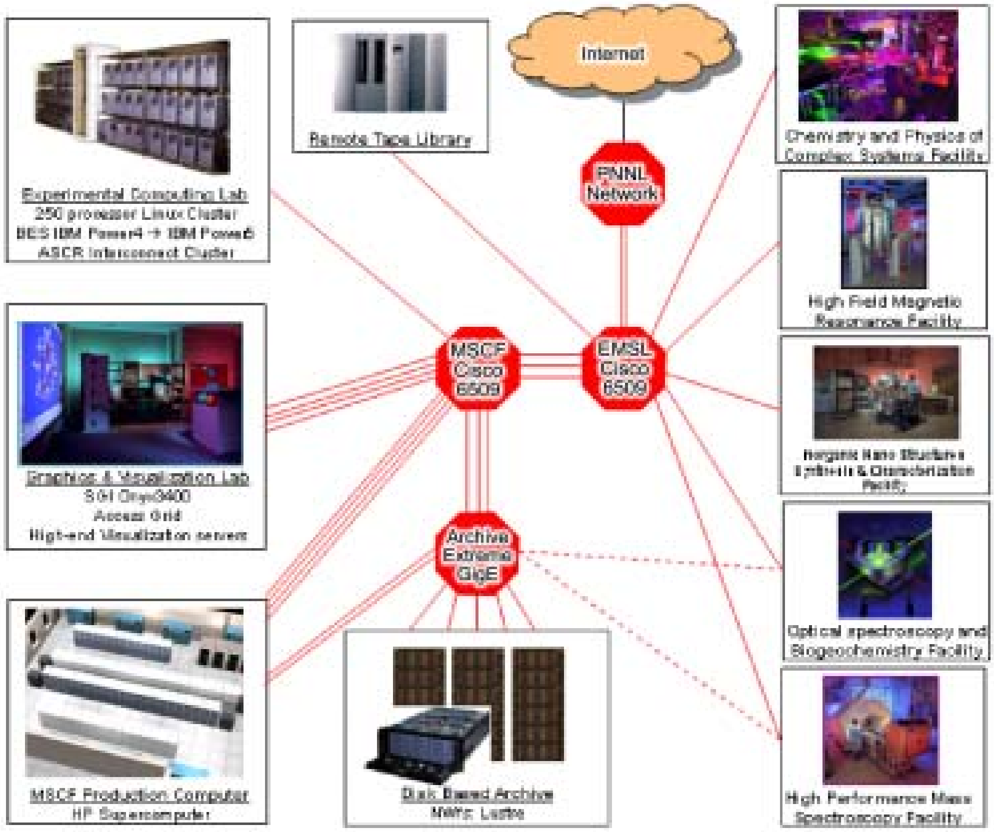


Figure 2. Upgraded MSCF with close ties to other EMSL instruments

- A Graphics and Visualization Laboratory (GVL) provides production graphics and visualization facilities for the display and analysis of complex data sets from both experiments and simulations. The GVL contains four high-performance graphics stations based on SGI technologies with high-speed gigabit Ethernet connections to the production supercomputers and to the EMSL archive system in the MSCF, a video system integrated with the workstations to facilitate the display and capture of scientific data, and video editing equipment for the preparation of scientific presentations. The video system also is connected to the EMSL seminar room as well as to the World Wide Web to facilitate data sharing. An IBM SGE (a joint collaboration with the IBM Watson Research Center) is connected to the MSCF IBM SP for the highest performance visualization capabilities.
- The Molecular Science Software Suite (MS3), a comprehensive, integrated set of tools, enables scientists to understand complex chemical systems at the molecular level. MS3 couples the power of advanced computational chemistry techniques running on existing and rapidly evolving high-performance, massively parallel computing systems with extensible problem solving capabilities. The suite consists of three components: 1) the Extensible Computational Chemistry Environment (Ecce), 2) the Northwest Computational Chemistry Software (NWChem), and 3) the Parallel Software Development Tools (ParSoft).

MSCF Organization and Staff

Three groups provide the user support, software support/development, and computer systems operations for the MSCF.

- The Scientific Consulting Group manages all user services for MSCF users. This effort includes handling all user proposal applications, following user progress of all computational projects, managing all proposal reviews for both GC and EMSL pilot projects, helping with user accounts, providing general consulting support for MSCF software packages, supporting and maintaining software, managing the Graphics and Visualization Laboratory, conducting training and user workshops, developing visualization software and high quality visualizations, and producing websites.
- The Molecular Science Software Group supports the Molecular Science Software Suite (MS^3, comprising NWChem, Ecce and ParSoft). This effort includes developing higher performance versions of the software, responding to user requests for additional features, supporting and maintaining the software, diagnosing MS^3 problems associated with computer vendor hardware and software, consulting on specific MS^3 problems, distributing MS^3 to remote sites, porting software to new architectures, and conducting training and user workshops.
- The Computer Operations Group runs the MSCF computer systems including all the large supercomputing systems, the experimental computing system, and the EMSL Archive, used for software development. The group is responsible for the installation and upgrades of the MSCF supercomputers, operating system support, improving the quality of the system software to improve the user environment, allocation tracking, direct interfacing with the vendor, system security, and system-level support to users.

User Computing in the MSCF

To optimally address the complex environmental problems facing DOE and the nation, as well as to best use limited staff resources, EMSL follows the recommendation of the MSCF Advisory Committee to use a computational GC approach to providing large blocks of resources to the user community. A Call for Proposals (CFPs) is issued, and teams of computational scientists respond with peer-reviewable proposals for system time allocations. Access for the GC teams is for periods of one to three years. The request for proposals is open to all researchers, regardless of their institution or source of funding.

The MSCF website (http://mscf.emsl.pnl.gov) is continuously updated to provide the MSCF users information they will need to know about activities of the MSCF. The website allows users to view the current status of jobs running on the MSCF computers. It also shows all current and past research activities of the MSCF.

MSCF Use

The MSCF provides computational resources for Computational Grand Challenge (GC) projects in environmental molecular science, basic and applied research areas that address the environmental problems and research needs facing the DOE and the nation. GC projects are usually funded for periods ranging from 1 to 3 years. The MSCF supported nine 3-year GC projects during FY02. The average annual computer time allocation for GC projects was 440,000 CPU hours. GC projects typically involve multiple investigators working collaboratively as teams. GC teams consist of computational scientists from universities, national laboratories, and industry.

The MSCF also supports smaller, short-term projects, which are called MSCF Pilot Projects. The Pilot Projects are limited to a maximum of 75,000 processor hours and for a one-year duration at most. The MSCF supported 39 Pilot Projects during FY02. The average allocation for Pilot Projects is 22,000 CPU hours. The Pilot Projects are typically directed at developing the capability to submit a GC proposal in the future (e.g., a combination of theory, method, or code development activities or doing calculations that provide the initial scientific basis of a GC proposal).

The MSCF high-performance computing systems had 387 users during FY2002, with 57% being external to PNNL and 42% being EMSL staff, PNNL staff, postdoctoral fellows or students. The distribution of the users by type of institution is 44% from universities, 42% from PNNL, 5% from industry, 8% from other DOE laboratories, and 0% from other government laboratories (Figure 3). The MSCF has 54% of its users located off-site.

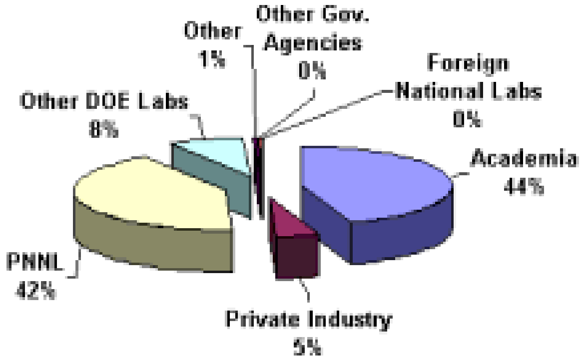


Figure 3. MSCF resource allocation by user affiliation.

Molecular Science Software Group

The Molecular Science Software Group consists of EMSL and PNNL matrixed staff focused on developing next-generation molecular modeling software for newly evolving computer technologies, especially massively parallel computers. This software consists of the user-friendly problem solving environment, Ecce, the massively parallel computational chemistry code, NWChem, and the parallel software tools, ParSoft. The following section outlines capabilities and EMSL activities related to these software developments.

Extensible Computational Chemistry Environment (Ecce)

Ecce is composed of a suite of distributed client/server UNIX-based X Window System applications seamlessly integrated together. The resulting environment enables research scientists to transparently use complex computational modeling software accessing high-performance compute resources from their desktop workstations. The major components:

- The Calculation Manager, which aids in the organization and manipulation of computational chemistry studies. This tool allows an at-a-glance overview of the status of every calculation, and easy access to key setup parameters and run statistics.
- The Builder enables the building, visualization, modification, and manipulation of 3-D images of chemical systems to be used in Ecce calculations.
- The Basis Set Tool enables the choice from over 230 predefined Gaussian basis sets or the ability to create new ones for use in ab initio electronic structure calculations.
- The Calculation Editor allows the user to choose input options in a point and click fashion and then generates the chemistry code specific input.

- The Job Launcher is a tool for submitting a calculation to a computer for processing. The user may submit a calculation to any computer that has been registered as a compute server where the user has an account.
- The Calculation Viewer provides convenient access to current information for a single calculation during execution or after completion. It has many features for viewing and visualizing chemical system properties.

The Ecce software architecture is based on an object-oriented chemistry data model to support the management of computational and experimental molecular data. The environment is built for extensibility to support new computational codes and capabilities as they become available. Ecce currently supports NWChem, Gaussian 98, and Amica. The Ecce application software currently runs on Linux, Sun and Silicon Graphics Inc. (SGI) workstations and is written in C++ using the Motif user interface toolkit and OpenGL graphics libraries. Ecce provides a sophisticated graphical user interface, scientific visualization tools, and an underlying data management framework, enabling scientists to efficiently set up calculations and store, retrieve, and analyze the rapidly growing volumes of data produced by computational chemistry studies.

NWChem 4.1 Functionality and Capabilities

Version 4.1 of NWChem was released in January 2002. Full documentation and information is available on the NWChem website (http://www.emsl.pnl.gov/pub/docs/nwchem). NWChem provides many methods to compute the properties of molecular and periodic systems by using standard Gaussian and pseudopotential based quantum mechanical descriptions of the electronic wavefunction or density. In addition, NWChem can perform classical molecular dynamics (MD) and free energy simulations. These approaches may be combined to perform mixed quantum-mechanics and molecular-mechanics simulations.

NWChem is available on almost all high-performance computing platforms, workstations, PCs running Linux or NT, and clusters of desktop or workgroup servers. NWChem development provides maximum efficiency on massively parallel processors; it has not been optimized for high performance on single processor desktop systems.

Molecular Electronic Structure

The following Gaussian-based quantum mechanical methods are available: self consistent field (SCF) or Hartree Fock (RHF, UHF, high-spin ROHF); Gaussian density functional theory (DFT), using many local and non-local exchange-correlation potentials (spin-restricted and unrestricted) with formal N^3 and N^4 scaling; spin-orbit DFT (SODFT), using RHF and UHF; MP2 including semi-direct using frozen core and RHF and UHF references and fully-direct with RHF reference; complete active space SCF (CASSCF); coupled cluster with single and double excitations (CCSD), coupled cluster with single and double excitations plus a perturbative corrections for triple excitations [CCSD(T), CCSD+T(CCSD)], with RHF reference; selected-configuration interaction (CI) with second-order perturbation correction; resolution of the identity integral approximation MP2 (RI-MP2), with RHF and UHF references; N-layered molecular orbital plus molecular mechanics (ONIOM) hybrid method of Morokuma and co-workers; electrostatic potential fit of atomic partial charges (CHELPG method with optional RESP restraints or charge constraints); COSMO energies for open and closed shell SCF and DFT. The spin-free one-electron Douglas-Kroll approximation is available for all quantum mechanical methods and their gradients; Dyall's spin-free modified Dirac Hamiltonian approximation is available for the Hartree-Fock method and its gradients; one-electron spin-orbit effects can be included via spin-orbit potentials.

Pseudopotential Plane-Wave Electronic Structure

Two modules are available to compute the energy, optimize the geometry, calculate numerical second derivatives, and perform ab initio molecular dynamics using pseudopotential plane-wave DFT.

- PSPW (Pseudopotenital plane-wave) – A gamma point code for calculating molecules, liquids, crystals, and surfaces.
- Band – A prototype band structure code for calculating crystals and surfaces with small band gaps (e.g., semiconductors and metals)

With Conjugate gradient and limited memory BFGS minimization; Car-Parrinello (extended Lagrangian dynamics); constant energy and constant temperature Car-Parrinello simulations; fixed atoms in Cartesian and SHAKE constraints in Car-Parrinello; pseudopotential libraries; Hamann and Troullier-Martins norm-conserving pseudopotentials with optional semicore corrections; automated wavefunction initial guess, now with LCAO; Vosko and PBE96 exchange-correlation potentials (spin-restricted and unrestricted); (gamma point) periodic orthorhombic simulation cells for calculating molecules, liquids, crystals, and surfaces; orthorhombic simulations cells with periodic and free space boundary conditions; modules to convert between small and large plane-wave expansions; interface to DRIVER, STEPPER, and VIB modules; polarization through the use of point charges; Mulliken, point charge, DPLOT (wavefunction, density and electrostatic potential plotting) analysis.

Molecular Dynamics

The following functionality is available for classical molecular simulations: single configuration energy evaluation, energy minimization, molecular dynamics simulation, and free energy simulation (multistep thermodynamic perturbation [MSTP] or multiconfiguration thermodynamic integration [MCTI] methods with options of single and/or dual topologies, double-wide sampling, and separation-shifted scaling). The classical force field includes effective pair potentials (functional form used in AMBER, GROMOS, CHARMM, etc.), first order polarization, self consistent polarization, smooth particle mesh Ewald (SPME), twin range energy and force evaluation, periodic boundary conditions, SHAKE constraints, and consistent temperature and pressure ensembles.

NWChem also has the capability to combine classical and quantum descriptions to perform

- mixed quantum-mechanics and molecular-mechanics (QM/MM) minimizations and molecular dynamics simulation
- quantum molecular dynamics simulation by using any of the quantum mechanical methods capable of returning gradients.

Parallel Software Tools and Libraries (ParSoft)

ParSoft includes high performance computing libraries and tools for applied parallel computing focused on interprocessor communications through the Aggregate Remote Memory Copy Interface (ARMCI), high-performance input/output through the Parallel I/O tools (ParIO), and programming models for hierarchical memory systems through the Global

Array (GA) and Memory Allocator (MA) tools. One of the primary goals for this work is to aid development of real scientific application codes.

The purpose of ARMCI is to provide general-purpose, portable, and efficient remote memory copy operations (one-sided communication) optimized for noncontiguous (strided, scatter/gather, I/O vector) data transfers. In addition, atomic and mutual exclusion operations are available. The development of this system is driven by the need to support the global address space communication model in the context of distributed regular or irregular distributed data structures including multidimensional arrays such as GA, communication libraries, and compilers. ARMCI is a standalone system that could be used to support other user-level libraries and toolkits that use MPI or PVM. Development of ARMCI has been supported by EMSL operations, the DOE2000 Global Arrays project, and recently by the Center for Programming Models for Scalable Parallel Computing.

The GA toolkit provides an efficient and portable "shared-memory" programming interface for distributed-memory computers. Each process in a MIMD parallel program can asynchronously access logical blocks of physically distributed dense multi-dimensional arrays without the need for explicit cooperation by other processes. Unlike other shared-memory environments, the GA model exposes to the programmer the non-uniform memory access (NUMA) characteristics of the high performance computers and acknowledges that access to a remote portion of the shared data is slower than to the local portion. The locality information for the shared data is available and a direct access to the local portions of shared data is provided. The GA toolkit has been in the public domain since 1994 and has been actively supported and employed in several large codes since then.

The ParIO toolkit provides a portable, high-performance I/O capability for use by applications. Three different models are available in ParIO:

- Disk Resident Arrays (DRA), which extend the GA NUMA programming model to disk. The library encapsulates the details of data layout, addressing and I/O transfer in disk array objects. Disk resident arrays resemble GAs except that they reside on disk instead of the main memory.
- Exclusive Access Files (EAF), which are generated and/or read by a single process of a distributed parallel application. Files are not shared between different processes. The library is an abstract high-performance files system which provides a common interface for a variety of architecture specific parallel storage systems.
- Shared Files (SF), which implement logically shared temporary files for parallel single program, multiple data (SPMD) applications. Any process can read and write at arbitrary locations in a file independently of other processes. Shared files must be created and destroyed collectively.

MA is a library of routines that make up a dynamic memory allocator for use by C, Fortran, or mixed-language applications. Fortran 77 applications require such a library because the language does not support dynamic memory allocation. C (and Fortran 90) applications can benefit from using MA instead of the ordinary malloc and free routines because of the extra features MA provides:

- both heap and stack memory management disciplines
- support for Fortran and C datatypes
- debugging and verification support including guards protecting allocated memory blocks, detection of overwritten ends of the allocated memory blocks, and the ability to assign names to allocated memory blocks

- usage statistics
- quantitative memory availability information.

MSCF Operations Group

The Computer Operations Group operates, maintains, and advances the capabilities of the MSCF scientific computing systems. The group is responsible for the operation and implementation of the various production supercomputers in the MSCF and has developed system management, monitoring, allocation management, and scheduling capabilities. The primary focus is on providing high-quality, reliable production computing cycles on the production computer in the MSCF to support very large parallel calculations for Grand Challenge projects. This year the infrastructure upgrade began with a replacement of the 512-node IBM SP with the first phase of the 11.4+ teraflop Hewlett-Packard supercomputer. The result is for an effective high-performance computing environment for MSCF users to run very large parallel calculations for their Grand Challenge projects. During 2002, roughly two-thirds of the full-year usage on the production supercomputers were in large parallel jobs each using over 12% of the systems resources (see Figure 4).

We achieved a very high up-time of greater than 99% for the production supercomputers. Overall annual usage for the MSCF supercomputers has remained very high at over 85%.

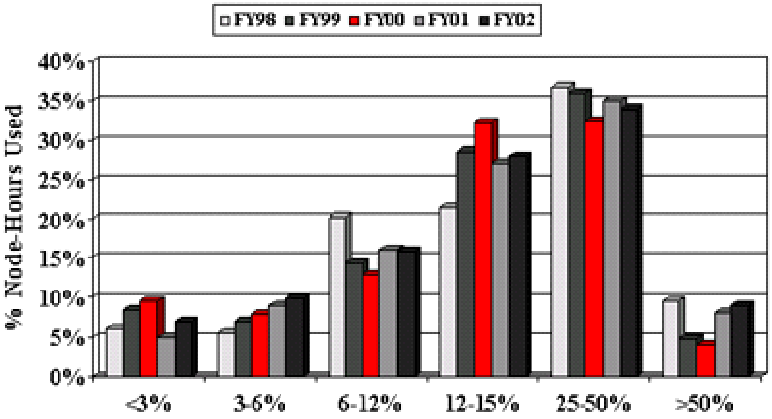


Figure 4. Percent of system used by a single job

Visualization and User Services (VisUS) Group

The MSCF Visualization and User Services Group provides support services for all users of the MSCF high-performance computers and the MSCF Graphics and Visualization Laboratory (GVL). The group provides an extremely diverse set of services. Scientists who need access to high-end computing equipment frequently have difficulties getting started. Problems range from logging in to getting user codes to run efficiently.

The group manages the Grand Challenge and Pilot Project proposal process for the MSCF. This includes receiving the proposals, preliminary review for EMSL mission, preparing proposals packets for external peer review, process peer review evaluations, granting project allocation, and managing the allocation by using the MSCF QBank.

During 2002, three scientific consultants responded to over 450 support requests received by e-mail, with about half as many requests handled over the phone or as direct office visits. Examples of questions fielded by our consultants include

- "How do I get an account?"
- "What resources are available?"
- "Can you acquire some software for me?"
- "Would you show me how to use this program?"
- "Is the computer running?"
- "Why did my job fail?"
- "What would be the best program to run for the project I'm working on?"
- "Will you help me analyze my results?"
- "How do I visualize my data?"

Consultants have various roles, including administrator, tutor, programmer, and research scientist. Specific services include helping new users with accounts and providing general scientific consulting support for MSCF software packages. Consultants perform technical consulting activities for high-performance computing software, parallel programming, and high-performance system testing. Consultants also work directly with Molecular Science Software Suite (MS^3) software development teams to give customer feedback and test functionality.

Information about the use and configuration of the MSCF computational resources is critical to the user base. MSCF information is provided efficiently to users over the Internet through the MSCF home page. The site contains all necessary information on how to get started and request accounts. The Website also provides details about the computer server configuration as well as documentation and Web-based tutorials for the MS^3. The consulting group also conducts MS^3 training workshops, MSCF workshops, demonstrations, and MSCF tours.

Scientists generate enormous amounts of data either from computational resources or from the EMSL instruments. This data is usually complex and difficult to understand. The MSCF GVL and support services help scientists visualize data. The GVL contains high-performance graphics compute servers for visualizing data. To help scientists present their findings, the GVL contains state-of-the-art multimedia equipment. The real-time digital video capture capability from the graphics computer servers allows quick yet high-quality video production. Users can generate presentation media in any form—from video (all international video standards) to web-based animations. The group also provides basic video production services.

Upgrades

Ecce

Many modifications and improvements in Ecce have been accomplished this year. The results of two major releases of the code, version 2.1 and 3.0, are briefly described below. A new Ecce website was designed and developed during 2002 for deployment in early 2003. This website is a tremendous marketing asset for the Ecce project as it provides an intuitive organization of the high-level capabilities of Ecce for prospective new users. It also allows current Ecce users access to information such as release notes, online help, and the frequently asked questions (FAQ) document. This new website is found at http://ecce.emsl.pnl.gov.

Version 2.1 of the Extensible Computational Chemistry Environment, Ecce, was released on February 11, 2002 with a patch release following on May 10. This version contains major enhancements to the code registration and job monitoring components of Ecce, including the integration of a new electronic structure code, Amica. On November 13, Ecce v3.0 was made available to external sites after several months of beta testing. This version supports Ecce applications running on the Linux operating system. This includes Intel Pentium, Intel Itanium, and AMD Athlon PCs running various Linux operating systems including Redhat versions 6, 7, and 8, and Mandrake versions 8 and 9. Ecce v3.0 was released to EMSL users on December 4th. See http://ecce.emsl.pnl.gov/using/releasenotes.shtml for more details about each version. The following outlines accomplishments of Ecce during this year.

- Linux is supported for the first time.
- LSF queue management system is now supported.
- The Amica code for performing high accuracy single-point energy calculations has been registered in Ecce (http://www.hec.utah.edu/gdanitz/amica/manual.pdf).
- The Calculation Manager now includes a prototype search capability.
- Globus toolkit v2.0, from Argonne National Laboratory, has been integrated for launching jobs.
- Online help has been updated.
- Basis sets have been updated based on Dr. David Feller's reference basis set library.
- Basis set names are now used for input decks and calculation importing.
- Advanced editing capability has been incorporated in the Basis Set Tool.
- Molecular orbitals containing d and f orbitals for spherical basis sets are available in the Calculation Viewer.
- Cube files can be displayed in the Calculation Viewer.
- The installation process has been streamlined.
- The registration of computational codes has been streamlined.
- The registration of compute servers has been simplified.
- Automatic restart capability has been built into the client-side job monitoring.
- All user interface widgets were reduced to a single 3rd party widget set.
- Inter-process messaging between Ecce applications has been redesigned in Java.
- Apache2 server is now used.

Ecce Molecular Dynamics

The two most significant accomplishments with respect to supporting Molecular Dynamics (MD) calculations in 2002 have been the complete restructuring of the force field clean operation and the initial development of tools for editing and manipulating force fields. The force field clean function was completely rewritten to support localized cleans so that small sections of a large molecule could be relaxed without having to do a minimization on the entire structure. The code restructuring also resulted in greater modularization of the underlying functionality that made it relatively straightforward to use in creating the force field and topology editor GUIs. These editors are used to complete the force fields necessary to run a simulation. For most calculations, at least part of the system is not covered by standard force fields and the remaining parameters need to be supplied by the user. The topology editor will make it much easier for the user to identify what portions of the force field still need to be parameterized and the force field editor will assist in creating the supplemental force field files used by other programs. Initial development of both these editors has been completed and it is expected that usable versions of these editors will be available in spring 2003.

NWChem

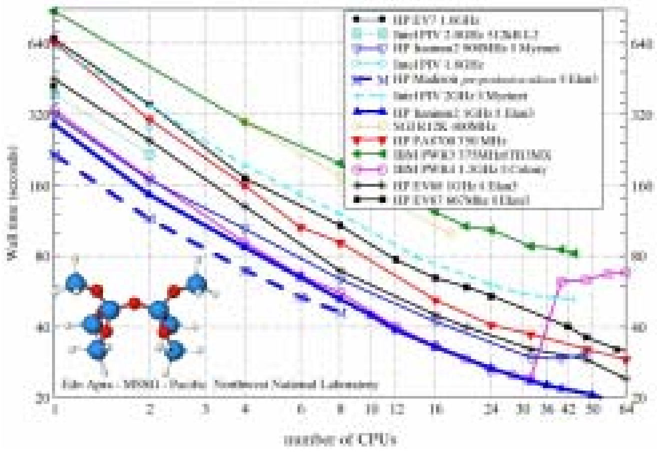


Figure 1. Wall clock time as a function of the number of processors used on various platforms for LDA calculations of 347 basis functions.

Porting NWChem and ParSoft to new architectures and development of new, scalable algorithms requires extensive benchmarking to obtain the best performance for many architectures. During this last year, we have spent extensive time understanding the architecture of the new HP system and porting our codes to that architecture. This has required us to develop multiple new, large benchmarks that are representative of the capabilities within NWChem. In Figures 1, 2, and 3, we show logarithmic plots of the wall clock time versus the number of processors used to execute the density-functional module of NWChem. In Figure 1, we show the time needed to compute the converged LDA wavefunction of the zeolite fragment (whose basis set is made of 347 Gaussian functions). The results of larger calculations are shown in Figures 2 and 3, calculations of the LDA wavefunction of the molecule (1687 basis functions) and $Si_{75}O_{48}H_{66}$ molecule (3554 basis functions), respectively. These results clearly show the excellent performance of the new HP system (HP Itanium 2 1GHz || Elan3) in the MSCF.

Figure 3 is an MP2 calculation on the $(H_2O)_7$ system using the aug-cc-pvdz basis set (287 AOs). This calculation requires extensive disk usage and, as with our other benchmarks, shows excellent performance on the new HP system.

The final example, shown in Figure 4, is a molecular dynamics simulation of haloalkanedehalogenase which has 41,259 atoms and was simulated using the AMBER force field and SPC/E water model, a cutoff of 1.0 nm, and included long range electrostatic energy and force corrections using the particle-mesh Ewald method on 64^3 grid. As can be seen from the figure, the time per step on the new HP machine is significantly less, so larger, longer simulations are possible with this machine.

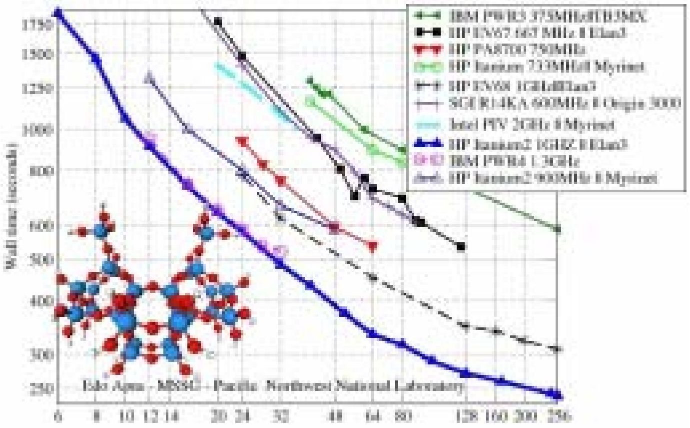


Figure 2. Wall clock time as a function of the number of processors used on various platforms for LDA calculations of 1687 basis functions.

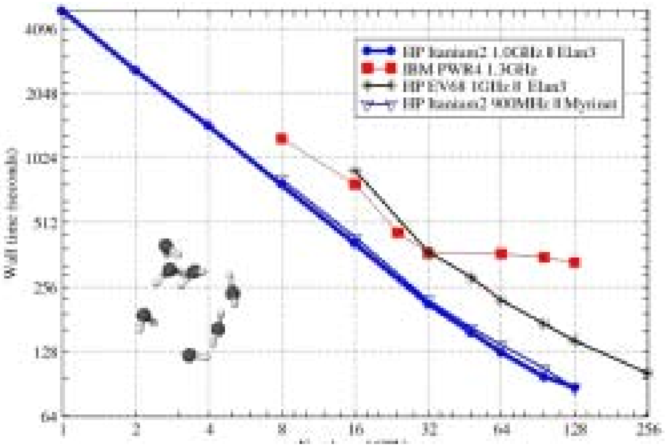


Figure 3. Wall clock time as a function of the number of processors used for an MP2 gradient calculation of $(H_2O)_7$ (287 basis functions).

Parallel Software Tools and Libraries (ParSoft)

Much of the EMSL work in ParSoft has concentrated on porting issues to the new MSCF machine and on new algorithms that are required for improved performance. One such improvement is the optimization of the performance of ARMCI accumulate operations on the MSCF HP Cluster:

The accumulate operation is important for chemistry calculations. It is defined as

Y = Y + a*X

where X and Y are distributed GA vectors, and "a" is scalar. The operation is atomic to allow multiple processes to contribute their updates to the vector Y while preserving consistency of the data. On parallel machines without Active Message interfaces, the accumulate operation is implemented as a series of the following operations: 1) lock, 2) get Y, 3) temp=Y+a*X, 4) put temp, 5) unlock. A novel technique was developed based on the Quadrics Elan Queues and Pthreads interface. It improves performance by shipping vector X to the remote side, where an additional thread executes the operation. The performance improvement is achieved by avoiding the cost of moving the vector Y in Step 2 and the locking operations. This is represented by bandwidth results in Figure 5.

The impact of this optimization on performance for sample Hartree-Fock (HF) and Density Functional Theory (DFT) calculations in NWChem is shown in Figure 6. The curves marked by "accQ" represent performance using the new implementation of the accumulate operations. In some cases a factor of two improvement is achieved.

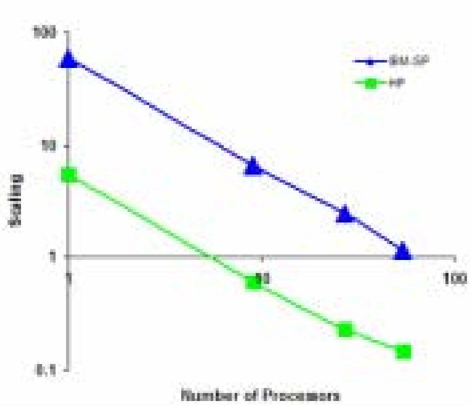


Figure 4. CPU time as a function of the number of processors used for a molecular dynamics simulation of haloalkanedehalogenase in an aqueous solution (41,259 atoms).

MSCF Supercomputer Upgrade

The 1TF of the new HP supercomputer in the Environmental Molecular Sciences Laboratory reached full operational status during the last part of 2002. The IBM supercomputer was decommissioned and removed during December of 2002. These accomplishments were finished on schedule and represent major milestone deliverables to DOE. The new supercomputer provides a wonderful new tool for the EMSL scientists. The EMSL staff has begun to build the 11.4 teraflop supercomputer in a remote HP facility. The next phase of the HP supercomputer will be installed in phases with full operations expected in summer of 2003.

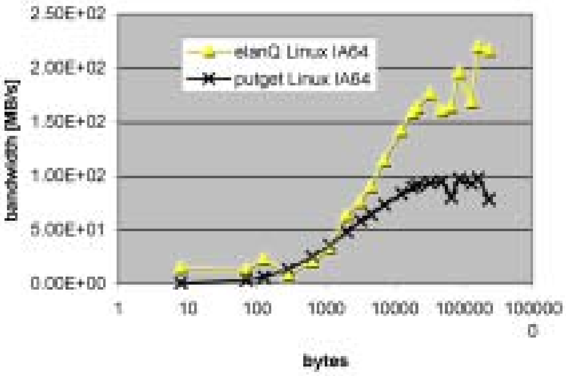


Figure 5. Bandwidth versus number of bytes of data using Linux/Itanium 2 processors with QsNet interconnect.

The computer system is being installed in phases to minimize disruptions and give early access to PNNL researchers. The ultimate system will have over 1,900 next-generation Intel Itanium Family Processors that are code-named Madison. Each of the 950 nodes will include a QSNet2/Elan4 interconnect from Quadrics.

The phased installation began May 1st 2002 with the installation of a 1/4 TFLOP/s prototype cluster that enabled EMSL researchers to tune NWChem and to develop operating system software to make the transition to the new production system seamless for the users. Phase I of the production system (Figure 7) was installed in August, 2002, and consisted of 128 HP Longs Peak nodes with 256 production-McKinley processors, corresponding to 1 TFLOP/s of computational power, and 1,572 GB of memory. The Phase I system used 1.0-GHz McKinley processors. The HP Longs Peak nodes include the HP zx1 chipset with the scalable memory expander, which provide four independent PCI-X busses, one at 1 GB/s, and a low-latency main memory with a bandwidth of 12.8 GB/s. All of the nodes

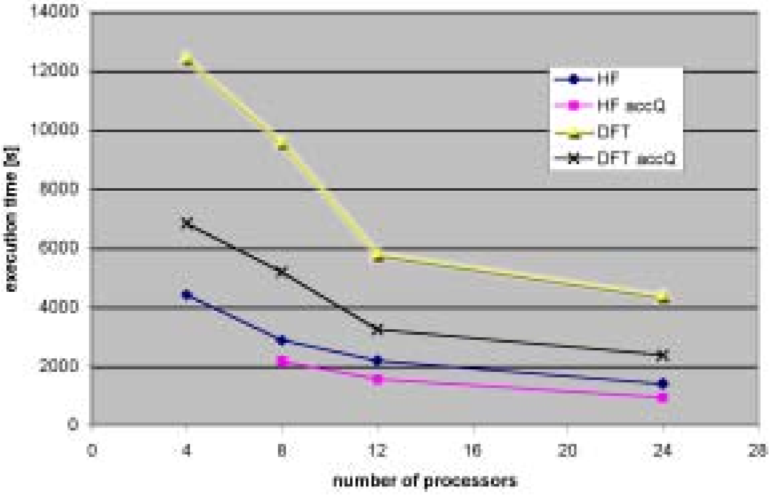


Figure 6. Impact of the new accumulation algorithm (accQ) on NWChem calculations. Time in seconds as a function of the number of processors.

will each have 190 MB/s sustained I/O rate to ~250 GB of local storage. The system also includes two rails of the Quadrics QSNet1/Elan3 interconnect, the lowest latency interconnect available at this time. The system also provides 16 TB of global storage.

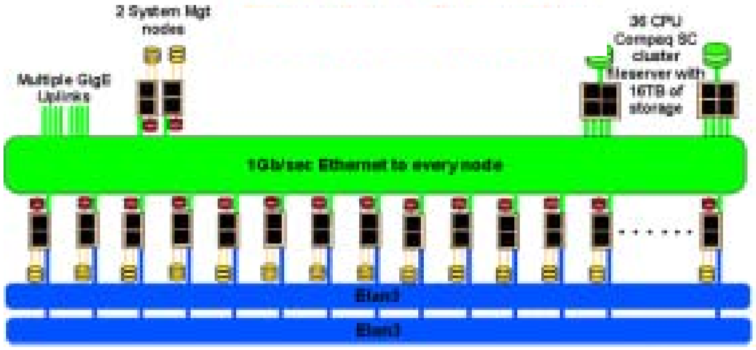


Figure 7. Phase I configuration.

Delivery of the final production system will begin in March of 2003 with the final upgrade scheduled to be completed by March 2004. The system was originally awarded as an 8.3 TF system with 1.8 TB of memory. The MSCF Operations staff has worked with HP to optimize the solution to better meet the needs of the EMSL users. The final system will have a total of over 1900 IA-64 processors with 6.8 TB of memory and 250 TB of aggregate storage. The final system (Figure 8) will include the QSNet2/Elan4 interconnect from Quadrics operating on an independent 1 GB/s PCI-X2 bus. Early access by external users to a section of this system is anticipated by summer 2003.

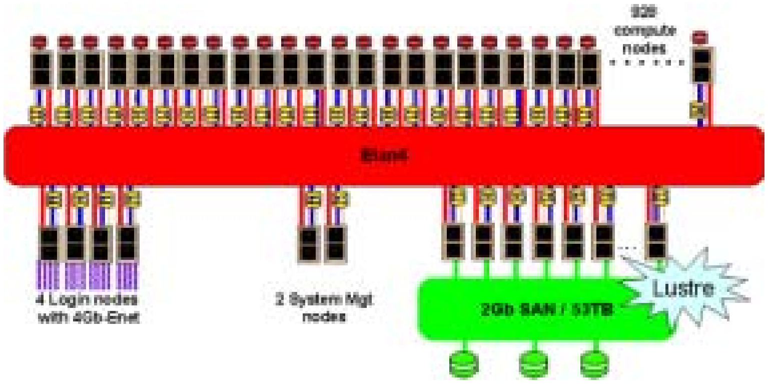


Figure 8. Final configuration.

The MSCF Operations group is working with HP to develop a scalable operating system, NW.Linux, for the final 1900+ processor cluster. NW.Linux includes all the necessary software to manage the cluster. The operations group of the MSCF has had to maintain and provide custom kernel modifications. The node installation is be done using multi-cast System Imager being developed by the MSCF operations group. Node health monitoring is being done with a multi-cast enabled version of Ganglia.

Lustre

The MSCF has been working to developing and improve the Lustre file-system for use on the HPCS2 computer system. Lustre is the global file-system that is cornerstone to the final phase of HPCS2. The design being developed will provide over 50 TB of storage and have performance of greater than 3.2GB/sec. Lustre is currently deployed in a testing environment as we continue to work with the open source community to enhance the code.

Multicast Cluster Installation

As we look forward to the next phase of the HPCS2 project we have begun to make enhancements to existing open-source system installation tools to allow quicker cluster installations using multicast. Without these enhancements the installation time for the final 950+ node cluster would take many days. Traditional install methods require the server to send the installation information individually to each of the client computers to be installed. Multicast IP protocols allow for multiple nodes to listen and send across a shared IP channel. This allows a single server to multicast installation information to many nodes at the same time, thus eliminating much of the network traffic needed in traditional installation methods. This allows multiple nodes to be installed in the same amount of time that it takes to install a single node.

Current tests show that network transmission speed from a server is the same whether it’s sending to a single client or to many clients. We’ve been able to demonstrate transmission of 550 MB of data in 30 seconds to multiple clients regardless of the number of clients.

Biomolecular Interfaces: Simulation of Sedimentation of Biomolecular Assemblies

AD Haymet,[(a)] BW Beck,[(a)] GC Lynch,[(a)] BM Pettitt,[(a)] TC Rogala,[(a)] and K Wong[(a)]
(a) University of Houston, Texas

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

Computationally Useful Bridge Diagram Series

In many-body theory, bridge diagrams are a set of topologically dense, large dimensional integrals of interest in the theoretical study of liquids. We calculated a new series for simple fluid mixtures of argon and krypton. Our results were both numerically encouraging for this system, and theoretically encouraging for the ongoing extension to more complex fluids and solvation studies (Figure 1) (Dyer et al. 2002, Perkyns et al. 2002).

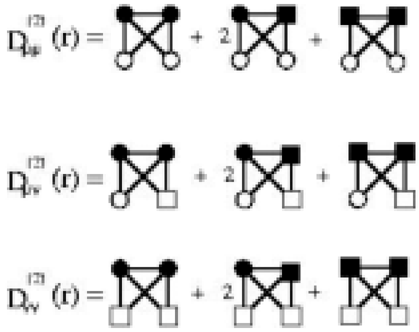


Figure 1. Theoretical Bridge diagrams for a set of topologically dense, large dimensional integrals of interest in study of liquids.

Base-pairing on Surfaces for Untethered Oligonucleotides

Recent designs of DNA microarrays use non-tethered DNA probes, which are bound to a positive charged surface by nonbonded, electrostatic interactions. They show specific, high affinity binding of large targets with relatively rapid kinetics. A 10 ns all-atom molecular dynamics simulation of the system was performed and found that the probe was bound to the surface primarily with the phosphates attracted to the surface. The bases remained stacked and pointed preferentially upward, available for hybridization with complementary target strand (Figure 2) (Belosludtsev et al. 2001, Wong et al. in press).

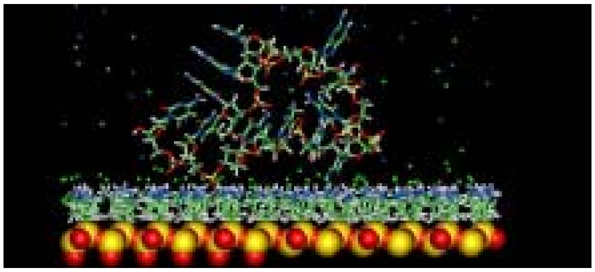


Figure 2. Molecular dynamics simulation of base-pairing on surfaces for untethered oligonucleotides.

Solvation of Biomolecular Interfaces: Role of Solvent in S. marcescens Endonuclease Dimerization

Simulations have revealed the presence of a solvent filled channel leading from the active site cavity of the enzyme to three distinct openings near the dimer interface. The presence of this channel has implications to both protein stability and regulation of enzymatic activity.

This channel appears to have an intermittent barrier midway down its length near residues Ile174, Ala186, and Phe188 that acts as a "gate" to control the diffusion of water between the active and the dimer interface.

The behavior of this gate appears to be important in the function of this enzyme and may help explain the evolution of the dimeric state for this system (Figure 3).

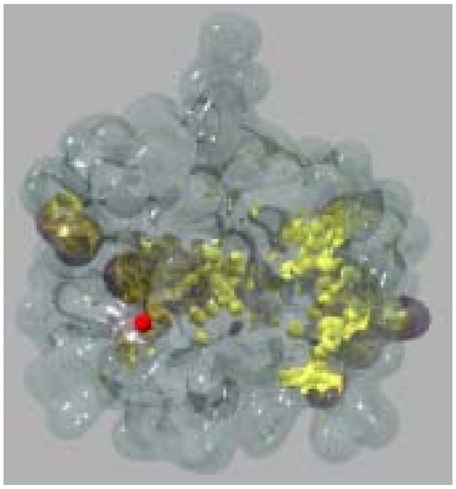


Figure 3. Average. Channel solvent density in enzyme monomer over 1ns.

References

Belosludtsev, Y, B Iverseon, S Lemeshko, R Eggers, R Wiese, S Lee, T Powdrill, and M Hogan. DNA Microarrays Based on Noncovalent Oligonucleartide Attachment and Hybridization in Two Dimensions. *Analytical Biochemistry,* **292***(2), 250-256* (2001).

Dyer, K, J, Perkyns, and BM Pettitt. Computationally useful bridge diagram series. III. Lennard-Jones mixtures. *J. Chem. Phys.,* **116**, 9413 (2002).

Perkyns, JS, KM Dyer, and BM Pettitt. Computationally usedful bridge diagram series. II. Diagrams in h-bonds. *J. Chem. Phys.,* **116**, 9404 (2002).

Computational Studies in Molecular Geochemistry and Biogeochemistry

AR Felmy,[a] DA Dixon,[a] M Dupuis,[a] TP Straatsma,[a] KM Rosso,[a] EJ Bylaska,[a] RD Lins Neto,[b] GA Voth,[c] JH Weare,[d] DA Yuen,[e] JW Halley,[e] DG Rancourt,[f] and PE Smith[g]

(a) Pacific Northwest National Laboratory
(b) Laboratory of Physical Science, Swiss Federal Institute of Technology, Zurich, Switzerland
(c) University of Utah, Salt Lake City
(d) University of California at San Diego
(e) University of Minnesota, Minneapolis
(f) University of Ottawa, Ontario, Canada
(g) Kansas State University, Manhattan

The ability to predict the transport and transformations of contaminants subsurface is critical for decisions on virtually every waste disposal option, from remediation technologies to evaluations of the safety of nuclear waste repositories, facing the Department of Energy (DOE). With this in mind, the DOE has recently sponsored a series of workshops on the development of a Strategic Simulation Plan (SSP) on applications of high performance computing to national problems of significance to the DOE. One of the areas selected for application was of subsurface transport and environmental chemistry. Within the SSP on subsurface transport and environmental chemistry, several areas were identified where applications of high performance computing could potentially significantly advance knowledge of contaminant fate and transport. 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.

Ab Initio Molecular Dynamic (AIMD) Simulations of Proton Exchange Around Highly Charged Ions in Solution

The properties of ions in aqueous solutions is of enormous importance to the interpretation of processes as diverse as the biological function of proteins, subsurface transport, mineral formation, and the cleanup of toxic waste.

In these systems the interaction of the ion with the surrounding water molecules leads to strong polarization, which impacts all aspects of aqueous complexation. Figure 1 illustrates proton motion seen in an AIMD simulation of the hydrolysis of the Al^{3+} ion. Very small barriers are seen for proton transfer from the first solvation shell to the second solvation shell. AIMD

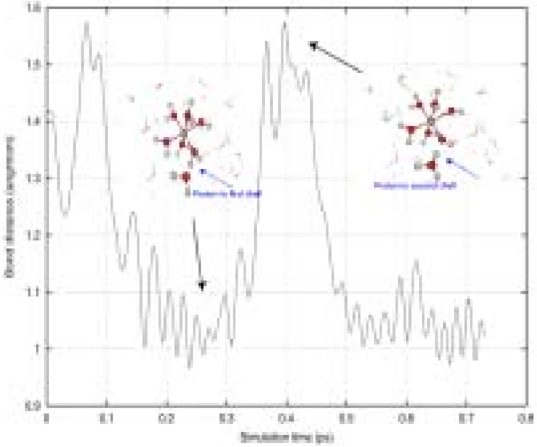


Figure 1. Proton motion seen in an AIMD simulation of the hydrolysis of the Al3+ ion.

simulations can provide critical insight into these types of speciation effects as well as for oxidation reduction reactions which require understanding the coupling of proton transport and electron transport (Lubin et al. 2000).

Development of Massively Parallel Ab Initio Plane-wave Molecular Dynamics Programs

Collaborators: Marat Valiev, Kiril Tsemekhman, Hannes Jonnson, James R. Rustad, Edo Apra

In this project, a state-of-the-art pseudopotential plane-wave Density Function Theory (PSPW) code has been implemented into the NWChem program package. NWChem is a computational chemistry package for parallel computers developed at PNNL. Both internal and external users are using this program, and the capabilities developed in this project have contributed to several DOE programs, including the Environmental Management Science Program, Basic Energy Science (BES) Nanoscale Science, Engineering, and Technology, BES Geosciences, and BES Material Sciences.

PSPW methods, unlike other Density Function Theory (DFT) or other quantum mechanical approaches, are not based on the traditional expansion of the wavefunction using basis functions centered on the atom, but rather the wavefunctions are expanded using plane-waves and pseudopotentials are used to represent atomic cores. The advantage of the PSPW method is that it has been shown to have accuracies close to chemical accuracy for many applications, yet is still fast enough to treat systems containing hundreds of atoms. Another significant advantage is its ability to simulate dynamics on a ground state potential surface directly at run-time. This method's efficiency and accuracy make it a desirable first-principles method of simulation in the study of complex molecular, liquid, and solid-state systems. Applications for this first-principles method include the calculation of free energies, search for global minima, explicit simulation of solvated molecules, and simulations of complex vibrational modes that cannot be described within the harmonic approximation.

Ab Initio Methods for Geochemical Electronic Structure Studies

Collaborators: Marat Valiev, James R. Rustad, Michel Dupuis

In this project we have developed a massively parallel code employing a state-of-the-art all-electron plane-wave DFT method, and have we developed plane-wave quantum mechanics/molecular mechanics (PW/MM), which can be used to study important geochemical processes (structures and reactivity) at solvated redox-active natural mineral surfaces and interfaces. These include changes in oxidation states, changes in coordination and oxidation states (i.e., bond breaking and bond formation), and redox reactions of pollutants on iron-oxides and other environmentally important surfaces. Figure 2 illustrates a coupled electron+proton transport mechanism in the reduction of chromate by biotite which can be studied using PW/MM methods (Bylaska et al. 2002).

Simulations of High Temperature (Pressure) Hydrothermal Solutions

Collaborator: Marat Valiev

Much of our molecular understanding of these hydrothermal solutions has been provided by X-ray adsorption fine structure experiments, which observe increasing ion-pair formation at elevated temperatures. However, the interpretation of these measurements usually requires some structural information about the solvation shell around the ions. Our AIMD calculations are being used to probe the chemistry of the metal ion solvation shell as well as ion-pairing kinetics and stability as a function of temperature, pressure and concentration (Lubin et al. 2000).

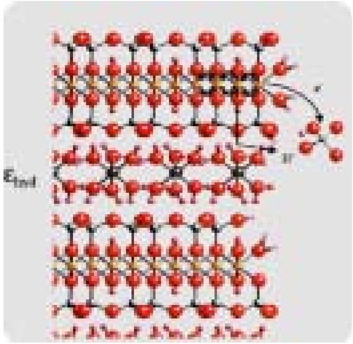


Figure 2. Exhibited: Mechanism of chromate reduction by natural

Classical molecular models of the goethite mineral were developed for both a periodically replicated bulk slab and for a finite fragment. Potential energy functions for the mineral models were developed within the AMBER force field, which facilitates biogeochemical applications. Intramolecular potential parameters were chosen to yield a semi-rigid mineral structure with flexible surface hydroxyl groups. The electrostatic potential for a goethite slab was determined from a periodic Unrestricted Hartree Fock calculation and used to assign point charges to the mineral atoms. The models were solvated in water and their effect on solution structure is shown to be very similar, predicting oscillations in solvent density near the mineral surface and structured orientation of water molecules (Figure 3) (Shroll and Straatsma in press a).

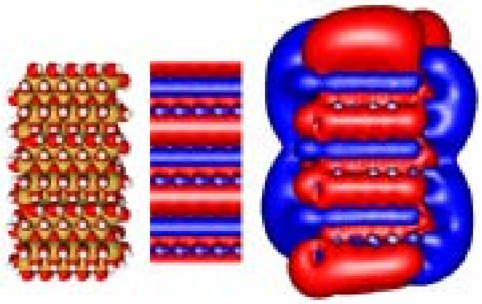


Figure 3. Electrostatic potential for a goethite slab.

Microbial Adhesion to Mineral Surfaces

The adhesion of *Pseudomonas aeruginosa* to the goethite mineral is investigated using classical molecular simulation. A fragment model for goethite has been integrated into a fully atomistic membrane model. Properties for the resulting system are evaluated for a 1.5 ns simulation in the isothermal-isobaric ensemble.

The response of the membrane to the presence of the mineral is investigated. Radial distribution functions are used to present an average picture of the hydrogen bonding. Analysis of orientational vectors, assigned to the saccharide groups, reveals the extent of the mineral's perturbations on the membrane.

Significant structural changes were observed for the outermost saccharide groups, several of which rotate to form hydrogen bonds with the mineral surface. The structure of the inner core, and the corresponding integrity of the membrane, is maintained. The mineral surface dehydrates slightly in the presence of the membrane as saccharide hydroxyl groups compete with water molecules for hydrogen bonding sites on its surface (Figure 4) (Shroll and Straatsma in press b, Shroll and Straatsma 2002, Lins and Straatsma 2001).

One-electron Reduction of Polychlorinated Solvents and Pesticides

Collaborators: Paul G. Traynyek, Dave A. Dixon

Currently in environmental chemistry, the predominant method for estimating reduction potentials is to use structure activity relationships, such as Benson's group additivity method. These classical methods are known to be reasonably accurate. However, it difficult to systematically improve the accuracy of a group additivity method since it is difficult to explicitly account for neighboring interactions and long-range through-bond effects.

As a consequence, these methods are best used for interpolating properties of a homologous series of molecular species. Another way to estimate reduction potentials is to use ab initio methods. Ab initio methods do not suffer from the neglecting neighboring interactions and have been shown to useful for calculating the thermodynamic properties for many types of species (Bylaska et al. 2002, 2000).

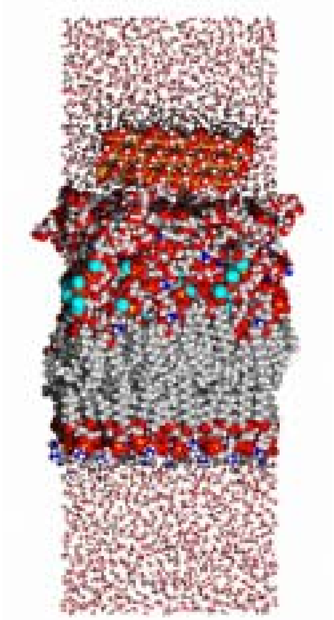


Figure 4. Adhesion of *Pseudomonas aeruginosa* to the goethite mineral.

References

Bylaska, EJ, DA Dixon, and AR Felmy. The Free Energies of Reactions of Chlorinated Methanes with Aqueous Monovalent Anions: Application of ab initio Electronic Structure Theory. *Journal of Physical Chemistry A*, **104**, 610-617 (2000).

Bylaska, EJ, DA Dixon, AR Felmy, and PG Tratnyek. One-Electron Reduction Of Substituted Chlorinated Methanes as Determined from Ab Initio Electronic Structure Theory. *Journal of Physical Chemistry A*, **106**(47): 11581-11593 (2002a).

Bylaska, EJ, M Valiev, R Kawai, and JH Weare. Parallel Implementation of the Projector Augmented Plane Wave Method for Charged Systems. *Computer Physics Communications.*, **143**, 11-28 (2002b).

Lins, RD, and TP Straatsma. Computer Simulation of *Pseudomonas aeruginosa* Adhesion to Goethite. *Biophysical Journal*, **81**, 1037-1046 (2001).

Lubin, MI, EJ Bylaska, and JH Weare. Ab initio molecular dynamics simulations of aliminum ion salvation in water clusters. *Chem. Phys. Lett.*, **322**, 447-453 (2000).

Shroll, RM, and TP Straatsma. Molecular Basis for Microbial Adhesion to Geochemical Surfaces. *Biophysics Journal*, (in press a).

Shroll, RM, and TP Straatsma. Molecular Structure of the Outer Bacterial Membrane of *Pseudomonas aeruginosa* via Classical Simulation. *Biopolymers*, **65,** 395-407 (2002).

Shroll, RM, and TP Straatsma. Molecular Dynamics Simulations of the Goethite-Water Interface. *Molecular Simulation* (in press b).

Simulation of Climate Forcing by Aerosols

SJ Ghan,[a] GI Fann,[a] X Bian,[a] EG Chapman,[a] RC Easter,[a] and RA Zaveri[a]
(a) Pacific Northwest National Laboratory

The largest source of uncertainty in estimates of the radiative force governing climate change is in the radiative forcing by anthropogenic aerosols. Current estimates of the global mean of the aerosol radiative forcing range from -0.3 watt per square meter (Wm^{-2})—a small fraction of the +2 Wm^{-2} forcing due to increasing greenhouse gases, to -3 Wm^{-2}—larger than the global mean greenhouse forcing, but opposite in sign. We have developed a global aerosol model that provides arguably the most detailed treatment of aerosols and their impact on the planetary radiation balance of any model, but our estimates of radiative forcing have been hindered by our lack of access to high performance computing resources. We propose to use the MSCF to conduct a series of simulations with and without emissions of a variety of aerosol particles and aerosol precursors. These extensive simulations will enable us to produce much more refined estimates of the impact of anthropogenic emissions on radiative forcing of climate change. To take full advantage of the parallelism available on the MSCF MPP1, we will apply the Global Array Toolkit to dynamically load balance the reactive chemistry component of our model. We will adapt our modifications of the serial NCAR Community Climate Model CCM2 to the parallel NCAR CCM3.6. We will also explore the application of parallel I/O using Chemio and MPI-I/O (e.g. ROMIO). In the area of reactive transport algorithms we will test the Exact Linear Part semigroup methods for advection-diffusion equations.

Application of a Subgrid Orography Scheme to a Global Climate Model

A subgrid orography scheme has been applied to the NCAR Community Atmosphere Model. The scheme produces climate predictions for each of a modest set of elevation classes within each grid cell. Climate fields can be distributed in postprocessing according to the high-resolution distribution of topography to produce global climate simulations with unprecedented detail. Most improved is the distribution of snow water, which is crucial for water resource management (Ghan, et al. 2002).

References

Ghan, SJ, X Bian, AG Hunt, and A Coleman. The thermodynamic influence of subgrid orography in a global climate model, *Climate Dynamics*, **208**, 31-44 (2002).

First Principles Methods for Predicting the Chemistry of Environmentally Relevant Systems

RA Friesner,[(a)] HA Stern,[(a)] BF Gherman,[(a)] BD Dunietz,[(a)] F Rittner,[(a)] E Batista,[(a)] MH Baik,[(a)] ME Wirstam,[(a)] Z Zhou,[(a)] WA Goddard,[(b)] DT Mainz,[(b)] WB Floriano,[(b)] RP Muller,[(b)] M Blanco,[(b)] YH Jang,[(b)] RJ Nielsen,[(b)] DM Philipp,[(b)] F Faglioni,[(b)] V Molinero,[(b)] MS Diallo,[(b)] M Beachy,[(c)] and RB Murphy[(c)]

(a) Columbia University, New York
(b) California Institute of Technology, Pasadena, California
(c) Schrodinger, Inc., Jersey City, New Jersey

The long-term objectives of this project are the development of integrated computational methods and software for modeling the chemistry of environmentally relevant systems. In order to model reactive chemistry in heterogeneous condensed media that characterizes environmentally relevant systems, a number of key core technologies will be developed and validated.

Hydroxylation of Camphor by Cytochrome P450cam

The enzymatic hydroxylation of camphor by cytochrome P450cam was studied using a novel computational method that allows for combining highly accurate Quantum Mechanical (QM) and computationally efficient Molecular Mechanics (MM) methods to construct a realistic computer model of the enzymatic reaction containing ~7,500 atoms. A reaction energy profile displaying a better agreement with experimental observations was obtained and the role of the protein environment for controlling key features of the hydroxylation reaction was elucidated (Figure 1) (Gullar et al. 2002, 2003).

Figure 1. An illustration of the computed transition state of the camphor hydroxylation reaction using a novel QM/MM method.

References

Guallar, V, BF Gherman, SJ Lippard, and RA Friesner. Quantum chemical studies of methane monooxygenase comparison with P450. *Curr. Opin. Chem. Biol.*, **6**, 236-242 (2002).

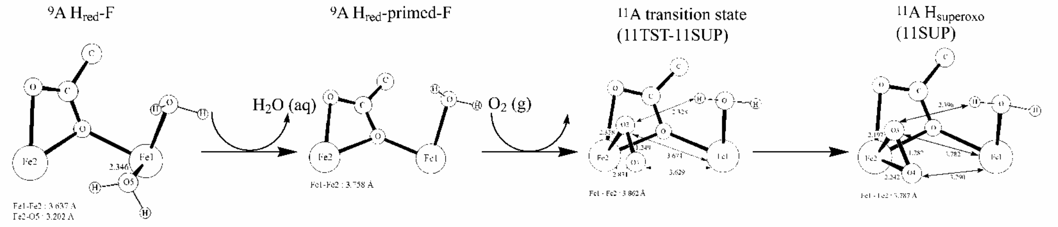
Guallar, V, MH Baik, SJ Lippard, and RA Friesner. Peripheral heme substituents control the hydrogen-atom abstraction chemistry in cytochoromes P450. Proceedings of the National Academy of Sciences of the United States of America, **100**, 6998-7002 (2003).

Ab Initio Quantum Chemical Study of Dioxygen Activation in Methane Monooxygenase (MMO)

Within the hydroxylase component (MMOH) of the soluble MMO system, the oxidation of methane to methanol by dioxygen is catalyzed at non-heme dinuclear iron active sites (Merkx et al. 2001). A large-scale theoretical model (~100 atoms) of the MMOH active site has been constructed (Dunietz et al. 2000) which reproduces the crystal structure for Hred (Rosenzweig et al. 1995) and is in accord with spectroscopic data (Merkx et al. 2001). Ab initio quantum chemical density functional (DFT) methods employing the B3LYP functional (Jaguar 2000) have been used to study the dioxygen activation phase of the MMOH catalytic cycle (Figure 2).

Ab Initio Quantum Chemical Study of Dioxygen Activation in Methane Monooxygenase (MMO)

The dioxygen activation process in MMOH begins with the loss of the weakly bound water ligand from Fe1 to solution in a nearly isoenergetic reaction. With both iron atoms now 5-coordinate (and, hence, with one open coordination site on each iron atom), dioxygen then reacts with MMOH at Fe_2 in a spin-allowed reaction to yield a multiplicity 11, ferromagnetically-coupled mixed-valent Fe(III)Fe(II) $H_{superoxo}$ intermediate. A weak hydrogen bonding interaction between O3 and the remaining water ligand stabilizes both the transition state and $H_{superoxo}$. Bidentate binding of the superoxo moiety to Fe_2 further stabilizes $H_{superoxo}$ (Figure 3).



bridging group. The possibility of two Hperoxo forms could explain the differing rates observed for Hperoxo formation (Figure 4).

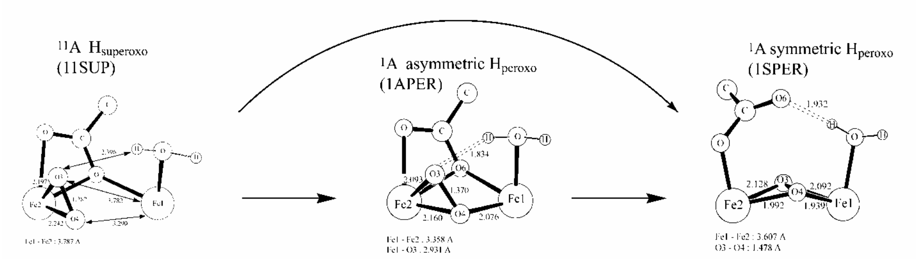


Figure 4. Mechanism of dioxygen activation: H_{seroxo} formation (only the core of the model is shown, dashed lines indicate hydrogen bonds).

Ab Initio Quantum Chemical Study of Dioxygen Activation in Methane Monooxygenase (MMO)

The final step in dioxygen activation, symmetric H_{peroxo} → Q, proceeds via homolytic cleavage of the peroxo O3–O4 bond. At the transition state, partial electron transfer from Fe1 to the dioxygen moiety has occurred. A mixed-valent Fe(III)Fe(IV) intermediate then forms and decays barrierlessly to give the high-valent Fe(IV)Fe(IV) Q, which has a bis-μ-oxo core in which the two irons have nearly equivalent coordination environments.

The relative energies for all the intermediates and transition states involved in dioxygen activation indicate that: a) H_{peroxo} is the first spectroscopically observable species following the decay of H_{red}, b) the overall reaction from H_{red} to Q is energetically downhill.

The proposed dioxygen activation mechanism suggests that both H_{peroxo} formation and H_{peroxo} decay are pH independent and without kinetic solvent isotope effects (Figure 5).

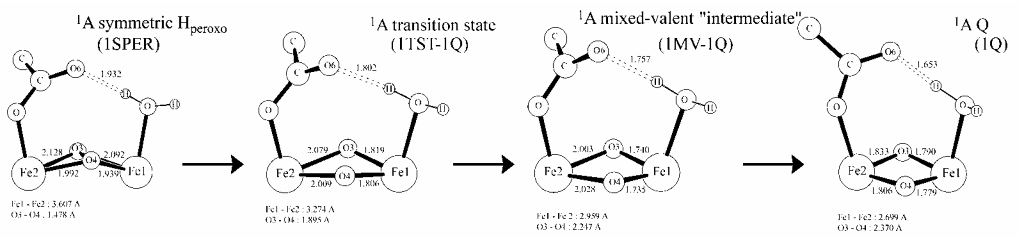


Figure 5. Mechanism of dioxygen activation: Q formation (only the core of the model is shown, dashed lines indicate hydrogen bonds).

A Self-Consistent Charge-Embedding Methodology for Ab Initio Quantum Chemical Cluster Modeling of Ionic Solids and Surfaces: Calculation of Work Functions of Metal Oxides

The understanding of the surface chemistry of metal oxides is one of the key issues in the study of environmental problems, such as the transport of organic material and heavy metals as they move through soil, and in many industrial applications ranging from catalysis to photoelectric cells. In performing ab-initio simulations of ionic crystals, the crystal is modeled using a cluster of ions embedded in an array of point charges. These point charges must be chosen to reproduce the crystal field at the surface of the crystal. A method for determining the right charges was developed. The point charges are obtained via a self consistent loop of fitting the charges to reproduce the field obtained by ab-initio calculations, using gradient corrected DFT, in the region of the cluster and its surroundings. This method was applied to the study of the work function of hematite (α-Fe_2O_3) and anatase (TiO_2) with results in reasonable agreement with experimental measurements (Figure 6).

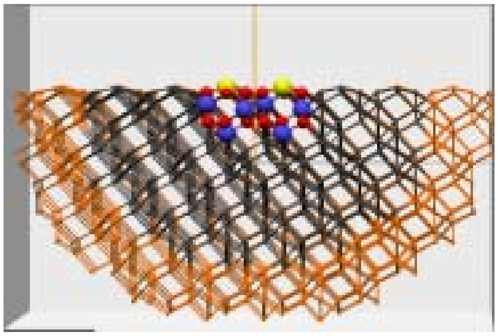


Figure 6. Study of the work function of Hematite (a-Fe_2O_3) and Anatase (TiO_2)

Work Function of Hematite (α-Fe_2O_3)

For a point charges approach, the electrostatic properties converge to crystal values as function of the size of the cluster for clusters of the order of 60 atoms.

The work function of this surface was calculated at 5.6 eV, in reasonable agreement with experimental measurements which range between 5.12 and 5.95 eV (Figure 7) (Gleitzer et al. 1991).

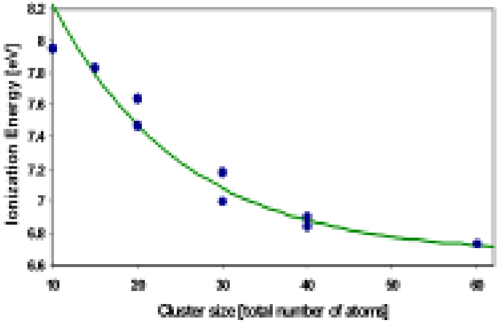


Figure 7. Ionization energy of the clusters as a function of the total number of atoms: Work Function of 1x1 anatase TiO_2 From Cluster Calculations.

Work Function of 1x1 Anatase TiO_2 from Cluster Calculations

We find that TiO_2 clusters converge to bulk behavior much sooner than iron oxide. A cluster of 18 atoms, embedded in point charges, has already converged to the work function of the larger clusters, 5.50 eV. This value agrees with experimental measurements on rutile TiO_2 which range from 4.6 to 5.9 eV (Figure 8) (Henrich and Cox).

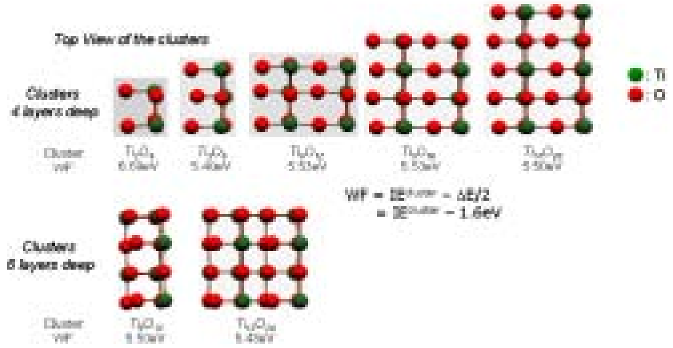


Figure 8. Work Function of 1x1 Anatase TiO2 From Cluster Calculations.

References

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

Gherman, BF, M-H Baik, SJ Lippard, and RA Friesner. Dioxygen activation in methane monooxygenase: a theoretical study. *J. Am. Chem. Soc.*, submitted for publication.

Gleitzer, C, J Nowotny, and M Rekas. Surface and Bulk Electrical Properties of the Hematite Phease Fe_2O_3. *App. Phys. A*, **53**, 310 (1991).

Henrich, VE and PA Cox. The Surface Science of Metal Oxides. New York: Cambridge University Press (1994).

Jaguar 4.1, Schrodinger, Inc., Portland, Oregon (2000).

Liu, KE, D Wang, BH Huynh, DE Edmondson, A Salifoglou, and SJ Lippard. Spectroscopic Detection of Intermediates in the Reaction of Dioxygen with the Reduced Methane Monooxygenase Hydroxylase from *Methylococcus capsulatus* (Bath). *J. Am. Chem. Soc.,* **116**, 7465 (1994).

Merkx, M, DA Kopp, MH Sazinsky, JL Blazyk, J Muller, and SJ Lippard. Dioxygen Activation and Methane Hydroxylation by Soluble Methane Monooxygenase: a Tale of Two Irons and Three Proteins. *Angew. Chem. Int. Ed.* **40**, 2782 (2001).

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

Valentine, AM, SS Stahl, and SJ Lippard. Mechanistic Studies of the Reaction of Reduced Methane Monooxygenase Hydroxylase with Dioxygen and Substrates. *J. Am. Chem. Soc.*, **121**, 3876-3887 (1999).

Reliable Electronic Structure Calculations for Heavy Element Chemistry: Molecules Containing Actinides, Lanthanides, and Transition Metals

WA de Jong,[(a)] *BP Hay,*[(a)] *DA Dixon,*[(a)] *J Li,*[(a)] *Z Zhang,*[(b)] *KG Dyall,*[(c)] *RJ Harrison,*[(d)] *BE Bursten,*[(e)] *MM Marino,*[(f)] *WD Ermler,*[(f)] *C Marsden,*[(g)] *K Hira,*[(h)] *L Visscher,*[(i)] *and SJA van Gisbergen*[(i)]

(a) Pacific Northwest National Laboratory
(b) Lawrence Berkeley National Laboratory, Berkeley, California
(c) NASA Ames Research Center, Moffett Field, California
(d) Oak Ridge National Laboratory, Oak Ridge, Tennessee
(e) Ohio State University, Columbus
(f) University of Memphis, Tennessee
(g) University of PaulSabatier/Laboratoire de PhysiquQuantique – IRSAMC, Toulouse, France
(h) University of Tokyo, Japan
(i) Vrijie Universiteit, Amsterdam, Netherlands

We propose to perform ab initio electronic structure calculations based on molecular orbital theory and density functional theory with the proper treatment of relativistic effects to study complexes of heavy elements in order to assist in understanding and predicting the chemistry of the actinides, lanthanides, and heavy transition metals—molecules critical to DOE missions including environmental management. The specific molecular species to be studied will be selected in conjunction with experimental and theoretical efforts at PNNL and elsewhere with emphasis on the environmental restoration needs of the DOE. A wide variety of relativistic and non-relativistic quantum chemical methods will be used to assist in the understanding and prediction of actinide, lanthanide, and transition metal chemistry with the goal of giving a firm theoretical basis to this area, and to extending expensive experimental results into new areas of parameter space. Information that can be provided for actinide containing molecules includes, but is not limited to:

- molecular structure and complex formation
- spectroscopic properties including electronic, vibrational and NMR
- complexation binding energies, redox chemistry, and solvation effects and energies.

This data will contribute to the characterization of the interaction of the actinide, lanthanide and heavy transition metal ions with organic complexing agents present in nuclear processing waste tanks and with anion ligands present in natural aqueous systems (e.g., carbonates) in order to better understand their fate and transport in the environment, as well as interactions with new materials such as phosphates and amides for the design of innovative in situ remediation technologies and novel separation systems.

Calculation of ^{17}O NMR Parameters in $UO_2(NO_3)_2{\cdot}3H_2O$

In this project, computational chemistry is partnering with NMR experimentalists to develop a new tool for speciation of nuclear waste tanks at Hanford using the unique NMR signals of molecules.

Our calculated NMR property tensors predicted a record setting (two orders of magnitude larger) chemical shift anisotropy for the ^{17}O NMR signal in uranyl nitrate crystals and a very broad NMR signal of around 1800 ppm.

Our predicted broad signal, which makes the experimental measurements difficult, has recently been confirmed in experiments by Herman Cho, using the NMR facility in EMSL. A good agreement was found between the experimental and simulated spectrum.

Actinide-Noble Gas Bonds: CUO(Ng) (Ng = Ar, Kr, Xe)

The CUO molecule has experimental U-C and U-O stretching frequencies at 1047 and 872 cm^{-1} in a neon matrix, and these frequencies red-shift by 195 and 68 cm-1 in the argon matrix. Density Functional Theory (DFT) calculations were used to understand the nature of these shifts.

The calculations (Figure 1) indicate that CUO formed in the neon matrix is a singlet (1S+), while it is a triplet (3F) in the argon matrix. The triplet becomes the ground state when Ar atoms bind to CUO molecule.

Experiments and the calculated CUO-Ng bond distances and bond energies reveal the formation of the first actinide-noble gas bonds. This discovery provides new insight and interpretation of some previous experimental results.

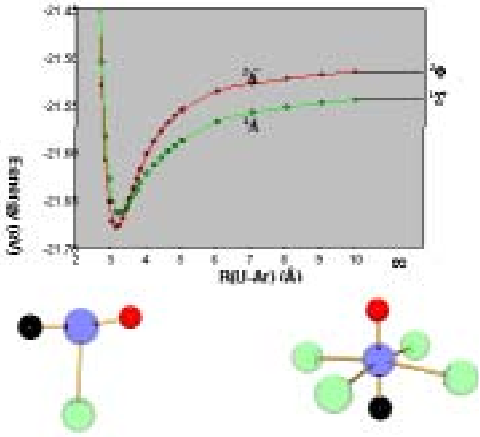


Figure 1. Potential energy curves of CUO in the argon matrix (top) and some structures of Ar-CUO complexes (bottom).

The results were published in *Science* (Li et al. 2002).

Analysis of Solvation effects of H_2O and CCl_4 on Uranyl by COSMO and DFT

In this study DFT and the conductor-like screening model (COSMO) were used to determine the effects of solvation on the structures and energetics of formation of uranyl and carbon tetrachloride complexes, and to calculate solvation energies.

These calcuated parameters will be used in Molecular Dynamics (MD) simulations to better understand the mechanism and kinetics of transport of uranyl between different solvent phases. This understanding would allow prediction of environmental effects of uranyl as well as provide insight into methods for remediation.

Energies of reaction were computed for
$UO_2^{2+} + nCCl_4 \rightarrow [UO_2(CCl_4)_n]^{2+}$
$[UO_2^{2+}(CCl_4)_n] + CCl_4 \rightarrow [UO_2(CCl_4)_{n+1}]^{2+}$

Solvation energies are computed for both carbon tetrachloride and water dielectrics (Figure 2).

Calculation of NMR parameters NH_4TcO_4

Computational chemistry calculations are used to calculate, simulate, and understand ^{99}Tc and ^{17}O NMR spectra measured by experimentalists at PNNL. The goal of this collaborative project is to provide new ways to analyze the rate of success in the extraction of technetium from nuclear waste, and its oxidation state.

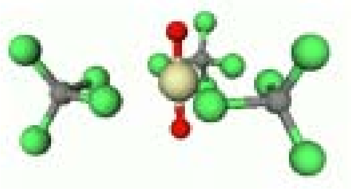


Figure 2. Optimized geometry for $[UO_2(CCl_4)_3]^{2+}$

The electric field gradient and chemical shift tensors of both ^{99}Tc and ^{17}O were calculated for various cluster sizes. Good agreement is found between the calculated and the measured spectra (Figure 3), the ^{99}Tc spectra are shown on the right. Calculated chemical shifts and spin-spin couplings are found to be in agreement with experimental observations. By analyzing both the Tc and the O data a complete picture of the bonding in TcO_{4-} is obtained.

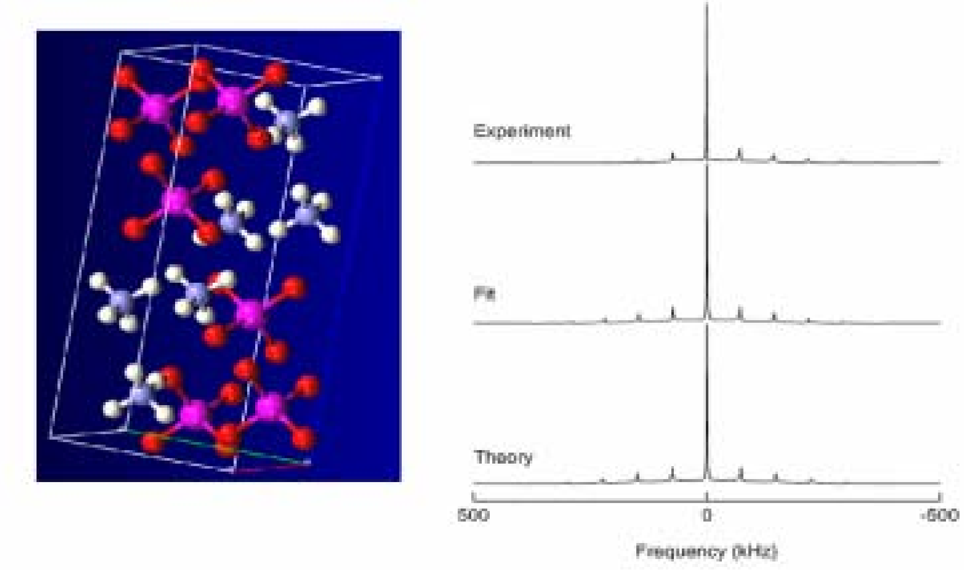


Figure 3. NH_4TcO_4 crystal unit cell and comparison between experimental and simulated ^{99}Tc NMR spectrum.

Structure and Energetics of the Hydrated Electron

A series of electronic structure calculations have been performed to determine possible stable structures for the hydrated electron and to determine its absolute hydration free energy, $\Delta G_{hyd}^{298}(e-)$. Based on the most stable structures, the absolute hydration free energy of the solvated electron in water has been determined using a reliable computational protocol of first-principles solvation-included electronic structure calculations.

The predicted $\Delta G_{hyd}^{298}(e-)$ value combined with our previously predicted $\Delta G_{hyd}^{298}(H^+)$ and $\Delta G_{hyd}^{298}(HO^-)$ (Zhan and Dixon in press, 2001) gives values for $\Delta G_{hyd}^{298}(e-) + \Delta G_{hyd}^{298}(H^+)$ and $\Delta G_{hyd}^{298}(e-) – \Delta G_{hyd}^{298}(HO^-)$ in excellent agreement with the corresponding values derived from available experimental data.

Actinide-Noble Gas Complexes

DFT calculations (Figure 4) suggest that the $CUO(Ng)_n$ ($n > 1$) species are more stable than CUO(Ng) (Ng = Ar, Kr, Xe) and that all the $CUO(Ng)_n$ species prefer the triplet ground state.

Subsequent experiments observed a series of substitution complexes $CUO(Ar)_{4-n}(Xe)_n$ [also $CUO(Ar)_{4-n}(Kr)_n$ and $CUO(Kr)_{4-n}(Xe)_n$]. Our theoretical predicted geometries and vibrational frequencies of $CUO(Ar)_{4-n}(Ng)_n$ molecules are consistent with the experimental assignments. The IR spectra show a slight red-shift of the U-C and U-O bands upon substitution by a heavier Ng atom. These are the first actinide-noble gas complexes and are the first neutral complexes with four Ng atoms bonded to one metal center (Figure 4).

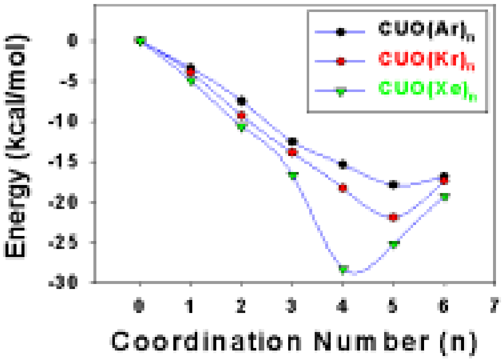


Figure 4. Binding energies of CUO(Ng)n.

Results have been published in *JACS* (Liang et al. 2002).

Using DFT to Explore the Structural Dichotomy Of Cp_3MX Compounds (M=Ti,Zr,Hf;X=Cl,CH_3)

Cp_3MX compounds exhibit two structure types (Figure 5). The Ti and Hf analogs prefer the 2-5,1-1 orientation (right) while the Zr analog prefers the 3-5 orientation.

DFT calculations were used to correctly predict the 3-5 isomer of Cp_3ZrCl to be more stable than the 2-5,1-1 isomer; they also correctly predict the 2-5,1-1 isomer of Cp_3HfCl to be more stable than the 3-5 isomer.

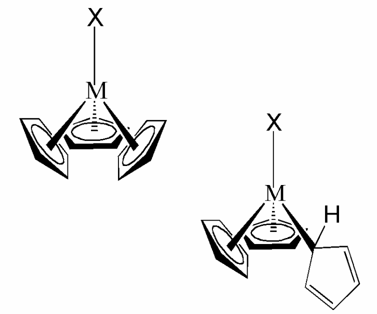


Figure 5. Two possible orientations of the Cp_3MX compounds, 2-5,1-1 (left), 3-5(right).

Preliminary results suggest the structural differences between the Zr and Hf compounds are relativistic in nature. Non-relativistic calculations show the 3-5 structure of Cp_3HfCH_3 to be more stable than the 2-5,1-1 isomer by 12.7 kcal/mol; however, inclusion of scalar relativistic effects correctly predict the 2-5,1-1 isomer of Cp_3HfCH_3 to be more stable by 6.5 kcal/mol.

Thorium-Water Complexes: Intriguing New Actinide Species

In laser-ablation experiments Th atoms react with H_2O to form a series of new Th, O, H species in argon matrix.

Relativistic DFT and ab initio MP2 calculations were performed on all proposed species and various possible isomers to assist in the interpretation of the experimental data and to assign frequencies and species.

A good agreement was found between the calculated and experimental vibrational frequencies, relative absorption intensities, and isotopic shifts, which provides strong support for these first identifications of Th, O, H molecules (Figure 6).

Results of this work were published in *JACS* (Liang and Bursten 2002).

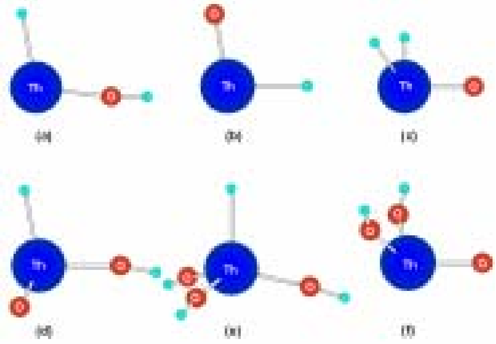


Figure 6. Thorium-water complexes: (a) HTh(OH), (b) HThO, (c) H_2ThO, (d) HThO(OH), (e) $HTh(OH)_3$, (f) $OTh(OH)_2$.

References

Liang, B, L Andrews, J Li, and BE Bursten. Noble Gas-Actinide Compounds: Evidence for the Formation of Distinct $CUO(Ar)_{4-n}(Xe)_n$ and $CUO(Ar)_{4-n}(Kr)_n$ (n=1, 2, 3, 4) Complexes. *J. Am. Chem. Soc.*, **124**, 9016 (2002).

Liang, B, L Andrews, J Li, and BE Bursten. Experimental and Theoretical Studies of the Products of Laser-Ablated Thorium Atom Reactions with H_2O in Excess Argon. *J. Am. Chem. Soc.*, **124**, 6723 (2002).

Zhan, C-G, and DA Dixon. First-Principles Determination of the Absolute Hydration Free Energy of the Hydroxide Ion. *J. Phys. Chem. A*, **106**, 9737 (2002).

Zhan, C-G, and DA Dixon. Absolute Hydration Free Energy of the Proton from First-Principles Electronic Structure Calculations. *J. Phys. Chem. A*, **105**, 11534 (2001).

Chemical Fate of Contaminants in the Environment: Chlorinated Hydrocarbons in the Groundwater

DG Truhlar,[a] CJ Cramer,[a] K Morokuma,[b] TH Dunning,[c] BC Garrett,[d] M Dupuis,[d] TP Straatsma,[d] and YA Borisov[e]

(a) University of Minnesota, Minneapolis
(b) Emory University, Atlanta, Georgia
(c) North Carolina State University, Raleigh
(d) Pacific Northwest National Laboratory
(e) Russian Academy of Science, Moscow

Chlorinated hydrocarbons (CHCs) are the most common contaminant found at hazardous waste sites and are the most prevalent contaminants on Department of Energy (DOE) weapons production sites. Many of the chlorinated hydrocarbons are either known or suspected carcinogens and thus pose health risks to the public and site workers. Chlorinated hydrocarbons, unlike simple hydrocarbons, are resistant to biodegradation, but can degrade by abiotic processes such as hydrolysis, nucleophilic substitution, and dehydrochlorination. Unfortunately, few studies of the reactions of chlorinated hydrocarbons have been reported in the literature, and disagreement still exists about the mechanisms and rates of many of the key reactions. We propose to model the reactions involved in the degradation of chlorinated hydrocarbons in the groundwater and the formation of toxic chlorinated compounds such as dioxin. The goals of the research proposed here are fourfold:

- development of a computational approach that will allow reaction pathways and rate constants to be accurately calculated
- development of more approximate approaches, evaluated against the more accurate approach, which will lay the groundwork for exploratory studies of more complex CHCs
- application of these approaches to study the degradation pathways of CHCs in aqueous liquids
- application of the more approximate approaches to study mechanism of forming complex chlorinated hydrocarbon polychlorinated benzene compounds and dioxins.

Over the past three years, significant progress has been made on developing a hierarchical computational approach that builds up from high level ab initio calculations on reaction in isolation to more approximate calculations in solvated clusters and bulk solvent. The primary focus of this development effort was high-level electronic structure calculations on gas-phase and microsolvated reactions. In the research program presented here, we propose to continue the development effort, with a focus on evaluation of approximate methods to include solvation effects. Layered electronic structure methods (e.g., ONIOM) have been tested against benchmark calculations and will provide the basis for evaluating more approximate hybrid quantum mechanical/molecular mechanics methods (e.g., QM/MM and IMOMM). The QM/MM methods treat solvent molecules explicitly and can be used to evaluate approaches that treat the solvent by a continuum model (e.g., dielectric continuum models). A major focus of this work will be the development of continuum solvation models, which are based upon the solvation models of Truhlar and Cramer, and which are appropriate for density functional calculations. Previous work was primarily focused on the initial elementary steps in reactions of OH^- with chlorinated methane molecules, $CHn\,Cl_{(4-n)}$, and studies of the reactions of OH^- with chlorinate ethylene molecules, $C_2Hn\,Cl_{(4-n)}$, were initiated. Future studies will extend calculations to study reactions that the products of these initial reactions can undergo. For example, the nucleophilic substitution reaction of OH^- with CH_2Cl_2 forms a chlorinated methanol molecule, which can undergo HCl elimination to form formaldehyde. These types of reactions will be the subjects of our future studies. In addition, we will employ more approximate, and less costly, computational methods to study reaction involving more complex chlorinated compounds such as polychlorinated benzene molecules and dioxins.

Reactions of Chlorinated Hydrocarbons in the Groundwater

We are developing continuum solvation methods in order to model reactions between simple CHCs and the hydroxide ion. In particular, we are using hybrid density functional theory (DFT) and a continuum solvation model based on the generalized Born (GB) approximation. It has been shown that hybrid DFT calculations of barrier heights, energies of reactions, and conformational energies are more accurate when diffuse basis functions are included (Lynch et al. 2000, Lynch and Truhlar 2001, Lynch et al. 2002). However, partial atomic charges, which are key components of the GB method, can become unphysical when diffuse functions are used (Figure 1).

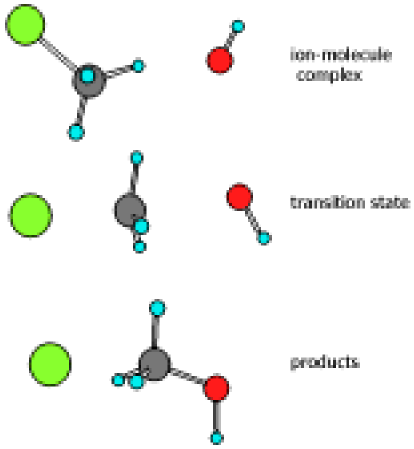


Figure 1. S_N2 reaction between methyl chloride and hydroxide ion.

We have therefore developed a method that alleviates some of the sensitivity to the inclusion of diffuse functions when calculating partial atomic charges from a Löwdin population analysis (Thompson et al. 2002). We have further improved these charges by mapping them to new charges so that they reproduce experimentally well-defined quantities, like the dipole moment (Figure 2) (Winget et al. 2002).

Benchmark Studies of Reactions of Chlorinated Hydrocarbons

A major challenge for computational chemistry is the accurate calculation of rate constants for reactions in aqueous solution, such as those involving environmental contaminants such as chlorinated hydrocarbons. We have developed a systematic approach to study the effects of molecular environments on the energetics and dynamics of chemical reaction in liquids. We have extended our previous benchmark calculations on the energetics for reactions of OH^- with methyl chloride, dichloromethane, chloroform, and carbon

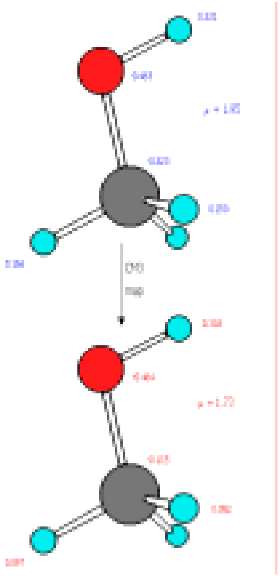


Figure 2. Redistributed Löwdin population analysis (Thompson et al. 2002) charges and dipole moment and CM3 (Winget et al. 2002) charges and dipole moment of methanol. The experimental dipole moment is 1.7 D.

tetrachloride (Borisov et al. 2001), to the studies of the effects of microsolvation (by a small number of water molecules) and approximate bulk solvation. We have systematically studied the competition between nucleophilic substitution (S_N2) and proton transfer reactions for these solvated reactions.

Benchmark Studies of Reactions of Chlorinated Hydrocarbons

The energies for bound reaction complex transition states, product complexes, and products of the reactions $OH^-(H_2O)_m$ + $CH_{(4-n)}Cl_n$ with m = 0, 1, and 2 and n = 1-4 are shown in Figure 3. The results presented in the figure are all at the MP2/aug-cc-pVDZ level of theory. The solid lines are for the nucleophilic substitution (S_N2) reaction and the dashed lines are for the proton transfer (PT) reaction. The circles, squares, and diamonds show the results with m = 0, 1, and 2 respectively. These results show that both the degree of chlorination and level of solvation on the reaction energetics have a profound effect on the reaction energetics.

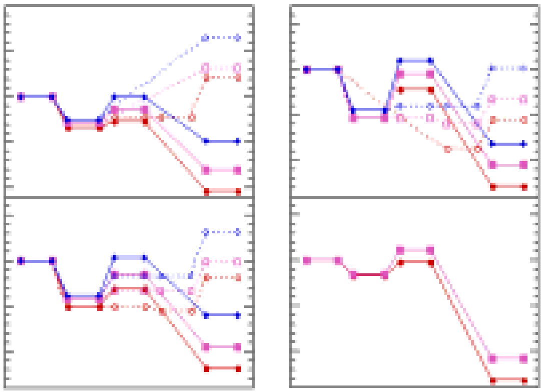


Figure 3. Energies for bound reaction complex transition states, product complexes, and products of the reactions $OH^-(H_2O)_m$ + $CH_{(4-n)}Cl_n$ with m = 0, 1, and 2 and n = 1-4.

References

Borisov, YA, EA Garcia, SL Meelke, RC Garrett, and TA Dunning. A Systematic Sludge of the Reactions of OH^- with Chlorinated Methanes. 1. Benchmakr Studies of Gas-Phase Reactions. *J. Phys. Chem.*, **105**, 7727 (2001).

Lynch, BJ, PL Fast, M Harris, and DG Truhlar. Adiabatic Connection for Kinetics. *J. Phys. Chem. A*, **104**, 4811 (2000).

Lynch, BJ, and DG Truhlar. How Well Can Hybrid Density Functional Methods Predict Transition State Geometrics and Barrier Heights? *J. Phys. Chem. A*, **105**, 2936 (2001).

Lynch, BJ, Y Zhao, and DG Truhlar. *J. Phys. Chem. A*, in press (2002).

Thompson, JD, JD Xidos, TM Sonbuchner, CJ Cramer, and DG Truhlar. *Phys Chem Comm.*, **5**, 117 (2002).

Winget, P, JD Thompson, JD Xidos, CJ Cramer, and DG Truhlar. Charge Model 3: A Class IV Charge Model Based on Hybrid Density Functional Theory with Variable Exchange. *J. Phys. Chem. A*, in press (2002).

A Computational Approach to Understanding Aerosol Formation and Oxidant Chemistry in the Troposphere

JS Francisco,[a] A Morita,[b] BC Garrett,[c] DA Dixon,[c] GK Schenter,[c] LX Dang,[c] SM Kathmann,[c] SS Xantheas,[c] JR Hynes,[d] R Bianco,[d] MA Miller,[e] and KA Peterson[f]

(a) Purdue University, West Lafayette, Indiana
(b) Kyoto University, Japan
(c) Pacific Northwest National Laboratory
(d) University of Colorado, Boulder
(e) University of Washington, Seattle
(f) Washington State University, Pullman

An understanding of the mechanisms and kinetics of aerosol formation and ozone production in the troposphere are currently a high priority because they are recognized as two major effects of energy-related air pollution. Atmospheric aerosols are of concern because of their effect on visibility, climate, and human health. Equally important, aerosols can change the chemistry of the atmosphere in dramatic fashion by providing new chemical pathways (in the condensed phase) that are not available in the gas phase. The oxidation of volatile organic compounds (VOCs) and organic sulfur compounds can form precursor molecules that nucleate aerosols. DOE's Atmospheric Chemistry Program has identified the need to evaluate the causes of variations in tropospheric aerosol chemical composition and concentrations, including determining the sources of aerosol particles and the fraction that are of primary and secondary origin. Tropospheric ozone is of concern primarily because of its impact on health. Ozone levels are controlled by NOx and by VOCs in the lower troposphere. The VOCs can either be from natural emissions from such sources as vegetation and phytoplankton or from anthropogenic sources such as automobiles and oil-fueled power production plants. The major oxidant for VOCs in the atmosphere is the OH radical. With the increase in VOC emissions, there is rising concern regarding the available abundance of HOx species needed to initiate oxidation. Over the last five years, there have been four field studies aimed at initial measurements of HOx species (OH and HO_2 radicals). These measurements revealed HOx levels that are two to four times higher than expected from the commonly assumed primary sources. Such elevated abundances of HOx imply a more photochemically active troposphere than previously thought. This implies that rates of ozone formation in the lower region of the atmosphere and the oxidation of SO_2 can be enhanced, thus promoting the formation of new aerosol properties. Central to unraveling this chemistry is being able to assess the photochemical product distributions resulting from the photodissociation of by-products of VOC oxidation. We propose to use state-of-the-art theoretical techniques to develop a detailed understanding of the mechanisms of aerosol formation in multicomponent (mixed chemical) systems and the photochemistry of atmospheric organic species. The aerosol studies involve an approach that determines homogeneous gas-to-particle nucleation rates from knowledge of the molecular interactions that are used to define properties of molecular clusters. Over the past few years the research team developed Dynamical Nucleation Theory, which is a significant new advance in the theoretical description of homogeneous gas-to-liquid nucleation, and applied it to gas-to-liquid nucleation of a single component system, water. The goal of the present research is to build upon these advances by extending the theory to multicomponent systems important in the atmosphere such as water-sulfuric acid. In addition, high level ab initio electronic structure calculations will be used to unravel the chemical reactivity of naturally occurring VOCs with the OH radical. Detailed mechanistic information regarding the degradation pathways and accurate reaction rate constants for key VOC reactions and will be obtained.

Sensitivity of Water Cluster Thermodynamics to Trace Impurities

Over the past few years our research team developed Dynamical Nucleation Theory - a novel theoretical formulation that allows the determination of homogeneous nucleation rates from a knowledge of the molecular interactions. Sensitivity analysis of the kinetic parameters shows that the underlying interaction potentials have a profound influence on the nucleation rate. This extreme sensitivity to the cluster energetics makes the study of aqueous impurities, such as sulfuric acid, ±ions, etc., imperative.

Immobilization of Radionuclides in Oxides

H Jonsson,[(a)] F Vila,[(a)] GA Henkelman,[(a)] K Tsemekhman,[(a)] A Rubio,[(b)] EJ Bylaska,[(c)] F Gao,[(c)] JA Nichols,[(c)] LR Corrales,[(c)] J Garza-Olguin,[(d)] and D Makarov[(e)]

(a) University of Washington, Seattle
(b) Donostia International Physics Center, Donostia-San Sebastian, Spain
(c) Pacific Northwest National Laboratory
(d) Universidad Autonoma Metropolitana, Iztapalapa, Mexico
(e) University of Texas, Austin

We will carry out a series of theoretical calculations related to the immobilization of radionuclides in oxides, both crystalline and amorphous. The principal questions that will be addressed are 1) How can long-lived excitons lead to degradation or annealing of the oxide network? 2) How do defects and long-lived excitons enable or hinder long-range migration of radioactive ions in the oxide, possibly causing extended structural transitions such as phase transformations, precipitation, or amorphization? 3) What factors affect the rate of dissolution of the oxide in water? The oxides that will be studied include amorphous silica (with inclusion of Al, Na and B), zirconia and zircon, with likely applications to pyrochlore and zirconolite structured ceramics. An essential component of this work is development of methodology and implementation to enable accurate computational studies of long-lived excited states, in particular self-trapped excitons and holes. These methods include density functional theory with self-interaction correction, Optimized Effective Potentials for DFT, time-dependent DFT and FW extension of DFT. The methods will be tested on systems where wavefunction-based NWCHEM density functional code and applied to extended systems subject to periodic boundary conditions. This is a grand challenge computational project in several ways: a) Atomic scale simulations will be carried out over a long time scale using a recently developed algorithm, reaching time scales that are many orders of magnitude longer than direct classical dynamics simulations can cover b) the dynamics in an excited electronic state will be simulated, a more challenging task than ground state dynamics, c) we will study the dynamics in amorphous solids, which are more complex that crystalline solids. The proposed work builds on the collaborative work of the Jonsson and Corrales groups on self-trapped excitons in quartz and silica glasses in a previous allocation period.

Exciton Induced Defect Formation in Fused Silica

- Initial 5.5 eV excitation leads to a delocalized exciton, as indicated by the excess spin on Figure 1a. In Figure 4, excitation is at 0 Å from dashed line to red line.
- Relaxation along the minimum energy path (MEP) on the triplet state, red line in Figure 4, leads to localization of the exciton, Figure 2a and a broken bond, Figure 2b.
- A curve crossing to the singlet state likely occurs at the minimum of the triplet state surface.
- Relaxation along the singlet state MEP, green solid line on Figure 4, leads to the formation of a defect pair consisting of a 5-fold and a 3-fold coordinated Si and O atoms, respectively, Figure 3a and 3b.
- The graph inset shows a path from the defect glass of Figure 3b to the original glass of Figure 1b with a barrier of 0.8 eV.
- Figure 4 shows a barrier to one other defect state that has a similar relative energy.

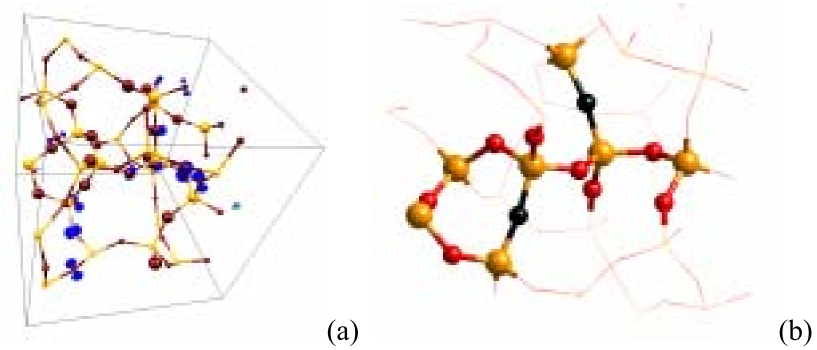

(a) (b)

Figure 1. Delocalized exciton.

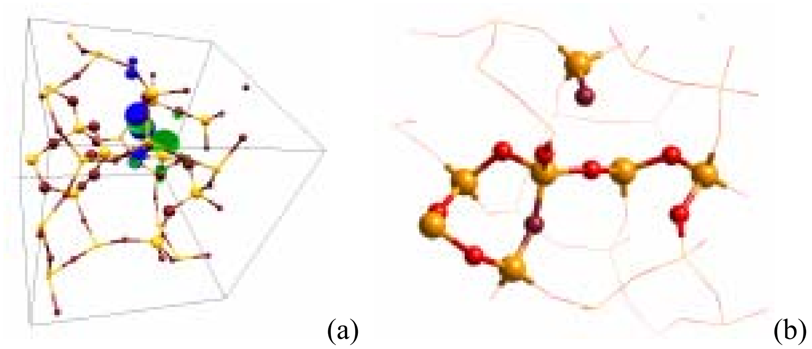

(a) (b)

Figure 2. Localization of the exciton.

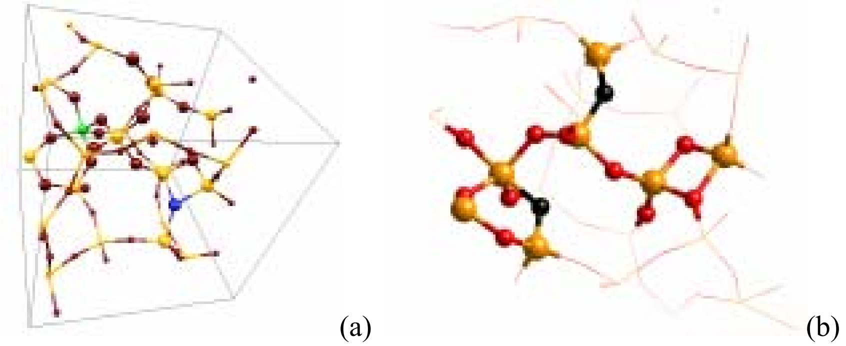

(a) (b)

Figure 3. Defect pair consisting of a 5-fold and a 3-fold coordinated Si and O atoms.

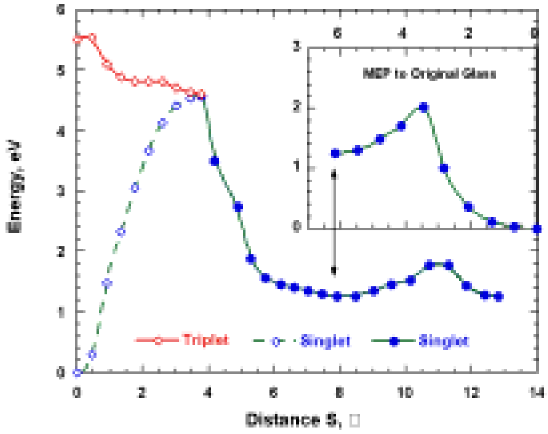


Figure 4. Barrier to one other defect state that has a similar relative energy.

Biomolecular Simulation of Base Excision Repair and Protein Signaling

TP Straatsma,[a] ER Vorpagel,[a] M Zacharias,[b] PE Smith,[c] CF Wong,[d]
JA McCammon,[d] JH Miller,[e] and K Wei[e]
(a) Pacific Northwest National Laboratory
(b) Institute for Molecular Biotechnology, Jena, Germany
(c) Kansas State University, Manhattan
(d) University of California San Diego
(e) Washington State University, Pullman

Environmental factors including ionizing radiation contribute to continuous damage of cellular DNA, in addition to endogenous sources. The damage resulting from oxidative stress and ionizing radiation is primarily in the form of oxidized bases, single strand breaks, and loss of bases. These are the targets of the Base Excision Repair (BER) mechanism enzymes, including polymerase-β. Failure to repair these damaged DNA sites correctly and in a timely mannercan result in cell death, carcinogenesis, or genetic diseases. Resulting mutations in cell signal transduction enzymes can lead to uncontrolled cell proliferation or differentiation. For example, mutations in Ras, the molecular switch in several growth-factor signaling pathways, have been found in about 30% of human tumors. These signaling pathways often involve a chain of protein kinases that activate or deactivate each other through phosphorylation reactions, eventually controlling the activation of transcription factors in the cell nucleus. This proposal seeks the massively parallel computing resources required to perform molecular modeling and simulation studies to enhance our understanding of the mechanism of human polymerase-β, one of the key enzymes in BER repair, and the cell signaling enzymes cyclic-AMP-dependent protein kinase and Ras. Specifically, this work will focus on the:

- dynamics of DNA and damaged DNA
- dynamics and energetics of base flipping in DNA
- co-solvent effects on biomolecular structures, including DNA mechanism and fidelity of nucleotide insertion by BER enzyme human polymerase-β
- mechanism and inhibitor design for cyclic-AMP-dependent protein kinase
- dynamics and energetics of Ras and its complex with effector molecule Raf for the molecular dynamics simulations and electronic structure calculations.

The proposed investigations will rely on NWChem, the massively parallel software for computational chemistry developed at EMSL. NWChem will also be used for the analysis of the generated molecular dynamics trajectories, and any additional analysis tools required for the proposed work will be developed within the analysis modules of NWChem.

Characterization of the Active Site of DNA Polymerase β

It is well established that the fully formed polymerase active site of the DNA repair enzyme, polymerase β (pol β), including two bound Mg^{2+} cations and the nucleoside triphosphate (dNTP) substrate, exists at only one point in the catalytic cycle just prior to the chemical nucleotidyl transfer step. The structure of the active conformation has been the subject of much interest as it relates to the mechanism of the chemical step and also to the question of

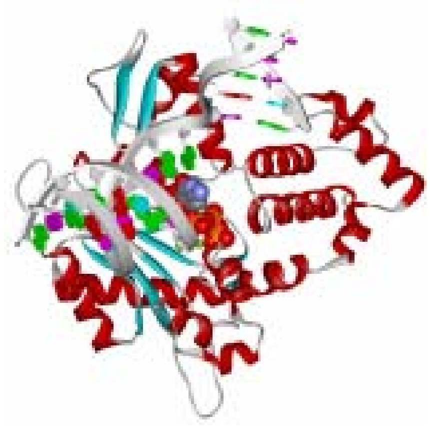


Figure 1. Polymerase active site of the DNA repair enzyme, polymerase β (pol β).

fidelity assurance. While crystal structures of ternary pol β – (primer-template) DNA-dNTP complexes have provided the main structural features of the active site, they are necessarily incomplete due to intentional alterations, e.g. removal of the 3'OH groups from primer and substrate, that were needed to obtain a structure from mid-cycle. Working from the crystal structure closest to the fully formed active site, (pdb: 1bpy), molecular dynamics simulations of the solvated ternary complex were performed, with the missing 3'OH's restored, via modeling, to the primer and substrate, and without restoration of the 3'OH's. The results of these simulations, taken together with ab initio optimizations on simplified active site models, indicate that the missing primer 3'OH in the crystal structure is responsible for a significant perturbation in the coordination sphere of the catalytic cation and allow us to suggest several corrections and additions to the active site structure as observed by crystallography. In addition, the calculations help to resolve questions that have been raised regarding the protonation states of coordinating ligands (Figure 1).

DOE Science Grid Operational Development and Testing

RS Studham,[a] D Cowley,[a] S Yabusaki,[a] A Coleman,[a] K Chanchio,[b] V Welch,[c] AA Rodriguez,[c] J Novotny,[d] K Jackson,[d] J Boverhof,[d] and W Johnston[d]

(a) Pacific Northwest National Laboratory
(b) Oak Ridge National Laboratory, Oak Ridge, Tennessee
(c) Argonne National Laboratory, Chicago, Illinois
(d) Lawrence Berkeley National Laboratory, Berkeley, California

"DOE Science Grid will lay the groundwork to support DOE Science applications that require, e.g., distributed collaborations, very large data volumes, unique instruments, and the incorporation of supercomputing resources into these environments" (Johnston et al. 2000).

MSCF is participating in the deployment of the DOE Science Grid and embracing grid middleware (embodied in the Globus toolkit) as a key technology to support large-scale computing and data transfer both inside and outside EMSL. The grid infrastructure will integrate EMSL instruments, computers, and data more seamlessly than ever before and allow a user to perform a single sign-on to authenticate to all grid resources.

Deployment on HPCS2

MSCF resources were deployed on the Grid as soon as it became operational. As currently configured, this allows users to submit parallel batch jobs and receive results without having to directly log in to the login nodes. It also allows users to move data on and off computer systems with GridFTP services, including third-party transfers where the user can initiate data movement directly between servers without logging in to either. In coming months, parallel FTP to large data stores will be available.

Ucode-Globus Proof of Concept

Ucode (http://www.mines.edu/igwmc/freeware/ucode) is an inverse modeling code used by PNNL's hydrologists to model Hanford groundwater flow. At PNNL, Ucode is run on dedicated Intel Architecture clusters that are configured to run multiple instances of the code in parallel. The individual Ucode instances are controlled by a PNNL-written program named taskr.pl. This works well in practice, but the hydrologists do not have enough compute cycles available and the design of taskr.pl requires that the cluster be dedicated to parallel Ucode.

We were able to leverage the features of the Globus toolkit to allow both Ucode and taskr.pl to use a Grid infrastructure to submit jobs and retrieve results. This has multiple benefits:

- Parallel Ucode can be run on generic Grid-enabled clusters, not just those that have been tailored for Ucode and taskr.pl.
- Globus can negotiate transparently with a cluster's resource management system; the application or user does not need to care what batch queuing system the cluster runs.
- The need for slave taskr.pl processes on compute nodes is eliminated.

We expect that this will make more compute cycles available to Ucode users by removing barriers to the use of other clusters. We proved during 2002 that Globus can successfully facilitate parallel Ucode jobs. Work will continue in 2003 on increasing robustness and enabling Ucode across DOE Science Grid sites.

Supercomputing Demonstrations

MSCF demonstrated its Grid capabilities at both the Supercomputing 2001 and 2002 conferences. In 2001, a joint Extensible Computational Chemistry Environment (ECCE)/Grid demo showed a user, working from a workstation on the exhibit floor, submitting an ECCE job remotely to PNNL's colony cluster, and then transferring data directly from colony to a mass storage system at NERSC. This was accomplished with a single sign on, and at no point did the user have to log into either colony or the NERSC system. The Globus middleware performed authentication, job submission, and data transfer on the user's behalf.

In 2002, MSCF sent a complete high performance IA64 Linux cluster to Supercomputing. The cluster had the Globus toolkit installed, and was configured to allow parallel job submission and file transfer on the exhibit floor. The cluster also participated in the Global Grid Testbed Collaboration's entry in the Supercomputing 2002 High Performance Computing Challenge. The Global Grid Testbed Collaboration entry won the "most geographically distributed" and "most heterogeneous application" challenges.

DOE Science Grid Helpdesk Infrastructure

PNNL is providing the infrastructure for DOE Science Grid helpdesk support and user training. To that end, we have set up a help desk server external to the PNNL firewall. Science Grid users may submit help requests via e-mail or a web interface. Science Grid help staff or administrators can then view and respond to the requests with the Java ESHQ Browser software developed at PNNL. Users and help staff can be anywhere on the internet and use the help system, provided they have network connectivity to the server (Johnston et al. 2000).

References

Johnston, W, R Bair, I Foster, A Geist, and B Kramer. The DOE Science Grid: Computing and Data Infrastructure for Large-Scale Science. http://doesciencegrid.org (2000).

Computational Quantum Chemistry

HF Schaefer,[(a)] NJ DeYonker,[(a)] CJ Barden,[(a)] NR Brinkman,[(a)] JM Gonzales,[(a)]
JP Kenny,[(a)] JD Larkin,[(a)] C Pak,[(a)] ND Koslap Petraco,[(a)] NA Richardson,[(a)]
KW Sattelmeyer,[(a)] MS Schuurman,[(a)] HL Woodcock,[(a)] and RJ Harrison[(b)]
(a) University of Georgia, Atlanta
(b) Oak Ridge National Laboratory

To gain experience in the use of the EMSL computational resources and in the development of programs for massively parallel computers in preparation for collaboration in the DOE BES SciDAC project. Calculations will be performed on a variety of small and large molecules using NWChem and other chemistry codes.

Ab initio Evaluation of S_N2 Identity Reactions: Evaluation of Marcus Theory

Classic gas-phase S_N2 reactions exhibit a double well potential with one transition state. We have computed the energetic quantities associated with the reaction $CH_3X + F^- \rightleftharpoons CH_3F + X^-$ for X = F, Cl, CN, OH, SH, NH_2 and PH_2.

It is our goal to expand this to the identity reactions of the type $CH_3X + X^- \rightleftharpoons CH_3X + X^-$ and use this data to compare explicitly evaluated results Marcus Theory predictions.

Classic gas-phase S_N2 reactions exhibit a double well potential with one transition state. We have computed the energetic quantities associated with the reaction $CH_3X + F^- \rightleftharpoons CH_3F + X^-$for X = F, Cl, CN, OH, SH, NH_2 and PH_2.

It is our goal to expand this to the identity reactions of the type $CH_3X + X^- \rightleftharpoons CH_3X + X^-$ and use this data to compare explicitly evaluated results Marcus Theory predictions.

Consider the two ion-molecule complexes (Figure 1). The top is a classic backside complex, the bottom a frontside complex. S_N2 reactions are formally thought to be backside inversions.

The central barrier, E^*, is the energy difference between these complexes and the transition state. If we use Marcus theory to predict this barrier, will we get the correct value for the frontside complex, or will it predict the value associated with the non-stationary point associated with a backside complex?

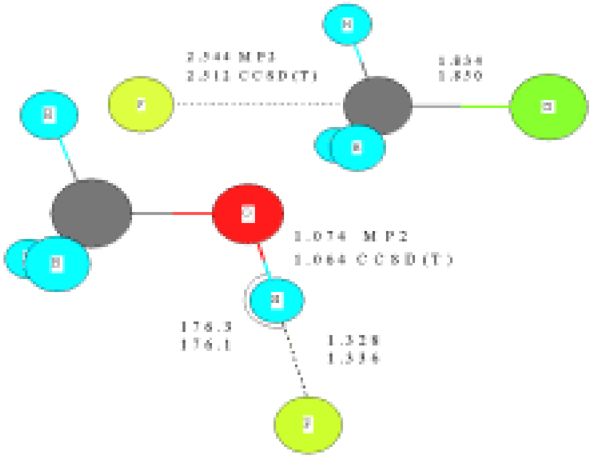


Figure 1. Two ion-molecule complexes. The top is a classic backside complex, the bottom is a frontside complex.

Computer Simulation of Single-Cell Irradiation Studies Using the PNNL Electron Microbeam

JH Miller[a] and F Zheng[a]
(a) Washington State University, Pullman

This project will perform computer simulations related to single-cell irradiation studies using the PNNL electron microbeam facility in the EMSL. These calculations are needed both to enhance the design of the irradiator and to aid in the interpretation of experimental data.

Approach: The basic computational tool is Monte Carlo simulation both for the probability that a cell absorbs energy from the microbeam and for the probability of detecting a response if energy is imparted to the cell. Energy absorption in single-cell irradiation studies will be modeled by the event-by-event detailed-histories track-structure simulator PITS. This code follows the slowing down in water of electrons with initial energies as large as 100,000 eV until they pass the threshold for electronic excitation of a water molecule (about 5 eV). The cellular responses of interests are those related to bystander effects in which cells that do not receive energy directly from the beam respond in a manner that is similar to the cell that was the target of microbeam irradiation. The code to model bystander responses is a Monte Carlo simulator of molecular transport from the targeted cell to other cells in the biological sample. The molecules of interests are peptides shed by the hit cell that activate receptors on neighboring cells. Both types of simulation require input data on cell morphology, which will be obtained by application of NWgrid to 3-D cellular images derived from confocal microscopy.

Modeling Radiation-Induced Bystander Effects

A model has been developed (Figure 1) to explain the induction of a calcium pulse when medium from irradiated tissue cultures is placed over unirradiated cells. That model assumes that the bystander effect is transmitted by shed peptides. Experiments to identify these peptides by mass spectrometry are in progress. The model predicts the concentration of shed peptides in the transferred medium. Experiments to measure these concentrations by ELISA techniques are in progress.

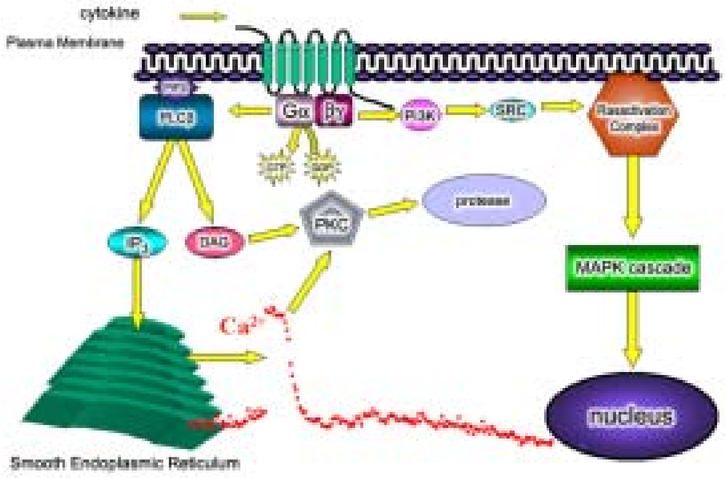


Figure 1. Model of the induction of calcium pulse when medium from irradiated tissue cultures is placed over unirradiated cells.

Development and Analysis of the Hybrid Quantum Mechanics–Molecular Mechanics Approach in NWChem—Why Catalytic Antibodies Cannot Achieve the Level of Catalysis of Enzymes

W Sudholt[a]
(a) University of California at San Diego

To deepen our understanding of natural and synthetical processes and to support current experimental progress, there is an increasing need for affordable theoretical calculation of chemical reactions in complex systems like solutions, biomolecules and materials. This task is addressed by hybrid quantum mechanical–molecular mechanical (QM/MM) methods, which build a bridge between accurate quantum chemical and efficient force field computations. Their treatment in NWChem was not satisfactory up to the latest release. However, very recently a new QM/MM model based on pseudobonds was implemented into the program. Thereby, it is intended to enhance the fidelity of the QM/MM module and to make its application more popular.

Hybrid quantum mechanics and molecular mechanics (QM/MM) methods combine the accuracy of the former to describe chemical reactions with the capability of the latter to treat large systems. A new approach based on the pseudobond parameterization of Zhang et al. (1999) was build into NWChem. In the first phase of the project, this implementation is tested for model peptides in water. In the second phase of the project, the technique will be applied to antibody catalysis (Figure 1).

Figure 1. Hybrid quantum mechanics and molecular mechanics (QM/MM) methods combine the accuracy of the former to describe chemical reactions with the capability of the latter to treat large systems.

Zhang, Y, TS Lee, and W Yang. A Pseudobond Approach to Combining Quantum Mechanical and Molecular Mechanical Methods. *J. Chem. Phys.*, **110**, 46 (1999).

Structure of Borate Glasses

LR Corrales[a] and RM Van Ginhoven[b]
(a) Pacific Northwest National Laboratory
(b) University of Washington, Seattle

A first step in developing borate interaction potentials is to carry out a structural study using Density Functional Chemistry Molecular-Dynamics (DFT-MD) methods to explore the persistence of boroxol rings in borate glasses. The persistence of boroxol rings in a borate melt can be studied by MD simulations using empirical potentials. However, the complexity of an environmentally dependent potential model does not lead to a straightforward generalization that makes the potential transferable to mixtures, such as boro-silicate glasses. Thus, alternative forms of empirical potentials must be considered. In an initial study, a subset of the boron atoms are placed in the boroxol ring structure. The percentage of boron atoms in boroxol rings are varied to create several initial configurations. These systems are then allowed to evolve by following a glass forming schedule and by high temperature annealing. The effects of initial configuration and the temperature history on the final structure will be determined. The desire is to use these calculations to modify and refine empirical potentials of borates. Then, boro-silicate and alumino-boro-silicate glasses can be simulated to validate the effectiveness of these modified potentials.

Energy Dependence of Vitreous B_2O_3 On Boroxol Ring Concentration

DFT planewave calculations are used to determine the relative energy of borate glasses as a function of the population of boroxol rings (the major superstructure unit in borate glass). At low and high concentrations there is little difference between two sets of distinct configurations. At the other concentrations, the differences are due to a wider possibility of configurations not sampled. Results suggest two minimum energy states for the number fraction.

The initial configurations (top set of images in Figure 1) are generated by fixing the concentration of boroxol rings (distributed randomly in a supercell) and then the remaining B and O atoms are added at random. The configurations are first relaxed using classical dynamics simulations. The range of boroxol ring fraction goes from 0% to 100%.

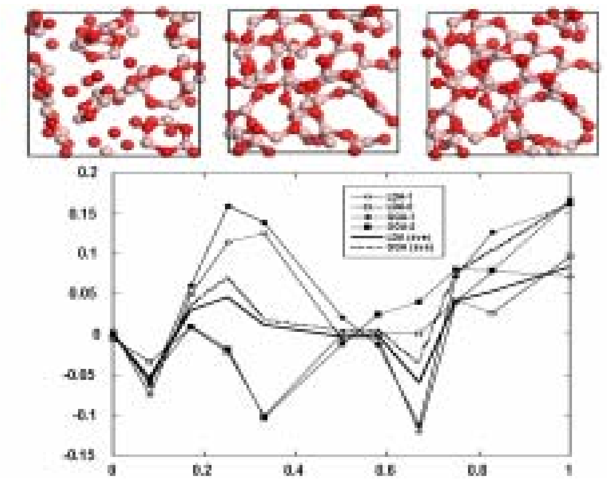


Figure 1. The relative system energy of two sets of samples are calculated using both LDA and GGA.

The relative system energy of two sets of samples are calculated using both LDA and GGA, with minor differences between the two. Their averages are also shown in the graph in Figure 1.

Ionization Energy Calculations of Deoxyribonucleotides

ER Vorpagel[(a)] *and X Wang*[(a)]
(a) Pacific Northwest National Laboratory

Gas-phase electron binding energies for 2'-deoxyribonucleotides have been measured using photodetachment photoelectron spectroscopy and electrospray ionization. Ion mobility experiments indicate only one conformation is present. Molecular mechanics calculations have been used to identify low energy conformations for these structures. We propose to use these starting structures for density functional theory (DFT) calculations using B3LYP and a TZVP basis set augmented with a set of diffuse functions to first optimize the structures and then calculate the vertical electron detachment energy (VDE) and the adiabatic electron detachment energy (ADE) for the five anions: $dAMP^-$, $dCMP^-$, $dGMP^-$, $dTMP^-$ and $dUMP^-$. The NWChem software will be used together with Ecce.

Ionization Energy Calculations of Deoxyribonucleotides

Experimental VDEs were measured for four 2'-deoxyribonucleic acid monoanions in the gas phase using photodetachment photoelectron spectroscopy (PES).

Theoretical VDE values were calculated using DFT to optimize the anions with the hybrid B3LYP exchange-correlation functional and the TZVP basis set with additional diffuse functions. VDE values are the energy difference between optimized anion and the neutral structure at the same geometry (Figure 1).

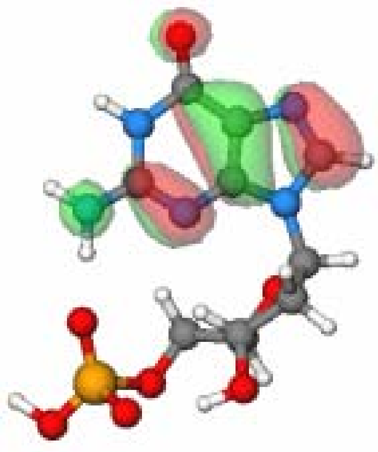


Figure 1. Optimized geometry of dGMP anion showing the molecular orbital from which the electron in removed.

The table below compares experimental VDE values with those calculated.

Molecule	VDE(exp)	VDE(calc)
dGMP	4.90 ± 0.1	4.90 eV
dTMP	5.80 ± 0.1	5.36 eV
dAMP	6.05 ± 0.1	5.23 eV
dCMP	5.80 ± 0.1	5.24 eV

Parameter Development and Initial Molecular Dynamics Simulation of Flavocytochrome Ifc3

ER Vorpagel,[(a)] DM Smith,[(a)] M Dupuis,[(a)] and TP Straatsma[(a)]
(a) Pacific Northwest National Laboratory

As part of the DOE Office of Science Program Notice 01-21 for Advanced Modeling and Simulation of Biological Systems, we have received funding for development of new modeling capabilities for the study of complex enzymatic reactions involved in the respiratory cycle of iron-reducing bacteria. The goal of this work is to provide the computational tools required for a detailed molecular-level characterization and understanding of the enzymatic reactivity of these bacteria through modeling and simulation, with a soluble flavocytochrome c3 of *Shewanella frigidimarina* as our target enzymatic system. Though the scope of the full project merits a grand challenge, we have broken up the initial research into several pilot projects that will provide guidance for a future grand challenge proposal. This pilot project involves developing the missing force field parameters for molecular dynamics simulation of *S. frigidimarina* flavocytochrome Ifc3 which contains four iron containing heme groups and flavin-adenine dinucleotide (FAD) plus the reactants, products and intermediates. We will expand the AMBER99 force field to include a consistent set of parameters for the heme groups, FAD and fumarate, using the recommended protocol, with geometries that we will determine from high-level (B3LYP/pVTZ) ab initio calculations and RESP-based partial atomic charges. We will develop the force field parameters for the porphyrin moiety with Fe in oxidation state III, as this is the appropriate state for the heme groups in the resting state of the enzyme. Flavins can exist in three different oxidation states (oxidized, semiquinone radical, and reduced) with multiple possible tautomeric forms for each oxidation state. X-ray crystal structures cannot be used to help deduce the oxidation state. Partial charges for the adenine moiety of FAD can be compared with the available parameters defined for DNA. Parameters will be developed for the two relevant isoalloxazine forms of FAD, the reduced anionic hydroquinone (activated) and the oxidized neutral isoalloxazine form. At physiological pH, the two carboxylic acid groups in the reactants and products (fumaric and succinic acid) are expected to be in the ionic state. We will determine partial atomic charges for the di-anionic, mono-protonated-anionic, and neutral forms. Torsion parameters will be adjusted to reproduce rotation barriers calculated from B3LYP/TZVP+ level theory. In addition, we will derive parameters for the partially reduced intermediate form of fumarate. Once all the parameters have been defined, we will carry out molecular dynamics simulations of the *S. frigidimarina* flavocytochrome Ifc3 in the open conformation starting with the 2.15 resolution crystal structure (PDB entry 1QO8) and in the closed form starting from the 1.80 resolution crystal structure (PDB entry 1QJD).

We will insert fumarate into the active site of the closed conformation structure to replace the substrate analogue found in the crystal structure (Figure 1). Molecular dynamics simulations will be carried out for these two systems with FAD in the inactive and in the activated state. For these simulations the enzyme system will be solvated in water, using the SPC/E water model (Berendsen et al. 1987). To account for electrostatic interactions beyond the cutoff radius, the particle-mesh Ewald method will be used. To complete such a large number of calculations will require the maximum 75,000 node-hours of computation time on MSCF Computer Resources. Since one of our objectives is to determine the influence of the clamping domain motion on the structure of the active site and the conformation of the substrate in the active site, we expect to have to extend these molecular simulations for multiple nanoseconds. Such long simulations are beyond the scope of this pilot project.

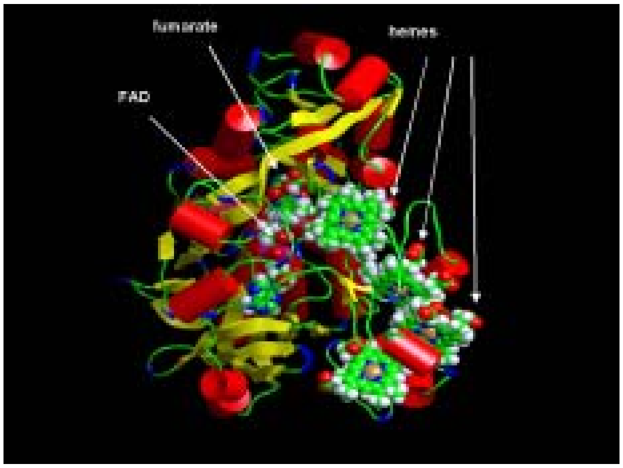


Figure 1. Simulation model of Flavocytochrome Ifc3.

- Non-protein group's partial atomic charges calculated from optimized structures using NWChem.
- For heme, both Fe(II) and Fe(III) charges were calculated. Missing Fe-N bond-stretch parameters were added to reproduce ab initio force constants.
- For flavin-adenine dinucleotide (FAD), both oxidized and reduced forms were calculated.
- Torsion parameters were assigned to reproduce the ab initio rotation barrier for the carboxylate group in fumarate.
- Full protein structure with Fe(III), oxidized FAD, fumarate, and counter ions was solvated (~70,000 atoms) and warmed to ambient temperature.

MD trajectory calculations in progress.

References

Berendsen, HJC, JR Grigera, and TP Straatsma. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.*, **91**, 6269 (1987).

Subsurface Multifluid Flow and Multicomponent Reactive Transport: Characterization of Processes and Properties

SB Yabusaki,[a] MD White,[a] DH Bacon,[a] AJ Valocchi,[b] RJ Lenhard,[c] K Pruess,[d] CI Steefel,[d] PC Lichtner,[e] JC Parker,[f] PM Jardine,[f] JP Gwo,[g] MF Wheeler,[h] TJ Arbogast,[h] and SL Bryant[h]

(a) Pacific Northwest National Laboratory
(b) University of Illinois, Chicago
(c) Idaho National Engineering and Environmental Laboratory, Idaho Falls
(d) Lawrence Berkeley National Laboratory, Berkeley, California
(e) Los Alamos National Laboratory, Los Alamos, New Mexico
(f) Oak Ridge National Laboratory, Oak Ridge, Tennessee
(g) University of Maryland, College Park
(h) University of Texas, Austin

This project targets the characterization of subsurface processes and properties in multidimensional porous media as the unifying theme for a team of researchers from six DOE labs, three universities, and one private contractor. The goal is to develop mechanistic process model representations and parameterizations that lead to enhanced understanding and scientifically defensible predictions of subsurface behavior. Long-term predictions of contaminant fate are critical to the management of the DOE missions of cleanup and long-term stewardship. The work scope for this project addresses multidimensional laboratory and field experiments as well as site-specific, problem-driven site characterization, including history matching. A common theme for the simulations is the use of large numbers of grid cells (~105-107) to resolve multiple scales of subsurface heterogeneity. By combining to form a single proposal, the team will use MSCF's resources efficiently, share ideas, and collectively benefit from the development and incorporation of new process models, robust parallel solvers, and high performance parallel libraries. An important aspect of supporting these activities will be the testing and evaluation of parallel programming tools, debugging environments, and visualization software.

The areas of research and expected results are summarized below.

Subsurface Multifluid and Reactive Transport Simulation Capabilities

- highly-resolved, three-dimensional depiction of spatially variable subsurface properties
- efficient parallel computation of long-term simulation periods
- incorporation of unique PNNL science: constitutive theory for
  - gas entrapment
  - salinity-dependent surface tension
  - saturation-dependent anisotropy.

Parallel Processing Simulators Target Hanford Vadose Zone

- resolves discontinuous, sloping sediment layers that control liquid migration in the Hanford vadose zone
- accounts for the interaction of sediments with the extreme physical and chemical properties of Hanford SX Tank Farm waste
- strengthens the scientific basis of long-term assessments of risk and waste management performance
- identifies realizable end state criteria as well as alternatives to baseline that reduce cleanup schedules and costs.

Patterning of Vinylferrocene on H-Si(100) via Self-Directed Growth of Molecular Lines and STM-Induced Decomposition

P Kruse, $^{(a)}$ ***ER Johnson,*** $^{(a)}$ ***GA DiLabio,*** $^{(a)}$ ***and RA Wolkow*** $^{(a)}$
(a) Steacie Institute for Molecular Sciences, National Research Council of Canada

In this work (Kruse et al. 2002), a self-directed growth process was used to grow ordered vinylferrocene molecular lines on the H-Si(100) surface. These metal-containing lines could be used as prepatterned catalysts for molecular processes such as carbon nanotube growth. A radical chain reaction mechanism on a cluster is examined using the density functional theory (DFT) capabilities within NWChem (Kendall et al. 2000).

The proposed mechanism is illustrated in Figure 1. Step (A) is the separated reactants which combine through a transition state (B) to form a Si-C bond with a carbon centered radical. This intermediate (C) then abstracts a hydrogen from the surface through the transition state (D) to leave a radical on the Si surface which can then propagate the reaction.

Calculations on small model systems are not sufficient for these systems since vinylferrocene is large and the optimal stacking distance of two ferrocene molecules due to π-π stacking interactions is approximately 7.1 Å. With a lattice spacing of 3.85 Å on the Si(100) surface, an alternating molecular arrangement during line growth is expected.

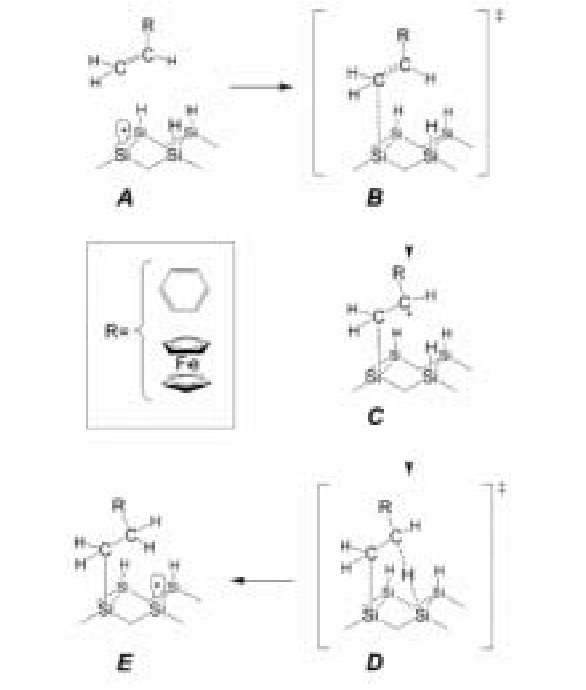


Figure 1. Radical chain reaction mechanism responsible for styrene and vinylferrocene line growth on the surface. (A) Separated reactants. (B) Addition transition state. (C) Radical intermediate. (D) H-atom abstraction transition state. (E) Final product.

A model system of vinylferrocene and a cluster of 15 silicon atoms was optimized using B3LYP DFT functionals and the 6-31G(d) basis set. The energetic results of these calculations are respresented in Figure 2. A very small barrier is found for the initial addition of the vinylferrocene on the surface with the radical intermediate 0.77 eV below the initial reactants. The barrier for the hydrogen abstraction is shown to be 0.95 eV above the intermediate and the final product is 0.16 eV below the intermediate energy. The calculations show that indeed vinylferrocene line growth will be observed.

Similar calculations with styrene (Lopinski et al. 2000) also show similar energetics to those of styrene. However, in that case the stacking distance is much smaller and the alternating pattern is not found. This supports the experimental observation that the vinylferrocene growth leads to widened line growth compared to styrene (Figure 2).

These studies, both experimental and theoretical, show that nanoscale structures of interesting functionality can be created with this growth process.

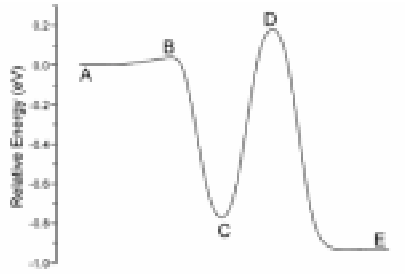


Figure 2. Calculated energies associated with the addition of vinylferrocene to the Si(100) surface corresponding to the mechanism illustrated in Figure 1.

References

Kendall, RA, E Apra, D Bernholdt, EJ Bylaska, M Dupuis, GI Fann, RJ Harrison, J Ju, JA Nichols, J Nieplocha, TP Straatsma, TL Windus, and AT Wong. High Performance Computational Chemistry: An Overview of NWChem a Distributed Parallel Application. *Computer Phys. Comm.,* **128**, 260-283 (2000).

Kruse, P, ER Johnson, GA DiLabio, and RA Wolkow. Patterning of Vinylferrocene on H-Si(1000) via Self-Directed Growth of Molecular Lines and STM-Induced Decomposition. *Nano Letters*, **2**, 807-710 (2002).

Lopinski, GP, DDM Wayner, and R Wolkow. Self-directed growth of molecular nanostructures on silicon. *Nature*, **406**, 48-51 (2000).

Electronic Structures and Energetics of [5,5] and [9,0] Single-Walled Carbon Nanotubes

J Cioslowski,[a] *N Rao,*[b] *and D Moncrieff*[a]
(a) Florida State University, Tallahassee
(b) Institue of Physics, University of Szczecin, Poland

Single-walled nanotubes (SWNTs) appear to be a solution for many problems, including storage of hydrogen, growth of crystals and sorption of toxic substances. However, the feature that seems to have caught the attention of many is their potential use in nanoelectronic devices. In this work (Cioslowski et al. 2002), NWChem (Kendall et al. 2000) is used to calculate standard enthalpies of formation, ionization potentials, electron affinities, and band gaps of finite length [5,5] armchair and [9,0] zigzag SWNTs capped with C_{30} hemispheres.

Schematics of the [5,5] airmchair and [9,0] zigzag SWNTs are shown in Figure 1. From the schematic, it is clear that structures and diameters of these tubes can be significantly different from one another. In this study, properties of the SWNTs are examined for the two systems with increasing numbers of carbons (and, therefore, length).

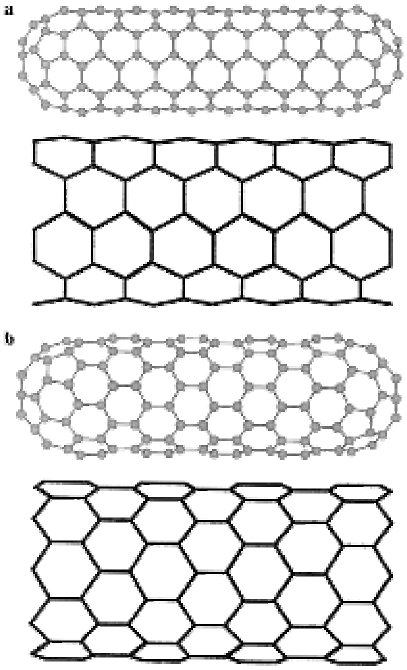


Figure 1. Carbon skeletons of (a) [5,5] armchair and (b) [9,0] zigzag SWNTs.

Geometry optimizations of the nanostructures were performed at the B3LYP/6-311G* level of theory using an accurate grid definition (57 shells in the Euler-MacLaurin radial quadrature and a Lebedev angular quadrature with 302 points. Isodesmic reactions are used to obtain standard enthalpies. One of the interesting conclusions from the optimizations revealed that the diameter of the capped ends is actually larger than the diameter of the rest of the tube (by as much as 0.08 Å in the [5,5] armchair systems).

The properties of SWNTs are found to be very dependent on the tube length, although there is slow convergence of these properties with respect to the tube length. The size of tubes studied ranged from C_{70} to C_{210}. It was shown that all of the properties of the [5,5] armchair system converge in an oscillatory manner. The band gap also oscillates with respect to length, and it is predicted that the transition to metal takes place in the bulk. However, the size of systems in this study was not sufficient to estimate the size of system where the metallic behavior dominates. It was observed that oscillation period of 3 is consistent with the wave vector of approximately 2π/3 at which the band crossing responsible for the metallic character of infinite length [n,n] armchair SWNTs occurs.

For the [9,0] zigzag structures there are two possible structures, D_3 and D_{3d} or D_{3h}. For all but the C_{78} system, the system with more symmetry is predicted to be more stable. The difference in the standard enthalpies of formation is estimated to be 4 kcal/mol for long tubes. An infinite length [9,0] zigzag SWNT is predicted to be a seminconductor rather than a metal in agreement with experimental evidence (Ouyang et al. 2001).

Future work will involve larger nanotubes of variable diameters.

References

Cioslowski, J, N Rao, and D Moncrieff. Electronic Structures and Energetics of [5,5] and [9,0] Single-Walled Carbon Nanotubes. *Journal of the American Chemical Society*, **124**, 8485 (2002).

Kendall, RA, E Apra, D Bernholdt, EJ Bylaska, M Dupuis, GI Fann, RJ Harrison, J Ju, JA Nichols, J Nieplocha, TP Straatsma, TL Windus, and AT Wong. High Performance Computational Chemistry: An Overview of NWChem a Distributed Parallel Application. *Computer Phys. Comm.,* **128**, 260 (2000).

Ouyang, M, JL Huang, CL Cheung, and CM Lieber. Energy Gaps in "Metallic" Single-Walled Carbon Nanotubes, *Science*, **292**, 702 (2001).

User Projects

First Principles Methods for Predicting the Chemistry of Environmentally Relevant Systems
RA Friesner, HA Stern, BF Gherman, F Rittner, BD Dunietz, E Batista, M-H Baik, ME Wirstam
Columbia University, New York City, New York
M Beachy, RB Murphy
Schrodinger, Inc., Portland, Oregon
DT Mainz, WA Goddard, MS Diallo, WB Floriano, RP Muller, M Blanco, YH Jang, RJ Nielsen, DM Philipp, F Faglioni, V Molinero
California Institute of Technology, Pasadena, California

Biomolecular Interfaces: Simulation of Sedimentation of Biomolecular Assemblies
AD Haymet, BW Beck, GC Lynch, BM Pettitt, TC Rogala, K-Y Wong
University of Houston, Texas

Computational Studies in Molecular Geochemistry and Biogeochemistry
AR Felmy, DA Dixon, M Dupuis, TP Straatsma, KM Rosso, EJ Bylaska
Pacific Northwest National Laboratory
GA Voth
University of Utah, Salt Lake City
JH Weare
University of California SanDiego
DA Yuen, JW Halley
University of Minnesota, Twin Cities
DG Rancourt
University of Ottawa, Ontario, Canada
P E Smith
Kansas State University, Manhattan
RD Lins Neto
Swiss Federal Institute of Technology, Zurich, Switzerland

Simulation of Climate Forcing by Aerosols
SJ Ghan, GI Fann, X Bian, EG Chapman, RC Easter, RA Zaveri
Pacific Northwest National Laboratory

Reliable Electronic Structure Calculations for Heavy Element Chemistry: Molecules Containing Actinides, Lanthanides, and Transition Metals

WA de Jong, BP Hay, J Li, DA Dixon
Pacific Northwest National Laboratory
MM Marino, WD Ermler
University of Memphis, Tennessee
RJ Harrison
Oak Ridge National Laboratory, Tennessee
K Hirao
University of Tokyo, Japan
Co Marsden
University of PaulSabatier/Laboratoire dePhysique Quantique – IRSAMC, Toulouse, France
L Visscher, SJA vanGisbergen
Vrijie Universiteit/ScientificComputing and Modelling, Amsterdam, The Netherlands
Z Zhang
Lawrence Berkeley National Laboratory, California
BE Bursten
Ohio State University, Columbus
KG Dyall
NASA Ames Research Center, Moffett Field, California
C-G Zhan
Columbia University, New York City, New York

Immobilization of Radionuclides in Oxides

H Jonsson, GA Henkelman, F Vila, K Tsemekhman
University of Washington, Seattle
JA Nichols, F Gao, EJ Bylaska, LRR Corrales
Pacific Northwest National Laboratory
D Makarov
University of Texas, Austin
J Garza-Olguin
Universidad Autonoma Metropolitana, Mexico City, Mexico
A Rubio
Donostia International Physics Center, Donostia-San Sebastian, Spain

A Computational Approach to Understanding Aerosol Formation and Oxidant Chemistry in the Troposphere

JS Francisco
Purdue University, West Lafayette, Indiana
KA Peterson
Washington State University, Pullman
R Bianco, JR Hynes
University of Colorado, Boulder
GK Schenter, SS Xantheas, LX Dang, BC Garrett, DA Dixon, SM Kathmann
Pacific Northwest National Laboratory
A Morita
Kyoto University, Japan
MA Miller
University of Washington, Seattle

Biomolecular Simulation of Base Excision Repair and Protein Signaling

TP Straatsma, ER Vorpagel
Pacific Northwest National Laboratory
M. Zacharias
Institute for Molecular Biotechnology, Jena, Germany
CF Wong, JA McCammon
University of California San Diego
PE Smith
Kansas State University, Manhattan
JH Miller, K Wei
Washington State University, Pullman

Chemical Fate of Contaminants in the Environment: Chlorinated Hydrocarbons in the Groundwater

DG Truhlar, CJ Cramer
University of Minnesota, Twin Cities
YA Borisov
Russian Academy of Science, Moscow
TH Dunning
NC Supercomputing Center, University of North Carolina, Chapel Hill
K Morokuma
Emory University, Atlanta, Georgia
TP Straatsma, M Dupuis, BC Garrett
Pacific Northwest National Laboratory

Staff

Molecular Sciences Visualization and User Services
Donald R. Jones, Technical Leader
(509) 375-2279, dr.jones@pnl.gov

Toni P. Quackenbush, Sr. Administrative Secretary
(509) 376-2767, toni@pnl.gov

Erich R. Vorpagel, Chief Scientist
(509) 376-0751, erich.vorpagel@pnl.gov

Wibe A. de Jong, Sr. Research Scientist
(509) 375-3973, wibe.dejong@pnl.gov

Jun Li, Sr. Research Scientist
(509) 375-2634, jun.li@pnl.gov

Alejandro Aceves-Gaono, LTE College Student
(509) 376-5006, alex.aceves@pnl.gov

Matrixed
Michael C. Perkins, Sr. Communications Specialist
(509) 372-6119, michael.perkins@pnl.gov

Kenneth A. Perrine, Sr. Research Engineer
(509) 375-6458, kenneth.perrine@pnl.gov

Molecular Science Computing Facility Computer Operations
R. Scott Studham, Technical Leader
(509) 376-8430, scott.studham@pnl.gov

Gary B. Skouson, Senior Research Scientist
(509) 376-5401, skouson@pnl.gov

Tim A. Witteveen, Senior Research Scientist
(509) 372-1363, timw@pnl.gov

David E. Cowley, Senior Research Scientist
(509) 376-9181, david.cowley@pnl.gov

Ralph E. Wescott, Senior Research Scientist
(509) 376-8314, ralph.wescott@pnl.gov

Scott M. Jackson, Senior Research Scientist
(509) 376-2205, scott.jackson@pnl.gov

Ryan P. Wright, Science & Engineering Associate
(509)376-3052, ryan.wright@pnl.gov

Evan J. Felix, Research Scientist
(509) 376-1491, evan.felix@pnl.gov

Nathan D. Tenney, Research Scientist
(509) 376-1493, nathan.tenney@pnl.gov

Zach A. Garrett, LTE College Student
(509) 376-3239, zachary.garrett@pnl.gov

Molecular Sciences Software

NWChem Staff
Theresa L. Windus, Technical Leader
(509) 372-6410, theresa.windus@pnl.gov

Edoardo Apra, Senior Research Scientist
(509) 375-2778, edoardo.apra@pnl.gov

Jessica Foreman, Adminstrative Secretary
(509) 375-2412, jessica.foreman@pnl.gov

Mahin Hackler, Science & Engineering Associate
(509) 375-2765, mahin.hackler@pnl.gov

So Hirata, Senior Research Scientist
(509) 375-3630, so.hirata@pnl.gov

Lisa Pollack, LTE Postdoc
(509) 375-6642, lisa.pollack@pnl.gov

Yuri Alexeev, LTE Postdoc
(509) 375-6433, yuri.alexeev@pnl.gov

Matrixed
Eric Bylaska, Senior Research Scientist
(509) 376-9114, eric.bylaska@pnl.gov

TP Straatsma, Technical Lead
(509) 375-2802, tps@pnl.gov

Ecce
Gary Black, Senior Research Scientist
(509) 375-2316, gary.black@pnl.gov

Brett Didier, Senior Research Scientist
(509) 372-6311, brett.didier@pnl.gov

Todd Elsethagen, Senior Research Scientist
(509) 375-4431, todd.elsethagen@pnl.gov

Sue Havre, Senior Research Scientist
(509) 375-6948, susan.havre@pnl.gov

Carina Lansing, Research Scientist
(509) 375-2482, carina.lansing@pnl.gov

Bruce Palmer, Senior Research Scientist
(509) 375-3899, bruce.palmer@pnl.gov

Michael Peterson, Science & Engineering Associate
(509) 372-4751, michael,peterson@pnl.gov

Karen Schuchardt, Staff Scientist
(509) 375-6525, karen.shuchardt@pnl.gov

ParSoft
Jarek Nieplocha, Technical Leader
(509) 372-4469, jarek.nieplocha@pnl.go