Center for High-Performance Computing

NEWS

THE UNIVERSITY OF UTAH

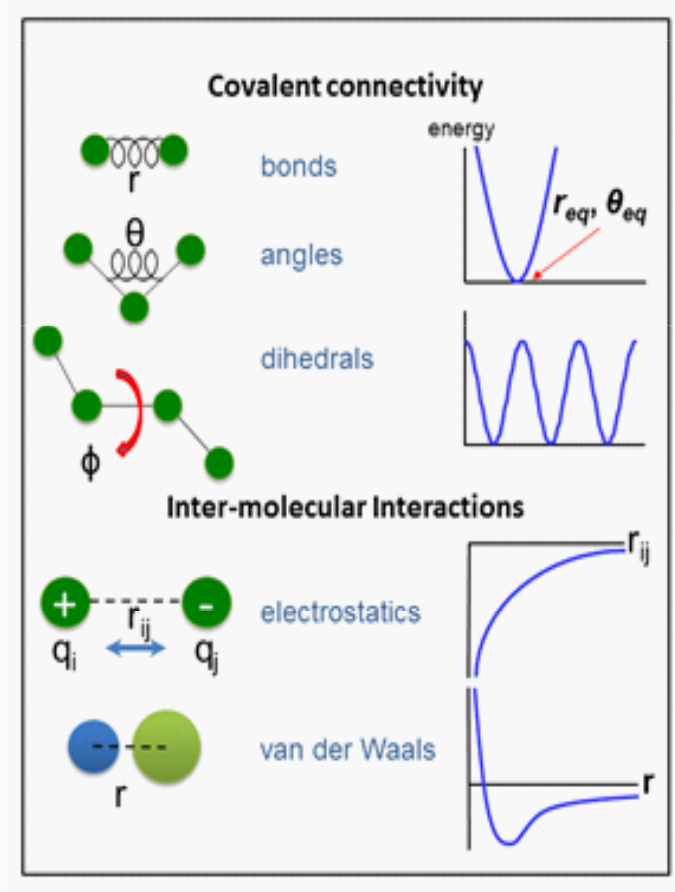
FALL 2013

Research

A Slice of Biomolecular Simulation at CHPC

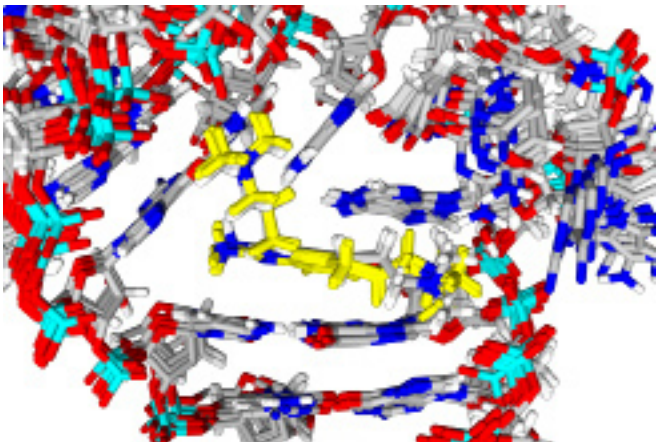
By Thomas Cheatham and Janet Ellingson

Remember when your teacher used a ball-and-stick model to demonstrate that molecules are made of atoms? While useful as introductory tools for understanding covalent connectivity and the conformations of molecules when teaching chemistry, these models are obviously too simple to study the complex biochemical processes of life—or are they?



Shown above are the interactions represented by molecular mechanics models. Various parameters in the "force field" define the appropriate covalent connectivity, inter-molecular interactions, and strength of interactions.

The processes of life are enabled by the dynamic behavior and interplay of biomolecular machines. These machines, including proteins, RNA and other biomolecules, function in a complex and crowded environment of water, salts, and other cellular components. Studying these systems experimentally can be challenging, as full insight requires an atomic level understanding of the structure and dynamics across time-scales ranging from femtosecond (10^{-15} sec) bond vibrations to large-scale conformational and/or chemical changes which can take seconds and beyond. Since the advent of modern computational simulation methods, a common question is whether these simulation methods can provide insight into biomolecular structure and dynamics. As indicated by the awarding of this year's Nobel Prize in Chemistry to Martin Karplus, Michael Levitt, and Arieh Warshel "for the development of multiscale models for complex chemical systems," the answer is yes! Moreover, the "models" used to study various biomolecular processes such as protein folding and drug-binding are conceptually not much more complicated than a ball-and-stick model, just with the addition of an equation describing how the spheres interact with each other (also known as the molecular mechanical "force field") and the means to move the model through time—via molecular dynamics simulations that follow Newton's equations of motion subject to the forces calculated from the force field.

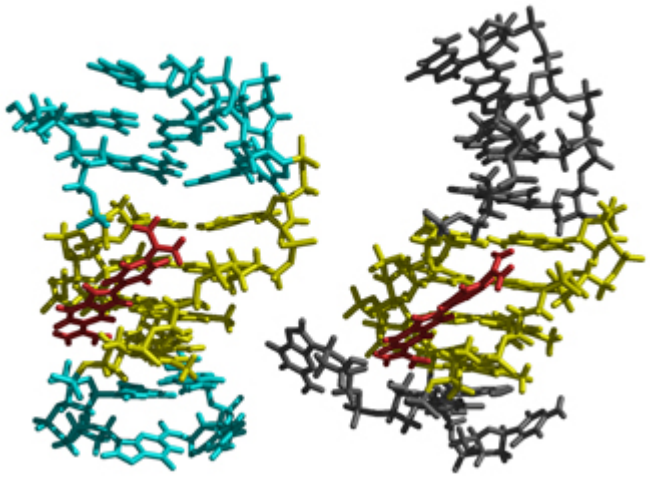


Shown above are molecular graphics representations of a portion of the refined NMR structures of two drug stereoisomers (yellow and gray) bound to a portion of the IRES RNA of the Hepatitis C virus. The work from the Davis and Cheatham labs in the Department of Medicinal Chemistry was published in PNAS 107, 7263 in 2010 and used extensive molecular dynamics simulation, in part on CHPC resources, to refine the structural models. Further simulations are underway to design alternative small molecules which bind with a higher affinity.

(Continued on Page 2)

(Continued from Page 1)

If you've run on CHPC machines in the past few years, Cheatham's name and those of his group members may be familiar to you as you fight them in the queues. Since arriving in Utah in 2000, his group has been an extensive user of CHPC resources. Now an associate professor in the Department of Medicinal Chemistry in the College of Pharmacy, Prof. Cheatham develops and applies biomolecular simulation methods to give insight into nucleic acid and protein structure and dynamics. The Cheatham lab is one of the core members of the team of AMBER molecular simulation software developers. AMBER, installed on CHPC machines and described in greater detail at http://ambermd.org, traces its roots and code back to the work of Karplus, Levitt, Warshel, and others. Development formally started in the mid-70s, and the suite of programs that make up AMBER have evolved into a powerful set of tools for understanding the structure, dynamics, interactions and free energetics of biomolecules. Cheatham himself started using and developing these tools around 1990 as a member of the Kollman lab at UCSF. Today, these tools are applied using very powerful large-scale HPC resources at the University of Utah and nationally in order to assess the reliability and validity of the models. Fully assessing the simulations requires the development of tools which enable faster and more complete analysis of the data. In addition, it is important to make this data available to the scientific community; to this end the group develops software to facilitate searching, displaying and dissemination of the raw and derived data.



Older MD simulation results of the dye DAPI bound to a B-DNA duplex in two simulation snapshots at 350 ns of MD simulation showing significant distortion due to force field problems which have since been fixed. These were run from 10/2002-3/2005

Are the simulation results reliable? When assessing the results of a computational simulation there are two main concerns: accuracy and convergence. Accuracy relates to how well the force field describes the interactions between atoms. Convergence means being able to fully follow the time-course of the conformational processes of interest. The two are intimately related; in order to properly assess accuracy of simulations, they must be close to converged. This is extremely challenging as many biological processes (such as protein folding) occur on the millisecond to second time-scale. While special-purpose machines that can reach millisecond time scales do exist, such as the Anton MD engine from D.E. Shaw Research, this is the exception rather than the rule; most simulations run on current hardware are just now reaching the microsecond time-scale.

Funded mainly by NIH grants, the key focus of the Cheatham lab at present is to validate and assess the force fields while also reaching convergence in the simulations of nucleic acids and proteins. As computer power continues to improve at a rapid pace, we will soon be able to reach biologically relevant time-scales, and in doing so expose problems that can lead to improvement in the force fields and simulation methods. However, before we get there, the Cheatham lab has repeatedly pushed current simulation methods to the limits on widely available resources and has demonstrated that (a) short simulations can stay near the experimental structure, (b) longer simulations tend to stray into areas not consistent with experiment, and (c) we have not reached convergence. For example, the previous figure shows the distortion of DNA helices due to force field artifacts from simulations on CHPC resources. However, thanks to advances in the force fields and sampling methods over the past 14 years, we are beginning to see the light at the end of the tunnel in both force field reliability and convergence.

Taking advantage of national resources: Where the rubber hits the road for large-scale simulation is the availability of our national cyberinfrastructure. One of the special features of AMBER is the outstanding performance of the PMEMD molecular dynamics engine on graphical processing units (GPUs) from nVIDIA. A single Titan or K20X GPU out performs 2-4 CPU nodes on the CHPC clusters at a fraction of the cost. At CHPC, a number of Ember nodes host Tesla 2090 GPUs. In addition there are a few Kepler GPUs that are mainly used for development and simulation setup. However, it is the access to significant GPU resources through very large allocations of computer time on the largest university machine in the world, Blue Waters at the University of Illinois (Cheatham heads one of ~33 NSF PRAC national science teams with access to this resource) and also through the NSF

(Continued on Page 6)

News

Julio Facelli Becomes Director of BMIC

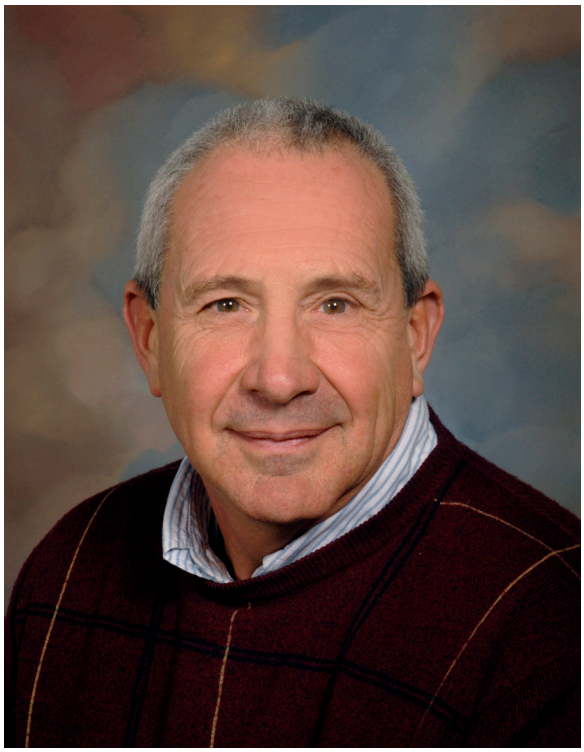
Effective October 1, 2013, Julio Facelli resigned his position as director of CHPC to accept the directorship of The Bioinformatics Core (BMIC) of the Center for Clinical and Translational Science (CCTS). Julio had been a part of CHPC since its beginning in 1989. He was appointed director in 1995. During those 24 years he moved the center's computing resources from the IBM 3090 (six CPUs and 256 megabytes of main memory) to the new Kingspeak cluster (90 Dell nodes each with 16 core and 64 gigabytes of memory), one of CHPC's three large clusters.

In his new position, Julio will continue to support research that requires high-performance computing. The BMIC provides information technology support to clinical and translational researchers (those who translate scientific findings into practical applications) through a variety of means, including Education, Innovation and Service.

The Education component of the BMIC uses a mix of formal courses and informal approaches to education in biomedical informatics and clinical research informatics for translational research. The formal approach includes course work, lectures and seminars from the University of Utah's Department of Biomedical Informatics and the CCTS Masters of Science for Clinical Investigation (MSCI) program. Informal training and presentations are always available upon request from research groups, divisions and departments.

The Innovation component of the BMIC aims to provide universal access to data from our partner institutions (Intermountain Healthcare, Veteran's Administration Medical Center and the Utah Department of Health) through the use of open source tools and applications, with customized development when required.

The Service component of the BMIC works across all CCTS cores to provide resources and services for the planning and implementation of translational research. This service focuses on a defined set of applications which are applied across multiple domains to more effectively support researchers.



Julio Facelli, Director, BMIC

CHPC works closely with the BMIC by providing the basic hardware and software infrastructure necessary for the BMIC applications. This collaboration allows a great deal of synergy between the two Centers allowing for significant economies of scale and opportunities to explore new research opportunities.

Main Contact:
Alisha M. Edison-Hair
Executive Secretary
Phone: 801-587-8001
Email: alisha.edison-hair@utah.edu

Core Administration:
Julio Facelli, Ph.D.
BMIC Director

Bernie LaSalle
BMIC Operations Director

Special Pricing on Kingspeak Nodes

CHPC has obtained special pre-authorized pricing from Dell for the purchase of nodes for the new kingspeak cluster. This special pricing is available through January 2014. If you are interested in purchasing nodes for your research please contact issues@chpc.utah.edu indicating your interest and the quantity of nodes, in multiples of 4.
Each node has:

- CPU: Intel Xeon (Sandybridge) e5-2670 2.6 Ghz (8 cores/ socket 2 sockets/node=16 Cores)
- Memory: 64 GB 1600Mhz RDIMMS. (4 GB/CPU-Core)
- Local Disk: 500 GB for swap and /scratch/local/
- Interconnect: Mellanox FDR Infiniband (56 Gbs)
- 5 Years Warranty

Kings Peak in the Uintah mountains is Utah's high-point. It is also the name of CHPC's newest cluster located at the University's Downtown Data Center.

CHPC
XSEDE HPC Workshops
INSCC Auditorium (Room 110)

Dec 4th and 5th - MPI

This workshop is intended to give C and Fortran programmers a hands-on introduction to MPI programming. Both days are compact, to accommodate multiple time zones, but packed with useful information and lab exercises. Attendees will leave with a working knowledge of how to write scalable codes using MPI – the standard programming tool of scalable parallel computing.

Tentative Agenda

Wednesday, December 4

9:00 Welcome
9:15 Intro to Parallel Computing
10:15 Computing Environment
11:00 Lunch break
12:00 Introduction to MPI
1:15 Scalable Programming: Laplace code
2:00 Laplace Exercises
3:00 Adjourn

Thursday, December 5

9:00 Laplace Exercises
10:00 Laplace Solution
10:30 Lunch break
12:30 Outro to Parallel Computing
1:30 MPI Debugging and Profiling
3:30 Adjourn

Jan 22nd - OpenMP

This workshop is intended to give C and Fortran programmers a hands-on introduction to OpenMP programming. Attendees will leave with a working knowledge of how to write scalable codes using OpenMP.

Tentative Agenda:

9:00 Welcome
9:15 Environment
9:30 It's a Multi-core World
10:00 Intro to OpenMP
11:00 Lunch Break
12:00 Exercise 1
12:30 More OpenMP
1:15 Exercise 2
2:00 Debugging & Profiling
2:30 OpenMP 4.0 Sneak Peek
3:30 Adjourn

The workshops are free, but you must register.
Please go to https://www.xsede.org/web/xup/course-calendar.
Choose from the list the course offered at the University of Utah.

News

Reorganization at CHPC

With the resignation of Julio Facelli as CHPC's director [see story on page 3] and the loss of several employees this summer, Steve Corbato, CHPC's interim director, has taken the moment to reorganize our staff to better serve our users.

A key change is the appointment of Anita Orendt as Assistant Director, Research Consulting & Faculty Engagement. In recognition that CHPC's relationship with faculty users is essential to successful HPC research, Anita will assume primary responsibility for overseeing the faculty collaboration efforts of all CHPC staff members and act as the point person for communication to the CHPC user community. She will also oversee several aspects of CHPC User Services activity, including coordination of the scientific staff efforts, user account maintenance, and the response to Jira issue reports. For our users who do molecular modeling, be assured that Anita will continue to provide support for those who use the chemistry packages at CHPC.

Anita Orendt, Assistant Director, Research Consulting & Faculty Engagement

Julia Harrison remains the associate director and Guy Adams continues as the assistant director of systems and networking.

As a recognition of their contributions to CHPC, several staff members are now idenitified as project leads: Brian Haymore is responsible for HPC and storage; Wayne Bradford will oversee security issues, especially with the HIPAA servers; Steve Harper is the lead on virtualization, including managing the VM farm and investigating other virtualization options; Sam Liston will take the lead on SuperComputing events with the current task of making a brilliant presentation at SC13 in Denver; and Joe Breen is the project lead on advanced networking.

Examples of Recent Research Using CHPC Resources

Abu-Zayyad, T., et al (2013). "The energy spectrum of ultra-high-energy cosmic rays measured by the Telescope Array FADC fluorescence detectors in monocular mode." Astroparticle Physics 48: 16 - 24.

Chen, D. P., J.H.; Hooper, J.B.; Klittnick, A.; Shen, Y.; Korblva, E.; Bedrov, D., Walba, D.M.; Glaser, M.A.; Maclennan, J.E., Clark, N.A. (2013). "A Twist-bend Chiral Helix of 8nm Pitch in a Nematic Liquid Crystal of Achiral Molecular Dimers." Proc. Natl. Acad. Sci.

Crosman, E. T., N.P. Lareau, Horel, J.D. (2013). Large-Eddy Simulations of Cold Air Pools in Utah's Salt Lake Valley. Western Air Western Air Quality Modeling Workshop – Technical Methods and Applications. Boulder, CO.

Henriksen, N. M., Davis, D.R., Cheatham, T.E. (2012). "Molecular dynamics re-refinement of two different small RNA loop structures using the original NMR data suggest a common structure." J. Biomol. NMR 53: 321-339.

Faber, S., Bartholomew, K., Li, X., Paez, A., Nurul Habib, K.M. (2013). Social Equity in Distance Based Transit Fares. GIS in Transit Conference. Washington, D.C.

Jensen, K. J. T., M.S.; Rost, S. (2013). "SPdKS analysis of ultralow-velocity zones beneath the western Pacific." Geophys. Res. Lett 40: 1 - 5.

Krizmanic, J., Bergman, D., Sokolsky, P. (2013). The Non-Imaging CHErenkov Array (NICHE): A TA/TALE Extension to Measure the Flux and Composition of Very-High Energy Cosmic Rays. The 33rd International Cosmic Ray Conference, Rio de Janeiro, Brazil.

Lehner, M., Rotunno, R., Whiteman, C. D. (2013). Initial simulations of flow over a small crater basin in preparation for an upcoming field experiment. 32nd International Conference on Alpine Meteorology. Kranjska Gora, Slovenia.

Schmerr, N. C., Kelly, B. M., Thorne, M.S. (2013). "Broadband array observations of the 300 km seismic discontinuity." Geophys. Res. Lett 40: 1 - 6.

Wei, L., Pu, Z., Wang, S. (2013). "Numerical simulation of the life cycle of a persistent wintertime inversion over Salt Lake City Boundary Layer Meteorology." Boundary Layer Meteorology.

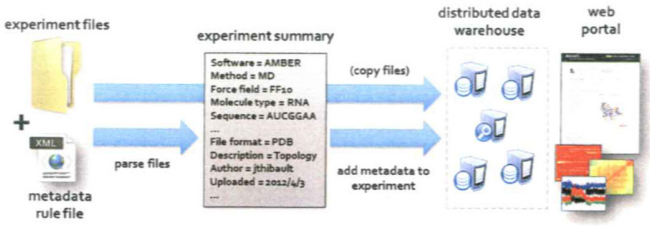
For a full bibliography go to: http://www.chpc.utah.edu/docs/research/CHPCBibliography.pdf

(Continued from Page 2)

Extreme Science and Engineering Discovery Environment (XSEDE) resources Keeneland at Georgia Tech and Stampede at the Texas Advanced Computing Center, the lab is able to push the limits of simulation. To make efficient use of the GPU accelerated code, ensembles of simulations can be run simultaneously on different GPUs with frequent exchanging of information to speed up the convergence in sampling the conformational distributions of biomolecules. The most common approach is replica-exchange MD (REMD) where periodically the independent simulations exchange temperature, pH, energy representations or force fields (Hamiltonian), or other properties. Thanks to access to Blue Waters, we are now able to efficiently run multi-dimensional REMD simulations, effectively reducing the time to solution from many months to days; however, this comes at the cost of significant data overload.

Data overload and the need to facilitate analysis: As MD simulations follow the motions of biomolecules and their surrounding environment, the main data explored is the MD "trajectory," effectively the time-course of the atomic positions, which is a "movie" of the molecular motion. Using millions of hours of computer time on a large-scale computational resource such as Blue Waters leads to the generation of an extremely large amount of data! At present, the Cheatham group has over 250 TB of disk space at CHPC used to host their data for further analysis. In fact so much data has been generated that oftentimes the backlog of complex analyses and data manipulations necessary to interpret the simulation results takes longer than the simulations themselves. To facilitate analysis, powerful domain-specific software has been developed (namely PTRAJ and CPPTRAJ, of which Cheatham and Roe are the primary authors) to extract meaningful insight from the series of numbers generated by the MD simulations. CPPTRAJ, a re-write of the original PTRAJ code built with big data and extensibility in mind, provides powerful tools to manipulate, analyze, and better understand the data. Moreover, both codes are not AMBER-centric and can be applied to the results from other popular MD simulation engines like Gromacs, CHARMM, NAMD and others. Both PTRAJ and CPPTRAJ are freely available in the AmberTools suite of programs, and if you perform MD simulations, you may consider investigating these analysis tools to explore the various processes occurring in the simulation. While these tools provide a means to dissect and discover what is going on in individual simulations, as we move to models where we are running ensembles of simulations and/or parameter scanning, management of the simulation process becomes an acute bottleneck.

Searching, displaying and sharing the data: How can we possibly manage and understand all this data? Moreover, this data—generated at considerable cost—could have unanticipated uses that may be enabled by wider dissemination. The question becomes how can we make this data available? The prototype for this is the iBIOMES environment developed by Thibault in the Facelli and Cheatham labs. A more detailed description of the prototype is available in J. Chem. Info. Modeling 53, 746 (2013). The iBIOMES software provides an automated way to annotate and parse the MD (and QM) simulation data to extract metadata—i.e. information that describes the simulations performed and available data—and tools to access the raw and derived simulation data and analyses. With capabilities to search, display and disseminate the data, users can more easily organize and share their simulation results. The data can be retained locally or stored in a distributed framework across other iBIOMES servers.



To facilitate exploring the data, a web interface is provided to "see" the simulation protocols and results. Importantly, this is customizable so users can tailor the web page to display their particular analyses of interest.

To learn more about iBIOMES and see the prototype, visit http://ibiomes.chpc.utah.edu/ibiomes-web/

For more information about the Cheatham lab, see http://www.chpc.utah.edu/~cheatham

Podcasts of CHPC Presentations

Each Spring and Fall CHPC staff offer presentations on the use of common tools in high performance computing. These discussions are recorded and posted on the CHPC website: http://chpc.utah.edu/docs/media/podcasts/
You can watch the following presentations:

Using Gaussian09 and Gaussview by Anita Orendt
Chemistry Packages for CHPC by Anita Orendt
Using Python for Scientific Computing by Wim Cardoen
Debugging with TotalView by Martin Cuma
Mathematical Libraries at CHPC by Martin Cuma
Introduction to Linux for HPC by Martin Cuma
Introduction to Parallel Computing by Martin Cuma
XSEDE Resource Support at CHPC by Julia Harrison
Overview of CHPC by Wim Cardoen
Intro to I/O in the HPC Environment by Brian Haymore
Intro to GPU Programming by Wim Cardoen

CHPC Staff Directory

Administrative Staff	Title	Phone*	Email	Location
Steve Corbató	Interim Director			
Julia D. Harrison	Associate Director			
Guy Adams	Assistant Director, Systems & Network			
Anita Orendt	Assistant Director, Research Consulting & Faculty Engagem			
Janet Ellingson	Administrative Manager & Newsletter Editor			

Scientific Staff	Expertise	Phone*	Email	Location
Wim Cardoen	Scientific Applications			
Martin Cuma	Scientific Applications			
Sean Igo	Natural Language Processing			
Jody Smith	Database Manager			
Albert Lund	Graduate Assistant			

Project Leads		Phone*	Email	Location
Wayne Bradford	Security			
Joe Breen	Advanced Network Initiatives			
Steve Harper	Virtualization			
Brian Haymore	HPC & Storage			
Sam Liston	Special Events/SC			

Technical Support Staff	Group	Phone*	Email	Location
Irvin Allen	Systems Admin			
Amanda Allen	Tech Assistant			
Nathaniel Ellingson	Tech Assistant			
Jake Evans	Network Engineer			
David Heidorn	Systems Admin			
Jimmy Miklavcic	Systems Admin			
David Richardson	Systems Admin			
Walter Scott	User Services			
Abby Tohinaka	Intern			
Alan Wisniewski	Network Support			

*All phone numbers are preceded by area code 801 unless otherwise noted.

What is CHPC?

The University of Utah's Center for High Performance Computing provides large-scale computer systems, networking, and the expertise to optimize the use of these high-end technologies. CHPC facilitates advancement in academic disciplines whose computational requirements exceed the resources available in individual colleges or departments. Since 1996 these collaborations have resulted in more than 829 technical publications. CHPC's purview is to support faculty and research groups whose main focus requires computing and advanced networking as core instrument(s) central to their research.

THE UNIVERSITY OF UTAH

Center for High Performance Computing
155 South 1452 East, RM #405
SALT LAKE CITY, UT 84112-0190

Welcome to CHPC News!

If you would like to be added to our mailing list, please fill out this form and return it to:

[redacted]
THE UNIVERSITY OF UTAH
Center For High Performance Computing
155 S 1452 E ROOM 405
SALT LAKE CITY, UT 84112-0190
FAX: (801)585-5366

(room 405 of the INSCC Building)

Name:
Phone:

Department or Affiliation:
Email:

Address:
(UofU campus or U.S. Mail)

Thank you for using our Systems!

Please help us to continue to provide you with access to cutting edge equipment.

ACKNOWLEDGEMENTS

If you use CHPC computer time or staff resources, we request that you acknowledge this in technical reports, publications, and dissertations. Here is an example of what we ask you to include in your acknowledgements:

"*A grant of computer time from the Center for High Performance Computing is gratefully acknowledged.*"

Please submit copies or citations of dissertations, reports, pre-prints, and reprints in which the CHPC is acknowledged to: Center for High Performance Computing, 155 South 1452 East, Rm #405, University of Utah, Salt Lake City, Utah 84112-0190