SCS
Integrating Advanced Computing with Science, Engineering and Liberal Arts

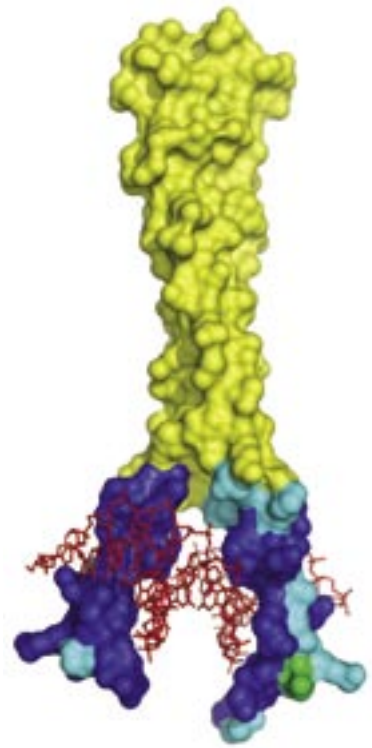
School of
Computational Science
at Florida State University

Trained Computer Models Find DNA Binding Sites on Proteins

INSIDE THIS ISSUE

Neural networks are a type of computer model, originally created to resemble a brain with its complexity of interacting cells – the neurons. While biological brains are still vastly more intricate than artificial neural networks, there are similarities. Like the brain, a neural network is made up of a large number of interconnected units. It can also be trained to perform a certain task.

Neural networks are particularly efficient tools for identifying patterns and finding meaning in a set of complicated data. To train them, you need a sufficiently large set of training data with the particular quality that you are looking for. Once trained, the network can look for the same quality in unknown material.



The image shows a surface representation of a protein (yellow, blue, cyan and green) binding to DNA, shown in red. The DNA makes contact with 32 amino acids of the protein. DISPLAR correctly predicts all of those shown in dark blue. Shown in cyan are the 12 amino acids that DISPLAR predicted, but which are in fact nearest neighbors of the DNA-contacting amino acids. One wrong prediction is in green. The protein is a "basic region leucine zipper transcription factor", involved in the cellular production on RNA.

GOAL IS TO PREDICT

Professor Huan-Xiang Zhou and his student Harianto Tjong use a method based on a neural network to identify specific sites on proteins for binding DNA. The goal is to be able to take any protein and predict whether it can bind to DNA, and if so, where on the protein molecule the binding site is located. As a first step they have tested the method on a set of known protein-DNA complexes.

Dr. Zhou and his group developed the method, named DISPLAR, based on their earlier method for protein-protein binding. Since that research was successful, Dr. Zhou put his student Harianto to work with the same technique to try to predict DNA-protein binding. A couple of intense months later, the two can show promising results – the method has proven very accurate.

BUILDING THE DATABASE

The first step of the work was to build a database with known protein-DNA complexes. Harianto explains that this

continued on page 2

The new academic year promises to be an exciting time for the SCS. The first students admitted to our innovative graduate degree programs will be arriving soon and new courses for those programs are being developed and will be offered over the upcoming year. You'll find some details about the new degree programs on page 3.

Our goal in establishing the new degree programs is to train the next generation of computational scientist. In response to the increasing interdisciplinary nature of research in the sciences and engineering, the SCS is particularly well suited for designing and implementing the type of interdisciplinary degree programs we are offering. The programs will fit hand and glove with our outstanding interdisciplinary research program. In fact, although there are other computational science degree programs being offered, the unique makeup of the SCS makes our programs stand out.

We will also be proudly welcoming four new faculty members to the SCS. This fall, Dr. Sachin Shanbhag will join the SCS and the Department of Chemical and Biomedical Engineering and Dr. Xiaoqiang Wang will join the SCS and the Department of Mathematics. Next spring, Dr. Ming Ye will join us and the Department of Geological Sciences. In addition, Dr. Anter El-Azab will transfer his primary home from the Department of Mechanical Engineering this fall as well. We are eagerly awaiting the arrival of all these faculty members; they will add immeasurably to the research and educational mission of the SCS. The new faculty members will be featured in future newsletters.

Max Gunzburger,
Director, SCS

was very time consuming. An original set of 1100 complexes was downloaded from an internationally available database. The number was reduced to 265, mainly by tedious, manual elimination of protein-DNA complexes that were too similar to each other.

The reason for this reduction was to create a balanced and representative training set for the neural network. It should contain all known types of binding mechanisms in fair proportions, and not be biased towards one kind of binding site.

LEARNING BY DOING

Over 200 of the complexes were used for training, and the actual test was run on 25 DNA-protein complexes. In this case it was obviously known from the beginning that the proteins actually bind to DNA, and where the binding sites are, but the question was if the model would also be able to predict that, based on its training.

The answer was yes, for a majority of the 25 tested binding proteins. It can be assumed that the same accuracy will be achieved with unknown proteins, but that has yet to be proven in another project for Dr. Zhou and his group.

CELLULAR WORKERS

DNA binding proteins play central roles in many biological processes. They take part in DNA replication and repair. They are also important for gene regulation, which is the cell's control system of on-and-off switches for the production of proteins.

harianto@sb.fsu.edu
zhou@sb.fsu.edu

Harianto Tjong is a graduate student in Physics. He is working in Dr. Zhou's biophysics lab, located in the Institute of Molecular Biophysics. Harianto, who got his B.Sc. in Indonesia, came to FSU in 2003, and plans to earn his doctoral degree in 2008.

New Degree Program in Computational Science Begins

The newly approved master's program in Computational Science will begin in the Fall 2006 semester and the Ph.D. in Computational Science is on track to begin in Fall 2007. Seven students, presented in this newsletter, will join our program in August of 2006; most of them seek a Ph.D. in Computational Science.

We expect our new degree programs to address the pressing need to train the next generation of computational scientists. Our graduates should be able to work in an interdisciplinary environment and successfully collaborate with scientists in other disciplines. Ideally, students should learn to develop and analyze new computational procedures which can be utilized in a variety of fields.

MASTER'S TRACKS

The M.S. program has several tracks. The major track is geared towards students who are eventually seeking a Ph.D. in Computational Science, while the other tracks are for students who want to terminate after a master's degree.

The latter tracks are modeled after the Alfred P. Sloan Foundation's Professional Science Master's (PSM) program. This is unique in its combination of technical courses with non-science courses such as project management and entrepreneurship. PSM programs have received attention in publications such as USA Today, The Wall Street Journal, and the Chronicle of Higher Education. In the July, 2004 USA Today article entitled "Add science, business, mathematics and stir" the prediction is made that the PSM degree "will become the 21st century's fastest ticket to the major leagues in business and government."

REQUIRED COURSEWORK

The coursework for the master's programs is based on 32 credit hours for the major track, and 36 for the PSM track. The course hours must be divided between three groups of courses: 1) required (core) courses in computational science, which cut across disciplines; 2) other computational science courses, such as scientific visualization, bioinformatics, numerical methods for earth and environmental sciences, etc. and 3) existing departmental courses in computer science, engineering, mathematics or an applied science.

In addition, a student following the PSM track must take 6 credit hours from non-science courses and complete a summer internship program or the equivalent for 6 credit hours. The master's programs are set up to be completed in two years.

The core computational science courses are new courses under development at SCS. These will provide the student with the programming skills, tools and algorithms necessary to tackle a range of real-world problems.

INNOVATION & FLEXIBILITY

A lot of flexibility has been built into the course requirements for the proposed Ph.D. since students' programs of study can be quite varied due to the interdisciplinary nature of the program. We will have a major track in addition to tracks with a specialization in a particular applied science or engineering area. We have proposed to establish tracks in atmospheric science, biochemistry, physics, biological sciences, geological sciences, and materials science.

For more information, please see our web site, www.scs.fsu.edu or contact Professor Janet Peterson at education@scs.fsu.edu

Number of required course hours for the SCS graduate programs

	MS	PSM	Ph.D.
Core computational science	7	7	11
Other computational science	9	9	9
Other departmental courses	6	6	9
Non-science courses		6	
Other required hours	10	8	25
Total number of credit hours	*32*	*36*	*54**

** 54 hours are required if starting from a B.S.*

Meet Some SCS Students

Sonali Joshi is known in the biology group for saying that Tallahassee is the coldest place in which she has ever lived, a statement that has caused some amusement among her friends at SCS. She was, however, born and raised in even warmer climate in India, where she got her B.S. in Engineering and later worked as a software programmer for four years. When she arrived in Tallahassee with her husband, she looked around for new career openings and found one in Computational Biology, with Professor Peter Beerli at SCS.

Sonali decided to switch tracks and use her programming skills in a different field. She had not studied biology since high school, but after taking biology classes for one semester she started as a graduate student in Biological Sciences in the Fall of 2004. While her daily work still involves programming, she now writes software that models and explores biological phenomena.

Her field is population genetics and her work is focused on the selection pressure on evolving genes. She is trying to find a method to measure the selection coefficient, which shows the intensity of selection. Gene mutations which have a drastic effect, good or bad, on the survival of the individual, will be subject to intense selection pressure. This selection will increase or decrease the frequency of the mutated gene in the population, and by looking at the variation among individuals in a population, one can determine the selection pressure on the gene. Biologists are particularly interested in positive selection, since this process is rare but drives evolutionary change.

Inspite of the cool Tallahassee winters, Sonali likes it here. The biology group is great, she says, and the city is a good a place to live.

Clayton Webster was born and raised in Brantford, Ontario, Canada, just like the great hockey player Wayne Gretzky. Clayton got his B.Sc. in Pure Mathematics from McMaster University in Hamilton, Ontario, and his M.Sc in Applied Math from the same university. For his Ph.D. studies, he applied to 12 graduate programs and was accepted to all. He chose SCS for one reason only – his future advisor Max Gunzburger. Clayton didn't know much about SCS, but was excited to join the diverse research group headed by Max Gunzburger and Janet Peterson.

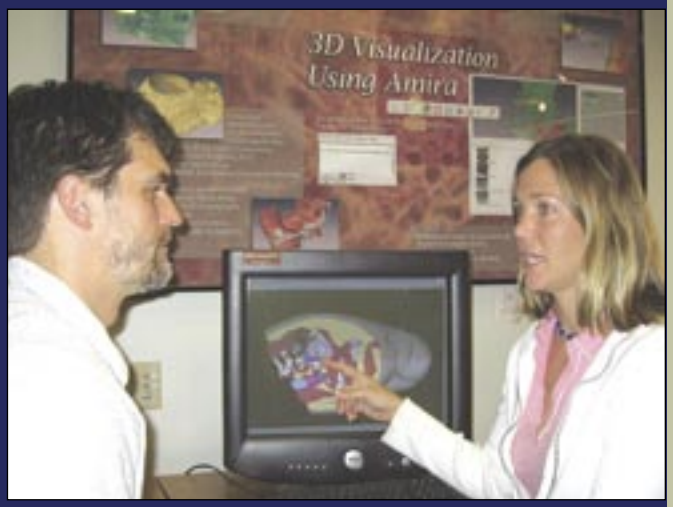
Even though moving this far south took some adjustment

Marie Chapla recently finished her Master's in Biological Oceanography under the guidance of Professor Douglas Nowacek. Though not an SCS student, Marie has done an important part of her work under the auspices of Dr. David Banks of SCS, who focuses his research on visualization of biological tissue.

In her thesis work, Marie investigated the hearing of the Florida manatee, *Trichechus manatus latirostris*. Every year, dozens of manatees are killed or hurt by collisions with motorboats, and to be able to prevent that kind of accidents, it is important to know what a manatee can actually hear.

Marie wanted to find the potential sound pathways through the manatee's head to the middle ears. She used both CT (Computerized Technology) and MRI (Magnetic Resonance Imaging). The CT scans of manatee heads provided information on the arrangement and density of soft tissues and bone. Using Amira visualization software and her CT scans, she was able to create 3D images of structures in the scanned heads without destroying any tissue.

Marie is a telling example of how SCS members collaborate with other departments, providing facilities and instruction for students and faculty across FSU.



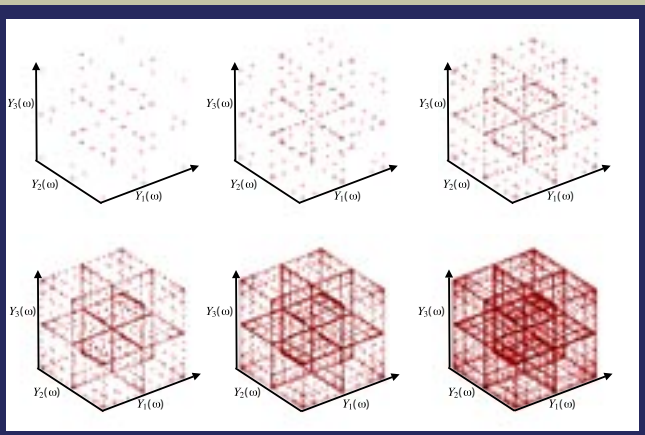
to both climate and culture, Clayton has not regretted his choice. He enjoys the area's beaches and weather, but more importantly, he met his fiancée in Florida. The two are getting married next March.

Clayton works on numerical solutions to stochastic partial differential equations, or SPDEs for short. Recently he started to collaborate with a newly hired faculty member, Raul Tempone, who is an expert in SPDEs.

SPDEs make up many mathematical models used to simulate physical and engineering problems. A SPDE model is a problem whose input data are usually affected by some uncertainty, due to variability in the system, or simply to measuring problems. An example is the study of groundwater flows, where the subsurface properties such as porosity and permeability in an aquifer have to be extrapolated from measurements taken only in few spatial locations.

To handle uncertainty, Clayton models input data as random variables, which adds complexity to the model. Since he works with problems with a large number of variables, he has had to deal with the *curse of dimensionality,* meaning that the volume of calculations increases very rapidly as we increase the number of random variables.

Clayton has developed techniques to efficiently treat problems that depend on a moderately large number of random variables, while keeping a high level of accuracy.



For a 3-dimensional finite probability space the construction of sparse grids using a method called the Smolyak algorithm can be seen in the figure. The goal of Clayton's analysis to predict statistical moments like mean value, variance, and covariance, or even the whole probability distribution of the quantities of physical interest, given the probability distribution of the input data.

Volkan Sevim is fascinated by complexity in general, and complex networks in particular. He is a Ph.D. student of Physics at SCS, but like his advisor Professor Per Arne Rikvold, he is much too curious to stay within traditional subject boundaries. They both see the potential of using methods from physics to explore other fields, which is why Volkan, with a background in statistical mechanics, is working on food webs.

A food web is basically the overall picture of who eats whom in nature. It is one of the most earthy and concrete systems imaginable. However, it can also be studied as a totally abstract and very complex system of connected on/off switches.

Networks in general are trendy, says Volkan, and food webs are no exception. Biodiversity is a hot topic, and many people raise questions about the stability of ecosystems. They want to know what happens to the food web if one species is removed from the system, or if the number of individuals within a species changes drastically. The models used so far have not been good enough to answer these questions, partly because of lack of biological data to test them against. The quantitative approach to biology is also relatively new, and many biologists still shy away from mathematical approaches to their problems. One exception is Professor Peter Beerli, a biologist at SCS, who is on Volkan's committee.

Volkan Sevim came to FSU from Istanbul, Turkey, where he got his B.S in physics. His advisor knew Per Arne Rikvold, and knew that FSU has a good physics department, which is why Volkan chose to came here. He likes Tallahassee, even if the night life here doesn't really measure up to Istanbul's. On the other hand he was happy to leave the Istanbul traffic behind.

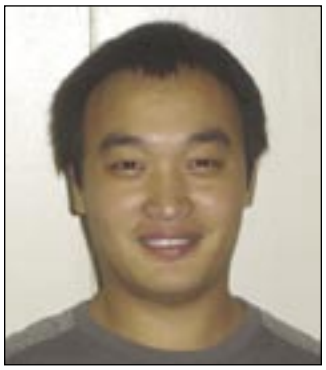
SCS Welcomes the First Students into New Computational Science Program

Evan Bollig will move to Florida from Minneapolis, where he recently finished his B.S. in Computer Science at the University of Minnesota Twin Cities. Evan has worked as a research assistant for the Department of Geology and Geophysics and the Minnesota Supercomputing Institute developing distributed visualization systems. There he began collaborating with Professor Gordon Erlebacher at SCS, who will be the advisor for his thesis work. Evan is co-author on two papers which appeared in Visual Geosciences in 2005. In addition to Computer Science, Evan completed a B.A. in German Studies, and spent a semester studying computer science at Blekinge Institute of Technology in Sweden. Coming from the northern Midwest, Evan is excited to leave behind snow and tornadoes for the chance to ride out a hurricane or two in Tallahassee.

Chris Harden has double Bachelor's degrees in Mathematics and Philosophy from the University of North Florida. He came to Tallahassee in the fall of 2005, where he started as a graduate student at the Department of Mathematics. During the spring semester he had a research assistantship at SCS under the direction of Professor Janet Peterson and is a co-author of a submitted paper which investigates properties of a novel parallel algorithm. His academic interests are centered around the various issues concerning numerical and computational solutions to partial differential equations. He has since transferred into the new SCS program. Outside of the academic world, his main interest is music from all over the world. He plays the mandolin and also enjoys recording and producing music.

Jakir Hossen will come to SCS to work with Professor Navon on four-dimensional variational data assimilation. Jakir has a Bachelor's degree in Mathematics and a Master's degree in Applied Mathematics, both with honors and both from the University of Dhaka in Bangladesh. After graduation he has spent 5 years teaching math at BRAC University in Dhaka.

Yusong Liu came to Tallahassee with his wife, who is a graduate student at FSU. Since his F visa does not allow him to do paid work, he has been volunteering in a couple of FSU labs, most recently with Professor Wei Yang at SCS. Yusong graduated from one of the top universities in China with a major in computer science and technology. He is planning on entering the field of structural biology by joining Professor Yang's lab group. Yusong also volunteers at the public library where he helps individuals with computing problems.



Wes Richardet was the first undergraduate student to contact SCS about the new graduate program, even before it was officially approved. Wes has the ideal background for an SCS student. He was a pioneer already as an

undergraduate, excelling in a new, interdisciplinary program called Scientific Data Analysis and Visualization at DePaul University in Chicago. The program requires students to study computer science, math and two different scientific fields, as well as data analysis and visualization. Wes chose Physics and Chemistry as his fields and rounded off his challenging schedule with classes in GIS (Geographical Information Systems) and remote sensing. Wes is a golfer, and thus likely to enjoy the climate change from Illinois.

Pablo Seleson is Sergio Fagherazzi's new graduate student. Born and raised in Argentina, where he finished high school, he later moved to Israel to study at the Hebrew University of Jerusalem. He has focused on, and excelled in, physics and philosophy, and is about to finish his Master's thesis on cosmology, in the topic of galaxy formation and dark matter. Pablo has also worked as a youth leader and as a volunteer involved in social work with children as well as organized educational projects in mathematics. Pablo will move to Tallahassee with his wife Esther, who has been working in business administration.

Geoffrey Womeldorff is a mathematics major with one year of graduate study from the University of North Florida (UNF). Geoff has varied work experience. As a student, he worked as a grader for mathematics courses and during the summer of 2005 he worked with a remedial mathematics program for incoming freshmen at UNF; prior to attending UNF, he held computer networking oriented positions. Geoff has a strong interest in computational science in general but has not decided the area where he will concentrate his studies. Outside of the academic realm, Geoff has musical interests and enjoys building synthesizers.

Congratulations to New SCS Doctors

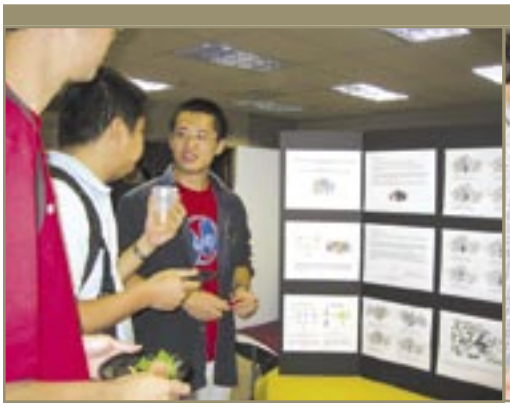
Two young SCS scholars recently received their doctor's degrees. By coincidence, they both defended their theses on the very same day in April. A few weeks later, the new doctors celebrated their birthdays – again on the same day. They were both born on May 5th, 29 years ago, **Santha Akella** in India and **Ibrahim Abou Hamad** in Lebanon.

The two students met at SCS, where they shared discussions, office space, late work nights and good times, but they worked in different scientific fields. Santha (to the right in the picture), whose undergraduate studies on ocean engineering and naval architecture led him into the field of fluid dynamics, worked with mathematics Professor Michael Navon. His field was optimal control, which broadly speaking is how to make physical systems work the way you want them to. In weather forecasting, optimal control is about how to include measurements in computer models for the most accurate results, a procedure called variational data assimilation.

In June, Santha moved to University of Pennsylvania for a post doc position. He will be working on a new type of model, which is small and can be used to make fast and accurate predictions of the spread of pollutants, pollen, radioactivity, etc.

Ibrahim (left in the picture) came to FSU to work on condensed matter physics. His doctoral work, done under the supervision of Professor Per Arne Rikvold, concerned surface science. The system in focus was a silver surface submerged in a bromine/chlorine solution. Ibrahim examined the ion exchange that takes place after a voltage is applied to the silver. He studied the system through a computer model, which allowed an impressively detailed study, down to single ions. Colleagues at other universities have done physical experiments in laboratories, with which Ibrahim's results were compared.

Science like this is important for fuel cell technology, biodetectors, and batteries. During his post doc stay at Mississippi State University with professor Mark Novotny, Ibrahim will work with lithium batteries.



SCS Student Expo 2006

The graduate student poster exhibition at SCS is on its way to becoming a tradition. This year's version featured 22 students, a significant increase compared to the first year.

SCS faculty and the students they are advising represent several different scientific disciplines, and in spite of the common core of math and computing, it is not always easy for a student in one area to explain his or her work to someone in another field. The X-po provides students with useful training for a future in interdisciplinary settings.

Although all the students participating in the 2006 X-po plan to receive their degrees in their home department, each has an affiliation with SCS. Since SCS will shortly receive its inaugural group of students into its new graduate program, future X-pos will include students working towards degrees in computational science as well as students from various departments who are advised by SCS faculty. The interdisciplinary discussions will most certainly prevail.

Left: Doctoral student Jiang Shao, SCS and Chemistry & Biochemistry, discusses his studies on how DNA repairing enzymes find their target. Right: Master's student Muriel Hannion, SCS and Geological Sciences, who graduated just after the X-po, explains her work on flow dynamics in salt marsh channels to SCS Professor Gavin Naylor.

Doctoral Student Yixin Shou Wins Prize

Doctoral student Yixin Shou received a best paper award at the ACM (Association of Computing Machinery) South East regional conference in March 2006 for her paper, which was selected from 100 accepted papers. Every year computer scientists and engineers from academia, industry, and government agencies come together to attend the ACM Southeast Conference. The ACM Southeast Conference is believed to be one of the oldest continuously held computer science conferences in the nation. Yixin is working in Robert Van Engelen's research group, and specializes in high-performance compiler optimizations. Co-authors were Professors Robert Van Engelen and Kyle Gallivan at SCS, and Yixin's fellow graduate student Johnnie Birch.

Yixin Shou and Johnnie Birch at the SCS student X-po where they presented their work to an FSU audience.

SCS — School of Computational Science

The mission of SCS is to be the focal point of computational science at Florida State University. The school supports and develops a variety of high performance computing facilities, accessible to the university community. SCS is designed to overlap with existing departments and schools to provide a venue for interaction among faculty and students across many disciplines.

Please visit our website at www.scs.fsu.edu.

SCS
Dirac Science Library
Florida State University
Tallahassee, FL 32306-4120
Telephone: 850 644-1010
Fax: 850 644-1593

Director: Dr. Max Gunzburger
850 644-7024

Editor: Eva Ronquist
850 644-0196
evaron@scs.fsu.edu

Newsletters from SCS are issued three times a year. Free subscriptions and single copies can be ordered from the editor.