"Computational science" involves the innovative and essential use of high-performance computation, and/or the development of high-performance computational technologies, to advance knowledge or capabilities in a scientific or engineering discipline. Please describe (in no more than 300 words) your specific research interest paying particular attention to how computational science will spur advances.

High-performance computing - membrane fluidity and atomic details – applications in key processes

Certain questions in biology are very difficult to answer through experimental means, either due of the short timescales involved or because the details of the system are too small to be adequately resolved. For my research on the interaction between membranes and the proteins associated with them, both the fast dynamics and the geometry of the problem hamper traditional experimental approaches in answering fundamental questions. In order to get around these limitations we use computational approaches, in particular all-atom molecular dynamics simulation, as our microscope to answer questions of biological relevance.

Our current research focus is on characterizing the conformational changes that peripheral proteins undergo in response to binding and insertion into biological membranes, key processes that often control the activity of the protein. Peripheral proteins are known to undergo conformational changes of various nature and magnitudes upon membrane binding, ranging from localized side chain reorientation at the membrane interface to wholesale reorganization of the protein, such as pore formation by certain viral proteins in the membrane of the host cell. The details of these changes, which are mostly induced by atomic-scale interactions between the protein and specific lipid molecules are however almost completely unknown. Designing experiments to probe aspects of such behaviors is difficult, as simple physical limitations prohibit enough photons from being collected in a sufficiently short timeframe. By contrast, simulating the dynamics of these membrane-protein systems using a high-performance computing environment offers a powerful approach in characterizing the essential interactions behind the observed phenomena at an atomic resolution. These witnessed interactions can then serve as a guide for further research, such as offering suggestions for mutations that modulate the interaction in a certain way, or modeling changes in the chemical environment that would disturb the interaction. These insights provided by simulations will further our understanding of the fundamental process of membrane-controlled protein activation in the years ahead.

The fellowship program of study requirement is designed to give you a breadth of competency in fields outside your own that will enhance your ability to perform computational science research. Please describe (in no more than 300 words) how you expect that the courses listed in your planned program of study outside your chosen discipline will contribute to your own research in the future. Describe why you chose these courses and how they will impact your research plans.

Undergraduate courses and research in mathematics and computer science introduced me to several challenges presented by tackling biological questions computationally. It taught me that numbers in a computer should not be too big or too small, subtraction of similarly sized numbers should be avoided, and that there are many ways of taking integrals and derivatives numerically. With this basis to build on, I can focus my attention towards classes that will help the most in working on molecular simulation and visualization.

Biological systems are large, and molecular dynamics simulations are generally run in parallel. While I have had a brief introduction to the computational issues these environments provide in introductory programming classes, I feel that CS 420 will provide useful insights regarding the scalability of my simulations.

Successful analysis of the results of molecular dynamics simulations performed on complex biological molecules also relies heavily on efficient visualization techniques: you see where the atoms are and what interactions are taking place at any given time. However, this is not the most useful representation of the data in a global sense. I hope to learn from CS 519 techniques for effectively conveying the information contained within a simulation in a way more effective than distilling the N dimensions of information down to a measly two-dimensional plot. However two-dimensional plots do have their uses, and STAT 428 is a course that should greatly improve the quality of such plots by introducing me a wide variety of concepts in statistics in a practical way.

While taking PHYS 550, I was introduced to the notion of treating biological and physical phenomena as Markov processes. While this assumption has delivered many fruitful models, I was never properly exposed to the mathematical formalism, an error I hope to correct by taking MATH 564.

The goal of this program is to support doctoral students in pursuit of novel scientific or engineering discoveries through the use of high performance computing resources. There are many reasons to consider migrating a simulation to, or hosting large data sets in a high-performance computing environment. Some common motivations discussed in the DOE report "A Science-based Case for Large-scale Simulation"("SCaLeS", vol. 1, 2003; vol. 2, 2004; http://www.pnl.gov/scales/) are:

1. Better resolve the full, natural range of length or time scales in a model.
2. Accommodate physical effects with greater fidelity.
3. Allow the model degrees of freedom in all relevant dimensions.
4. Better isolate artificial boundary conditions or better approach realistic levels of dilution.
5. Solve an inverse problem, or perform data assimilation.
6. Perform optimization or control.
7. Quantify uncertainty.
8. Improve statistical estimates.
9. Operate without models.

For more information on these topics see the SCaLeS report cited above. By explicitly discussing one or two of the topics above or a different one of your own choosing describe in 300 words or less how migrating to a high performance computing environment would advance your research beyond what is possible with a modest sized cluster.

Simulating a biological process can be done at different levels of detail, ranging from the exquisite detail in quantum-mechanical treatments to detailed all-atom molecular dynamics simulations to the far fuzzier coarse-grained simulations. Selecting the appropriate level of detail generally is a balance between capturing the essential details of the system while not overtaxing the computational power available. For researching membrane-protein interaction, all-atom simulations are ideal, as they capture the essential dynamics of the system. In principle, a single all-atom simulation would require nothing more than a single processor, but a single processor could not simulate both the scope (hundreds of thousands to millions of atoms) as well as the timescale (nanoseconds-seconds) of the process. To counter this limitation, some groups remove details from the simulation, e.g., by using an implicit solvent or simplifying the system by removing degrees of freedom. This is unsatisfying, as now instead of simulating the *in vivo* conditions, one has to depend on the correctness of a model whose veracity is unknown. The personally preferred alternative is to scale the computing power used to the problem at hand and continue with the detailed approach. In practice, the scale of computing power that this requires is much greater than what is available on a single lab cluster, where 2 nanoseconds can easily take a day for a system of around 200,000 atoms. Only when the transition is made to a larger super-computing center is it possible to reach the timescales where the most interesting interactions take place without having to use an overly-simplified model.

Application

Name: Joshua Vermaas

Program of Study

Listed are the courses in science and engineering, applied mathematics, and computer science that you agreed to take on your proposed Program of Study.

University: University of Illinois at Urbana-Champaign

Course number	Course Title	Credit hours	Term and Year	Grade	Academic Level
Science/Engineering					
BIOP550	Biomolecular Physics	4S	Spring 2011		G
PHYS554	Nonequilibrium Stat Mechanics	4S	Fall 2010	A	G
Mathematics and Statistics					
MATH564	Applied Stochastic Processes	4S	Fall 2011		G
STAT428	Statistical Computing	4S	Spring 2012		B
Computer Science					
CS420	Parallel Programming: Science & Engineering	4S	Fall 2011		B
CS519	Scientific Visualization	4S	Spring 2012		G

I have read this program of study and affirm that, in my opinion, it satisfies the fellowship program requirements. This POS has been approved by my advisor, **Emad Tajkhorshid**, and I understand that, if offered a fellowship, my advisor and I are required to sign this page and send it to the Krell Institute.

Student's signature _____ Date _____

Graduate Advisor: **Emad Tajkhorshid**

Graduate Advisor's Institute: **University of Illinois at Urbana-Champaign**

Graduate Advisor signature _____ Date _____

Krell Institute (Office use only) _____

DOE Computational Science Graduate Fellowship Program
Krell Institute
Attn: DOE CSGF Coordinator
1609 Golden Aspen Drive, Suite 101
Ames, IA 50010
Phone: 515-956-3696
Fax: 515-956-3699
csgf@krellinst.org

Course Description

BIOP550: Biomolecular Physics

Physical concepts governing the structure and function of biological macromolecules; general properties, spatial structure, energy levels, dynamics and functions, and relation to other complex physical systems such as glasses; recent research in biomolecular physics; physical techniques and concepts from theoretical physics emphasized.

PHYS554: Nonequilibrium Stat Mechanics

Introduction to the mathematical description of classical and quantum stochastic systems, thoroughly addressing the tools and the mode of thinking of non-equilibrium statistical mechanics. Review of equilibrium statistical mechanics; Einstein and Smoluchowski diffusion equation; generalized moment expansion of correlation functions; noise-induced limit cycles; time series analysis; diffusion-controlled reactions; classical dynamics under the influence of stochastic forces; observables connected with brownian transport, echoes, and hysteresis; spin-boson model. Examples from biological physics and theoretical condensed matter physics.

MATH564: Applied Stochastic Processes

Introduction to topics such as spectral analysis, filtering theory, and prediction theory of stationary processes; Markov chains and Markov processes.

STAT428: Statistical Computing

Examines statistical packages, numerical analysis for linear and nonlinear models, graphics, and random number generation and Monte Carlo methods.

CS420: Parallel Programming: Science & Engineering

Fundamental issues in design and development of parallel programs for various types of parallel computers. Various programming models according to both machine type and application area. Cost models, debugging, and performance evaluation of parallel programs with actual application examples.

CS519: Scientific Visualization

Visualization techniques useful in analysis of engineering and scientific data. Physical models; methods of computational science; two- and three-dimensional data types; visual representation schemes for scalar, vector, and tensor data; isosurface and volume visualization methods; visual monitoring; interactive steering.