Vibrations may be Key to Protein Functions

P.K. Agarwal,* Oak Ridge National Laboratory

Computational methods have long been used to extend the reach of experimental biology through data analysis and interpretation. But the real power lies in biomolecular simulations that explore areas of research which are impossible to reach through experiments. One area where simulations are beginning to make an impact is how biologists think about the function of proteins.

Proteins hold the key to understanding the molecular basis of life. Previously, protein complexes have been viewed as static entities, with biological function understood in terms of direct interactions between components. Based on computational simulations, a new view is now emerging, where proteins are considered as very efficient molecular machines which are dynamically active and where internal protein dynamics are closely associated with their structure and function. This emerging view has broad implications for protein engineering and drug design.

At ORNL, Dr. Pratul K. Agarwal has been investigating the dynamics of protein-protein and protein-DNA complexes at multiple time scales. Using biomolecular simulations and advanced visualization techniques, he has been able to identify a network of protein vibrations in the enzyme *cyclophilin A*. The discovery of this network is based on investigation of protein dynamics at picosecond to microsecond-millisecond time scales. This network plays a vital role in the function of this protein as an enzyme. *Cyclophilin A* is involved in many biological reactions including protein folding, intracellular protein transport, and is required for the infectious activity of HIV-1. The modeling of *cyclophilin A* took more than 2 months of time on the ORNL's Cheetah supercomputer. Currently work is in progress to harness the power of next generation supercomputers to better understand protein dynamics.

Pratul Agarwal (left) and Stewart Dickson (right) discuss the impact of protein vibrations in the enzyme Cyclophilin A (background). The molecule is displayed in ORNL's visualization facility, "EVEREST" (Exploratory Visualization Environment for REsearch in Science and Technology). Visualization of the protein dynamics at the molecular level is revealing a new aspect of protein function -- a network of protein vibrations which plays a crucial role in understanding the function of these tiny molecular machines.

Dr Agarwal is working on parallelization and optimization of molecular dynamics (MD) code for supercomputers. Parallelization of MD codes is of wide interest to biological community, because with the current computational resources MD modeling falls short of simulating biologically relevant time scale by several orders of magnitude. The ratio of desired and simulated time scale is somewhere between 100,000-1,000,000. In the past, developers of the AMBER code were successful in achieving good speed on Cray T3E with up to 256 processors.[1] This enabled 1 microsecond MD simulation of a small protein in water (12,000

* 865-574-7184, agarwalpk@ornl.gov

atoms), over a period of about 3 months. Today's biological systems of interest consist of millions of atoms, which will require substantially more computing power of hundreds to thousands of processors for extended periods of time.

Dr Agarwal, in collaboration with Jim Maltby of Cray Inc., has ported the popular MD code AMBER onto the Cray X1 supercomputer. Currently, work is underway to optimize AMBER for parallel execution on several hundreds of processors by evaluating each loop for vectorization in the commonly used Particle Mesh Ewald (PME) subroutines. The initial benchmarking results on 128 processors of X1 show good speed up, with vectorization reducing the run time significantly. These benchmarking tests also indicated that speed up continues to improve for higher number of processors as number of atoms in the simulation grows. The most time consuming part of MD runs is the calculation of the non-bonded interactions, therefore, other non-bonded interactions subroutines (Generalized Born) are also being optimized for vector processors of X1.
NAMD is another popular parallel MD program developed by the Theoretical and Computational Biophysics Group at University of Illinois at Urbana-Champaign. Work is also under progress to port NAMD to X1 and optimize it for longer simulation periods.

Dr Agarwal is also developing a new algorithm for efficiently searching the conformation space of a protein for local and global minima. This method is designed to be high-performance and high-throughput. Combined with parallel programming strategies and the IBM's Blue Gene cellular architecture this method will search the minimum energy conformation of a protein very rapidly. This project is currently at development and initial application stage.

The enzyme human cyclophilin A (CypA) is involved in many biological reactions including protein folding and intracellular protein transport.[2] It is a ubiquitously expressed cytosolic protein.
CypA belongs to a class of enzyme called prolyl-peptidyl cis/trans isomerase (PPIase), because it catalyzes the isomerization of peptidyl-prolyl amide bonds that are N-terminal to proline residues in a wide variety of peptides and protein substrates.[3] Recent nuclear magnetic resonance (NMR) relaxation experiments indicate that protein dynamics is linked to PPIase activity of CypA.[4] These experiments detected motions for the backbone of several amino acid residues (including catalytically important Arg55, and Asn102 of the active site) only during substrate turnover.

Dr Agarwal's pioneering research on biomolecular modeling of CypA is based on performing a series of computational studies designed to investigate the role of dynamics of various protein parts in the detailed mechanism of cis/trans isomerization catalyzed by human CypA. Computational modeling of PPIase activity of CypA has discovered a network of protein vibrations promoting catalysis and provided insights into the detailed reaction mechanism.[5,6] The results support the structural studies,[7] which proposed a PPIase reaction pathway requiring a minimum deviation from the ground state. Dr Agarwal's research has also defined new ways to identify important interactions at the binding interface of a protein-protein complex, based on visualization of protein-protein interaction energy map (PPIEM) and dynamical cross-correlation map (DCCM).

[1] http://www.psc.edu/science/kollman98.html
[2] Gothel, S. F.; Marahiel, M. A. (1999) Cell. Mol. Life Sci. 55:423-436
[3] Fischer, G. (2000) Chem. Soc. Rev. 29:119-127
[4] Eisenmesser, E. Z.; Bosco, D. A.; Akke, M; Kern, D. (2002) Science
295:1520-1523
[5] Agarwal, P. K.; Geist, A.; Gorin, A. (2004) Biochemistry, (2004) 43, 10605-10618.
[6] Agarwal, P.K. Proteins: Structure, Functions and Bioinformatics (2004) 56, 449-463.
[7] Howard, B. R.; Vajdos, F. F.; Li, S.; Sundquist, W. I.; Hill, C. P. Nature Struc Biol
(2003) 10:475-481.

For further information on this subject contact:
Dr. Fred Johnson, Program Manager
Mathematical, Information, and Computational Sciences Division
Office of Advanced Scientific Computing Research
Phone: 301-903-3601
fjohnson@er.doe.gov