thesis of steroids and lipids, xenobiotic degredation, and procarcinogen activation. A better understanding of their function is thus desirable from a pharmacological point of view. The stereo specific hydroxylations they perform also make them interesting subjects for protein engineering. Here we present results of a recently developed method for investigating the mechanisms of substrate egress from a buried active site.

Crystallographic structures of one mammalian and four bacterial P450s have been determined (Williams et al., 2000, Molecular Cell 5:121-131, Hasemann et al., 1995, Structure 3:41-62). For P450BM-3, there is crystallographic and simulation evidence for opening of a substrate access channel, whilst for P450eryF such conformational changes have not been observed either in different crystal structures or by standard molecular dynamics methods.

Here, an artificial expulsion force on the substrate is applied in addition to the standard molecular dynamics force field. This allows the observation and investigation of expulsion events on much smaller time scales than they would normally be observed. This random expulsion MD method (CREMD) has been tested in simulations of the substrate-bound P450BM-3 and applied to P450cam and P450eryF. A common exit pathway is identified for all three proteins. However, different mechanisms of substrate exit are observed; these are due to different substrate size and flexibility, protein substrate interactions and also protein flexibility in the exit channel regions.

CAREER COLUMN

This essay is the first of an occasional career column to be included in News & Views. Our first column is an essay by Klaus Schulten, based on his experience at the MGMS Society meeting in April, 2000.

To Students Contemplating Research in Biomolecular Modelling

Klaus Schulten

3147 Beckman Institute, University of Illinois, 405 N. Mathews, Urbana, IL 61801. kschulte@ks.uiuc.edu
www.ks.uiuc.edu/~kschulte/

Students with good backgrounds in physics, mathematics, and computing, with a strong motivation to discover the molecular organization of life, and the interest to work in a team of physicists, chemists, biologists, and computer scientists can find great opportunities in biomolecular modelling. The intellectual mastering of life at the level of biomolecules and their assemblies is a main research goal of modern biophysics. In pursuing this goal, the field follows the example of physics, a discipline that has been so successful in carrying the torch of mathematics into the natural sciences.

Many challenges for theoretical and computational work exist in modern biology. Most pressing is the need for concepts and algorithms to handle and analyze the rapidly increasing genetic databases. The complete genomes of many biological species have recently become available and many more will become available soon. For many proteins, variants for numerous biological species are known with rich and still unearthed information underlying the conservation and variability of amino acids. The problem of predicting the structure of proteins from genetic sequences, the so-called protein folding problem, begs a solution with immeasurable opportunities once a solution is at hand. A further challenge is the structure-function relationship of proteins, a problem with extremely wide scope and characterized both through universality and diversity. Nearly half a decade after the first discovery of the atomic structure of a protein, scientists have not yet unveiled the mechanisms by which proteins achieve their many functions: to catalyze reactions, to receive and generate signals, to endow cells with shape and elasticity, to move cells and their internal cargo, and to control genetic expression. The complexity of proteins still demands the utmost respect from theoretical scientists.

Much of the present research focuses on proteins. Theoretical studies of proteins proliferated with the advent of sufficiently powerful computers to simulate large particle systems and with the explosive increase of atomic resolution structures of proteins. The latter structures, though necessary prerequisites, in themselves are not sufficient for any physical theory of protein function; the motions in a protein play an equal role. On the methodological side, the field is concerned with providing accurate, yet simple, force fields that govern the atomic motion of proteins. Ultimately, force fields will be determined in combined classical (for the nuclear motion)/quantum chemical (for the valence electrons) calculations.

Researchers are investing currently strong efforts in developing efficient computational methods for classical dynamics of proteins involving tens to hundred thousands of atoms; a serious hurdle, for example, is the calculation of Coulomb forces since they need to be evaluated for all pairs of atoms for all time steps of the classical motion. Suitable integration schemes can economize the costly update of forces, in particular the Coulomb forces, with a resulting boost in computational efficiency. At present a practitioner of the theory of proteins needs to be extremely competent in scientific computing with an understanding of massively parallel computing holding a particular promise for further success.

An entry into protein dynamics is provided by studying normal modes. Since the classical Hamiltonian describing atomic motion is significantly non-harmonic and extremely heterogeneous, conventional normal mode analysis as applied, e.g., for crystals, is not suitable. A quasi-harmonic description derives normal modes from a principal component analysis, i.e., from a diagonalization of the covariance matrix of all atomic positions, averaged over time. However, a gliding average with a, say, 100 ps window, reveals that modes derived in such a way vary in time due to significant conformational transitions and disorder in proteins. Protein motion also needs to be characterized on spatial scales involving multi-atom segments of proteins. In fact, many proteins exhibit conformational changes that can be described as rotations of segments around hinges or as motions of flaps formed by secondary structure elements, e.g., α-helices or loops between α-helices. Finally, the structure of proteins can undergo melting-type transitions when the environment changes, e.g., through binding of a charged substrate, a feature that is exploited by proteins involved in cellular signaling.

The abstraction of functional properties from molecular dynamics simulations remains an important challenge. Following established approaches, one can identify correlation functions and susceptibilities that provide essential characteristics of protein dynamics and can be related to observation and function. Examples are:

- A correlation function called the dynamic structure function is the Fourier transform of the motion of a protein's constituents and is observable through neutron diffraction or Mössbauer spectroscopy.
- The correlation function of the energy difference between two quantum states with diagonal coupling to the protein matrix accounts for the transition rate between the two states, e.g., for the rate of electron transfer.
- The dielectric susceptibility and thermal susceptibility, determined through monitoring dipolar or energy fluctuations account for dielectric properties and the degree of order of water associated with proteins.

The theory of proteins, as a relatively young field, can benefit tremendously from related and already established fields. The closest relative is the theory of liquids since solvent molecules, though not connected into a polymer and much more homogeneous in structure, are subject to similar forces and to similar disorder phenomena. On larger length and longer time scales, molecular hydrodynamics can provide much guidance to gain understanding of low frequency motion of protein segments encompassing many atoms. Condensed matter theory of disordered materials likewise deals with systems, e.g., glasses, of great conceptual similarity. Condensed matter theory can also serve as a reminder that the primary role of theory is not quantitative description, but rather qualitative understanding; anybody suspecting that not much useful can come of such a role should have a close look at the triumphs of condensed matter theory.

The beauty of theoretical protein science stems from its rapidly increasing treasure of new structures and functions; one could hardly imagine a science with more relevance to the existence and well-being of humanity, and with greater riches in new discoveries and new challenges.

The greatest such challenge considers the interplay of many proteins in biological cells. This interplay is the core attribute of living systems, one may even say that it answers the question of the origin of life. Living systems are all made of many molecular components that self-assemble, control each other, and self-replicate, and dealing with such systems has been a problem for scientists who were good at taking the molecular machines in living cells apart and

learning what they are made of. But how are these machines assembled? How can it be that proteins, describable by the laws of physics, assemble themselves into cellular machines and structures, these into complete living cells, and the latter into whole organisms that require a whole new language for their description, that of biology.

Where in this hierarchy of assemblies does the step from inorganic matter to living systems occur? Certainly, single cells must be considered alive, and half a cell does not remain alive. What is the simplest living cell made of?

Is the cell the smallest unit that typifies a living system? Or can one glean the key characteristics of a living system from its constituent parts? Would such attempt run counter to understanding life? Or could it provide a more gradual route to develop such understanding? The answer cannot be known beforehand, but going this path scientists need to guide their search and analysis by keeping in mind the eventual goal of understanding how life comes about through the self-organization of innate matter.

Are we aware of principles that govern such self-organization? We definitely can state three principles. First, the sustenance of the order of systems in living cells requires a constant consumption of energy. Living systems cannot be static and in keeping alive consume resources. It is important that this principle holds down to any level of order in cells. Second, self-organization is based on interactions of many components. Mathematically, this is reflected by the fact that nonlinear dynamical systems that account for interactions of components exhibit emergent properties in developing order from homogeneous initial conditions. Biologically, this is reflected in the many control components found in living cells: genes are controlled often through numerous other genes and molecules; the cell also has developed many signaling loops for the control of its metabolism and movement. Third, one needs to account for the empirical fact that living systems are robust against minor perturbations such that they seek to employ networks of interactions that allow operation under perturbations as opposed to using network types that function only in a narrow range of physical parameters.

The principles of self-organization have been established during the past decades, but the real being of cells is molecular and there is still a wide gap in our knowledge. We need to build bridges between the molecular level of cells and higher organizational forms. To see where such bridges can be built is the most difficult part in this endeavor.

Suitable bridges have been identified. An example is a cellular membrane in archaebacteria (purple membrane) that absorbs light and converts its energy into a membrane potential. The structure of this system is essentially known completely at atomic level detail, posing a great challenge to the next generation of researchers: can we take this structure and explain based on pure physics the biological function. Another example is the photosynthetic unit that exists in the cell membrane of certain photosynthetic bacteria and is more complex than the purple membrane. This apparatus contains six types of multi-protein complexes: the photosynthetic reaction center, the light harvesting complex 1, the light harvesting complex 2, a cytochrome, a bc1 complex, and an ATPase. All proteins are structurally known today so that the whole machine can be assembled and its assembly as well as synergistic function explained.

More examples that reflect the research opportunities available to students who choose a career in biomolecular modelling can be given, e.g., an explanation of the molecular factory in cells that manufacture proteins as instructed by genetic information, the ribosome, the structure of which is nearly solved.

UPCOMING MEETINGS

July 2000

The Annual Meeting of the American Crystallographic Association

July 22-27, 2000
RiverCentre
175 West Kellogg Blvd.
St. Paul, MN, USA
www.rivercentre.org

The ACA Transactions Symposium will be on using crystallography to understand biological mechanisms, encompassing molecules of all sizes. Two full-day workshops are planned for Saturday, July 22; SHELX for twins and macromolecular structures; QUEST; and one half-day workshop on Making Technical Presentations. Technical sessions will begin on Sunday, July 21, with hot new structures, refinement at ultra-high resolution, battery materials, and advances in small-angle scattering. Sessions continue through Thursday, with new science from neutron sources, protein-nucleic acid interactions, high throughput crystallization, service crystallography at synchrotrons, crystal engineering, science at long-length scale, nuclear industry materials, problem structure determination, and network glasses.

Contact

American Crystallographic Association
P.O. Box 96, Ellicott Station
Buffalo, NY 14205-0096 USA
Tel.: 716-856-9600 ext. 379
Fax: 716-852-4846
aca@hwi.buffalo.edu

4th World Multiconference on Systemics, Cybernetics, and Informatics

July 23-26, 2000
Sheraton World
Orlando, FL, USA

The purpose of the Conference is to bring together academic and professional leaders, consultants, scientists, engineers, theoreticians, and practitioners to discuss the areas of Systemics, Cybernetics, and Informatics (SCI). These three areas are being increasingly related to each other and to almost every scientific discipline and human activity.