Answering Fundamental Questions in Biology with Bioinformatics

Although biological advances depend on experiments, emerging theory-based disciplines also expand our understanding of life and its origins

Milton H. Saier, Jr.

For many decades molecular biology has been mainly an empirical science, greatly dependent on experiments and data analysis to provide insights into biological phenomena. However, genomics is causing a revolution, one that will dramatically change this picture once molecular biologists become more adept at deciphering the reams of data now being provided by these genomic sequencing efforts.

Consider Escherichia coli, perhaps the best-understood organism on earth. I maintain that we understand less than 1% of the information encoded within its genome, even though more than 50% of its gene products have been identified. To truly understand E. coli and scores of other organisms, we need better ways of extracting and analyzing genomics information. This immense deciphering task belongs to the realm of bioinformatics, and the even greater task of rendering that information intelligible to the human brain falls into the discipline of biosystematics.

How can we design better means for analyzing genomic data? First, any systematic approach to classifying biological entities needs to take evolution into account. Because molecular phylogeny reflects the evolutionary process, it provides a reliable guide to structure, function, mechanism, metabolism, and physiology. In short, phylogenetic analyses provide a rational approach for studying a plethora of biological phenomena and thus can help us to create disciplines that will allow seeming analytic impossibilities of today to become commonplace practices of tomorrow.

Bioinformatic Tools Applied to Macromolecular Evolution

During the past two decades, my colleagues and I have studied the origins and evolutionary histories of proteins. Our studies indicate that many different protein families, defined on the basis of sequence similarities, arose independently of one another at different times in evolutionary history, following different routes.

When sequence data for microbial genomes first became available, we adapted available software and also designed new programs for analyzing these data. This approach allowed us quickly to identify probable transmembrane proteins, estimate their topologies, and determine the likelihood that they function in transport, a topic of particular interest to our research group. This work allowed us to expand previously recognized families of such proteins and to identify dozens of new families.

We next attempted to design a comprehensive system for classifying the transport systems found in all living organisms. The classification system that we devised is based primarily on mode-of-transport and energy-coupling mechanisms, secondarily on molecular phylogeny, and lastly on substrate specificities of individual per-
meases within the protein families. Using this system, we have attempted to answer fundamental questions in biology and gain insight into the evolutionary origins of transporters found in vastly different organisms.

Extensive databases and websites are available describing the genomic analyses and transporter classification (TC) system that we developed (http://www-biology.ucsd.edu/~msaier/transport/). The Transporter Classification Database (TC-DB) provides (i) names, (ii) abbreviations, (iii) transporter classification numbers, (iv) descriptions of the families, (v) primary references, and (vi) representative well-characterized family members, including their names, substrate specificities, organismal sources, and database accession numbers. Search tools are available to facilitate use of the TC database (see http://tcdb.ucsd.edu/tcdb/). Interested readers are invited to browse these resources.

Using Bioinformatics To Address Fundamental Biological Questions

Combined bioinformatic and biosystematic approaches allow us to address several fundamental questions about transmembrane transport systems and to develop probable answers based on systematic phylogenetic analyses. This analytic approach could be applied to virtually any molecular biological subdiscipline.

- Question 1: Did integral membrane transport proteins arise as an independent protein class or from other types of proteins?

Among the more than 400 families of transport systems in our TC system, very few include integral membrane homologues that function in a capacity other than transport. Moreover, almost all these exceptions are receptors. In some cases, transporters gained receptor functions while retaining their transport functions, but in other cases they gained this function while losing the capacity to transport. A loss of transport capacity typically is accompanied by the gain or loss of specific protein domains; in a few cases, another protein with high affinity for the transport protein homologue accounts for the loss of function. Dissociation of such protein-protein complexes or losing the extra domains may restore the lost transport functions.

Integral membrane channel-type proteins typically consist of simple peptides with one, two, or three transmembrane segments (TMSs). However, secondary carriers that catalyze uniport, symport, or antiport are more complex and typically consist of larger multispansing polypeptide chains. In many cases, sequence analyses of the carriers reveal repeat units that are the sizes and topologies of known channel-forming peptides. For example, many transport carriers with 12 TMSs arose by successive intragenic duplications. Despite documented variations, we find no evidence suggesting a common origin for carriers and enzymes, leading us to believe that transport carriers evolved independently of enzymes and other protein functional types.

- Question 2: Without three-dimensional structural data, can independent origins be established for two families of transport systems having no sequence or motif similarities?

Many integral membrane transport proteins contain 6 or 12 TMSs. For example, both mito-
chondrial carrier (MC) family members and the aquaporins and glycerol facilitators within the major intrinsic protein (MIP) family contain six TMSs per polypeptide chain. Sequence analyses reveal that MIP family members arose by duplication of a three-TMS-encoding genetic element, while members of the MC family arose by triplication of a two-TMS-encoding element. Moreover, MC family members are found exclusively in eukaryotes, while MIP family members are distributed widely among both prokaryotes and eukaryotes.
Based on such findings, we conclude that the MIP family arose before the three domains of life (bacteria, archaea, and eukaryotes) diverged from each other, whereas the MC family arose late within eukaryotes, after endosymbiotic α-proteobacteria became permanent denizens of eukaryotic cells. The advent of mitochondria evidently required a new mode of communication between the mitochondrial matrix and the cytoplasmic compartment of such cells. Thus, mitochondrial carriers depend on a distinctive solute:solute exchange mechanism rather than on the cation symport mechanisms that bacteria use for ingesting nutrients.

- Question 3: Did bacteria, archaea, and eukaryotes exchange transporter genes appreciably during the past two billion years?

After analyzing dozens of large transport protein families, we find several examples of horizontal transfer between distinct kingdoms within these domains, such as between gram-positive and gram-negative bacteria. However, we find little evidence of horizontal transfer of transporter genes among the three domains occurring any time during the past two billion years. Thus, although hundreds of members of the MC family are found in eukaryotes, not a single such member is found in a prokaryote. Moreover, of the hundreds of sequenced homologues of the phosphoenolpyruvate:sugar phosphotransferase system (PTS), every one is in a bacterium, without a single example in an archaeon or a eukaryote.

If transporter gene transfer occurred early during the divergence of archaea and eukaryotes from bacteria, we could not recognize such an event on the basis of our phylogenetic analyses. Moreover, many proposed examples of lateral transfer between domains can more readily be explained by early gene duplication events that occurred prior to the Great Split. Selective transmission of some of these early homologues to various organismal types often looks like lateral gene transfers. Thus, although horizontal transfer of genetic material among the domains may have occurred very early during evolutionary time, such transfers were exceptionally infrequent during the past 2 billion years.

- Question 4: For multicomponent transport systems such as the ATP-binding-cassette (ABC) or complex protein secretion systems, did shuffling of protein constituents occur between systems?

After analyzing numerous multicomponent transport systems phylogenetically, we found little evidence for shuffling of protein constituents during their late evolutionary divergence. We included several protein secretion systems, such as types I, II, III, and IV as well as ABC-type solute uptake systems. The protein secretion systems consist of many proteins that were often
transferred laterally among gram-negative bacteria. Although several such multicomponent systems may be found within a single bacterial cell, the systems apparently did not exchange protein constituents, even though they can be exchanged experimentally by genetic manipulation. One important caveat: phylogenetic analyses may overlook constituent shuffling between closely related systems.

These observations suggest to us that complex multicomponent transport systems depend on extensive protein-protein interactions, which probably arose through coevolution of the protein constituents. Once any two systems have diverged appreciably in sequence, the constituents in one system no longer can interact properly with those in another, effectively preventing shuffling between the two.

● Question 5: For ancient families that arose before bacteria, archaea, and eu-karyotes diverged, have the proteins of a family generally acquired distinctive properties within each of these three kingdoms?

We examined proteins from 20 large ubiquitous families represented in all three domains to see if proteins from each domain exhibit distinctive characteristics. We found that the archaeal integral membrane proteins are consistently smaller than their bacterial homologues, while the eukaryotic homologues are much larger. Moreover, among transporters in the three major eukaryotic kingdoms of plants, animals, and fungi, the animal and fungal homologues are of comparable size, whereas the plant homologues are substantially smaller. Although these surprising observations presumably reflect evolutionary pressures during protein sequence divergence, we do not know what those pressures were.

● Question 6: In eukaryotes, are all members of transporter families confined to a particular organelle or to the plasma membrane, or are some family members found throughout cellular compartments?

Mitochondrial carriers are restricted to certain eukaryotic organelles. Not a single one appears to be a constituent of the plasma membrane, endoplasmic reticulum, Golgi apparatus, or nuclear membrane. Instead, they are found in a subset of organelles, including mitochondria, peroxisomes, amyloplasts, and hydrogenosomes. Are other families similarly restricted? Some of them, such as members of the cystine transporter family, are probably restricted to the intracellular vesicular membranes of eukaryotes, whereas other families can be found in the plasma membrane as well as in various organelles.

● Question 7: In bacterial pathogens, does multidrug resistance (MDR) arise by activation of stable genes encoding drug efflux pumps or by mutations of genes encoding other types of transporters?

MDR efflux pumps began causing clinical problems relatively recently, in parallel with the extensive use of antibiotics in medicine and as supplements in animal feeds. However, our analyses indicate that these MDR efflux pumps did not arise through recent mutations in genes encoding transporters that changed their substrate specificities.

Instead, such MDR pumps are encoded within the genomes of virtually all microorganisms, so these genes are present and thus need...
only to be activated to become problematic. Moreover, lateral transfer of genes among bacteria has occurred frequently, particularly for plasmid-encoded systems, suggesting that such genes can be acquired fairly readily even if they are not initially present. Finally, although mutations that enable transporters to act on different types of substrate are rare, experiments and phylogenetic analyses indicate that simple point mutations can readily narrow or broaden a particular transporter’s specificity toward a single class of compounds (e.g., sugars, amino acids, or drugs). These findings provide clues for developing strategies to control MDR among bacterial pathogens.

Prospects for Genomics-Based Predictive Biology

Bioinformatics provides a powerful tool for addressing evolutionary issues. Using simple computational approaches, we have addressed a series of fundamental questions about transport systems, and, surprisingly, the answers frequently prove contrary to those we expected based on intuition and our knowledge of scientific doctrine.

Some problems in biology are likely to yield to experimental approaches, while others will depend on alternative computational approaches. For example, we have answered evolutionary questions about biological macromolecules that could never have been addressed using classical molecular biological approaches.

Although this genomics-based approach to theoretical biology is only emerging, soon no major university will be without such theoreticians. We can anticipate that growing numbers of imaginative bioinformaticists will make valuable contributions and open new arenas. The possibilities are unlimited.

SUGGESTED READING


Bacterial Protein Toxins
EDITORS: Drusilla L. Burns, Joseph T. Barbieri, Barbara H. Iglewski, Rina Rappuoli

Designed for newcomers to the field of toxins, this important volume is intended to show how these proteins work while providing an up-to-date review of the field. Bacterial Protein Toxins describes all aspects of the biology of toxins, including their synthesis and secretion from the bacterial cell, their travels to and into the target host cell, and their modes of attacking the host cell machinery. It illustrates how bacterial toxins, each of which has distinct individual properties, often share mechanisms of secretion, membrane transport, and enzymatic action.

Condensed Contents
I. Genetics and Regulation
II. Toxin Biogenesis: Crossing Bacterial Membrane Barriers
III. Toxin Delivery into Eukaryotic Cells
IV. Toxin Action
V. Bacterial Toxins: Friends and Foes

List price: $115.95; ASM member price: $105.95

Adopters—request your free CD-ROM containing all art from the text. Please email requests to books@asmusa.org

ASM accepts VISA, Mastercard, American Express, EuroCard, purchase orders, and checks drawn on U.S. banks in U.S. currency. Canadian orders must include appropriate GST/HST.
WRITE TO: ASM Press, P. O. Box 605, Herndon, VA 20172, USA
CALL: 1-800-546-2416 or 703-661-1593
FAX: 703-661-1501 • WEB: www.asmpress.org