A complex system that works is invariably found to have evolved from a simple system that worked.
John Gall, *Systemantics* (1977)

STANDARDS

Toxicogenomics: Roadblocks and New Directions

The toxicogenomics research community may be "building a Tower of Babel" unless it devises ways to communicate and share data in a meaningful way among assorted research groups and across research platforms. Or so Bennett Van Houten, acting science advisor for the NIEHS Toxicogenomics Research Consortium, told a newly created National Research Council (NRC) panel that convened 6 February 2003 to discuss focused efforts to generate useful information in toxicogenomics.

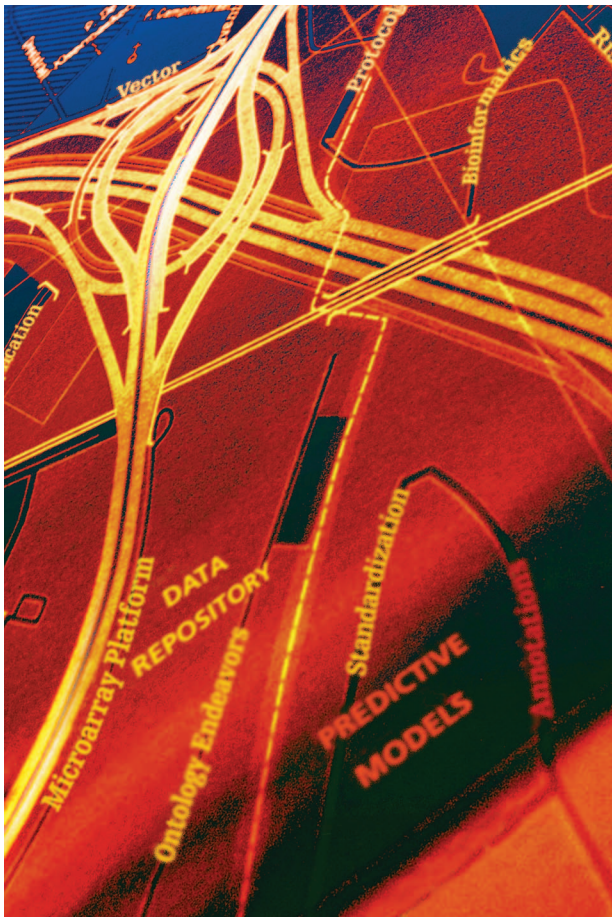
Established at the behest of NIEHS director Kenneth Olden and deputy director Samuel Wilson, the Committee on Emerging Issues and Data on Environmental Contaminants provides a public forum for communication among stakeholders about new evidence and concerns not just in toxicogenomics but also in environmental toxicology, risk assessment, exposure assessment, and related fields.

Standardization of experiments, vocabularies, and other activities within and across DNA microarray platforms is critical to toxicogenomics, said consortium coordinator Brenda Weis. "Currently, there are no standard protocols for toxicogenomics," she said, adding that gene annotation is one of the biggest challenges facing the field. For example, the pAC3 gene, a vector commonly used to clone DNA fragments, has been cited 59 different ways in assorted research papers. [For more on the topic of standardization, see "Data Explosion: Bringing Order to Chaos with Bioinformatics," p. A340 this issue.]

Although researchers must be trained in any new standards before they can be implemented, the effort will be worth it: such standards will ensure that results are credible, that full data sets and annotations are usable, that data from public repositories are accessible, and that data sets are permanently available, said speaker Chris Stoeckert, an associate professor of genetics at the Penn Center for Bioinformatics at the University of Pennsylvania. Projects such as Minimum Information About a Microarray Experiment, a workgroup of the Microarray Gene Expression Data Society (**http://www.mged.org/**), are already working on standardization.

Standardization poses many complex challenges, however. For example, microarray users must integrate such efforts with existing ontology endeavors. Standardizing protocols across a single platform even within an individual company can be difficult when labs are scattered across the country, said speaker Donna Mendrick, vice president and scientific director for toxicology at Gene Logic.

Ultimately, researchers may be required to provide images along with data to enhance quality control, Stoeckert said, although the microarray community is divided on this point, because the images are valuable but expensive to store and distribute. No decisions have yet been made as to whether the NIEHS consortium will reject data that don't meet whatever standards are eventually enacted, Van Houten reported.



The road ahead. Committees of scientists are working to chart a course for the progress of toxicogenomics.

Other issues spring from standardization problems. A federal liaison group organized by the NIEHS has identified four potential roadblocks to the optimal use of toxicogenomics, Wilson reported: premature use of information without strong scientific justification, communication of flawed interpretations of data by the scientific community, failure to educate stakeholders (including the public) on the new science, and failure to fill information gaps. "We need to ensure experiments are in line with guidelines [being developed by the toxicogenomics community] for evolution of the field," Wilson said. He discussed the need for a "roadmap" of the field so that progress can be evaluated against expected outcomes. Risk assessment–oriented evaluations of new chemicals and drugs are also a priority.

Challenges exist for virtually all conceivable applications of toxicogenomics technologies, be they industry efforts to build robust predictive models for specific agents across a single platform or academic endeavors to collect comprehensive amounts of data. There are substantial concerns about how toxicogenomics data are going to be used and shared, how proprietary databases will be managed, and how potential privacy issues will be handled.

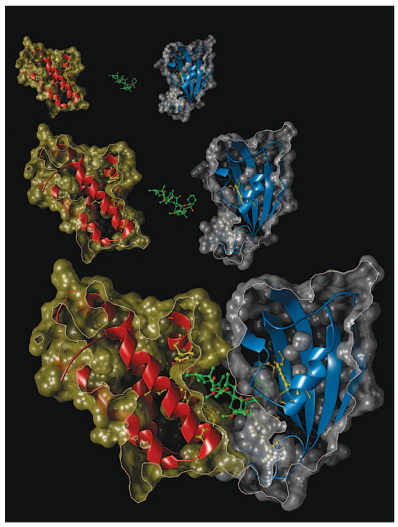
To address many of these issues, the NRC committee is evaluating three project proposals generated by committee working groups. These studies would examine the potential impacts and limitations of emerging technologies—genomics, proteomics, toxicogenomics, and bioinformatics—on risk assessment, environmental decision making, toxicology research, and public health, as well as whether current knowledge bases and tools fit the needs of scientific researchers and public health policy makers and workers. The standing committee itself will not conduct these studies but will recommend them for separate NRC approval and funding. **–Julie Wakefield**

PhotoDisc, Christopher G. Reuther/EHP

CANCER

Rapamycin Throws a Master Switch

Research on the potential anticancer drug rapamycin has revealed a possible new mechanism for suppressing large numbers of genes simultaneously, rather than each gene individually. Normally, genes are individually activated or inactivated by proteins targeted to the specific genes. But recent research led by principal investigator X. F. Steven Zheng, an assistant professor of pathology at the School of Medicine of Washington University in St. Louis, Missouri, shows that a protein named "target of rapamycin," or TOR, acts on many different genes simultaneously, producing a stress response that can stop cancer cells from reproducing. The study is published in the December 2002 issue of *Molecular Cell*.



Zheng and colleagues studied the molecular action of rapamycin, which currently is used to suppress the immune system after kidney transplant. The drug is derived from the soil bacterium *Streptomyces hygroscopicus*, native to the island of Rapa Nui (Easter Island). Rapamycin regulates a myriad of diverse cellular functions at the level of transcription and translation by inhibiting TOR. Clinical studies show that rapamycin also appears to both inhibit the formation of tumors and suppress tumor angiogenesis (the development of the blood vessels a tumor needs to obtain nutrients from its host), thus taking double-barreled aim at human cancers. These unique properties have led physicians to test its use as an anticancer drug.

In an approach known as "chemical genomics," Zheng and colleagues used rapamycin to inactivate TOR in yeast in a collection of mutant yeast strains, one for each gene in the yeast genome and each lacking one gene. This enabled them to measure how TOR interacts with each yeast gene, using rapamycin sensitivity (how much growth is inhibited by the drug) as a gauge.

Zheng and his team found about 300 yeast genes to be associated with TOR-related activities. The product of one such gene is a protein known as silent information regulator 3, or Sir3, which clings to the genes responsible for stress proteins, thereby inactivating them and keeping them silent. Sir3 appears to be the key to TOR's multigene activity: When rapamycin inactivated TOR, Sir3 molecules began detaching themselves from the chromatin regions carrying stress protein genes. This triggered a stress response; cells started producing stress proteins, their walls thickened, and they stopped proliferating. This is likely one of several mechanisms that contribute to shutting down cancerous cells. Michael McDaniel, a professor of pathology and immunology at the Washington University in St. Louis School of Medicine, calls it "a novel transcriptional mechanism that may further enhance the use of rapamycin as an anticancer agent."

Moreover, the researchers found that when rapamycin suppressed TOR, it also interrupted nutrient processing pathways, thereby preventing yeast cells from using glucose to produce energy and amino acids to make new proteins. Zheng and his team suggest that when rapamycin inhibits TOR, it works by eliciting a number of responses such as those of stress and starvation. Such responses are believed to cause cells to stop proliferating.

McDaniel notes that a key feature of rapamycin's overall mechanism of action—its ability to block cellular growth and proliferation—extends to normal, healthy cells as well as cancerous ones. He suggests that these potential adverse effects of rapamycin may be minimized by short-term use and the optimization of drug doses. He adds that Zheng's genomic study in yeast should provide a detailed map of the pathways by which the drug works, which will help in devising better therapeutic interventions for relevant human diseases.

The knowledge developed from the study of rapamycin's action needs to be verified in human cells, particularly tumor cells. If these results are borne out and the side effects are not intolerable, the result may be a new approach to causing malignant tumors to go into remission. If certain types of cancer are not fully cured, they might at least be upgraded from fatal diseases to manageable chronic diseases. **–Julian Josephson**

Origin of hope? New research shows that the immunosuppressant drug rapamycin, isolated from a bacterium native to Rapa Nui (Easter Island), acts as a link to bring together two immune system proteins. This association halts cell division, which may make the drug useful in the treatment of cancer.

PhotoDisc; inset: Richard Gilliland/Cornell Theory Center

MEETING REPORT

TXG at SOT

Since the first seminal paper on genomics appeared in the 20 October 1995 issue of *Science*, coverage has grown rapidly to some 3,000–4,000 reports yearly, many on gene–environment interactions, according to NIEHS deputy director Samuel Wilson. The influx of new data is revealing many surprises, many described at the 42nd annual meeting of the Society of Toxicology, held 9–13 March 2003 in Salt Lake City, Utah.

Just 15 years ago, researchers could study how toxicants alter only individual genes, and they spent years dissecting single genes. This approach was grossly inadequate, because "genes do not act in isolation, but by talking to other genes," says Kenneth Ramos, chairman of the Department of Biochemistry and Molecular Biology at Kentucky's University of Louisville School of Medicine and toxicogenomics editor of *EHP*. Today, DNA microarrays capture the expression of thousands of genes in response to environmental stressors.

Ramos and colleagues study molecular and genetic impacts of environmental contaminants—including the polycyclic hydrocarbon benzo[*a*]pyrene (BaP)—on heart disease and cancer. In a study of mouse vascular cells exposed to BaP, 1,383 of 9,000 genes were altered. Many affected genes regulate cell growth and differentiation; the big surprise, says Ramos, was that BaP also affects genes involved in immune modulation, such as those of the class I histocompatibility complex. "Our findings link immune cell activation as a key molecular event in the BaP-induced atherogenic response," he says. This fits well with *in vivo* observations that immune cells infiltrate the arterial wall in the early stages of atherosclerosis.

Leona Samson, a professor of toxicology at the Massachusetts Institute of Technology in Cambridge, described her work with *Saccharomyces cerevisiae*. She used *S. cerevisiae* mutants, each missing one gene, to identify genes that help cells recover from damage by the alkylating agent methylmethane sulfonate (MMS). Of the 6,000 yeast genes, about 400 are sensitive to MMS, including genes related to cell death and DNA repair. But most of the recovery genes Samson identified are involved with functions such as cytoskeleton remodeling, protein degradation, RNA synthesis, and lipid metabolism. The new data suggest that to recover from DNA damage and avoid cell death, "cells have a lot of other things to repair," says Samson. She and her colleagues are examining recovery pathways for other alkylating agents and ultraviolet light, and each shows a unique pattern. She says the goal is to "predict whether a cell, organism, or even person will recover from damage."

Yet another surprise came from research at the NIEHS National Center for Toxicogenomics (NCT), where the gene *Dss1*—previously associated only with a developmental abnormality of the hands and feet—was found to play a role in skin cancer. In mouse skin cells treated with phorbol ester tumor promoters, *Dss1* was expressed during early stages of tumor formation. "There is no previously recognized basis for predicting a role of *Dss1* in skin tumorigenesis," says NCT director Raymond Tennant.

The latest. New findings and a new *EHP* section on toxicogenomics were unveiled at SOT.

Scientists are uncovering many new genes and pathways never before imagined to be associated with gene–environment responses. A key next step is "phenotypic anchoring," connecting specific gene changes to markers of toxicity [see NCT Update, p. A338 this issue]. In a proof-of-concept experiment on phenotypic anchoring, researchers in Richard Paules's NCT lab monitored liver toxicity in rats induced by the drug methapyrilene. They verified that the expression of a group of genes corresponds to histological changes such as liver necrosis and periportal inflammation.

In addition to the data presented at the annual meeting, another source of information was unveiled—*EHP*'s new Toxicogenomics Section. This section will appear quarterly and will present news, research, and perspectives in toxicogenomics and related disciplines. **–Carol Potera**

BIOINFORMATICS

Putting Proteins in One Place

The growing wealth of information about the human proteome—the hundreds of thousands of proteins at work in the human body—is useful only if scientists can get their hands on it. To give researchers faster worldwide access to high-quality protein data, the National Human Genome Research Institute (NHGRI) and five other institutes and centers of the NIH have awarded $15 million to create a comprehensive, public data bank of protein sequences.

The United Protein Database, or UniProt, will combine three existing databases—Swiss-Prot, TrEMBL, and the Protein Information Resource (PIR). By the end of the three-year grant, UniProt should contain annotated entries on more than 2 million proteins, including information on protein sequences, functions, modifications, and other characteristics.

UniProt brings together scientists and resources that complement one another, says Peter Good, program director for genome informatics and computational biology at the NHGRI. The database will combine 830,000 entries from TrEMBL, 123,000 entries from Swiss-Prot, and 283,000 entries from PIR.

TrEMBL contains more entries because it is a computational database—computer programs use the protein sequences to make predictions of protein function. Swiss-Prot uses the more time-consuming hand-annotation method, which means that a scientist reads articles that mention a particular protein, extracts the relevant information, then adds it to the database. PIR, operated by Georgetown University Medical Center and the National Biomedical Research Foundation in Washington, D.C., contains both computer-annotated and hand-annotated entries. The PIR will cease to be updated, and its staff will assist with hand-annotating the TrEMBL records.

PIR will also contribute its "protein family" method of classification, which groups proteins by function based on sequence similarity. If two proteins fall into the same family, scientists can infer that the proteins may have similar functions. This method—created by one of the pioneers of protein sequence databases, Margaret Dayhoff—has been developed further by Cathy Wu, director of bioinformatics for PIR and one of the principal investigators of UniProt.

PhotoDisc, Arnold Greenwell/EHP, Matt Ray/EHP

Other UniProt principal investigators are Rolf Apweiler, who is head of the Sequence Database Group at the European Bioinformatics Institute, and Amos Bairoch, who is group leader of the Swiss-Prot Group at the Swiss Institute of Bioinformatics.

Any biomedical scientist interested in protein function will benefit from the new database, Good says. Drug discovery in particular involves pinpointing proteins that are altered in diseased tissue, then exploring these proteins further to determine if they will make good targets for new drugs.

A typical proteomics experiment uses mass spectrometry to identify proteins and parts of their sequences. "The way to add meaning to those sequences is to search databases, which contain links to essentially all human knowledge surrounding those protein targets," says Tim Haystead, an associate professor of pharmacology and cancer biology at Duke University in Durham, North Carolina, and founder of the drug discovery company Serenex. Right now, scientists have to search several different databases to find all existing information on a sequence. "A unified database will make it easier for us," Haystead says.

William Pearson, a professor of biochemistry and molecular genetics at the University of Virginia and a member of PIR's oversight and scientific advisory board, agrees that scientists who work with protein sequence data have been frustrated both by the need to search multiple databases and by the sometimes contradictory information arising from their provenance in different methods of annotation. "When these [databases] all get put together, they're going to have much more consistent ways of referencing data and giving names to things," he says. "It will be much more efficient."

Funded in October 2002, UniProt is still a work in progress. "Part of the challenge is getting three groups that have different [organizational] cultures to interact," Good says. For continuity, the three groups will maintain their current search interfaces, each of which will eventually access the entire UniProt database. UniProt will also be accessible via a central website, **http://www.uniprot.org/**. A basic version of that site will be up this year, according to Apweiler.

UniProt will be freely available to all, but not until after the expiration of a license with industry covering access to Swiss-Prot records by commercial users. The license for Swiss-Prot records allowed the European Bioinformatics Institute and the Swiss Institute of Bioinformatics to continue developing Swiss-Prot in the absence of government funding. With support from the NIH, the entire UniProt database will be available free of charge for both academic and commercial users by January 2005. **–Angela Spivey**

txgnet Bioinformatics Organization

Bioinformatics—the rapidly evolving science of managing and analyzing biological data using advanced computing programs—is a vital tool for evaluating the volumes of data generated by genomics research. Bioinformatics Organization is a nonprofit international group with more than 5,500 members that is working to promote the free and unrestricted exchange of bioinformatics resources and data among all scientists, including those with little funding or at small institutions, who may not be able to afford access to cutting-edge resources through the usual channels. As part of this mission, the group has established a forum for information and data exchange at **http://bioinformatics.org/**.

The Bioinformatics FAQ page includes a brief discussion of how the Human Genome Project has impacted bioinformatics and related fields such as computational biology and pharmacogenomics. The page also provides overviews of the technologies currently being used in bioinformatics, lists of books, links to university bioinformatics programs around the world, and portals to web directories and tutorials. This section also includes practical advice for performing a number of common bioinformatics functions and a glossary of commonly used terms.

Group members are currently involved in nearly 100 projects, which are listed under the Hosted Projects header on the homepage. Examples include ALiBio, a free online library of algorithms, and BioQuery, which allows visitors to search multiple biomedical databases simultaneously and automatically informs users when new data matching a search query become available. Multi-Genome Navigator, or MuGeN, is a software package that allows users to explore multiple annotated genomes simultaneously. And GUI Blast gives Windows users a graphical user interface for using Basic Local Alignment Search Tool (BLAST) software.

The Research Laboratory section of the site contains databases of, among other things, expressed sequence tag (EST) clusters. Users of the EST database can clusterize publicly available EST sequences and contigate them using a specific assembler known as zEST. The site's developers hope visitors will thus help build the repository of EST clusters, which can be used in the discovery of new genes, splice variants, and gene polymorphisms. Also available in this section are databases of immigrant genes, leukemia genes, and pancreatic cancer genes.

The main page lists current news items from sources including journals, newspapers, government agencies, and software developers. Archived news items dating back to 2000 provide a look back at the discipline's progress. Registered users can post items of interest to the rest of the bioinformatics community.

Since 2002, Bioinformatics Organization has awarded the annual Benjamin Franklin Award to an individual its members feel has "promoted freedom and openness in the field of bioinformatics." The 2003 award went to Jim Kent of the University of California, Santa Cruz, who used his own GigAssembler program to assemble the public fragments of the human genome before Celera Genomics was able to assemble their private human genome sequence. This helped keep these data in the public domain. **–Erin E. Dooley**