IN-51

After the Genome IV

Envisioning Biology in the Year 2010

The After the Genome meetings were started in 1995 to help the biological community think about and prepare for the changes in biological research in the face of the oncoming flow of genomic information. The term "After the Genome" refers to a future in which complete inventories of the gene products of entire organisms become available.

The organizers, Roger Brent, Susan Burgess, Thomas Meade, and Chris Sander, intend to distinguish this meeting by bringing together intellectuals from subject fields far outside of conventional biology with the expectation that this will focus researchers beyond the immediate future. Hence the subtitle for this year's meeting, "Envisioning Biology in the Year 2010".

Accordingly, the organizers are bringing together a broadly multi-disciplinary group of thinkers and working scientists. The participants include biological researchers and information workers as well as some visionary intellectuals and executives from the computer industry, high ranking officials from government agencies, reporters who chronicle biology with a long-term perspective, representatives of philanthropic organizations who are in a position to spend money flexibly to catalyze organizational change, and eminent intellectuals from other disciplines.

This year, the meeting will deal with four themes:

- new technologies relevant to post-genomic data collection in the first session chaired by Thomas J. Meade of the California Institute of Technology.
- their integration into deducing events at the level of the single in the second session chaired by Roger Brent of the Molecular Sciences Institute.
- unconventional informatic techniques for extracting meaning from this data in the third session chaired by Chris Sander of Millenium Pharmaceuticals.
- and an attempt to envision how these approaches will impact organismic biology, and more speculatively, ecology and evolution in the final session chaired by Susan Burgess of Structural Bioinformatics, Inc.

NAG2-6025

CASI

After the Genome IV
gratefully acknowledges the support
of our Sponsors

Academic Press

California Institute of Technology

Clinical Micro Sensors, Inc.

Entelos, Inc.

Genetics Institute

Millenium Informatics

Molecular Sciences Institute

Monsanto

National Cancer Institute

NASA Ames

NeXstar

Pharmacia & Upjohn

Perkin-Elmer/Genscope

Smith-Kline Beecham

Stratagene, Inc.

Structural Bioinformatics, Inc.

and acknowledges

Agile, Inc. (Ben Seibel) for our website

Saturday
10 October 1998

Plenary Session

4:00 PM	Registration	*Trappers Room*
5:00 PM	Social Hour (Registration Open)	*Explorers Room*
6:00 PM	Dinner Buffet	*Explorers Room*
	Saturday, October 10: Session I	*Explorers Room*
7:30 PM	Thomas J. Meade, California Institute of Technology Susan K. Burgess, Structural Bioinformatics, Inc. *Welcome.*	
7:45 PM	Roger Brent, Molecular Sciences Institute *Introduction to ATG IV.*	
8:00 PM	Dan Janzen, University of Pennsylvania *Biodiversity Development of a Large Tropical Conserved Wildland: Area de Conservacion Guanacaste.*	
8:45 PM	Stephen Grand, Cyberlife Technology *The Biology of an Alien Species: How Cells Make Souls.*	
9:30 PM	Reception	
10:00 PM	Discussion Session	

Sunday
11 October 1998

Cross-Disciplinary Technologies for Post-Genomics
Thomas J. Meade, Session Chair

7:00 - 8:25 AM	Breakfast Buffet	*Mural Dining Room*
8:00 - 8:25 AM	Registration for Late Arrivals	*Trappers Room*

	Sunday, October 11: Session 1	*Explorers Room*
8:30 AM	Opening Remarks by Session Chair Thomas J. Meade, California Institute of Technology	
8:45 AM	Jeff Byers, Naval Research Labs *Controlling Biosynthetic Pathways with Magnetoelectronics.*	
9:30 AM	Chad Mirken, Northwestern University *Ultraselective DNA Detection Methods Based Upon Novel Nanoparticle Probes.*	
10:15 AM	Allen Northrup, Cephiad, Inc. *Next Generation of Gene Amplification and Detection Technologies.*	

11:00 - 11:45	Break	

	Sunday, October 11: Session 2	*Explorers Room*
11:45 AM	Rob Carlson, Molecular Sciences Institute *Doing Biology with Microfabricated Microfluidic Devices.*	
12:30 PM	David Soane, Alnis, Inc. *Synthetic Polymer Complements and their Applications.*	

Sunday
11 October 1998
continued

Cross-Disciplinary Technologies for Post-Genomics
Thomas J. Meade, Session Chair

1:15 - 2:30 PM	Luncheon Deli Buffet	*Mural Dining Room*
2:30 - 5:00 PM	Afternoon Activities	
5:00 PM	Social Hour - Micro Brew Bar	*Explorers Patio*
6:00 PM	Western BBQ Dinner	*Explorers Room*
	Sunday, October 11: Session 3	*Explorers Room*
7:30 PM	Tom Tullius, Boston University *Making Large-Scale Maps of DNA Structure in Gene Control Regents.*	
8:15 PM	Raoul Kopleman, University of Michigan at Ann Arbor *Using Nanopebbles to Interrogate Living Cells.*	
9:00 PM	Posters and Reception	
10:00 PM	Discussion Sessions	

Monday
12 October 1998

Cellular Components and Pathway Analysis/Modeling
Roger Brent, Session Chair

7:00 - 8:25 AM Breakfast Buffet *Mural Dining Room*

Monday, October 12: Session 1 *Explorers Room*

8:00 AM Introduction by Session Chair
Roger Brent, Molecular Sciences Institute

8:15 AM Bill Hutchens, Ciphergen
Extracting Information from Biological Systems with Solid Phase Fractionation Time of Flight Mass Spectrometry.

9:00 AM Dennis Bray, University of Cambridge
Spatical Organization and the Logic of Signal Transduction Pathways.

9:45 AM Drew Endy, University of Wisconsin at Madison
Experimental Characterizations of Altered Genetic Element Order on the T7 Growth Cycle.

10:30 - 10:45 AM Short Break

Monday, October 12: Session 2 *Explorers Room*

10:45 AM John Weinstein, NIH, NCI
Information-intensive Drug Discovery: Genomics, Preteomics, and Bioinformatics.

11:30 PM Yehoshua Bruck, California Institute of Technology
Gene Regulation, Asynchronous Computing and Stochastic Competition.

Monday
12 October 1998
continued

Cellular Components and Pathway Analysis/Modeling
Roger Brent, Session Chair

12:15 PM	Boxed Lunches To Go	*Explorers Room*
	Depart for Yellowstone National Park Excursion	*Chartered Bus*
12:15 - 5:00 PM	Afternoon Tour of Yellowstone	*Yellowstone National Park*
5:00 PM	Yellowstone National Park Dinner	*Old Faithful Inn*
6:30 PM	Depart Old Faithful Inn	*Chartered Bus*
	Monday, October 12: Session 3	*Explorers Room*
8:00 PM	Katheryn Resing, University of Colorado *Cellular Function Information from Protein Mass Spectrometry.*	
8:45 PM	Guri Giaever, Stanford University *Drug-Induced Haploinsufficiency: A Genomic Approach to Drug Target Identification.*	
9:30 PM	Posters and Reception	
10:00 PM	Discussion Sessions	

Tuesday
13 October 1998

Syntax for Intracellular Signalling and Integrated Bioinformatics
Chris Sander, Session Chair

7:00 - 8:25 AM	Breakfast Buffet	*Mural Dining Room*

	Tuesday, October 13: Session 1	*Explorers Room*
8:30 AM	Introduction by Session Chair Chris Sander, Millenium Pharmaceuticals	
8:45 AM	Tom Patterson, Entelos, Inc. *Modelling Physiology at the Organ-system Level.*	
9:30 AM	Jeff Skolnick, Scripps University *Large Scale Structure Prediction.*	
10:15 - 11:00	Break	
	Tuesday, October 13: Session 2	*Explorers Room*
11:00 AM	Barry Honig, Columbia University *Structure-based Analysis of Sequence and Function.*	
11:45 AM	Jeff Saffer, Battelle-Pacific Northwest *Use of Context Vectors Analysis of Large Volumes of Biological Data.*	
12:30 - 1:45 PM	Fish Fry Luncheon	*Explorers Patio*

Tuesday
13 October 1998
continued

Syntax for Intracellular Signalling and Integrated Bioinformatics
Chris Sander, Session Chair

1:45 - 6:00 PM	Afternoon Activities	
6:00 PM	Dinner	*Mural Dining Room*
	Tuesday, October 13: Session 3	*Explorers Room*
7:30 PM	Anna Tsao, DARPA *Mathematical Approaches to Dealing with the Curse of Dimensionality in Department of Defense Signal Processing Applications.*	
8:15 PM	Tom Schneider, NIH, NCI *Molecular Information Theory: From Clinical Applications to Molecular Machine Efficiency.*	
9:00 PM	Posters and Reception	
10:00 PM	Discussion Sessions	

Wednesday
14 October 1998

Biodiversity/trans-species Inference Leading to an Enlightened Biology

Susan Burgess, Session Chair

7:00 - 8:25 AM	Breakfast Buffet	*Mural Dining Room*
	Wednesday, October 14: Session 1	*Explorers Room*
8:30 AM	Opening Remarks by Session Chair Susan K. Burgess, Structural Bioinformatics, Inc.	
8:45 AM	James Shapiro, University of Chicago *Natural Genetic Engingeering of Genome System Architecture in Evolution.*	
9:30 AM	Lee Eiden, NIH *Cross-species Pathway Comparison, Pharmacological Dissections.*	
10:15 AM	Greg Benford, University of California at Irvine *Fixing the Greenhouse.*	
11:00	Boxed Lunches Available	
11:00	Meeting Adjourned	

AFTER THE GENOME IV

Abstracts

for

Saturday,

10 October 1998

BIODIVERSITY DEVELOPMENT OF A LARGE TROPICAL CONSERVED WILDLAND: AREA DE CONSERVACION GUANACASTE.

D. Janzen and W. Hallwachs
University of Pennsylvania, Philadelphia, PA 19104
Djanzen@sas.upenn.edu whallwac@sas.upenn.edu

The only chance that tropical wildlands have of surviving into perpetuity is through integration with the human genome, through their being peacefully positioned somewhere within the three major human activities - sex, shelter and food. The Area de Conservacion Guanacaste in northwestern Costa Rica is a working pilot project aimed at this integration, otherwise known as Biodiversity Development or the gardenification of nature. We discuss the technical, sociological, economic and political aspects of such an act.

BIOLOGY OF AN ALIEN SPECIES: HOW CELLS MAKE SOULS.

S. Grand
Cyberlife Technology
Quayside Bridge Street, Cambridge, CB4 8AB UK
Stephen.grand@cyberlife.co.uk

In general, Biology at the macro level has been obliged to try and understand complex, ready-built organisms, for which no instruction manual exists. Genes and outward behavior can be correlated, but mechanisms by which the former lead to the latter are usually opaque, especially when the behavior is high level. An alternative approach is to play at being a god. Armed with enthusiasm and a small selection of basic, biologically plausible building blocks, can we synthesize a complete organism that exhibits high-level behaviors such as learning, courtship, and curiosity? Can such learning by putting things together offer insights that learning by taking things apart has so far failed to do?

Will the future creation of intelligent artificial life forms constitute a new branch of biotechnology? A useful discussion in the space of forty-five minutes is a tall order, so to avoid disappointment you might wish to know the answers in advance. They are: yes, perhaps, and probably.

AFTER THE GENOME IV

Abstracts

for

Sunday,

11 October 1998

CONTROLLING BIOSYNTHETIC PATHWAYS WITH MAGNETOELECTRONICS

J. M. Byers
Materials Physics, Code 6344, Naval Research Laboratory Washington, D.C. 20375
and Department of Physics, George Washington University, Washington, DC
Tel. (202) 767-6147 FAX (202) 767-1697 E-Mail byers@foucault.nrl.navy.mil

Rapid development of small-scale magnetic structures for use as sensors, nonvolatile memory elements and localized magnetic field sources has provided a new technology within the last ten years called magnetoelectronics. This new form of electronics is being adapted to more traditional silicon-based electronics in the semiconductor industry. Our aim is to adapt this technology for use in biological applications. Using magnetoelectronics to couple to biological systems is compelled by the ability of magnetism to function within a saline solution but not interfere with biochemical processes. The talk will show how magnetoelectronics has been used in biosensors and could control biosynthetic pathways via biologically functionalized magnetic nanoparticles. The ultimate goal is explore how to create a kind of biomolecular electronics that could exist in a layered IC geometry or 'chip' package without the need for optical coupling but instead a magnetoelectronic interface. The role of genomics in creating biomolecular electronics will also be addressed.

This work is supported by the Office of Naval Research (ONR) and the Defense Advanced Research Projects Agency (DARPA).

ULTRASELECTIVE DNA DETECTION METHODS BASED UPON NOVEL NANOPARTICLE PROBES.

C. A. Mirkin
Northwestern University, Department of Chemistry, 2145 Sheridan Road, Evanston, IL 60208
Camirkin@chem.nwu.edu

New detection methods for DNA and RNA, which are based upon novel oligonucleotide functionalized nanoparticles, will be described. These methods rely on the difference in properties between dispersed particles and those assembled into extended periodic materials by hybridization with complementary target molecules. The extraordinary selectivity exhibited by one such system based upon Au nanoparticles will be described. This system allows one to colorimetrically detect a DNA sequence with near perfect selectivity at femtomole target levels. The origin of this selectivity and relationship to the novel approach of using nanoparticle-based materials for DNA detection will be discussed. The prospect of using these systems for high selectivity and sensitivity point-of-site assays will be addressed.

NEXT GENERATION OF GENE AMPLIFICATION AND DETECTION TECHNOLOGIES.

M. A. Northrup
Cepheid, 1190 Borregas Avenue, Sunnyvale, CA 94089-1302
Northrup@cepheid.com

Future nucleic-acid-based diagnostic instruments need improvement over the current state-of-the-art. Increasing the speed and sensitivity of the assays, while reducing costs are clear goals. Recently, it has become possible to microminiaturize fluidic and sensing components using micromachining and precision injection molding. There has been a large amount of interest and effort in the area of miniaturization of such systems, yet not all of the properties of fluidics and sensing methods improve as they are drastically reduced in size. It is clear that implementing miniaturized diagnostic instruments is not a matter of simply "shrinking" their conventional counterparts, nor of automating existing manual procedures. What is required to harness the full potential of scaling technologies is the use of design methods that take into account scaling effects and the development of completely new processing approaches. In particular, we will discuss sample preparation and detection approaches in this context.

DOING BIOLOGY WITH MICROFABRICATED MICROFLUIDIC DEVICES.

R. Carlson
The Molecular Sciences Institute, 2168 Shattuck Avenue, Berkeley, CA 94704
Rcarlson@sequence.molsci.org

As biology proceeds beyond sequencing genomes to exploring phenotypic implications of expression and engineering DNA, it will be useful to physically manipulate single cells and their contents. This style of measurement and experiment requires tools on the same length scale as the cells themselves. Towards this end, several decades worth of experience in building integrated circuits can be utilized to control geometries at the micron level, providing for conception and construction of a new microfluidic toolbox. Selecting a single cell from a population and probing its biochemical state or genome are ideal applications of microfabrication.

SYNTHETIC POLYMER COMPLEMENTS AND THEIR APPLICATIONS.

D. Soane
Alnis, LLC. 1933 Davis Avenue, Suite 266, San Leandro, CA 94577
soane@ix.netcom.com

A new approach has been developed to synthesize molecular complements that precisely map the size, shape, and surface characteristics of selected molecules or entities (the targets). Such synthetic polymer complements (SPCs) are unique in that they possess molecular dimensions, thus offering several opportunities worthy of detailed scientific exploration. Alnis has studied the potential of SPCs for enzyme stabilization in non-aqueous and elevated-temperature environments via molecular scaffolding. Specially prepared SPCs may be used to recognize and sequester contaminants or infectious agents from mixed fluids. SPC-target conjugates also hold the promise for the delivery of therapeutic agents that are intrinsically fragile or difficult to uptake. Finally, the SPCs themselves may exhibit biological and pharmacological activities as a result of their affinity toward selected targets.

MAKING LARGE-SCALE MAPS OF DNA STRUCTURE IN GENE CONTROL REGENTS.

T. Tullius
Boston University, Department of Chemistry, 590 Commonwealth Avenue,
Boston, MA 02215

USING NANOPEBBLES TO INTERROGATE LIVING CELLS.

R. Kopelman, M. Philbert, H. Clark, and M. Hoyer
University of Michigan, Departments of Chemistry and Toxicology,
Ann Arbor, MI 48109-1055
Kopelman@umich.edu

Traditional investigations into the role of physiologic influences on normal function and disease have been superseded by "cracking the genetic code." Since then, quantal leaps in understanding of genetic mechanisms of disease, nucleic acid interactions in transcription/translation, and the plethora of signal transduction mechanisms which initiate these processes have moved to the forefront of modern biological research. However, molecular biological research has always had as its backdrop traditional physiology, electrophysiology and more recently patch clamping and fluorescent microscopic approaches. Probes Encapsulated By Biologically-Localized Embedding (PEBBLEs) and Nano Opto-Chemical Systems (NOCS) constitute the next generation of nano-physiological tools which combine applied physics and photochemistry with molecular biology and optical fluorescence microscopy. These new tools may be used to gather targeted three-dimensional physiologic data in vivo, in real time. PEBBLEs have proven useful in the detection of minute shifts of pH, nitric oxide and calcium concentrations in compartments as small as the phagosome of a macrophage. The incorporation of molecular recognition into NOCS hold the promise of providing precise targeting of these sensors to specific organelles, cytoskeletal elements or membranous boundaries of the cell. Combinations of fluorescent indicators which emit in the visible spectrum in conjunction with reporters of enzymatic action in NOCS (e.g., green fluoresent protein, luciferase, etc.) will provide novel methods for precise nano-localization of cellular critical molecular and ionic processes in the maintenance of normal cellular funciton or the initiation and progression of disease.

AFTER THE GENOME IV

Abstracts

for

Monday,

12 October 1998

EXTRACTING INFORMATION FROM BIOLOGICAL SYSTEMS WITH SOLID PHASE FRACTIONATION TIME OF FLIGHT MASS SPECTROMETRY.

W. Hutchens
Ciphergen Biosystems, 490 San Antonio Road, Palo Alto, CA 94306

SPATIAL ORGANIZATION AND THE LOGIC OF SIGNAL TRANSDUCTION PATHWAYS

D. Bray and R. Bourret
Zoology, University of Cambridge CB2 3EJ UK
Microbiology and Immunology, University of North Carolina, Chapel Hill NC 27599.

We have been investigating the pathway of intracellular signals used by coliform bacteria in the detection of chemotacticstimuli. The function and formation of this pathway were examined by means of computer-based models based on physiological data collected from single tethered bacteria of over 60 mutant genotypes. Both deterministic and individual-based stochastic programs have been used. Quantitative discrepancies between computer models and experimental data throw a spotlight on areas of uncertainty in the signal transduction pathway, highlighting the importance of spatial organization to the logical operation of the pathway. In particular they emphasize the function of a specific, well-characterized, cluster of proteins associated with the chemotaxis receptors which acts like a self-contained computational cassette.

Most recently we examined the failure of conventional models to reproduce the very high gain and dynamic range of the chemotactic response. This led us to propose a mechanism in which signal amplification takes place by the probabilistic spread of conformations in clusters of receptor complexes on the surface of the bacterium. This mechanism, analogous to the behavior of magnetic dipoles in a spin glass, can account quantitatively for the remarkable sensitivity and bandwidth of chemotactic detection. Conformational spread in clusters of receptors is likely to be employed widely by cells other than bacteria and for purposes other than chemotaxis.

EXPERIMENTAL CHARACTERIZATIONS OF ALTERED GENETIC ELEMENT ORDER ON THE T7 GROWTH CYCLE.

D. Endy and I. Molineux
University of Wisconsin at Madison, 1415 Engineering Drive, Madison, WI 53706
University of Texas at Austin, Department of Microbiology, Austin, TX 78712
Drew@t7.che.wisc.edu molineux@mail.utexas.edu

INFORMATION-INTENSIVE DRUG DISCOVERY: GENOMICS, PROTEOMICS, AND BIOINFORMATICS.

J. N. Weinstein
Laboratory of Molecular Pharmacology, National Cancer Institute (NCI), NIH,
Building 37, Room 5D-02, Bethesda, MD 20892
Weinstein@dtpax2.ncifcrf.gov

The current revolution in drug discovery is based largely on developments in molecular biology and informatics. A case in point is provided by the drug discovery program of the NCI, which has profiled more than 65,000 compounds for their activity against 60 human cancer cell lines. The resulting patterns of activity (pharmacological fingerprints) have proved rich in information on mechanisms of drug action and resistance (Paull, et al., JNCI 81:1088,1989; Weinstein, et al., Science 258:447, 1992). To characterize the 60 cell types (and selected transfectants) with respect to molecular markers, we have done careful parallel harvests of DNA, RNA, and protein for what I have termed "omic" analysis -- including protein expression profiling by 2-D gels and mRNA expression profiling by high density cDNA and oligonucleotide microarrays. The data complement those of the NCI's Cancer Genome Anatomy Project in that the 60 cell line "patients," unlike most human ones, have extensive, well-defined treatment histories -- i.e., they have been treated with >65,000 agents one at a time and independently (Weinstein, et al., Science 275:343, 1997).

(Many others have contributed to this work, including U Scherf, M Waltham, TG Myers, WC Reinhold, L Smith, L Tanabe, JK Lee, D Andrews, J Buolamwini, W van Osdol, G Li, DA Scudiero, NL Anderson, DT Ross, M Eisen, PO Brown, D Botstein, D Shalon, E Lashkari, R Simon, L McShane, E Lander, T Golub, H Coller, P Tamayo, D Slonum, KW Kohn, Y Pommier, EA Sausville, and the late KD Paull.)

GENE REGULATION, ASYNCHRONOUS COMPUTING AND STOCHASTIC COMPETITION.

J. Bruck and M. Gibson
California Institute of Technology, Mail Code 136-93, Pasadena, CA 91125
Bruck@paradise.caltech.edu

As more data become available about genetic regulatory networks, it is becoming apparent that ad-hoc methods of describing biological systems are insufficient for analyzing complex biochemical networks. There are two main classes of models of gene regulation. The first class consists of very high level, qualitative models. The problem is that it is hard to get quantitative predictions from it. The second class of models consists of physical biochemistry. These models use low level quantitative data, such as, binding constants, rate constants, etc. This sort of models should, in principle, give nearly exact quantitative predictions. However, in practice, such models tend to involve a number of thermodynamic and kinetic parameters, many of which are not known. We will argue that these low-level details are important and can be utilized as part of our proposed hierarchical methodology that consists of four levels of abstraction. The lowest level is the biochemical details of expression of an individual gene. The second level is a stochastic competition, which is basically the short term interaction between two or more genes being expressed simultaneously. The third level of abstraction is a probabilistic finite state machine, which puts together short term interactions into a long term framework. The final level is the organism level, where we use the probabilistic finite state machines to make predictions about the organism under various manipulations. We will present theoretical results of applying the new modeling methodology to the lysis/lysogeny decision process in Lambda phage.

CELLULAR FUNCTION INFORMATION FROM PROTEIN MASS SPECTROMETRY.

K. Resing and N. Ahn
University of Colorado, Boulder, Department of Chemistry and Biochemistry,
Campus Box 215, Boulder, CO 80309
Resing@stripe.colorado.edu

The development of electrospray and matrix assisted laser desorption ionization for introduction of biological samples into a mass spectrometer has greatly facilitated analysis of proteins. The first obvious advance has been in analysis of post-translational modifications, because it is now possible to sequence peptides in complex mixtures. More recently, the development of protocols for data base searching using sequence from peptides generated from in-gel digests has opened new areas of cell biology, particularly in the new field of proteomics, the characterization of the protein compliment of a given cell. We are using this technique to approach signal transduction from a systems analysis viewpoint. Finally, it is seldom appreciated that mass spectrometry can provide information on dynamic motion or conformational change in proteins by analysis of deuterium exchange at backbone amides. Analysis of the changes detected in the signaling kinases ERK2 (extracellular regulated kinase 2) and its activator, MKK (MAP kinase kinase), shows that these two enzymes show fundamental differences in the pattern of changes and that these reflect different activation strategies of each.

DRUG-INDUCED HAPLOINSUFFICIENCY: A GENOMIC APPROACH TO DRUG TARGET IDENTIFICATION.

G. N. Giaever
Stanford University, School of Medicine, Stanford, CA 94305-5307
Ggiaever@cmgm.stanford.edu

Alterations in gene dosage can be used as a tool for drug target identification. We have shown that by lowering a single geneís dosage from two copies to one copy in diploid yeast, the resulting heterozygote is sensitized to a drug that acts on this geneís product. This haploinsufficient phenotype thereby identifies the gene product of the heterozygous locus as the likely drug target.

This new finding is exploited in a genomic approach to drug target identification. Genome sequence information is used to generate molecularly tagged heterozygous yeast strains that are pooled, grown competitively in drug, and analyzed for drug sensitivity using high-density oligonucleotide arrays. This approach makes it feasible to screen all potential drug targets in yeast in a single assay.

AFTER THE GENOME IV

Abstracts

for

Tuesday,

13 October 1998

MODELLING PHYSIOLOGY AT THE ORGAN-SYSTEM LEVEL.

T. Patterson
Entelos, Inc. 4040 Campbell Avenue, Suite 200, Menlo Park, CA 94025
Paterson@entelos.com

LARGE SCALE STRUCTURE PREDICTION.

S. Skolnick
Scripps Research Institute, 10550 N. Torrey Pines Road, TPC5, La Jolla, CA 92037
Skolnick@scripps.com

STRUCTURE-BASED ANALYSIS OF SEQUENCE AND FUNCTION.

B. Honig
Columbia University, 221 Black Building, New York, NY 10032
Honig@tranto.bioc.columbia.edu

USE OF CONTEXT VECTORS FOR ANALYSIS OF LARGE VOLUMES OF BIOLOGICAL DATA

J. D. Saffer
Battelle-Pacific Northwest Labs, Richland, WA 99352

Advances in the genome era has led to many ëlarge-scaleí biological approaches including high-throughput sequencing, gene array technology, and automated methods for defining protein-protein interactions. Effective use of the resulting large volume of data presents new challenges. The key issues in dealing with this data flood are (1) the need interrogate all the data simultaneously, (2) the need for rapid analyses, and (3) the need to integrate multiple data types. Context vectors provide one approach that addresses these issues.

Practical use of context vectors involves three basic steps. First, the vector has to be created based on meaningful characteristics. In this vector creation process, the type of question being asked will determine how to create the vector. Usually, the number of possible characteristics and iterations leads to an overwhelming computational problem. This problem can be greatly simplified through the use of feature extraction methods. Second, the vectors have to be visualized. This can be accomplished using a variety of approaches, often including various clustering methods. Third, sufficient tools are needed to interact with the visualization. It is through these interactions that discoveries are made.

We have successfully applied context vectors to the analysis of newly discovered proteins of unknown function. For example, a 15 minute analysis of the sequence of predicted proteins from Methanococcus jannaschii identified a large number of putative transport proteins that were not recognizable as such using BLAST and that other computational methods took weeks to find. This type of analysis has been extended to very large data sets, such as SwissProt, and demonstrates the ability of the method for large-scale comparisons. In addition, visualization of protein sequence context vectors was shown to provide a rapid means for comparing whole genomes.

We have also applied context vectors for the analysis of gene expression array data. Traditional methods of analyzing such data rely primarily on Boolean queries, which allow assessment of expected associations. In contrast, we have found that visualization of array data allows discovery and exploration of unexpected relationships.

MATHEMATICAL APPROACHES TO DEALING WITH THE CURSE OF DIMENSIONALITY IN DEPARTMENT OF DEFENSE SIGNAL PROCESSING APPLICATIONS.

Anna Tsao
DARPA, Defense Sciences Office, 3701 N. Fairfax Drive, Arlington, VA 2203
Atsao@darpa.mil

In the last few years, dramatic strides have been made in several signal processing applications that can be characterized as having large numbers of degrees of freedom. The experience has highlighted several principles for approaching information extraction problems that appear to have broad generality. In this talk, these principles will be discussed and illustrated using examples from DoD-sponsored research, primarily in harmonic analysis-based methods.

MOLECULAR INFORMATION THEORY: FROM CLINICAL APPLICATIONS TO MOLECULAR MACHINE EFFICIENCY.

T. D. Schneider
National Cancer Institute, Frederick Cancer Research and Development Center,
Laboratory of Experimental and Computational Biology,
P. O. Box B, Frederick, MD 21702-1201
toms@ncifcrf.gov http://www-lecb.ncifcrf.gov/~toms/

Information theory was introduced by Claude Shannon in 1948 to precisely characterize data flows in communications systems. The same mathematics can also be fruitfully applied to molecular biology problems. We start with the problem of understanding how proteins interact with DNA at specific sequences called binding sites. Information theory allows us to make an average picture of the binding sites and this can be shown with a computer graphic called a sequence logo (http://www-lecb.ncifcrf.gov/~toms/sequencelogo.html). Sequence logos show how strongly parts of a binding site are conserved, on a scale in bits of information. They have been used to study a variety of genetic control systems. More recently the same mathematics has been used to look at individual binding sites using another computer graphic called a sequence walker. Sequence walkers are being used to predict whether changes in human genes cause mutations or are neutral polymorphisms. It may soon be possible to predict the degree of colon cancer by this method. Information theory can also be used to understand the relationship between the binding energy dissipated when two molecules stick together and the amount of sequence conservation of the molecules measured in bits. Using the Second Law of Thermodynamics, this relationship can be expressed as the efficiency of the molecular interaction. Surprisingly, many molecular systems including genetic systems, visual pigments and motility proteins have efficiencies near 70%. A purely geometrical explanation of this result shows that although biological systems are selected to have the highest efficiency, it is restricted to 70% because having precisely distinguishable molecular states is more important.

AFTER THE GENOME IV

Abstracts

for

Wednesday,

14 October 1998

NATURAL GENETIC ENGINGEERING OF GENOME SYSTEM ARCHITECTURE IN EVOLUTION.

J. Shapiro
University of Chicago, 920 E. 58th Street, Chicago, IL 60637
Jsha@midway.uchicago.edu

Molecular genetics teaches three lessons relevant to thinking about the nature of genetic change during evolution:

(1) Genomes are organized as hierarchies of composite systems (multidomain protein-coding sequences; functional loci made up of regulatory, coding, processing and intervening sequences; multi-locus regulons and replicons) interconnected and organized into specific "system architectures" by repetitive DNA elements.
(2) Genetic change often occurs via natural genetic engineering systems (cellular biochemical functions, such as recombination complexes, topoisomerases, and mobile elements, capable of altering DNA sequence information and joining together different genomic components.
(3) The activity of natural genetic systems is regulated by cellular control circuits with respect to the timing, activity levels, and specificities of DNA arrangements (e.g. adaptive mutation, Ty element mobility, P factor insertions).

These three lessons provide plausible molecular explanations for the episodic, multiple, non-random DNA rearrangements needed to account for the evolution of novel genomic system architectures and complex multilocus adaptations. This molecular genetic perspective places evolutionary change in the biologically responsive context of cellular biochemistry.

CROSS-SPECIES PATHWAY COMPARISON, PHARMACOLOGICAL DISSECTIONS.

L. Eiden
NIMH, 5600 Fishers Lane Room 11-103 Rockville, MD 20857
Eido@codon.nih.gov

FIXING THE GREENHOUSE.

G. Benford
University of California at Irvine, Physics & Astronomy, 4176 FRH, Irvine, CA 92697
Gbenford@uci.edu

No issue holds more profound possible consequences for the next century that greenhouse warming. Yet so far the debate and hand-wringing have been both angry and unimaginative. There may very well be fairly simple fixes, even cheap ones -- but the tone of discussion never makes this clear. Could we intervene to offset the warming? Accept that greenhouse gases will rise and find ways to compensate for them? Surprisingly, some schemes appear possible to deploy now, and at reasonable cost. They could be turned on and off quickly, if we get unintended effects.

The simplest way to remove carbon dioxide from the air is to grow plants, preferably trees. About half the U.S. CO2 emissions could be captured if we grew tree crops on economically marginal croplands and pasture. This would cost about five billion dollars a year. Soaking up the world's present CO2 increase would take up an Australia-sized land area, i.e., a continent. One should compare the cost of achieving this same end by suppressing fuel use: roughly $500 billion a year to reduce global CO2 emissions by 30%, a number easily debatable within a factor of two.

An even more direct approach promises similar savings. What could be more intuitively appealing than simply reflecting more sunlight back into space, before it can be emitted in heat radiation and then absorbed by greenhouse gases? People can understand this readily enough; black T-shirts are warmer in summer than white ones. We could compensate for the effect of all greenhouse gas emission since the Industrial Revolution by reflecting one percent more of the sunlight. Our upper atmosphere is a good place to intervene, because much sunlight gets absorbed in the atmosphere on its way to us. We already add a perfectly good reflecting area to the upper atmosphere as part of everyday flying -- aircraft exhausts. Making the fuel mixture in a jet engine burn rich can leave a ribbon of fog behind for up to three months, though as it spreads it becomes invisible to the eye. These motes would also come down mostly in rain, not troubling the brow of the EPA. Fuel costs about fifteen percent of airlines' cash operating expenses, and running rich increases costs only a few percent. This means that for about ten million dollars this method would offset the 1990 U.S. greenhouse emissions, quite a cheap choice. Adding this to the cost of an airline ticket would boost prices perhaps one percent. These ideas envision doing what natural clouds do already, as the major players in the total albedo picture. A four percent increase in stratocumulus over the oceans would offset global CO2 emission. Land reflects sunlight much better than the wine-dark seas, so putting clouds far out from land, and preferably in the tropics, gets the greatest leverage. Making clouds is an old but still controversial craft. Clouds condense around microscopic nuclei, often the kind of sulfuric acid droplets geoengineers could spread in the stratosphere. Clouds cover about thirty-one percent of our globe already, so a four

percent increase is not going to noticeably ruin anybody's day. Tinkering with such a mammoth natural process is daunting, but about four hundred medium-sized coal-fired power plants give off enough sulfur in a year to do the job for the whole Earth.

At first these should operate as regional experiments, to work out a good model of how the ocean-cloud system responds. Cost: about two billion dollars per year. Simply adding sand or glass to ordinary asphalt ("glassphalt") doubles its albedo. A 1997 study showed that Los Angeles is five degrees F. warmer than the surrounding areas, mostly due to dark roofs and asphalt. White roofs, concrete-colored pavements and about ten billion new shade trees could cool the city below the countryside, cutting air conditioning costs eighteen percent. It might take only a few billion dollars to mitigate the U.S. emission of CO2. Any greenhouse fix must accept that our prosperity is built upon cheap, handy energy, and the developing nations will not give it up.

TOOLS AND DATABASES FOR THE ANALYSIS OF BIOCHEMICAL REACTION NETWORKS

A. Arkin
Lawrence Berkeley National Laboratory,
Physical Biosciences Division, Berkeley, CA 94720
E-mail: aparkin@lbl.gov Web: http://www.lbl.gov/~aparkin

Biological regulatory networks are the circuitry that control cellular function and malfunction. The chemistry underlying the function of these networks is extraordinarily complex and difficult, if not impossible to understand knowing only a list of the parts (genes, proteins, and other chemicals) and a list of which parts react with which other parts. Just as when analyzing and diagnosing complex electronic circuits, mathematical models and computational tools for analysis and simulation are necessary if we are to understand, control and even design our own biological and genetic reaction networks. In order to build such tools many different types of data will have to be databased in addition to the genomic parts lists and 'network topology' specifications that are now the most commonly used information. Physical data such as detailed mechanisms, kinetic constants and data, thermodynamic parameters have to be readily available. Also databases of phenotypic data such as gene chip data under numerous conditions, cell morphology, cell fate maps, etc. must also be kept in order tovalidate the model/simulations. We are developing a suite of network deduction, simulation and analysis tools analogous to the Spice simulation/analysis tool ubiquitously used by electrical engineers. The tool interfaces to genome databanks, pathway databases such as EcoCyc and Kegg; and specialized databases of mechanisms, physical constants and behavioral data in order to predict the biochemical network structure and predict and analyze the resultant genetic/biochemical pathway kinetics as a system. Applications of the tools to metabolic systems, microbial pathogenic systems will be briefly discussed.

COMPUTER-AIDED RESOLUTION OF AN EXPERIMENTAL PARADOX IN BACTERIAL CHEMOTAXIS.

W. Abouhamad*, D. Bray#, M. Schuster*, K. Boesch*,
R. Silversmith*, & R. Bourret*
*University of North Carolina, Department of Microbiology & Immunology,
Chapel Hill, NC 27599-7290
(Telephone) 919-966-2679 (FAX) 919-962-8103 email: bourret@med.unc.edu

Escherichia coli responds to its environment by means of a network of intracellular reactions which process signals from membrane bound receptors and relay them to the flagellar motors. Although characterization of the reactions in the chemotaxis signaling pathway is sufficiently complete to construct computer simulations that predict the phenotypes of mutant strains with a high degree of accuracy, two previous experimental investigations of the activity remaining upon genetic deletion of multiple signaling components yielded several contradictory results [Conley et al., J. Bacteriol. 171, 5190 (1989); Liu & Parkinson, Proc. Natl. Acad. Sci. USA 86, 8703 (1989)]. For example, "building up" the pathway by adding back CheA and CheY to a gutted strain lacking chemotaxis genes resulted in counterclockwise flagellar rotation, whereas "breaking down" the pathway by deleting chemotaxis genes except cheA and cheY resulted in alternating episodes of clockwise and counterclockwise flagellar rotation. Our computer simulation predicts that trace amounts of CheZ expressed in the gutted strain could account for this difference. We tested this explanation experimentally by constructing a new deletion of the che genes that cannot express CheZ and verified that the behavior of strains "built-up" from the new deletion does in fact conform to both the phenotypes observed for "break-down" strains and computer-generated predictions. Our findings consolidate the present view of the chemotaxis signaling pathway and highlight the utility of molecularly-based computer models in the analysis of complex biochemical networks.

[This work was very recently published in J. Bacteriol. 180, 3757 (1998)]

LACK OF STANDARDIZATION OF ANYALYTICAL TECHNIQUES AND INFORMATION HANDLING AFFECTS THE ADOPTION OF GENETIC DATA IN CLINICAL TRIALS AND THE CLINIC.

M. Bywater
Perkin-Elmer/Genscope, 50 Danbury Road, Wilton, CN 06897
Email: Mbywater@genscope.com

Although the promise of molecular biology has been the era of Molecular Medicine, the translation of research observations into clinical facts has taken longer than many scientists have optimistically predicted.

With the advances made in Genomics, identification of inherited traits associated with susceptibility to disease have increased awareness in the medical field of the power of genetic information. Inherited genetic markers as well as genetic mutations are assigned the potential of either being disease causing or indicators of disease progression or response to therapeutic agents.

The acceptance of these research results into clinical trials is hampered by the lack of robust results now generated by a variety of non-standardized technologies, and the lack of links to give function of biological pathways.

This results in data, which is not acceptable as parameters in clinical trials. Confusing evidence also hampers the clinical adoption of genetic markers as new standards of care for the management of disease.

A common language and standardized interpretation of results will provide a basis to simulate clinical trials using computed information. Several sources ultimately resulting in the use of molecular information to streamline protocols. These demands impose new challenges for companies providing tools for such studies and impact the business models currently available.

INTEGRATING BIOINFORMATICS, DATA, AND BIOLOGISTS.

J. M. Cherry
Stanford University School of Medicine, Department of Genetics, Medical Center,
Room M341, Stanford, California 94305-5120
Voice: 650-723-7541 FAX: 650-723-7016 email: cherry@genome.stanford.edu

We are working to build better databases, enhance the predicted information with experimental results from the literature, provide access to the large scale results but also provide summaries of these results to facilitate greater availability within the greater scientific community. The ability to make appropriate inferences from the wealth of systematic genomic sequence, expression, two-hybrid and other large scale projects requires a solid base of biological information. The data must be maintained and distributed in a manner that enhances scientific discovery. Computed predictions of biological function provide a wealth of useful information. These hypothetical statements are even more powerful if associated with experimental results. Within the Department of Genetics at Stanford a diverse group of experts has been assembled. Here is a brief list of projects I am managing. 1) The Saccharomyces and Arabidopsis databases are collecting a variety of information from individual laboratories and from large functional genomic projects. 2) A database and associated software is being created for the analysis of DNA microarray results. 3) A shared classification or grouping system is being developed with other database groups to facilitate querying in a uniform manner across different species databases. This network of biological processes will include a common vocabulary and be used by the databases to categorize an organism's gene products. 4) Finally, we are analyzing the integrated data from DNA microarray experiments from Saccharomyces to identify expression control elements.

BUILDING UPON MOLECULAR PROFILING: CREATING TECHNOLOGY PLATFORMS THAT WILL REVOLUTIONIZE CANCER DETECTION, DIAGNOSIS AND TREATMENT.

C. A. Dahl
National Cancer Institute, Building 31, Room 11A03,
31 Center Drive, MSC 2590, Bethesda, MD 20892-2590
Phone: (301)496-1550 Email: carol_dahl@nih.gov

The Challenge: It has become clear that cancer is a set of diseases that result from changes in the genome and the expressed products of the genome. The pathway of technological opportunity resulting from this fundamental observation can have profound impact on the management and prevention of cancer.

The first step in this pathway requires defining the molecular profiles of cancer. To this end the National Cancer Institute (NCI) has launched an initiative, the Cancer Genome Anatomy Project (http://www.ncbi.nlm.nih.gov/ncicgap/), which will provide both the information and technology infrastructure needed to uncover the molecular profiles of cancers. Correlation of the molecular profiles with critical characteristics of cancers, such as prognosis and response to therapy, will provide a foundation for improved decision-making relative to needs and strategies for intervention and prevention.

The NCI is now interested in looking beyond these near term goals. Building upon molecular profiling, we wish to create technology platforms that will revolutionize cancer detection, diagnosis and treatment. The NCI is interested in identifying technology systems or components that will enable sensing of molecular alterations in the body in a way that is highly sensitive and specific, yet non-intrusive. Molecular profile information would then be transmitted to external monitoring devices that would provide input to the physician. The technology system should additionally serve as the platform for, or have a seamless integration with, capabilities for intervention specific for the detected molecular profile. The capability for intervention should allow for control and monitoring of the intervention that will ultimately be under the supervision of the physician.

Achieving this ambitious objective will require the development and integration of a series of capabilities including highly specific molecular recognition, signaling capability, controllable intervention capabilities, methods for monitoring intervention release and impact, and biotolerance. This will require the input and collaboration of investigators from a variety of disciplines, many of which have not traditionally engaged in cancer research.

Current approaches to cancer detection and diagnosis tend to be highly invasive and inadequately informative with regard to the underlying molecular basis of the

specific disease. Therapies range from dramatically invasive procedures, such as surgery, to the administration of relatively toxic agents, such as chemotherapy and radiation. Ultimately patients would benefit from the availability of non-invasive approaches to cancer detection and diagnosis linked to minimally debilitating treatments that are tailored to target the precise molecular alterations in the individual tumor.

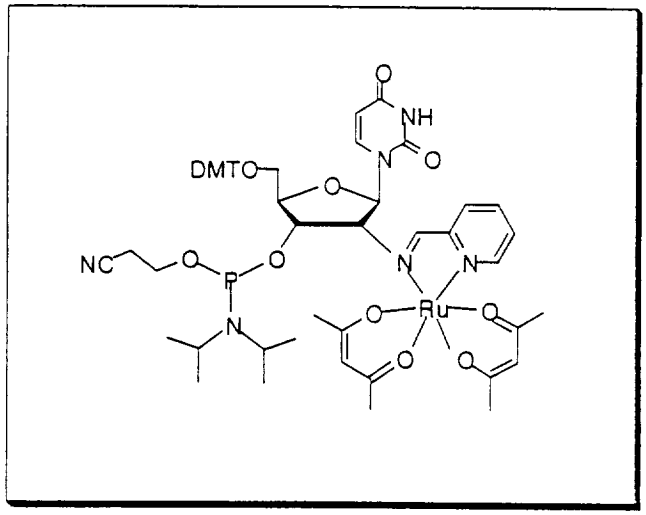
The development of platform technologies that enable non-intrusive sensing and intervention in the individual will revolutionize the clinical approach to detection, diagnosis and treatment. Such systems could allow detection and elimination of cancer cells in their earliest stages. They would forestall the development of very large tumors and would move meaningful intervention to a much earlier point in the path of progression. This will minimize patient inconvenience and incapacitation, and allow detection, diagnosis, and treatment to be closely coupled, enabling effective administration in a single, seamless process. Send Your Ideas!

The NCI is inviting members of the academic, government and industrial research communities to provide input that will contribute to defining the general scope of a Broad Agency Announcement request for contracts in late 1998 or early 1999. Input is encouraged in the form of white papers, but will also be accepted in alternative informal written formats. Interested parties are encouraged to provide information on new areas of technological opportunity that could speed progress toward the scientific goals stated above.

ELECTRON TRANSFER IN DNA? THE DESIGN OF A STRUCTURALLY DEFINED SYSTEM

N. L. Frank, E. J. Krider, J. J. Rack, T.J. Meade*

Incorporation of metallated nucleosides into DNA at fixed and known locations is crucial to examination of electron transfer rates and the variation of these rates with structure, distance, and position. Electron transfer through DNA may be central to important biological processes such as DNA damage/repair and synthesis, information transcription, and DNA-drug interactions. The mechanism, however, of energy and electron transfer central to these processes is unclear. Interpretation of the results is complicated by the distribution of distances examined, as well as uncertainty in the mechanism of charge migration. Recent studies have attempted to examine the rates of electron transfer in DNA with varying distance by direct intercalation of donor/acceptor molecules into DNA. The results of such studies have shown a remarkable dependence of electron transfer rates on the *design* of the experiment, suggesting that a detailed and systematic study is needed. We are currently investigating the mechanism of long-range electron transfer processes in modified oligonucleotides with covalently fixed donors and acceptors. Novel synthetic methodology has been developed for the site-specific incorporation of transition metal acceptor-donor complexes into oligonucleotiodes by solid phase synthesis; in particular, site specific labeling of the ribose moiety.



DATA INTEGRATION:
THE ISSUE OF STRUCTURE VERSUS FUNCTION.

M. N. Liebman
Bioinformatics, Wyeth-Ayerst Research, R3037, 145 King of Prussia Road,
Radnor, PA 19087
liebmam@war.wyeth.com

The issue of data integration has played a significant role in the development and collection of genome sequence data and in raising issues about its integration with functionally related biological information. This post-genome data integration issue has typically focused on the use of relational data models because of the size of component data sets, but we have noted that this can be significantly limiting in providing the access to complex data inquiries and relationships. A data model which is more functional than structural is being implemented to integrate genome sequence, protein structure and function, with higher order data including pathway models, cellular processes, toxicology, pharmacokinetics and clinical data.

SEEING IS BELIEVING: MONITORING IN VIVO GENE EXPRESSION BY MAGNETIC RESONANCE IMAGING

A. Y. Louie and T. J. Meade
California Institute of Technology, Division of Biology, Biological Imaging Center, and the Beckman Institute, Pasadena, CA 91125

MRI offers a non-invasive means to map brain structure and function by sampling the amount, flow or environment of water protons in vivo. This intrinsic contrast can be augmented by the use of paramagnetic contrast agents; however, these agents are little more than anatomical reporters which can at best label individual fluid compartments or distinguish tissues that are magnetically similar but histologically distinct. To permit more direct imaging of the physiological state of cells or organs, we have prepared and tested a new class of "smart" contrast agents that are activated by β-galactosidase. The complexes were tested in Xenopus laevis embryos and provide the first example of direct, three-dimensional visualization of gene expression by MRI.

GENE EXPRESSION DATA MINING: MAKING THE CONNECTIONS

G. Michaels
George Mason University, Fairfax, VA 22030-4444
Email: gmichael@gmu.edu

Several large-scale approaches for generating gene expression data are currently competing for broad acceptance. We have concentrated on developing informatics and visualization tools for working with thousands of genes. An analysis of a yeast whole genome, gene expression experiment will be presented.

DEVELOPING DRUGS TO TREAT ISCHEMIC DISEASES

G. Miller
Galileo Laboratories, Inc., 935 East Arques Avenue, Sunnyvale, CA 94086
Email: gmiller@galileolabs.com

Galileo Laboratories, Inc. mission is to develop drugs to treat ischemic diseases such as heart attack and stroke. The Company's technical focus is on developing investigative tools to enable the discovery of drugs to treat both the ischemic and reperfusion components of ischemic injury. Galileo has constructed cell-based model systems that are representative of how vital organs make and regulate oxygen dependent energy synthesis. Galileo's cell-based model system approach is inexpensive, fast, robust in data, and lends itself to complementary drug discovery tools such as genomics, proteomics, combinatorial chemistry and high throughput screening techniques. The Company is applying its proprietary Metabolic Target Array Queryä [MTAQä] to identify target sets and therapeutic candidate relevant to ischemic disease, and is leveraging its MTAQä technology in PKADME-Toxicology and human medical nutritionals applications.

Biographical Sketch

Guy Miller, MD, PhD, Chairman and Chief Executive Officer. Dr. Miller is a founder of Galileo Laboratories, Inc. Prior to founding Galileo, Dr. Miller was an Assistant Professor at Johns Hopkins University, School of Medicine. Dr. Miller obtained his PhD in Chemistry under the direction of Professor Sidney Hecht, John Mallet Professor of Chemistry at the University of Virginia, and his MD at the Medical College of Pennsylvania. After completing an internship in Surgery at the University of Chicago, Dr. Miller completed a residency in Anesthesiology & Critical Care Medicine at Johns Hopkins Hospital, followed by a fellowship in Multidisciplinary Critical Care Medicine. Currently, he holds an appointment as Clinical Instructor, Stanford University School of Medicine where he attends patients in the Medical Surgical Intensive Care Unit.

GENE CLUSTERS: AN APPROACH TO FUNCTIONAL CLUSTERING FOR PROKARYOTICS

R. Overbeek
Argonne National Laboratory
Argonne, IL 60439
Email: overbeek@mc.anl.gov

The presence of gene clusters on prokaryotic genomes is well recognized. The origins of these clusters and the role they play is still a topic of discussion. What we have found is that one can use such clusters to accurately predict functional coupling between genes. The technique requires the use of numerous genomes to produce reliable coupling data. Once one has access to 15 – 20 prokaryotic genomes containing a functional subsystem (e.g., a pathway) it it possible to predict which genes ar3 functionally coupled (although it is often the case that a complete set of genes for the pathway will not exist within a single cluster in any of the genomes, and about 65% of the relevant individual genes within the set of genomes will not be included within clusters). We can now couple hundreds of "hypothetical proteins" to specific functional subsystems, or in many cases simply assert that a set of hypothetical proteins together form a functional subsystem.

A rudimentary analysis reveals that the coupling information grows as the square of the number of genomes included in the analysis: below about ten genomes, the number of reliably, inferrable couplings is quite limited, but by the time twenty complete genomes exist including a given functional subsystem (i.e., pathway) the complement of genes related to the system is revealed clearly.

We believe that this technology will play a key role in characterizing the genes in prokaryotes.

INFERENCE OF GENETIC PROGRAMS IN MAMMALIAN CNS DEVELOPMENT AND INJURY

R. Somogyi
Incyte Pharmaceuticals, Inc.; 3174 Porter Dr.; Palo Alto, CA 94304
Phone: 650-845-4210 Fax: 650-845-4255 email: rsomogyi@incyte.comBi

Large scale gene expression mapping is motivated by the premise that biological information is transmitted from gene sequence to gene activity patterns, and, through a hierarchy of inter- and intracellular signäling functions, back to the regulation of gene expression. This process can be conceptualized as a genetic network. In an effort to understand the output of the genetic network, we have conducted extensive surveys of the dynamics of gene expression in mammalian CNS development and injury. Analysis of this data suggests modules of genetic programs that are conserved among CNS regions (spinal cord and hippocampus), and can be reactivated following injury (kainic acid induced seizures). These results indicate that more detailed ìtop-downî studies of this kind, coupled with advanced inference techniques, may help resolve the distributed molecular processes underlying complex genetic signaling networks. Such understanding will be critical for the discovery and validation of drug targets in CNS pathology.

After the Genome IV Participant List

Adam Arkin
Lawrence Berkeley National Laboratory
1 Cyclotron Road MS 3-144
Phical Biosciences Divisions
Berkeley, CA 94720 USA
aparkin@lbl.gov

M.J. Finley Austin
Merck Genome Research Institute
Sumneytown Pike
PO Box 4, WP44L=206
West Point, PA 19486 USA
finley_mgri@merck.com

Cindy Bamdad
Clinical Microsensors, Inc
101 Waverly Drive
Pasadena, CA 91105 USA
bamdad@microsensor.com

Joel Bellenson
Pangea
1999 Harrison St Ste 1100
Oakland, CA 94612 USA
joelb@pangeasystems.com

Gregory Benford
University of California at Irvine
Department of Physics & Astronomy
4176FRH
Irvine, CA 92697 USA
gbenford@uci.edu

Robert Berwick
Massachusetts Institute of Technology
NE 43-765
Cambridge, MA 02139 USA
berwick@ai.mit.edu

Bob Bourret
University Of North Carolina
Department of Microbiology and Immunology
Chapel Hill, NC 27559-7290 USA
bourret@med.unc.edu

Dennis Bray
University of Cambridge
Department of Zoology
Cambridge CB2 3EJ
United Kingdom
d.bray@zoo.cam.ac.uk

Roger Brent
Molecular Sciences Institute
2168 Shattuck Ave. 2nd Floor
Berkeley, CA 94704 USA
brent@molsci.com

Christoph Brockel
Hoescht Marion Roussell
Hoechst-ARIAD Genomics Center
26 Landsdowne Street
Cambridge, MA 02139-4324 USA

Jehoshua Bruck
California Institute of Technology
MS 136-93
Pasadena, CA 91125 USA
bruck@paradise.caltech.edu

Soren Brunak
Center for Biological Sequence Analysis
DTU Building 208
Lyngby DK-2800
Denmark
brunak@cbs.dtu.dk

Susan Burgess
Structural Bioinformatics
10929 Technology Place
San Diego, CA 92127 USA
sburgess@strubix.com

Jeff Byers
Naval Research Laboratory
Materials Physics, Code 6340
4555 Overlook Ave., SW
Washington DC, 20375 USA
byers@foucault.nrl.navy.mil

Margaret Bywater
Perkin-Elmer/Genscope
50 Danbury Road
Wilton, Conn 6897 USA
mbywater@genscope.com

Judith Campbell
California Institute of Technology
M/C 147-75
Pasadena, CA 91125 USA
jcampbel@cco.caltech.edu

Lynn Caporale
Consultant
1 Sherman Square
New York, NY 10023 USA
caporale@usa.net

Rob Carlson
The Molecular Sciences Institute
2168 Shattuck Ave.
Berkeley, CA 94704 USA
rcarlson@sequence.molsci.org

Mike Cherry
Stanford Medical Center, Room M341
Stanford University School of Medicine
Stanford, CA 94305-5120 USA
cherry@genome.stanford.edu

David Cohen
University of Texas
Austin, TX 78712 USA
dcohen1@mail.utexas.edu

Robert Cohen
The Molecular Sciences Institute
2168 Shattuck Ave.
Berkeley, CA 94704 USA

Carol Dahl
NCI
Building 31, Room 11A03
31 Center Drive, MSC 2590
Bethesda, MD 20892-2590 USA
carol_dahl@nih.gov

David Dolinger
Stratagene
11011 North Torrey Pines Road
La Jolla, CA 92037 USA
ddolinger@microgene.com

Shawn Dunnick
California Institute of Technology
Beckman Institute 139-74
Pasadena, CA 91125 USA
adunnick@gg.caltech.edu

Lee Eiden
NIMH
5600 Fishers Lane Room 11-103
Rockville, MD 20857 USA
eido@codon.nih.gov

Emelyn Eldredge
Academic Press
525 B Street Suite 1900
San Diego, CA 92101-4495 USA
eeldredge@acad.com

Drew Endy
University of Wisconsin at Madison
1415 Engineering Drive
Madison, WI 53706 USA
drew@t7.che.wisc.edu

Richard Fine
Paradygm Technologies, Inc.
420 Bedford Rd.
Ridgewood, NJ 07450
rmfine@aol.com

Natia Frank
California Institute of Technology
Beckman Institute 139-74
Pasadena, CA 91125 USA
nfrank@gg.caltech.edu

Rainer Fuchs
Ariad Pharmaceuticals
26 Landsdowne Street
Cambridge, MA 02139-4234 USA
rainer.fuchs@ariad.com

Jill Fujisaki
Entelos, Inc
4040 Campbell Ave., Suite 200
Menlo Park, CA 94025 USA
fujisaki@entelos.com

Guri Giaever
Stanford University
School of Medicine
Stanford, CA 94305-5307 USA
ggiaever@cmgm.stanford.edu

Dean Goddette
Structural Bioinformatics Inc
10929 Technology Place
San Diego, CA 92127 USA
dgoddette@strubix.com

Stephen Grand
Cyberlife Technology
Quayside Bridge Street
Cambridge CB5 8AB
United Kingdom
stephen.grand@cyberlife.co.uk

Winnie Hallwachs
University of Pennsylvania
Department of Biology
Philadelphia, PA 19104 USA
whallwac@sas.upenn.edu

Pat Hess
Quest Diagnostics
35608 Ortega Highway
San Juan Capistrano, CA 92690-6130
USA
hessp@questdiagnostics.com

Barry Honig
Columbia University
221 Black Building
New York, NY 10032-3702 USA
honig@trantor.bioc.columbia.edu

Tim Hunkapiller
University of Washington
8247 E. Mercer Way
Mercer Island, WA 98040 USA
tim@mbt.washington.edu

Bill Hutchens
Ciphergen Biosystems
490 San Antonio Road
Palo Alto, CA 94306 USA
bhutchens@ciphergen.com

Russel Jacobs
California Institute of Technology
Beckman Institute 139-74
Pasadena, CA 91125 USA
rjacobs@caltech.edu

Dan Janzen
University of Pennsylvania
Department of Biology
Philadelphia, PA 19104 USA
djanzen@sas.upenn.edu

Cynthia Kenyon
University of California, San Francisco
513 Parnassus
San Francisco, CA 94143-0554 USA
ckenyon@biochem.ucsf.edu

Raoul Kopelman
University of Michigan
4744 Chemistry
Ann Arbor, MI 48109 USA
kopelman@umich.edu

Glenn Larsen
Genetics Institute
87 Cambridge Park Drive
Cambridge, MA 02140 USA
glarsen@genetics.com

Michael Liebman
Wyeth-Ayerst Research, R-3037
140 King of Prussia
Road Radnor, PA 19087 USA
liebmam@war.wyeth.com

Mary Weir Lipton
Battelle Pacific Northwest National Lab
PO Box 999 MS K8-98
Richland, WA 99352 USA
mary.lipton@pnl.gov

Larry Lok
The Molecular Sciences Institute
2168 Shattuck Ave.
Berkeley, CA 94704 USA
lok@molsci.org

Angelique Louie
California Institute of Technology
Beckman Institute 139-74
Pasadena, CA 91125 USA
alouie@gg.caltech.edu

E.D.(Sonny) Maynard
Defense Advanced Research Projects Agency
Information Technology Office
3701 North Fairfax Drive
Arlington, VA 22203-1714 USA

Tom Meade
California Institute of Technology
Beckman Institute 139-74
Pasadena, CA 91125 USA
tmeade@gg.caltech.edu

George Michaels
George Mason University
Institute for Computational Sciences and Informatics
Fairfax, VA 22030-4444 USA
gmichael@GMU.edu

Guy Miller
Galileo Laboratories
935 East Arques Avenue
Sunnyvale, CA 94086 USA
gmiller@GalileoLabs.com

Chad Mirkin
Northwestern University
Department of Chemistry
2145 Sheridan Road
Evanston, Il 60208-3113 USA
camirkin@chem.nwu.edu

Ian Molineux
Department of Microbiology
University of Texas
Austin, TX 78712-1095 USA
molineux@mail.utexas.edu

John Moult
CARB (caspII)
9600 Gudelsky Drive
Rockville, MD 20850 USA
jmoult@indigo5.carb.nist.gov;
tunc@iris4.carb.nist.gov

M. Northrup
Cepheid
1190 Borregas Avenue
Sunnyvale, CA 94089-1302 USA
northrup@cepheid.com

Ross Overbeek
Argonne National Laboratory
Argonne, IL 60439 USA
overbeek@mc.anl.gov

Luis Parodi
Pharmacia & Upjohn
luis.a.parodi@eu.pnu.com

Ljiljana Pasa Tolic
Battelle Pacific Northwest National Lab
PO Box 999, MS K8-98
Richland, WA 99352 USA
ljiljana.pasatolic@pnl.gov

Thomas Paterson
Entelos, Inc.
4040 Campbell Ave., Suite 200
Menlo Park, CA 94025 USA
paterson@entelos.com

Dhiraj Pathak
Glaxo Wellcome Inc
5 Moore Drive
Research Triangle Park, NC 27709
dp33010@glaxowellcome.com

Adrienne Regard
Structural Bioinformatics
10929 Technology Place
San Diego, CA 92127 USA
regard@strubix.com

Kathryn Resing
University of Colorado
Campus Box 215
Boulder, CO 80309 USA
resing@stripe.colorado.edu

Isidore Rigoutsos
IBM TJ Watson Research Center
PO Box 704
Yorktown Heights, NY 10598 USA
rigoutso@us.ibm.com

Jeff Saffer
Battelle Pacific Northwest National Labs
PO Box 999 MS P7-58
Richland, WA 99352 USA
jd.saffer@pnl.gov

Chris Sander
Millennium Pharmaceuticals
620 Memorial Drive
Cambridge, MA 02139 USA
sander@mpi.com

Tom Schneider
NCI
Laboratory of Computational and
Experimental Biology
Frederick, MD 21702-1201 USA
toms@ncifcrf.gov

James Schwaber
Du Pont Merck
DuPont Central Research
Experimental Station E-328/B31
Wilmington, DE 19880-0328 USA
schwaber@EPLRX 7.es.dupont.com

Vicki Seyfert
NIAID
Solar Building, Room 4A21
6003 Executive Boulevard
Bethesda, MD 20852 USA
Vseyfert@mercury.niaid.nih.gov

James Shapiro
University of Chicago
920 E. 58th Street
Chicago, IL 60637-4931 USA
jsha@midway.uchicago.edu

Robin Silva
Clinical Microsensors, Inc. and
Flehr Hohbach Test
4 Embarcadero Center Suite 3400
San Francisco, CA 94111 USA
rsilva@flehr,com

Jeffrey Skolnick
The Scripps Research Institute
10550 N. Torrey Pines Road
TPC5
La Jolla, CA 92037 USA
skolnick@scripps.com

David Soane
Alnis, L.L.C.
11933 Davis Street, Suite 258
San Leandro, CA 94577 USA
soane@ix.netcom.com

Roland Somogyi
Incyte
3174 Porter Drive
Palo Alto, CA 94304 USA
rsomogyi@incyte.com

Joseph Sorge
Stratagene
11011 North Torrey Pines Road
La Jolla, CA 92037 USA
joesorge@stratagene.com

Diana St. James
California Institute of Technology
Beckman Institute 139-74
Pasadena, CA 91125 USA
chem108@cco.caltech.edu

Evi Strauss
921 Fulton Street
San Francisco, CA 94117 USA
estrauss@cmgm.Stanford.edu

Bijal Trivedi
Nature Biotechnology
235 W. 22nd Street Apt 5-0
New York, NY 10011 USA
bpt202@is8.nyu.edu

Anna Tsao
Defense Advanced Research Projects
Agency
Defense Sciences Office
3701 N. Fairfax Drive
Arlington, VA 22203-1714 USA
atsao@darpa.mil

Tom Tullius
Boston University
Department of Chemistry
590 Commonwealth Avenue
Boston, MA 02215 USA
tullius@bu.edu

Joan Valentine
University of California, Los Angeles
405 Hilgard Avenue
Los Angeles, CA 90095-1569 USA
jsv@chem.ucla.edu

John Weinstein
NCI
Bldg 37 Rm 5C-25 NIH
9000 Rockville Pike
Bethesda, MD 20892 USA
weinstein@dtpax2.ncifcrf.gov

Jeff Wiseman
SmithKline Beecham Pharmaceuticals
709 Swedeland Road
PO Box 1539, Mail Code UW 2940
King of Prussia, PA 19406-0939 USA
jeffrey_s_wiseman@sbphrd.com

Barbara Wold
California Institute of Technology
MS 156-29
Pasadena, CA 91125 USA

MEETING AFTER THE GENOME IV

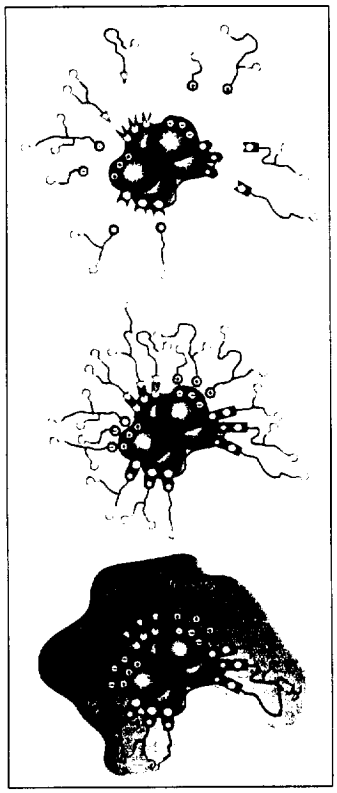
New Ways to Probe the Molecules of Life

JACKSON HOLE, WYOMING—Almost 200 years after Lewis and Clark first glimpsed the Grand Tetons, a posse of about 80 scientists gathered here from 10 to 14 October for an exploration of their own. At the annual "After the Genome" meeting, they discussed how to get from genomic information to an understanding of biology. Highlights include powerful computer programs for modeling human diseases and new techniques for protein analysis.

Making Coats for Molecules

For humans, Halloween is over, and the witches, Monicas, and Bill Clintons have taken home their prizes for best costume and packed their gear away until next year. But a team at the biotech start-up company Alnis, in San Leandro, California, has devised ingenious costumes for proteins and other molecules that they could wear all year long.

Alnis's scientific founder, David Soane, and his colleagues have found a way of trapping individual molecules inside polymer coatings a single molecule thick. Although the method is in its infancy, researchers can envision a wealth of applications for it. "This is a clever idea and the method has real scope," says Alexis Bell, a chemical engineer at the University of California, Berkeley, and an adviser to the company. The coatings could stabilize nature's biological catalysts, the enzymes, enabling them to tolerate organic solvents and high temperatures, and protect therapeutic proteins, such as insulin, from digestion so that they could be taken by mouth.



Entrapment. After the coat molecules bind to specific sites on the target, they are linked together.

CREDIT: ALNIS

Other investigators are also devising stabilizing coats for proteins, but the Alnis method boasts the advantage that it can be adapted to a wide variety of biological molecules and solvent systems. What's more, because the polymer coats retain the impression of the target molecule's shape even after it is removed, they could be used for molecular detection both in the body and in biological samples, such as blood or biopsy materials. "It's potentially a very exciting technology, particularly for detecting small molecules and perhaps even proteins," says Frances Arnold, a chemical engineer at the California Institute of Technology in Pasadena.

Soane's method is a twist on molecular imprinting, a technique that has been around for several decades. In molecular imprinting, the target molecules are embedded in a material that polymerizes around them to produce a three-dimensional block bearing the targets' impressions. The block can be used for a variety of applications. By breaking it into chunks, for example, researchers can generate a chromatographic material that grabs onto the target molecules, allowing their isolation from complex mixtures.

Instead of forming a polymer block, Soane generates a molecular glove that perfectly fits a protein or other molecule. He accomplishes this by exposing the molecule to custom-designed, polymerizable building blocks with distinctive heads and tails. The heads, for example, may carry positive or negative charges that allow them to bind to oppositely charged amino acid residues in the protein, while the tails, which are hydrophobic and tend to congregate with each other, are designed to link together.

Once the heads of the chemical have bound to the target molecule, Soane uses treatments such as ultraviolet light to link the chemicals into a shell, dubbed a synthetic polymer complement (SPC), around the protein. It's also possible to construct the SPC coat in such a way that an enzyme protein retains its catalytic activity. One way of doing this is to protect the enzyme's catalytic site with a molecule, such as one of the enzyme's own substrates, that binds to the site and can be removed once polymerization is accomplished. As a "proof of principle" test, Soane has shown that an SPC coat made the enzyme chymotrypsin far more durable at high temperatures in an organic solvent while still allowing it to be active.

Soane says that the SPC covering can also be released from its target molecule, although he won't say how because the technique is proprietary. If the empty shells then encounter the molecule again, they can bind it. He's shown, for example, that empty SPCs can recognize a small molecule called esculin that contains a sugar. Eventually, the chemical molds might be used for molecular detection—in effect serving as artificial antibodies that are more stable, cheaper, and quicker to make than the real thing. For example, SPCs linked to tracers that can be detected by ultrasound might help with early, noninvasive diagnosis of cancer.

"We have the beachhead successes for the recognition and binding aspects," Soane says. Still, he adds, "it will be a long time between now and when a diagnostic or therapeutic discovery is made." But costumes as good as these seem likely to win a prize or two eventually, for utility if not for beauty.

Chips for Protein Analysis

For the past several years, the fluorescent glow of DNA chips has signaled a revolution in researchers' ability to detect nucleic acids and monitor gene activity in living cells. But developing ways to keep track of the many different proteins in a cell has been much more difficult. Although techniques like the polymerase chain reaction can amplify scarce DNA into detectable amounts, the tiny concentrations of proteins in cell extracts, blood, and other biological samples can't be boosted so easily. But a new tool might help with protein analysis: the ProteinChip technology developed by scientists at Ciphergen Biosystems Inc. of Palo Alto, California.

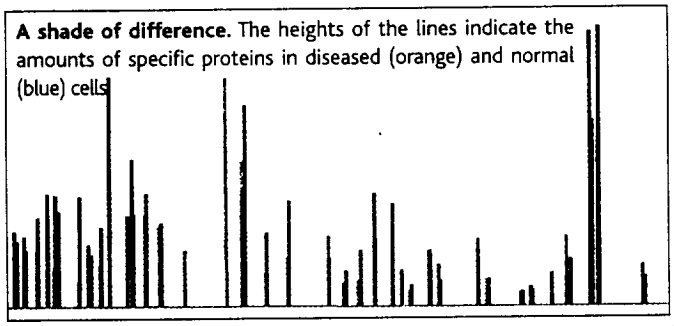
Because the Ciphergen method combines a tiny chip with a "sticky" surface with the sensitive analytic capabilities of mass spectrometry, it doesn't require an amplification step. Consequently, it is not only very fast but can be used with small samples—microliters instead of the milliliters of conventional methods. "They're tackling one of the core problems of analyzing proteins: looking at proteins that are present in very low abundance," says Jeff Wiseman of SmithKline Beecham Inc. in King of Prussia, Pennsylvania. The method should allow

CREDIT: CIPHERGEN BIOSYSTEMS INC.

scientists to discover new proteins, purify them, measure their quantities, and discern their chemical and biological properties, all with one chip.

The technology is the brainchild of William Hutchens of Ciphergen and his colleagues. The chip, which is about a millimeter across, holds some kind of molecular bait—antibodies, carbohydrates, receptors, or any of a wide variety of smaller synthetic chemicals—that can trap many different proteins at once. To perform an analysis, a researcher applies a sample to a chip, lets the proteins adhere to it, and then washes away anything that doesn't stick.

In the next step, a laser zaps the chip surface with just enough energy to break noncovalent bonds and release the proteins. An electric field shoots these proteins to the detector of a mass spectrometer, which reads out their molecular weights. (The company calls the process Surface-Enhanced Laser Desorption/Ionization or SELDI.)



A shade of difference. The heights of the lines indicate the amounts of specific proteins in diseased (orange) and normal (blue) cells.

CREDIT: CIPHERGEN BIOSYSTEMS INC.

Knowledge of the chemical nature of the molecular bait combined with the molecular weights of the proteins permits one particularly useful analysis: producing fingerprints of the protein composition of samples containing hundreds or thousands of proteins. By comparing closely related samples—blood serum from a healthy person and from someone with a disease, for example, or extracts of dividing and nondividing cells—scientists can detect changes in the amounts and types of proteins.

According to John Quackenbush, a molecular biologist at The Institute for Genomic Research in Rockville, Maryland, such changes can provide valuable clues to which chores the proteins are performing. In one set of experiments, for example, he compared the protein content of dividing and nondividing cells of the bacterium *Haemophilus influenzae*. The analysis picked up 600 of the 1740 proteins thought to be encoded in the *H. influenzae* genome, and for about 30 to 60 of them, the amount varied under the two conditions. "It's pretty extraordinary to be able to sit down and, in the course of a few hours, get information about 600 distinct proteins," Quackenbush says.

Once an interesting protein is identified—say, one that is made in large amounts in dividing cells but not in quiescent cells—the chips can be used to characterize, isolate, and even sequence it. By systematically testing different combinations of wash conditions and chip materials, researchers can use SELDI to explore the properties of a protein, such as how strongly it adheres to a surface with positive charge and whether it binds a particular metal ion. Furthermore, Ciphergen's computer programs can identify the combination of surface and wash conditions under which the protein of interest has the fewest neighbors, opening the way to purifying the protein.

Scientists can sequence the protein on the chip by exposing it to enzymes that release its peptides one by one, so that they can be analyzed by mass spectrometry. Conventionally, Hutchens says, different tools had to be used for each step from discovering a protein to isolating and characterizing it. But with SELDI, he adds, "you use exactly the same chip [for all three steps]."

Computer Models for Drug Testing

Companies developing new drugs usually face a great leap into the unknown when they move from test tube and animal studies into clinical trials. Given that it takes at least $20 million just to get a drug into human efficacy tests, failures can be expensive. One critical choice comes early: which of the many disease-related molecules should be targeted. "If you don't make the right choice of drug target at the beginning, you can really have a big mistake at the end," says Robert Dinerstein of Hoechst Marion Roussel Inc. in Bridgewater, New Jersey. Now scientists at Entelos Inc. in Menlo Park, California, are trying to reduce the guesswork by simulating diseases—and the molecular interactions that underlie them—in a computer.

At the meeting, Tom Paterson of Entelos reported that the company had so far built models for three common diseases: asthma, obesity, and HIV/AIDS. Each one seeks to combine what's known about the molecular and cellular changes leading to the disease with the symptoms it causes. The Entelos system "links the basic processes to their consequences in the entire system," says Dinerstein, who has used the asthma program in his work on respiratory diseases. "That hasn't been done before."

Using these programs, researchers can conduct virtual experiments to pretest drugs, modeling what happens when a drug alters the activities of a specific molecule. So far the models have helped pharmaceutical companies develop new hypotheses about mechanisms of disease and evaluate existing and novel therapeutic approaches.

To construct the models, the Entelos team formulates mathematically based hypotheses about how all the components in the disease system interact. With asthma, for example, they incorporate what is known about the role of inflammatory cells and the factors they make and respond to in constricting the respiratory airways. The researchers then tune the math and the relationships between the different parts of the model until it accurately reflects the way the disease behaves. The simulation can then show what happens to any one component of the system in response to a change in another part of it—caused, say, by administering a drug or exposing the airways to allergens. "There's nothing quite this comprehensive," says one of Entelos's scientific advisers, bioengineer Douglas Lauffenburger of the Massachusetts Institute of Technology.

Dinerstein tested the asthma simulation by seeing how it responds to existing drugs. He found in the model exactly what companies had learned from clinical trials: Effective drugs decrease airway resistance, while ineffective drugs, including some that companies had pursued quite aggressively, don't.

Using the asthma program, Dinerstein's group also carried out a virtual experiment in which they blocked the activity of a certain inflammatory factor to see if it might be a good target for an inhaled form of therapy. The next asthma attack was worse because another part of the body was compensating for the decreased inflammatory response. "We hadn't really thought about the rebound effects," says Dinerstein.

In addition, the software provides information management tools with quick connections to the literature references and the mathematics on which a given part of the model rests. Researchers can also incorporate their results into the program and chronicle the evolution of their thinking.

Different parts of the model vary in reliability, depending on the information available. But as Dinerstein notes, even the gaps can help because they point out where researchers should direct their studies. Now that scientists are investing a large effort toward finding the sequence and function of all the human genes, such models are badly needed, says Lauffenburger. "The promise of this whole new field of molecular medicine requires that we get an idea of the consequences of molecular alterations. Until you can put together models like this, it's all pretty much guesswork."

–EVELYN STRAUSS

Evelyn Strauss is a writer in Berkeley, California.

ATTACHMENT 3

induces a different feeding structure named syncytium. While *Mi* resistance results in localized tissue necrosis, no such hypersensitive reaction is observed with $Hs1^{pro\text{-}1}$. As judged from sequence analysis, $Hs1^{pro\text{-}1}$ is also is part of a signal cascade, but seems to be located in the plasma membrane whereas Mi is cytoplasmic.

These findings have two major implications for future breeding strategies and on our understanding of host–pathogen interactions. At first glance, it seems puzzling that one gene causes resistance to species that obviously have nothing in common except that both make use of their stylets to feed from the host tissue. However, the aphid inserts its stylet intercellularly to feed from the phloem, whereas the nematode penetrates the cell membrane with its stylet to take up nutrients from inside the cell. As it is extremely unlikely that the Mi protein reacts with the same ligand on nematodes and aphids, it cannot be the primary receptor, and probably partakes in an intracellular signal transduction cascade that starts with ligand binding to yet unknown receptors and ends with the activation of defense-related genes within the nucleus that ultimately results in localized cell death.

The prospects for breeding new disease resistant tomato varieties are immediate, and it is tempting to think of transforming unrelated crop species that are attacked by aphids and root knot nematodes. Some caveats, however, would urge caution in the implementation of this attractive strategy. It seems unlikely that *Mi* will be effective against cyst nematodes because of the lack of homology to the $Hs1^{pro\text{-}1}$ gene. Also, breeders must be aware of resistance breaking aphids frequently evolving from the huge number of individuals in the field. One such isolate has already been shown to break the *Mi* resistance in Motelle and transgenics[2]. As *Mi* resistance follows a one-to-one relationship, in which a virulence gene of the pathogen matches a resistance gene from the host single, mutational events can result in virulent isolates that pose a threat to all varieties.

1. Vos, P. et al. *Nat. Biotechnol.* **16**:1365–1369.
2. Rossi, M. et al. *Proc. Natl. Acad. Sci. USA* **95**:9750–9754.
3. Milligan, S.B. et al. *Plant Cell* **10**:1307–1319.
4. Kaloshian, I. et al. *Mol. Gen. Genet.* **257**:376–385.
5. Messeguer, R. et al. *Theor. Appl. Genet.* **82**:529–536.
6. Jones, J.D.G. *Curr. Opin. Biotechnol.* **7**:155–160.
7. Cai, D. et al. *Science* **275**:832–834.

Modeling the oddities of biology

Bijal Trivedi

During the opening session of "After The Genome IV"*—a conference recently convened to discuss future directions for research following completion of the human genome sequence—Stephen Grand, chief technology officer of CyberLife Technology (Cambridge, UK), told the audience about his perfect pets or "norns." These are cute, furry, haploid, computerized "creatures," with about 400 genes, reproducing via meiosis, with chromosomes that align and crossover, creating duplications, deletions, and mutations. After random destruction of one genome, they have viable offspring.

Norns are one of the more sophisticated attempts to model life processes based on information at the genetic level. They are able to learn as their neural networks become reinforced by neurotransmitters—which punish and reward as they interact with their environment—and they eat, sleep, age, and suffer from disease similar to their biological brethren. They are also an extreme example of an unconventional interdisciplinary approach to biology "after the genome."

The conference of the same name attempted to initiate and foster collaborations between researchers from academia and industry, and to introduce them to government officials and representatives from philanthropic organizations who hold the purse strings and the power to trigger change at the institutional level. Although representatives from the latter two groups were few, the eclectic mix of researchers—biologists, physicists, mathematicians—interacted with each other with the initial awkward anticipation of

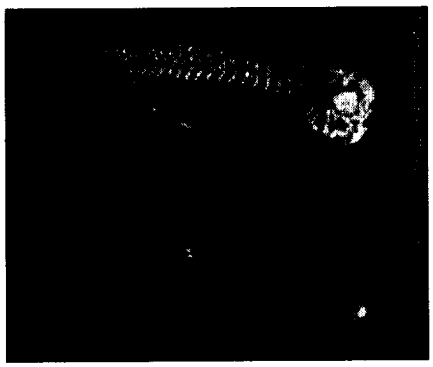


Figure 1. An MRI image tracing gene expression in the *Xenopus* embryo, as reported at the meeting.

courtship: Few expected a "grand unified project" to emerge because of the diverse spectrum of expertise, but many found the exposure to different fields refreshing. Several initiated new collaborations.

Three presentations, which were universally well received, all emphasized modeling. Thomas Paterson of Entelos (Menlo Park, CA) presented a software package for simulating complex disease states such as asthma and obesity. This method of disease analysis takes a "top down" approach, pinpointing the most debilitating clinical aspects of a disease and then searching for the cellular, molecular, and environmental mechanisms involved.

The most ambitious aspect of the software is that it attempts to integrate clinical, physiological, biochemical, and genotypic data into a graphical interface of bubbles and arrows. The interface masks an underlying mathematical infrastructure that specifies precise relationships between molecular species and dynamic processes.

The models assume that cellular processes constantly strive to keep the cell in a homeostatic state. When certain parameters are exceeded, disease may ensue. A global overview of a disease makes it possible to uncover synergies that otherwise remain invisible when pathways and processes are studied in isolation. The software is designed to help pharmaceutical companies determine drug targets and simulate the downstream ramifications of specific drug mechanisms. This should allow them to simulate drug interactions within a healthy system, before they conduct extensive clinical or laboratory experiments.

In another talk, Drew Endy, now at the Molecular Sciences Institute (Berkeley, California), and Ian Molineux of the University of Texas at Austin, described work attempting to simulate a relationship between the genetic architecture and fitness of an organism, and then testing the results in the laboratory. Endy chose as his study system the venerable bacteriophage T7—a virus for which there is now available full sequence data on 122 genes and regulatory regions and extensive information on the bacterial metabolic pathways in which they participate.

Taking the 20 most essential T7 genes, Endy "slid" them back and forth along the genome and looked for the simulated effects of gene position on the viability of the phage. After simulating the effects of moving the

RNA polymerase to three different positions in the genome, he then went back to Molineux, and created the corresponding strains of T7 in the laboratory.

The model predicted that two of these strains of phage would grow more slowly than wild type, whereas the third would grow faster. Laboratory data showed that the model was indeed correct for the first two strains, but incorrect for the third, as the last strain grew slower than wild type. Thus, comparison of T7 phage's simulated growth rate with that of the actual strains provided a measure both of the strength of the model and of current understanding of T7 biology. When the experimental biology does not match the simulation, it telegraphs that a piece of the biological puzzle remains to be found.

Further work should be directed toward ascertaining whether such information can be used to design "fitter" organisms or different genomes for different functions. An obvious application would be to design better vectors for gene/drug therapy.

More immediate in its application was an impressive demonstration by Thomas Meade from the California Institute of Technology (Pasadena, CA) of real-time, three-dimensional visualization of gene expression using magnetic resonance imaging (MRI). This MRI technology uses "smart" contrast agents that provide anatomical information and also reveal the metabolic status of the cell.

In MRI, a magnetic field orders the majority of hydrogen nuclei within an organism to align with, rather than against, its direction. Pulsing radio frequencies then induce a spin flip in the nuclei of hydrogen atoms. When the radio pulse is removed, the nuclei return to their original state, emitting a radio wave. The time taken to return to the unexcited state is called the relaxation time, which is inversely proportional to the intensity of the image created by the computer. An agent that can decrease the relaxation time thus increases the local intensity of the image. Atoms with unpaired electrons are able to reduce the relaxation time; this is why gadolinium, an ion with seven unpaired electrons, is commonly used as a clinical contrast agent.

Meade's group has created a compound in which the gadolinium ion is masked by an enzyme substrate. When the enzyme is presented, the substrate is cleaved and the gadolinium ion is exposed and free to interact with nearby hydrogen nuclei. These substrate masks are enzyme-specific and can trace the activity of individual genes during the development of an organism.

Meade's team presented data obtained after injecting the gadolinium contrast agent into both halves of the two-cell stage of a *Xenopus* embryo. At the 16-cell stage, mRNA for β-galactosidase was injected into one cell that was known from developmental fate maps to differentiate into the left dorsal notochord. An MRI image of the live organism showed that only the left dorsal notochord was lit up, indicating that the gadolinium had been "turned on" only in cells containing the mRNA (see Fig. 1.)

The ability of the technology to differentiate metabolic states of cells suggests many potential clinical applications. The contrast agents may also provide a tool for understanding the pattern of development at the level of temporal gene expression.

Perhaps by 2010, biology will, in fact, have reached the end of the reductionist road, and efforts will be largely directed at reassembling the pieces that took more than half a century to dissect into a more comprehensible whole. In such an environment, meetings like this will become more frequent, especially if the various computer simulations promising to predictively integrate genetic sequence, expression, and physiological data begin to make good on their promises.

**Jackson Lake Lodge in Jackson Hole, WY, October 10–14, 1998.*