Genomes to Life

Michael L. Knotek
ASC-AC
May 2, 2001

GENOMES to
LIFE
ACCELERATING
BIOLOGICAL
DISCOVERY

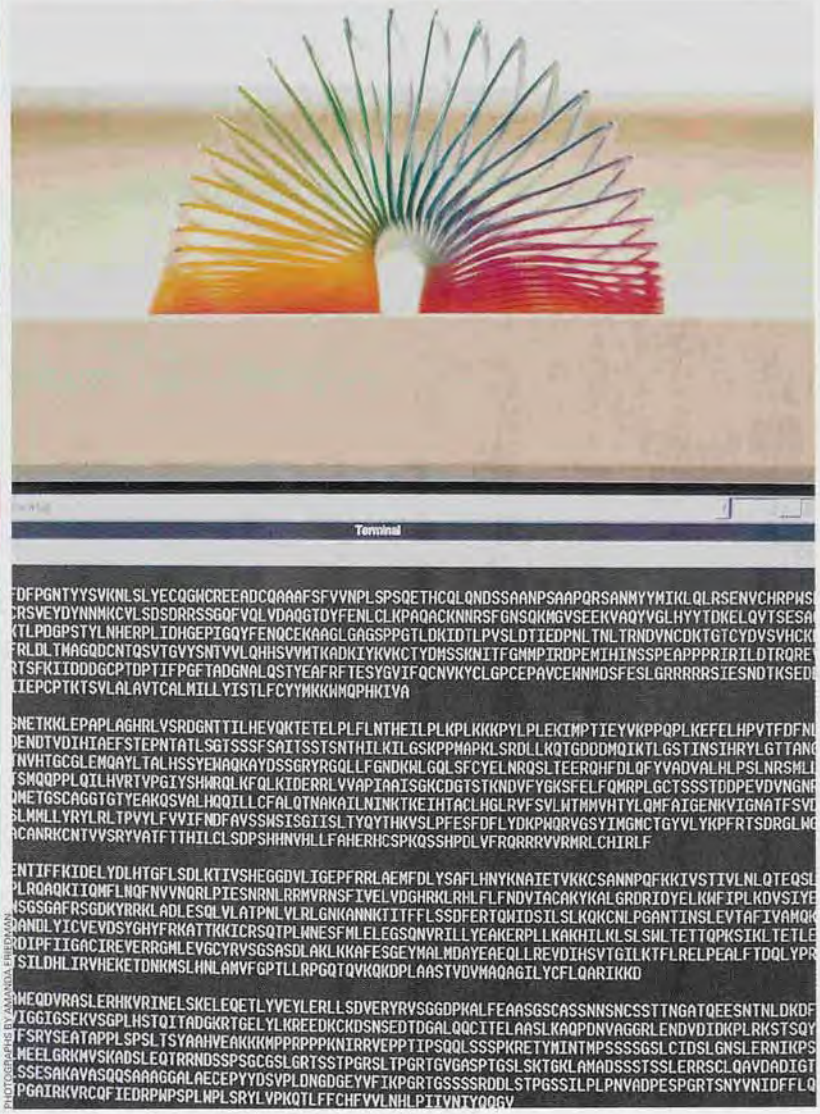
Jobs of the Future

To build better drugs, the exploding field of bioinformatics is looking for highly trained workers comfortable with supercomputing and biology. By Brad Stone

Wanted: Hot Industry Seeks Supergeeks

CRAIG BENHAM HAS A problem. As a professor at Mount Sinai School of Medicine in New York, he trains students in the exploding new field of bioinformatics—the fusion of high-powered computing and biology that is aimed at revolutionizing the health-care industry. But Benham can't keep a postdoctorate researcher for more than a year. They keep leaving for jobs that pay up to $100,000 at bioinformatics start-ups, giant pharmaceutical companies or technology giants like Motorola and IBM that are targeting the rapidly growing life-sciences field. "These companies need a whole new class of biologists who have training in the computational and mathematical methods," Benham says. "I've got one former student who has been hired four times in three years, increasing his salary 30 percent each time. There's huge demand for these skills." Benham knows of what he speaks: this summer he will join the University of California, Davis, heading up its new $95 million bioinformatics program.

Bioinformatics encompasses its more widely known offspring, genomics, the study of genes and their function. It involves the use of supercomputers, vast databases and complex software to analyze the mountain of data that has emerged from the sequencing of the human genome, whose mapping was completed last summer. The Human Genome Project was a decadelong effort that cataloged 3 billion DNA base pairs—the chemical components that together spell out the genes that make us who we are. But the data that were supposed to change the world are merely three gigabytes of raw letters, a jumble of A's, T's, G's and C's (which stand for the chemical ingredients



PHOTOGRAPHS BY AMANDA FRIEDMAN

FLY WITH SLINKY: Genetic code of the fruit fly on the computer of a bioinformatics student

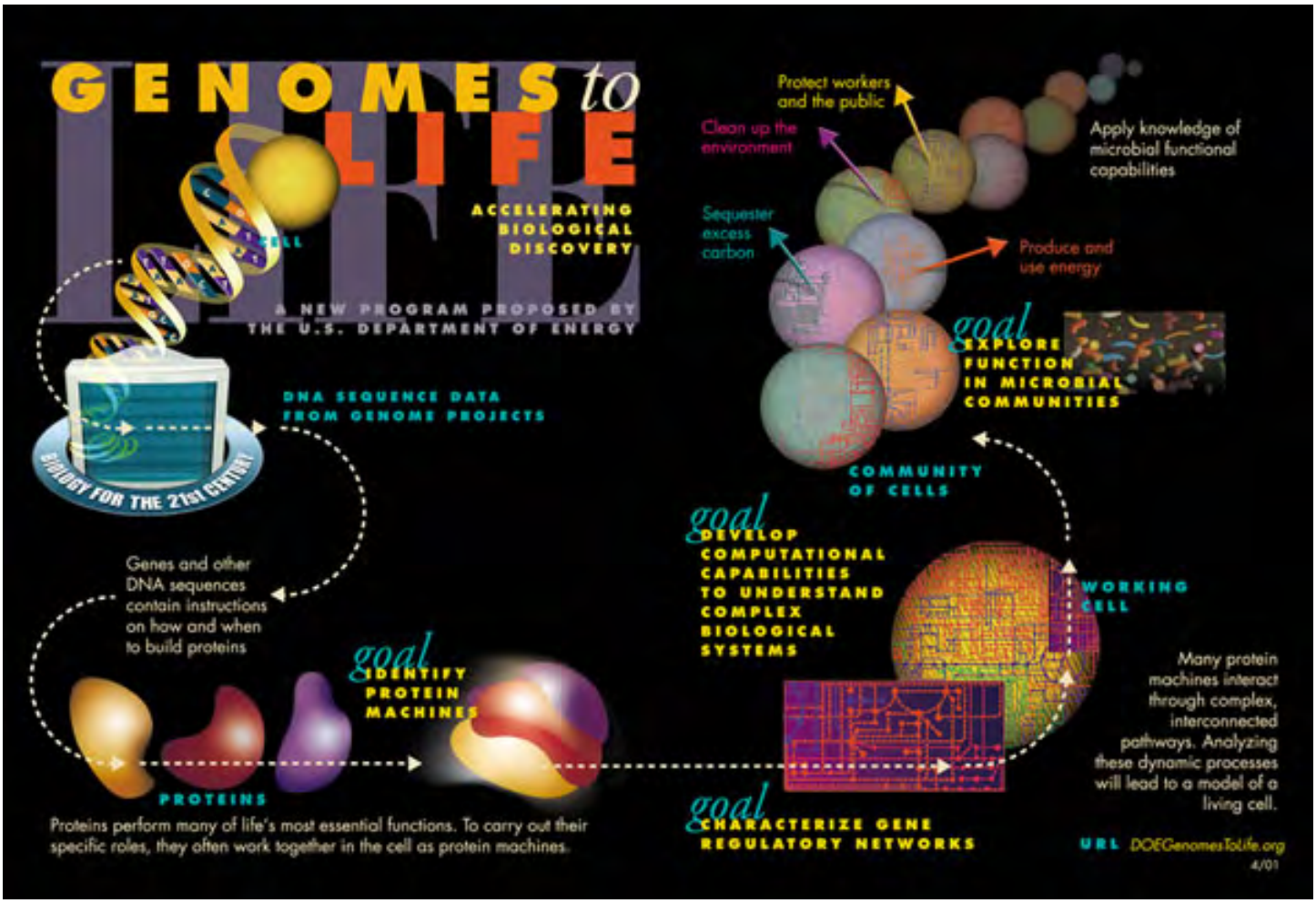
DOE/SC-0036

GENOMES *to* LIFE

ACCELERATING BIOLOGICAL DISCOVERY

Program proposed by the
Office of Biological and Environmental Research
and
Office of Advanced Scientific Computing Research
of the
U.S. Department of Energy
April 2001

DOEGenomesToLife.org


GENOMES to LIFE
ACCELERATING BIOLOGICAL DISCOVERY
A NEW PROGRAM PROPOSED BY THE U.S. DEPARTMENT OF ENERGY
DNA SEQUENCE DATA FROM GENOME PROJECTS
BIOLOGY FOR THE 21st CENTURY
Genes and other DNA sequences contain instructions on how and when to build proteins
goal IDENTIFY PROTEIN MACHINES
PROTEINS
Proteins perform many of life's most essential functions. To carry out their specific roles, they often work together in the cell as protein machines.
Protect workers and the public
Clean up the environment
Apply knowledge of microbial functional capabilities
Sequester excess carbon
Produce and use energy
goal EXPLORE FUNCTION IN MICROBIAL COMMUNITIES
COMMUNITY OF CELLS
goal DEVELOP COMPUTATIONAL CAPABILITIES TO UNDERSTAND COMPLEX BIOLOGICAL SYSTEMS
WORKING CELL
Many protein machines interact through complex, interconnected pathways. Analyzing these dynamic processes will lead to a model of a living cell.
goal CHARACTERIZE GENE REGULATORY NETWORKS
URL DOEGenomesToLife.org
4/01

Genomes to Life

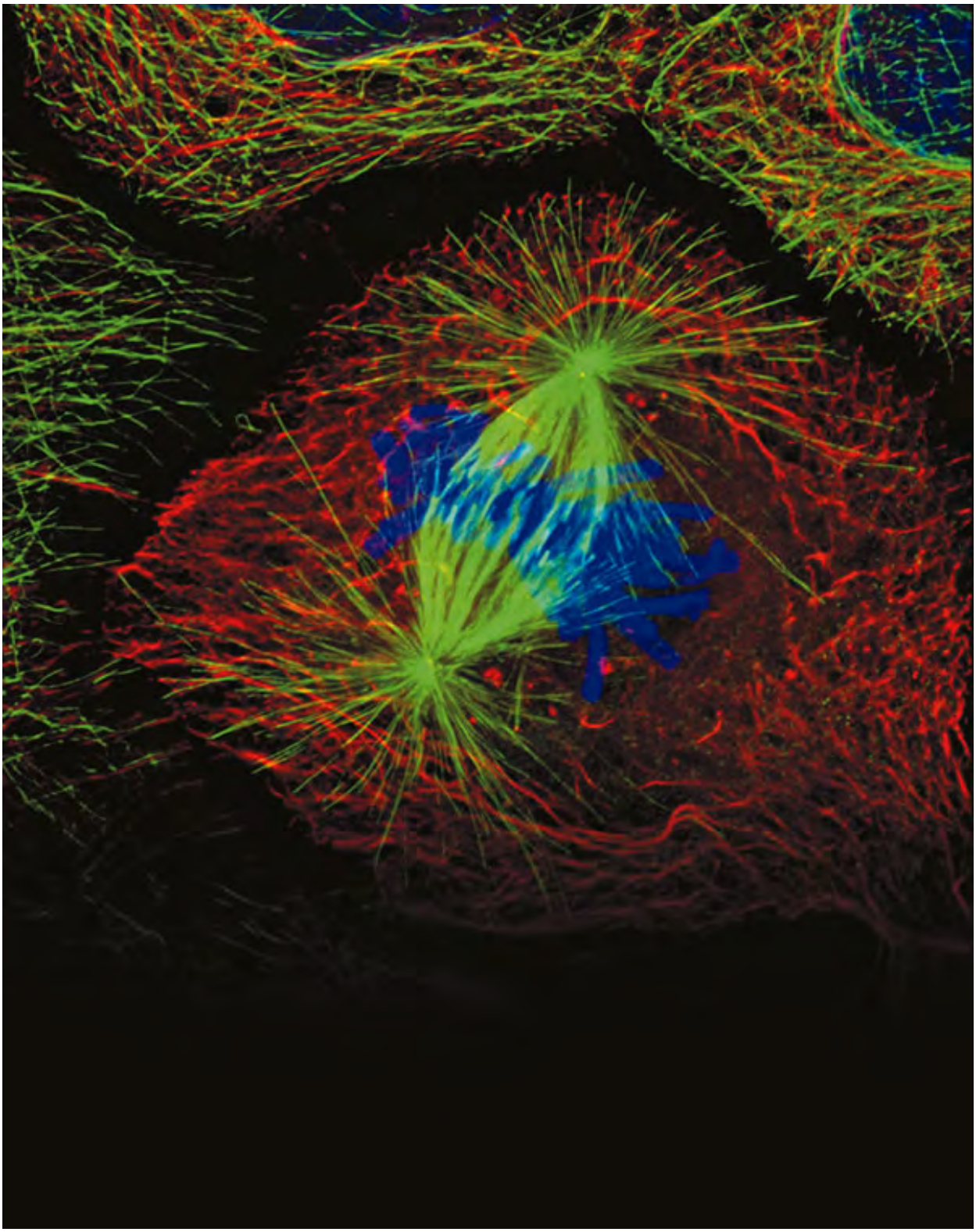
- Genomics and 21st Century Biology: ***Achieving a fundamental, comprehensive, and systematic understanding of Life***
  - Comprehensive Strategy
  - High Throughput
  - Data Intensive
  - Genomics - Expression (Cis-acting + transcription) - Proteomics - Complexes - Pathways - Networks - Processes - Cells - Communities and Organisms

COMPUTING!

"The Killer APP!!"

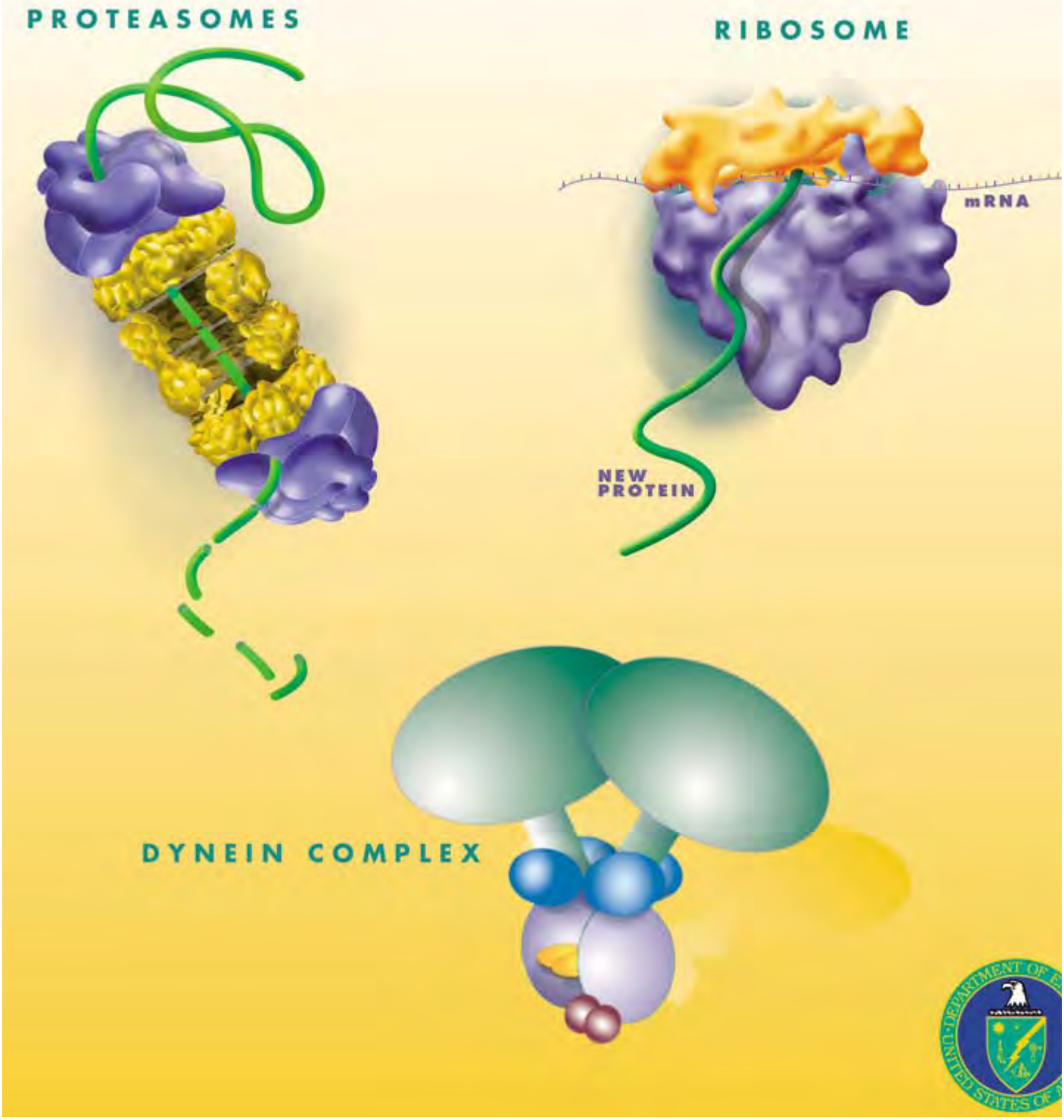
Genomes to Life

- The Microbial Cell Project
  - From Genomes to Cells
  - First Installment on GTL
  - Comprehensive View of Biology on Targeted Systems of Importance to DOE
  - End Goal of Modeling Realistic Simple Cell
  - Intellectual Litmus Test of GTL



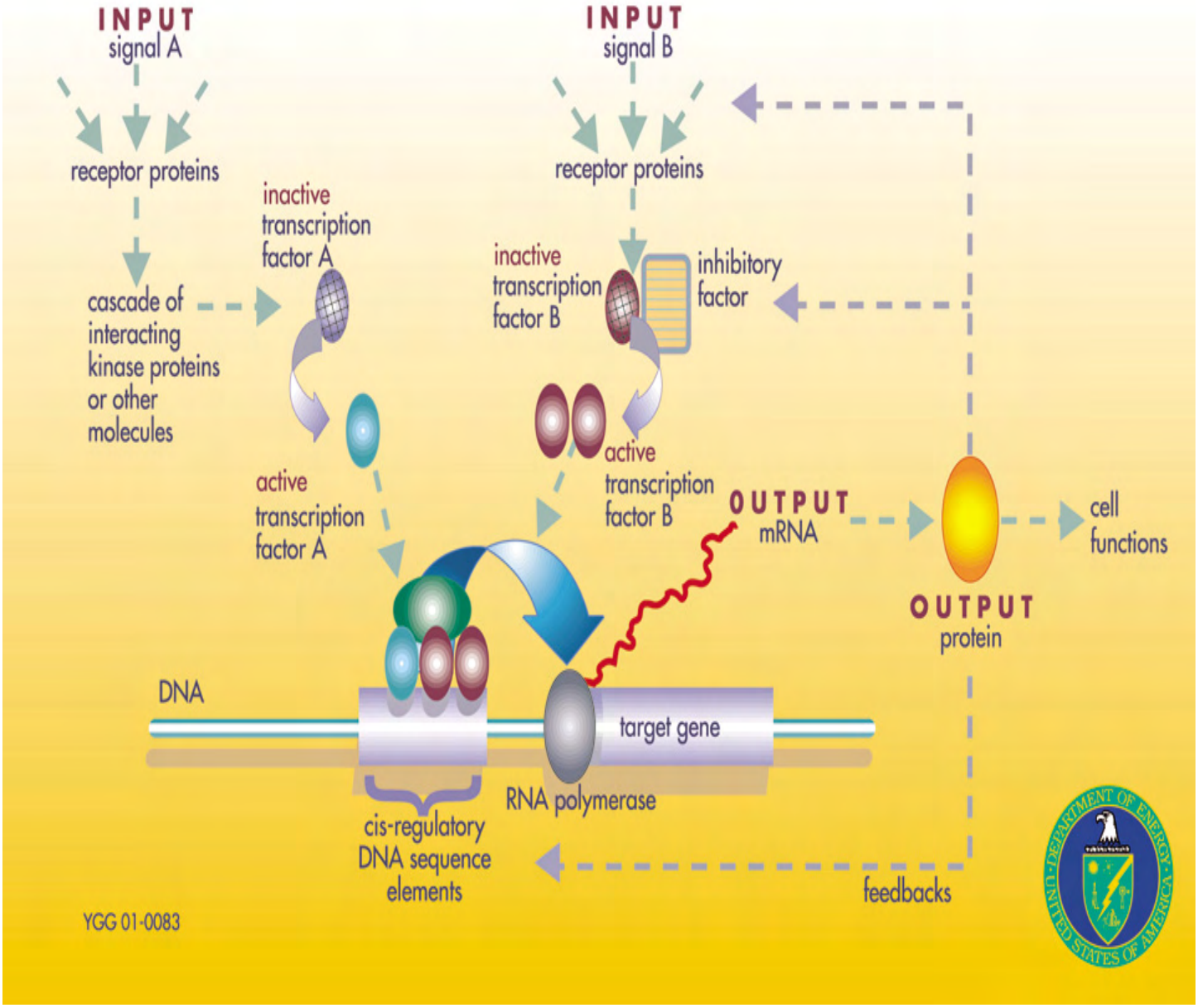
GENOMES to LIFE

MOLECULAR MACHINES OF LIFE

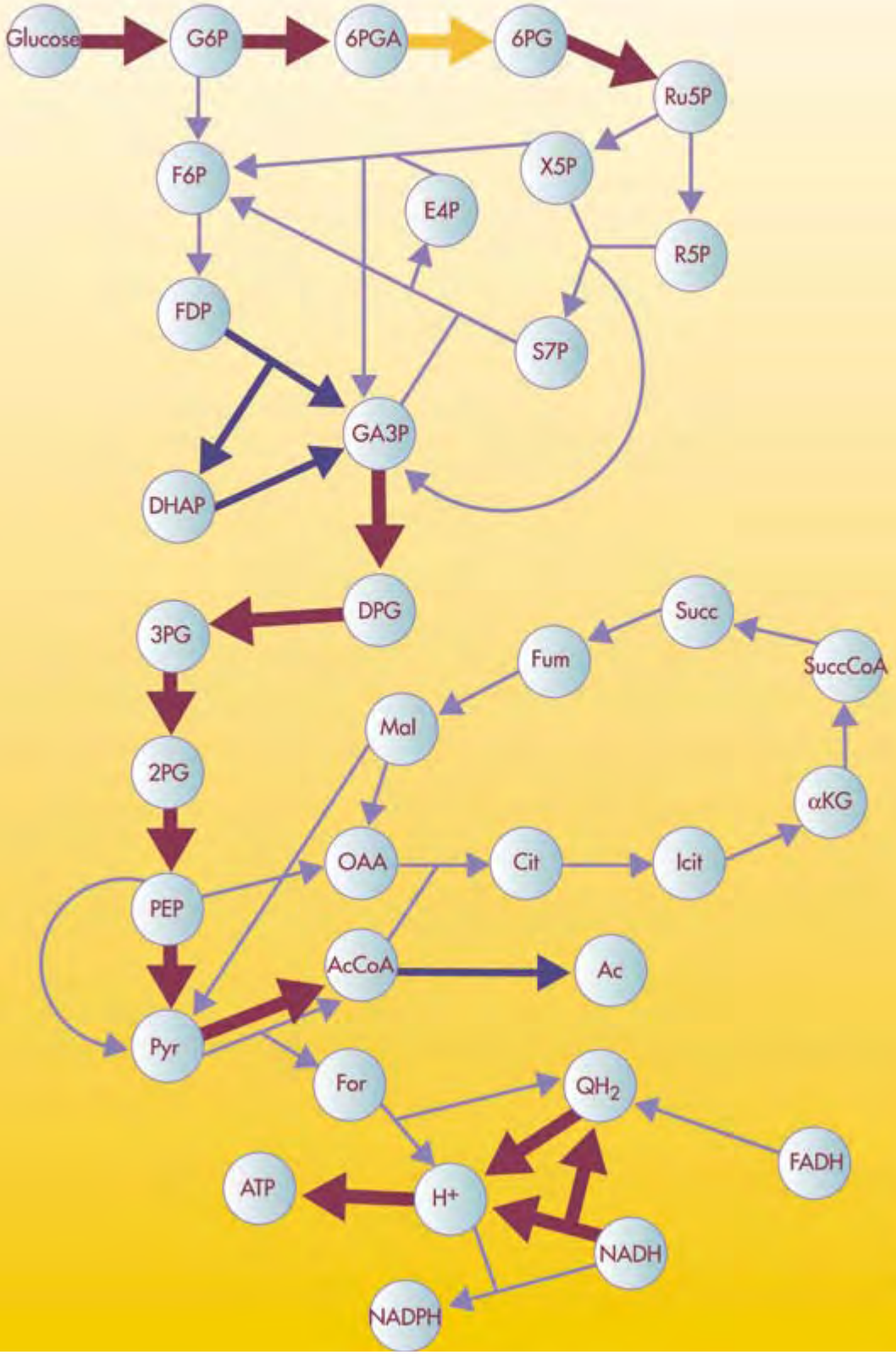


GENOMES to LIFE

A GENE REGULATORY NETWORK



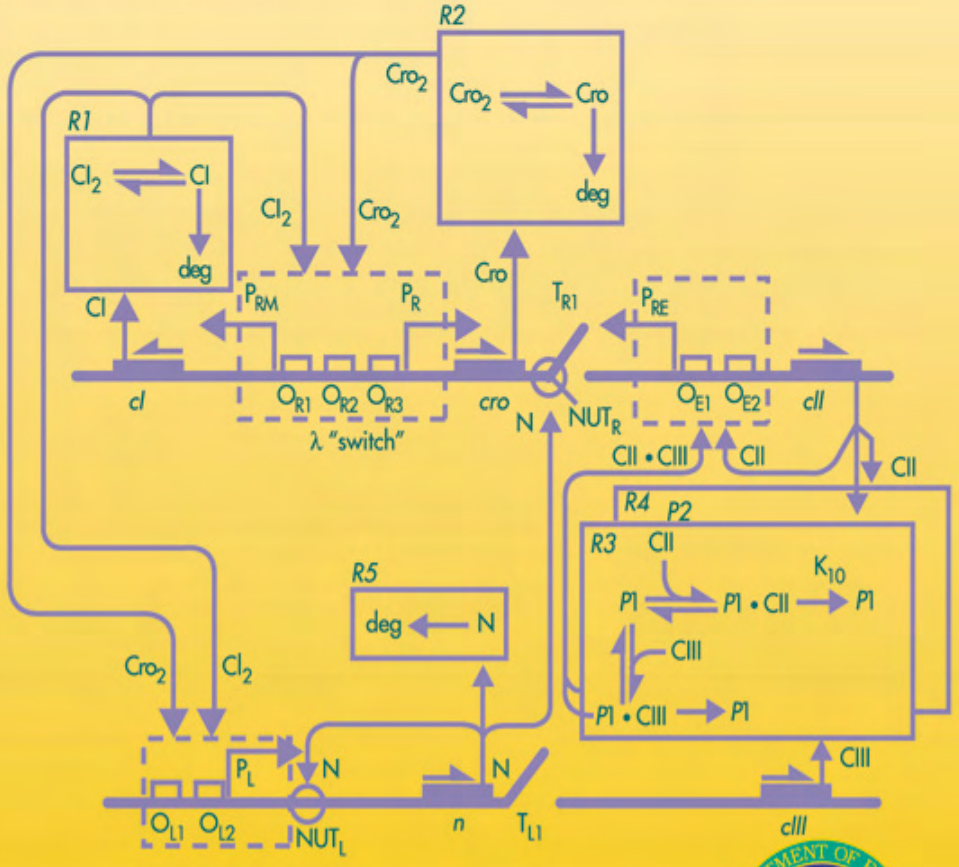
GENOMES to LIFE



GENOMES to LIFE

COMPUTATIONAL CAPABILITIES TO UNDERSTAND AND PREDICT COMPLEX BIOLOGICAL SYSTEMS

IBM SP supercomputer at Oak Ridge National Laboratory



Virus pathway kinetics model

GENOMES to LIFE

IDENTIFY AND CHARACTERIZE THE MOLECULAR MACHINES OF LIFE

goal 1

GENOMES to LIFE

CHARACTERIZE GENE REGULATORY NETWORKS

goal 2

GENOMES to LIFE

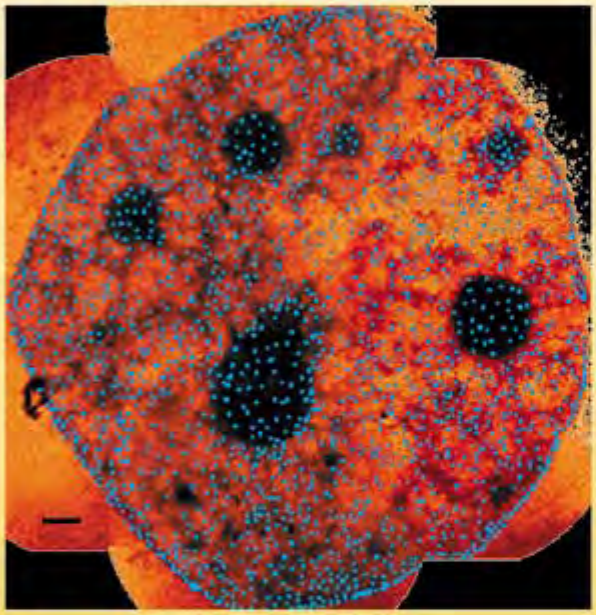
CHARACTERIZE THE FUNCTIONAL REPERTOIRE OF COMPLEX MICROBIAL COMMUNITIES IN THEIR NATURAL ENVIRONMENTS AT THE MOLECULAR LEVEL

goal 3

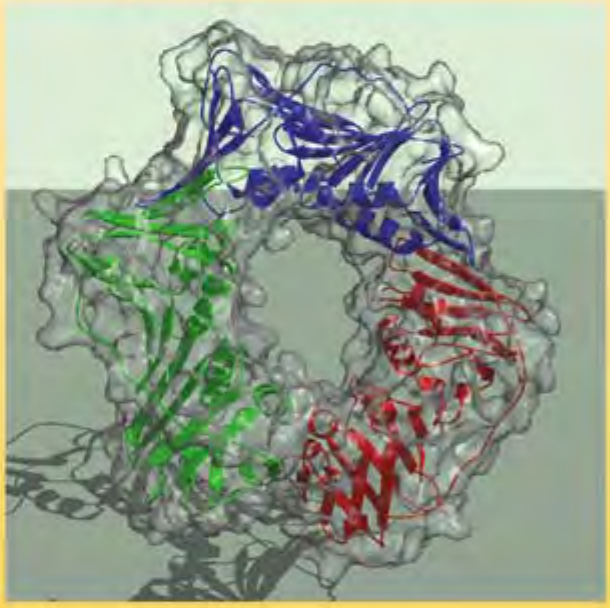
GENOMES to LIFE

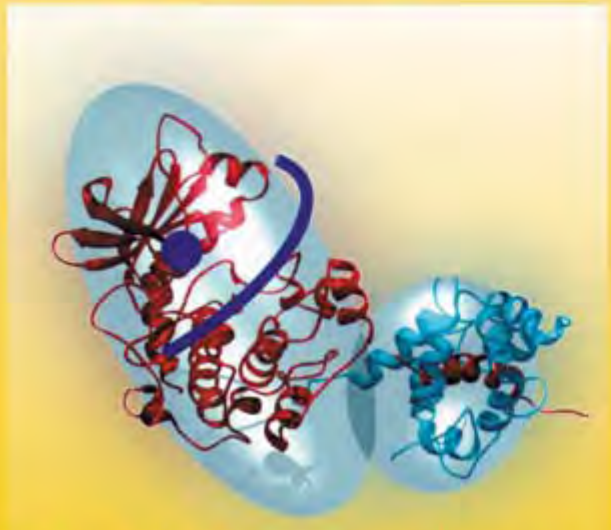
DOE STRENGTHS AND CAPABILITIES

Soft X-ray image of a human mammary cell from Lawrence Berkeley National Laboratory. Blue dots label proteins of the nuclear pore complex.



The role of the Rad checkpoint complex was inferred from the 3-D structure predicted by comparative modeling at Lawrence Livermore National Laboratory.





Details on the binding and dynamics of an enzyme and its activator calmodulin were revealed using a combination of technologies at Los Alamos National Laboratory.

GENOMES *to* LIFE

DOE STRENGTHS AND CAPABILITIES

Environmental Molecular Sciences Laboratory's 800-MHz nuclear magnetic resonance spectrometer at Pacific Northwest National Laboratory

Advanced Photon Source at Argonne National Laboratory

Site plan of the Spallation Neutron Source being built at Oak Ridge National Laboratory in collaboration with ANL, BNL, LBNL, and LANL

GENOMES to LIFE

DOE STRENGTHS AND CAPABILITIES

Advanced Light Source at
Lawrence Berkeley National Laboratory

Mass spectrometer in the
Environmental Molecular Sciences
Laboratory at Pacific Northwest
National Laboratory

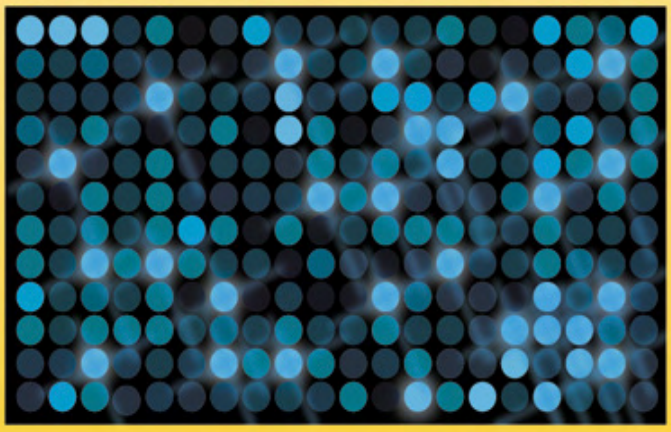
A beamline at the National
Synchrotron Light Source at
Brookhaven National Laboratory

GENOMES to LIFE

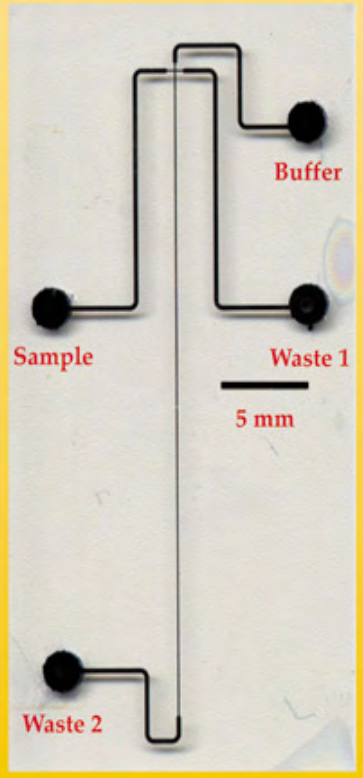
DOE STRENGTHS AND CAPABILITIES

Next generation DNA sequencing technology from University of California, Berkeley

Production Sequencing Facility at DOE's Joint Genome Institute



Microarrays for simultaneous analyses of tens of thousands of samples at Oak Ridge National laboratory

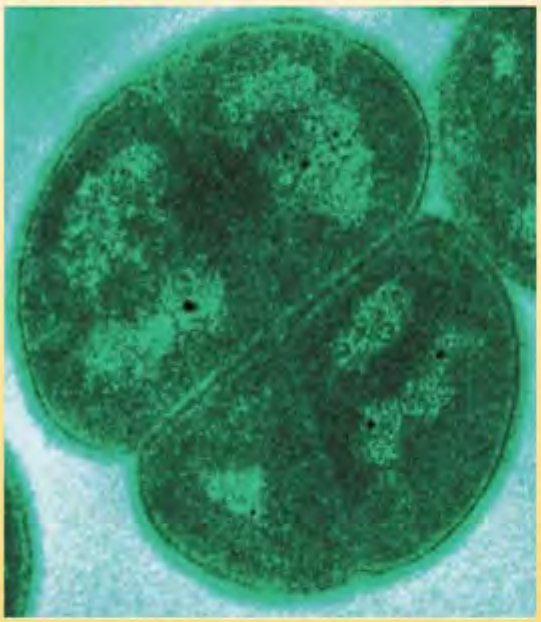


Microfabricated electrophoresis device ("lab on a chip") at Oak Ridge National Laboratory

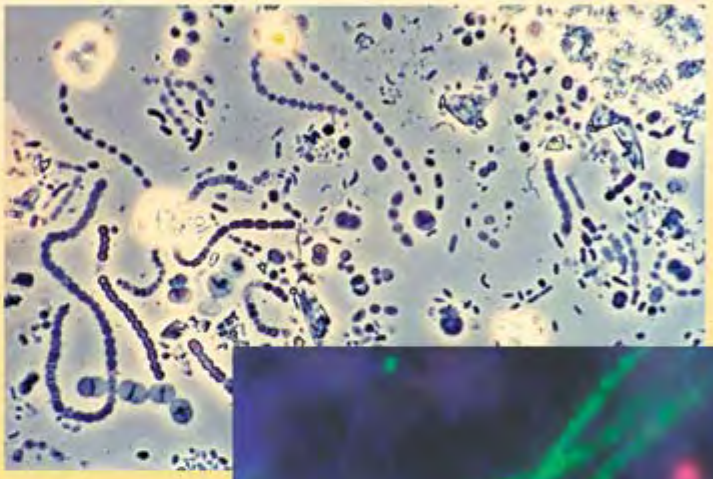
GENOMES to LIFE

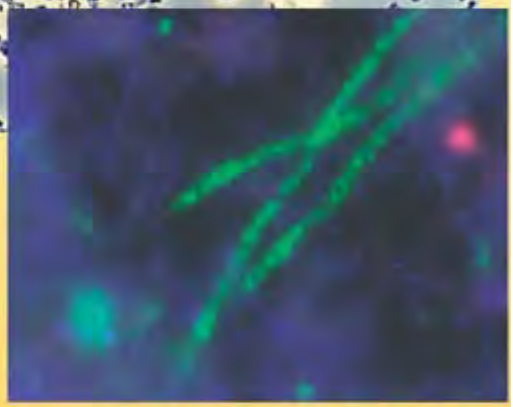
FUNCTIONAL REPERTOIRE OF NATURAL MICROBIAL COMMUNITIES

Deinococcus radiodurans



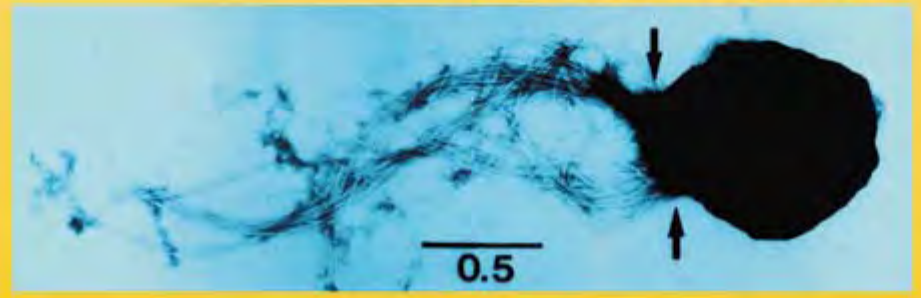
Microbial community





Uncultured microbes

Methanococcus jannaschii



GENOMES to LIFE

DEVELOP THE COMPUTATIONAL METHODS AND CAPABILITIES TO ADVANCE UNDERSTANDING OF COMPLEX BIOLOGICAL SYSTEMS AND PREDICT THEIR BEHAVIOR

goal 4

- Assemble and annotate genomes
- Analyze protein-expression and protein-complex data
- Derive and model metabolic pathways and regulatory networks
- Model microbial cell functions (Microbial Cell Project)
- Model and simulate microbial community actions (Microbial Cell Project)

INFRASTRUCTURE FOR THE NEW BIOLOGY

- Databases and data integration
- High-performance computing tools
- Modeling and simulation codes and theory
- Visualization and user interfaces