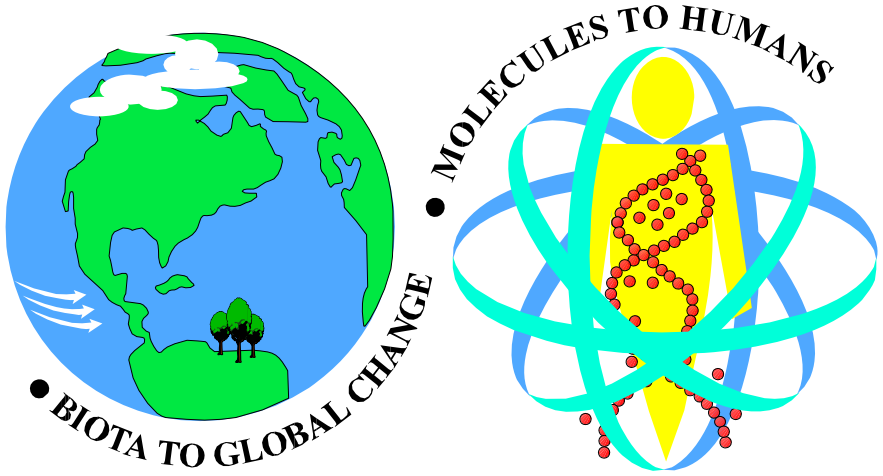
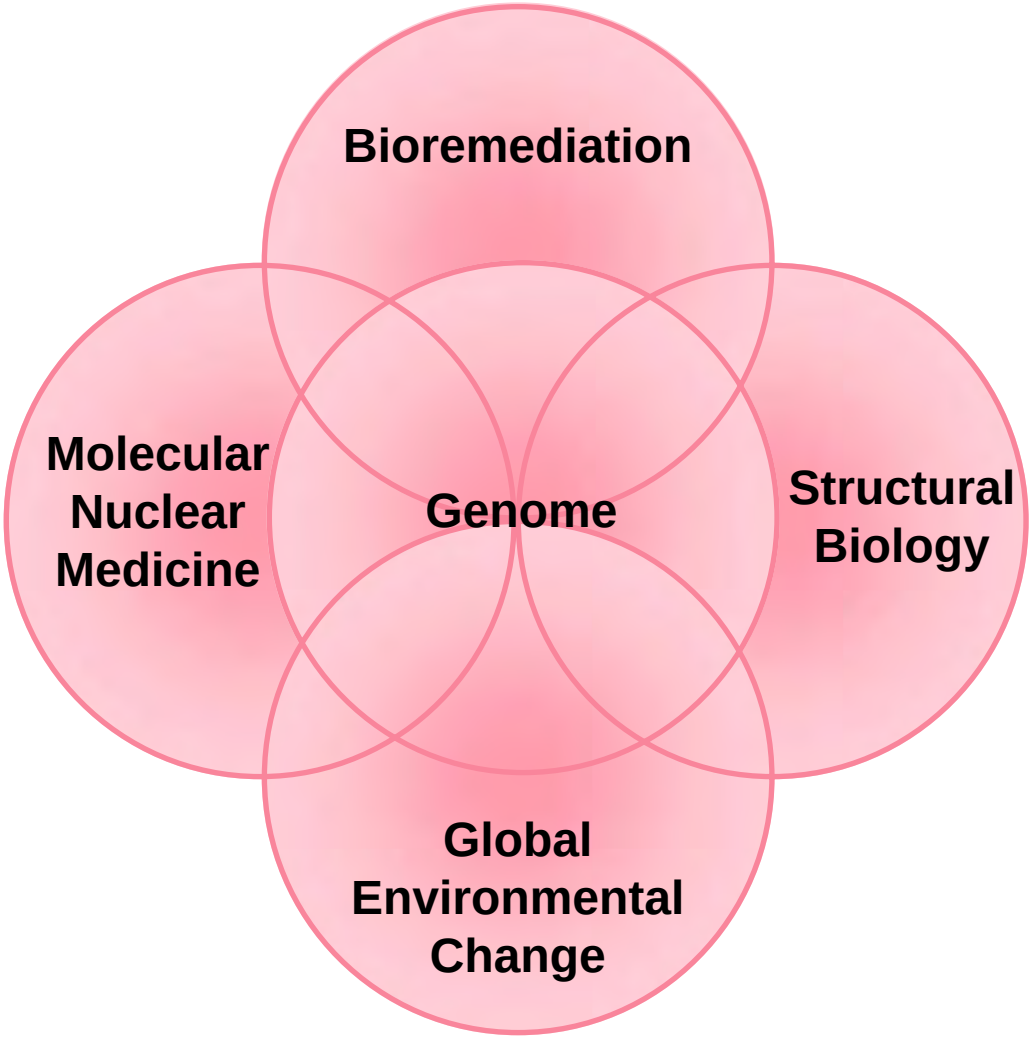
Computational Needs and Challenges in the Age of Genome-Scale Biology

Meeting of the Advanced Scientific Computing Advisory Committee

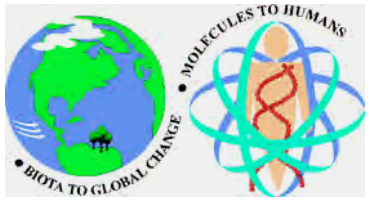


Ari Patrinos, Ph.D.
Office of Biological and Environmental Research
Office of Science
May 2-3, 2001

BER Program



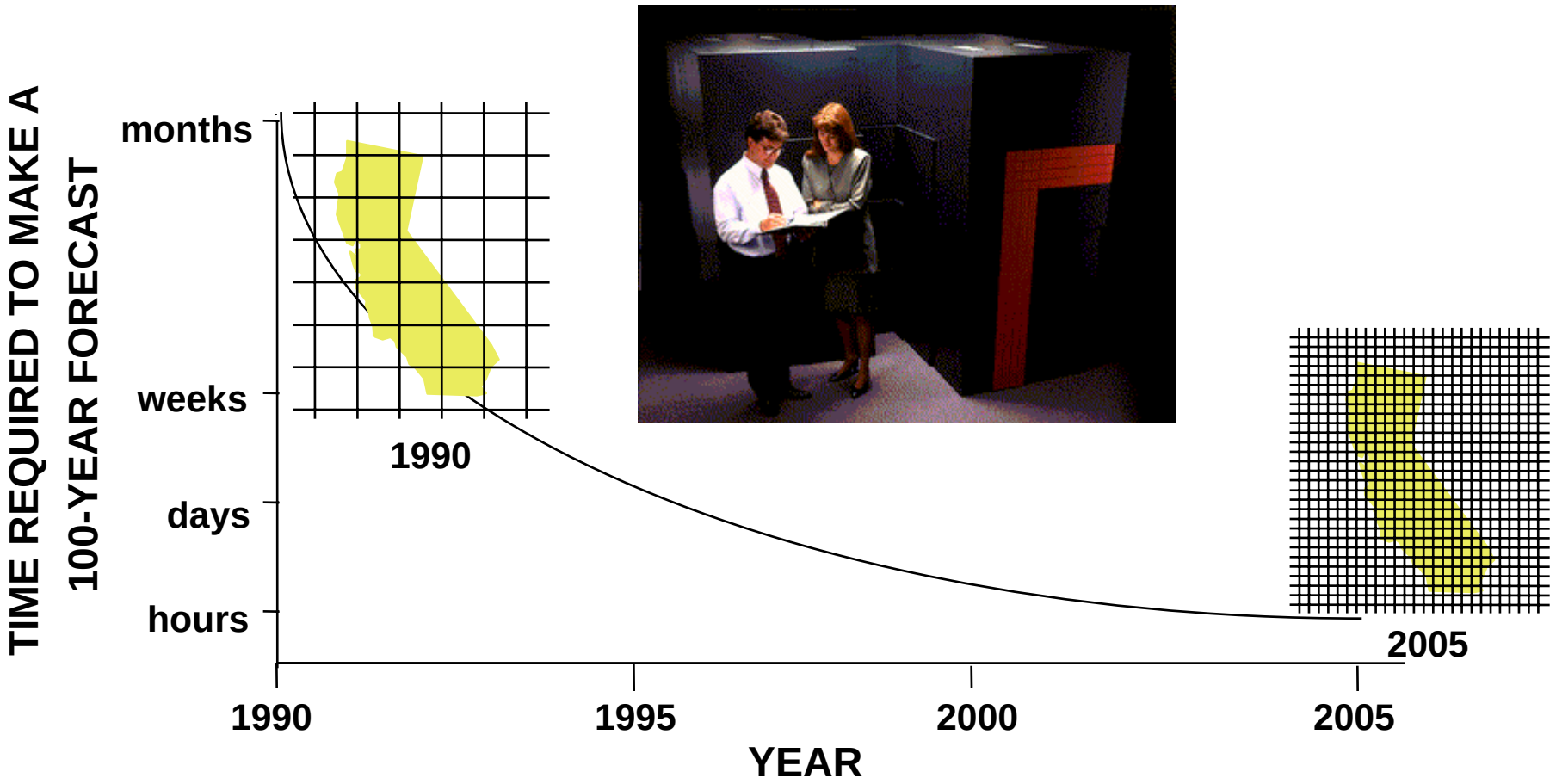
A diverse research portfolio driving science at the interface



BER and ASCR

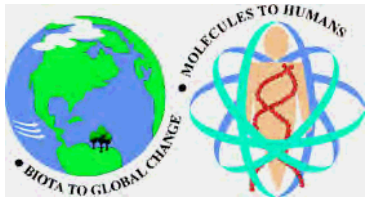
- an ongoing partnership addressing scientific challenges at the interface of computing and global change research
- a new partnership meeting the needs of tomorrow's biology

The DOE Climate Change Prediction Program Links Climate Research to TERASCALE Computing



Higher resolution, more accurate climate models will replace uncertain global trends with accurate regional forecasts

Industry Meeting the Computational Challenges in Biology

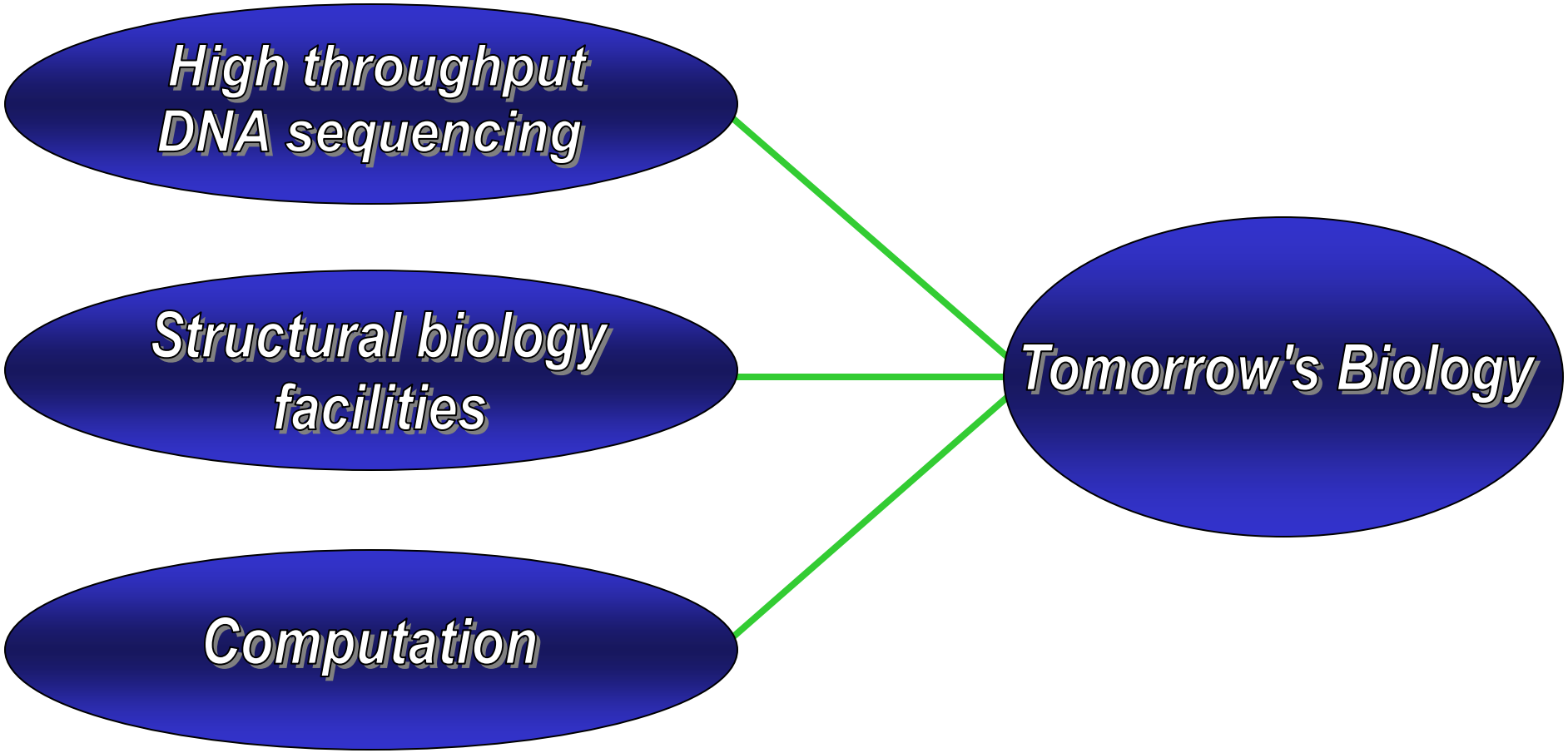


Big Blue Aims to Crack Protein Riddle
Science, December 17, 1999

IBM plans $100M biotech research unit
USA Today, August 16, 2000

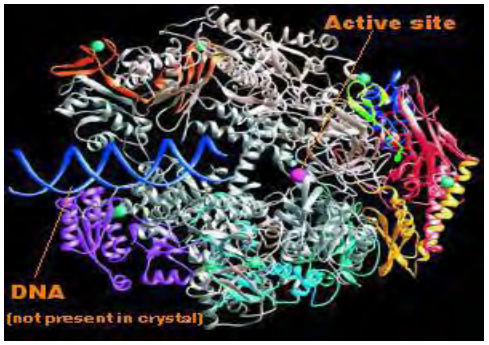
Compaq, Celera Genomics & Sandia National Lab
Announce Research & Development Alliance
To Design 100 TeraOPS Supercomputer Specifically for Life Sciences Use
January 11, 2001

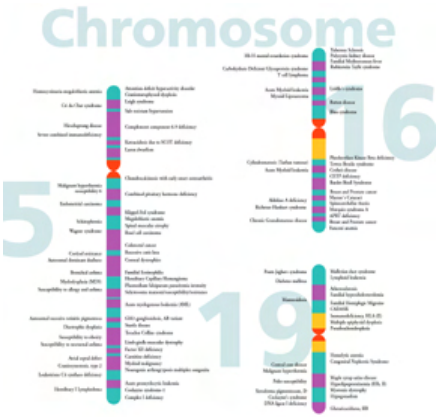
DOE Provides Technology for Tomorrow's Biology

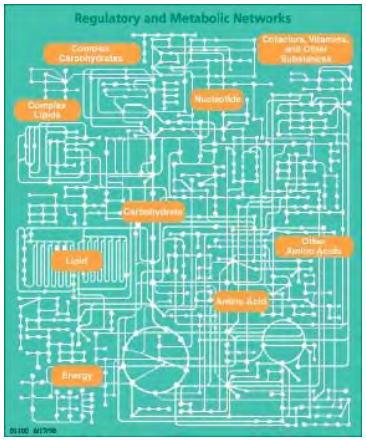


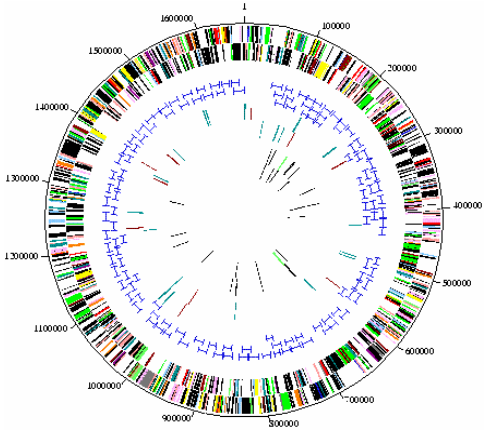
Computing – a key component of BER's biology program

- Structural biology
- Genomics
- Genomes to Life – the next step in biology

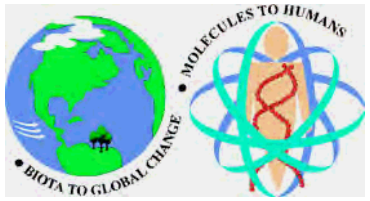








Needed - A new relationship between biology and mathematics/ computer science



- Traditional relationship - service
  - what tools can the mathematicians/ computer scientists give me to solve my problems?
- New (for biologists) paradigm - partnership
  - what are the fundamental underlying mathematical/computational principles of living systems?

High Performance Computers Will be Needed for Each Component of Cell-level Simulations

Genome assembly:

>10 TeraFlops sustained speed required to keep up with expected sequencing rates

Protein structure prediction (using threading):

>100 TeraFlops speeds required to model all proteins in a microbial genome in a day

Classical Molecular Dynamics (molecular mechanics force field):

100 TeraFlops for 1 day to simulate DNA-protein interaction (20000 atoms for 2 ms)

First Principles Molecular Dynamics (Quantum mechanical force field):

1 PetaFlops for 1 day to simulate reaction in enzyme active site (200 atoms for 1 ns)

Simulations of biological networks:

>10 TeraFlops speeds required to perform simple correlation analysis for small biological network (100 genes, 500 molecular components,10 compartments)

Computational biology is characterized by its need for continuous high performance computing, rather than periodic large scale simulations

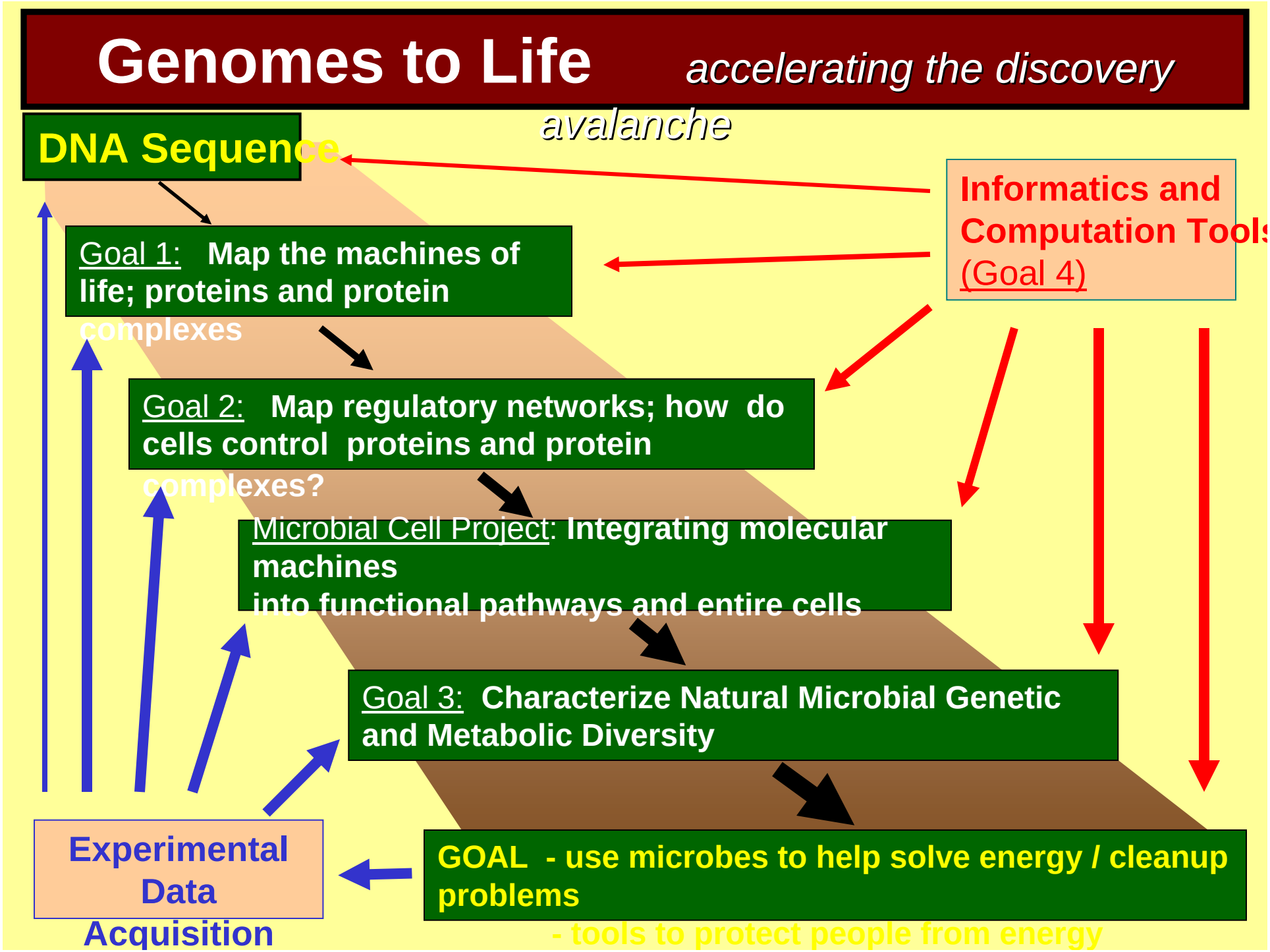
DOE/SC-0036

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 the discovery
avalanche
DNA Sequence
Informatics and Computation Tools (Goal 4)
Goal 1: Map the machines of life; proteins and protein complexes
Goal 2: Map regulatory networks; how do cells control proteins and protein complexes?
Microbial Cell Project: Integrating molecular machines into functional pathways and entire cells
Goal 3: Characterize Natural Microbial Genetic and Metabolic Diversity
Experimental Data Acquisition
GOAL - use microbes to help solve energy / cleanup problems
- tools to protect people from energy

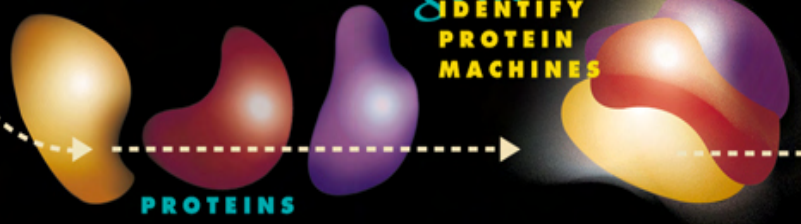
GENOMES to LIFE

ACCELERATING BIOLOGICAL DISCOVERY

A NEW PROGRAM PROPOSED BY THE U.S. DEPARTMENT OF ENERGY

DNA SEQUENCE DATA FROM GENOME PROJECTS

Genes and other DNA sequences contain instructions on how and when to build proteins

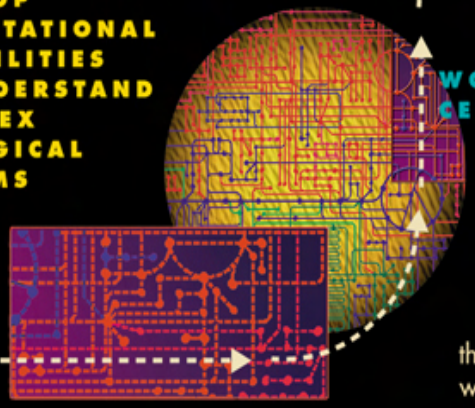


PROTEINS

goal
IDENTIFY PROTEIN MACHINES

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.

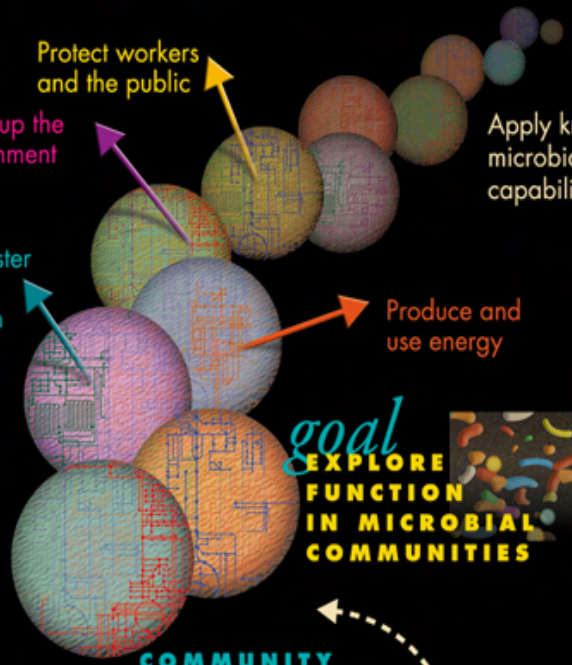
goal
DEVELOP COMPUTATIONAL CAPABILITIES TO UNDERSTAND COMPLEX BIOLOGICAL SYSTEMS



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



COMMUNITY OF CELLS

goal
EXPLORE FUNCTION IN MICROBIAL COMMUNITIES

URL DOEGenomesToLife.org

4/01

Bringing the Genome to Life Requires Many Areas of Computational Biology

Assembly and analysis of genome sequences

Prediction of protein structure

Molecular modeling of biochemical reactions

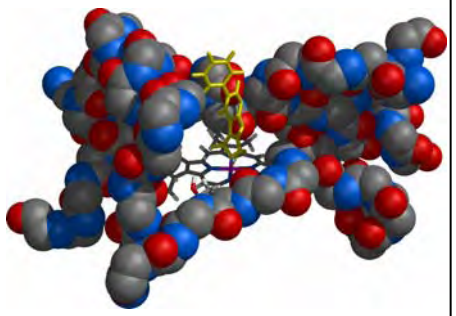
Process simulation of biological systems

ATCTCCGC
CGCTAGCTAA
AATATG
GCTAGCTAATA
TCTCCGCTA

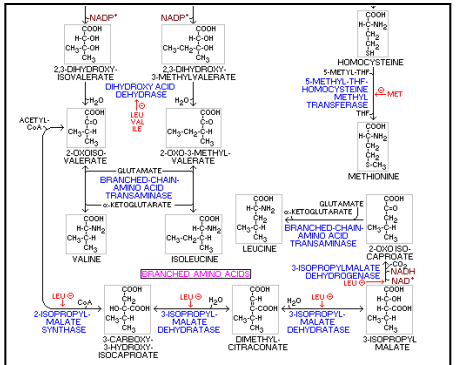
ATCTCCGCTAGCTAATATG

Consensus DNA sequence from shotgun sequencing

Predicted structure for DNA clamping protein



Enzymatic activation of environmental mutagen



Portion of amino acid synthetic pathways

Each of these types of modeling requires advances in simulation techniques and computer hardware