

Overview of CS Research at ORNL

Al Geist

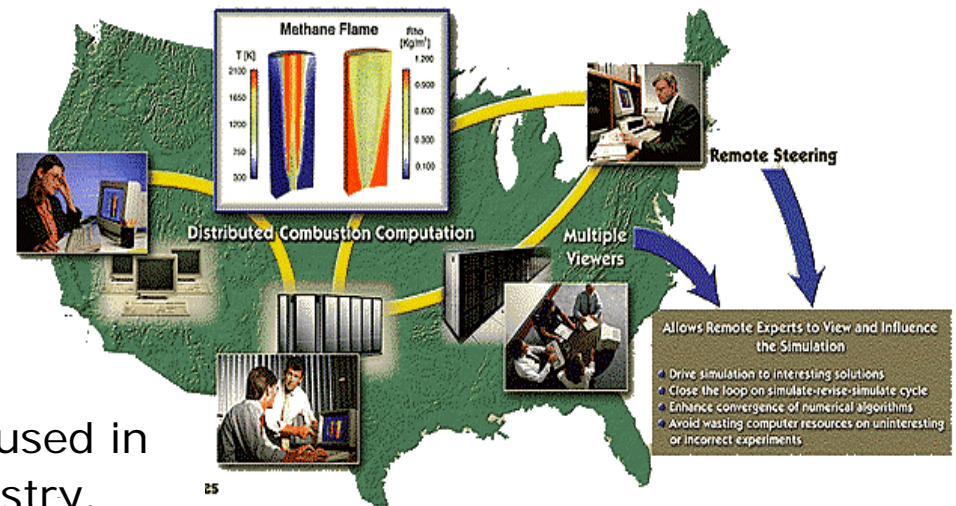
**Computer Science and Mathematics Division
Oak Ridge National Laboratory**

**December 12, 2005
RAMS Workshop
Oak Ridge, TN**

ORNL Enabling Technology Software

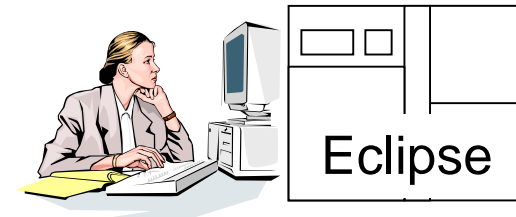
Significant impact and world-wide influence on Cluster computing and the Science enabled by it.

- Track record of developing very popular software.
 - PVM – 400,000
 - OSCAR – 100,000
 - enote – 8000
- Influencing Standards
 - MPI, BLAS, LAPACK
- Enabling Science
 - OSCAR and Enote are widely used in education, research, and industry.



Goal is to accelerate the process of Scientific Discovery

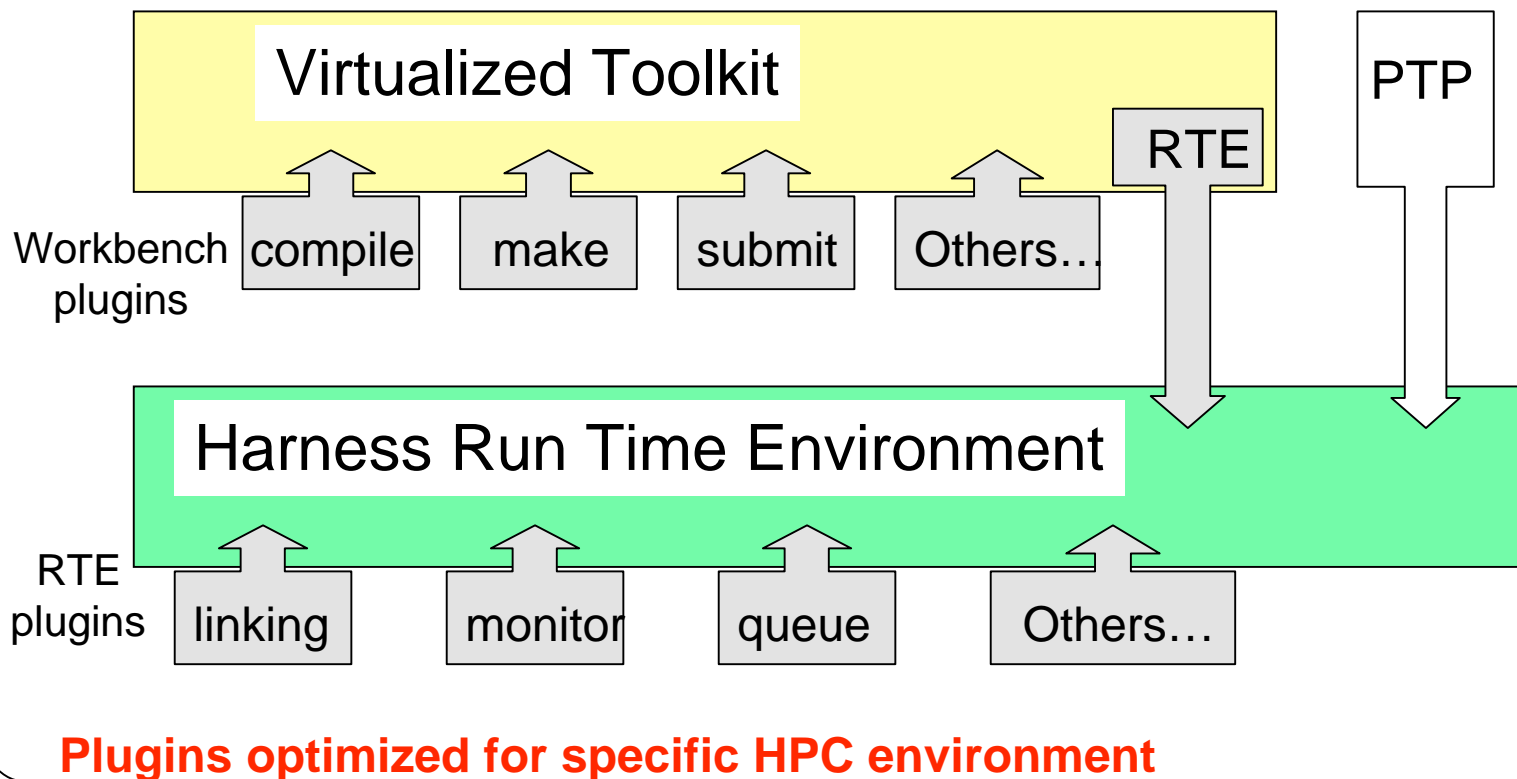
Harness Workbench Unified Programming Environment



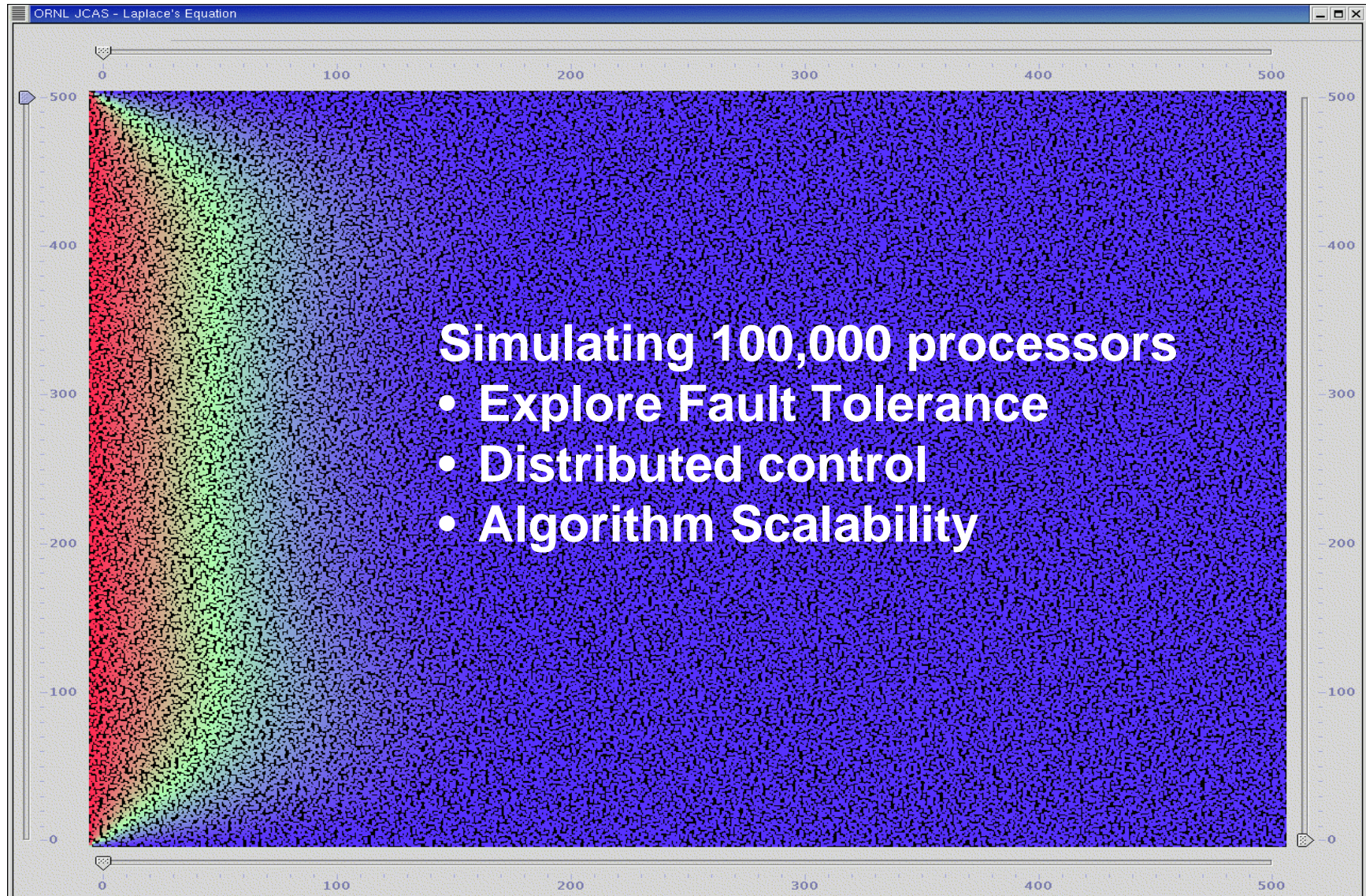
Harness Workbench

Common view across diverse HPC environments

Connection to other environments



Developed a Simulator for Peta-Scale System





Scalable Systems Software for Terascale Computer Centers



www.scidac.org/ScalableSystems

Problem

- Computer centers use incompatible, ad hoc set of systems tools
- Present tools are not designed to scale to multi-Teraflop systems

Solution

- Collectively (with industry) define standard interfaces between systems components for interoperability – XML, WS
- Create scalable, standardized management tools for efficiently running our large computing centers

Impact

- Reduced facility mgmt costs.
- More effective use of machines by scientific applications.

Resource
Management



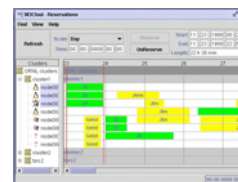
Accounting
& user mgmt



System
Monitoring



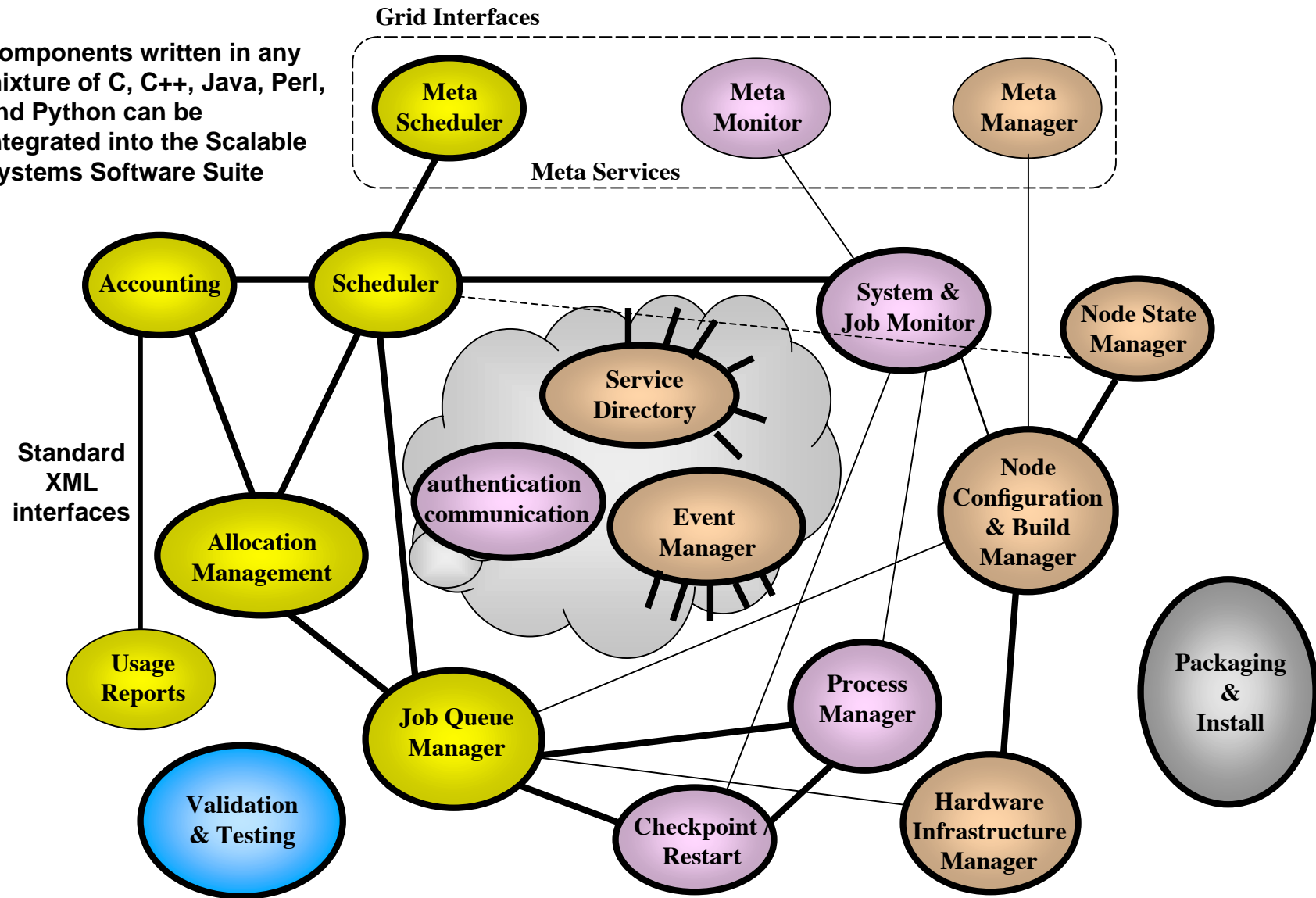
System
Build &
Configure



Job management

Scalable Systems Software Suite

Components written in any mixture of C, C++, Java, Perl, and Python can be integrated into the Scalable Systems Software Suite



Management of Scientific Data Sets Drives Algorithmic Breakthroughs

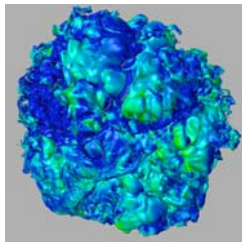
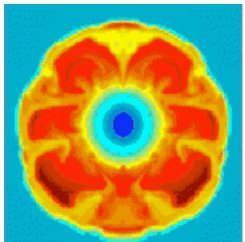
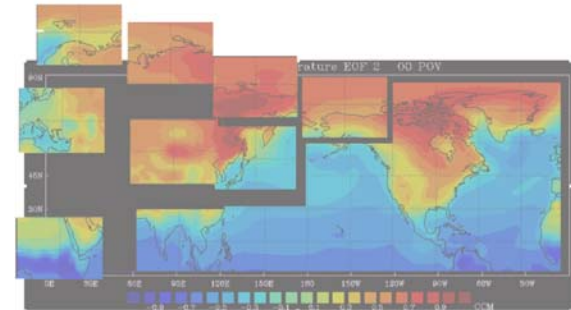
Tera&Petabytes

Existing methods do not scale in terms of time and storage

Develop effective & efficient methods for mining scientific data sets

Distributed

Existing methods work on single centralized dataset. Data transfer is prohibitive



Supernova Explosion:

1-D simulation: 2GB
2-D simulation: 1TB
3-D simulation: 50TB

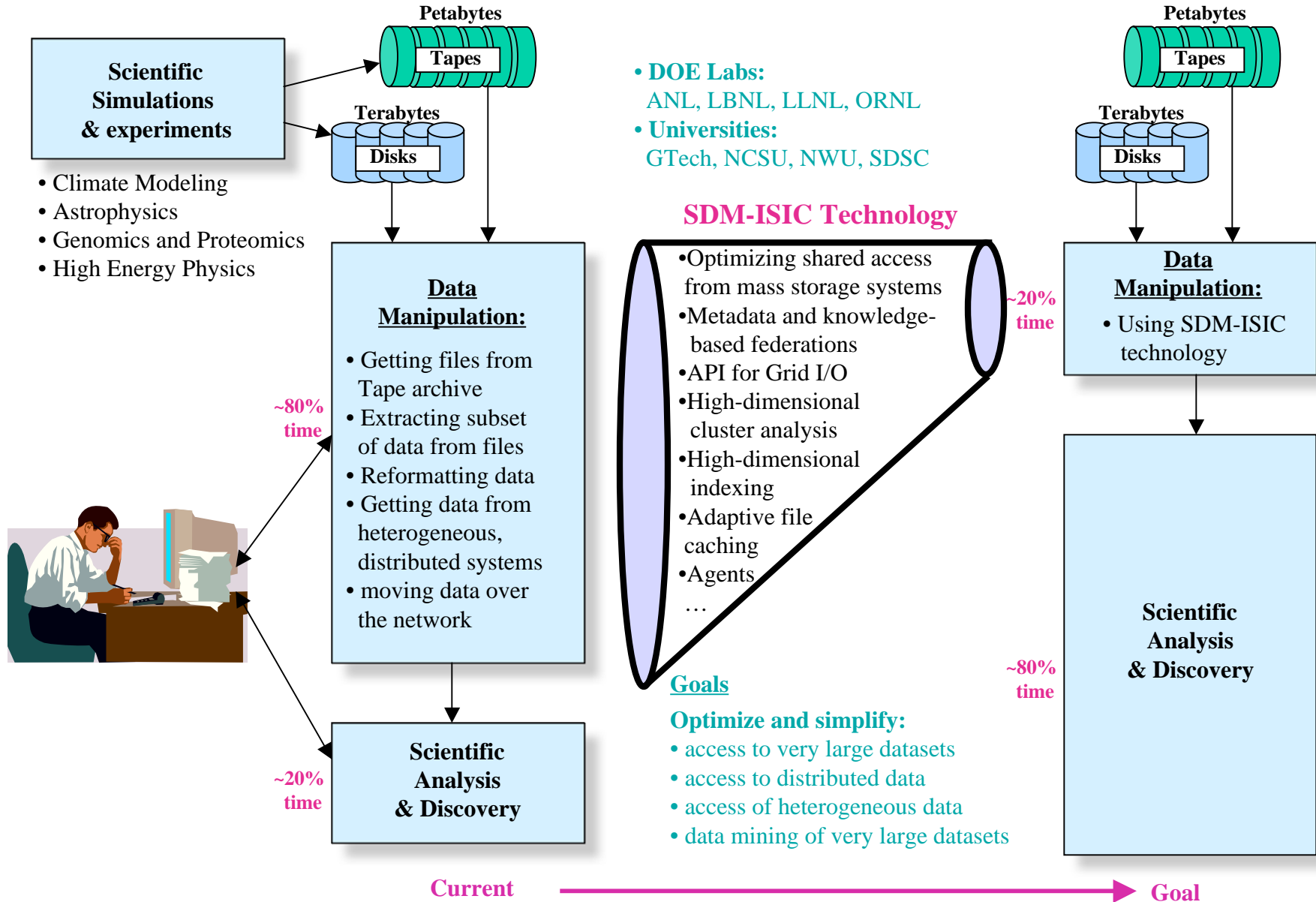
High-dimensional

Existing methods do not scale up with the number of dimensions

Dynamic

Existing methods work w/ static data. Changes lead to complete re-computation

Scientific Data Management



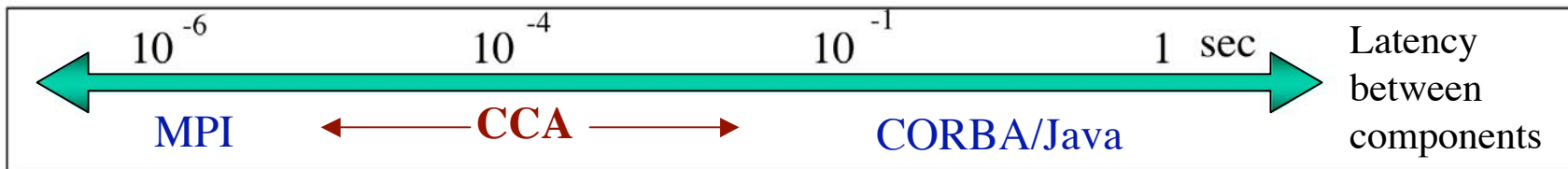
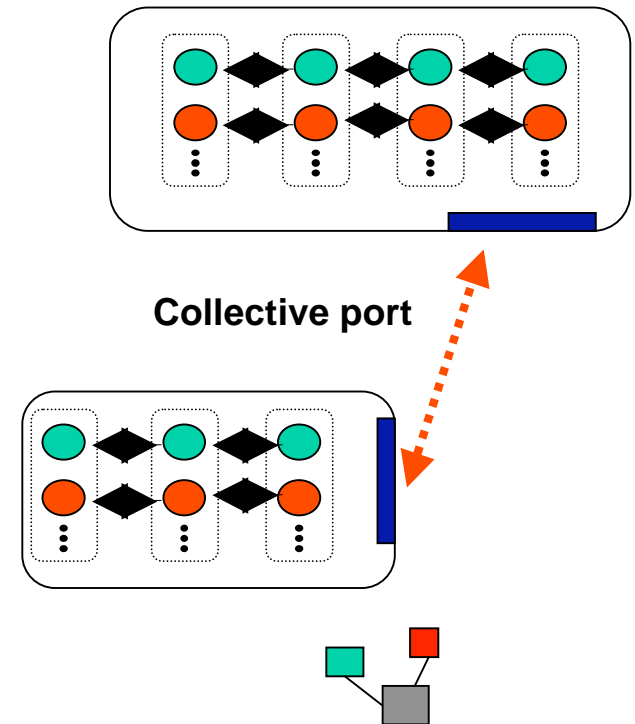
Common Component Architecture (CCA)

CCA is a multi-lab effort to provide a standard for **interoperability of high performance components** developed by many different groups in different languages or frameworks.

- Efficient coupling of SPMD components running on SMP nodes.
- Coupling of parallel components (collective ports)

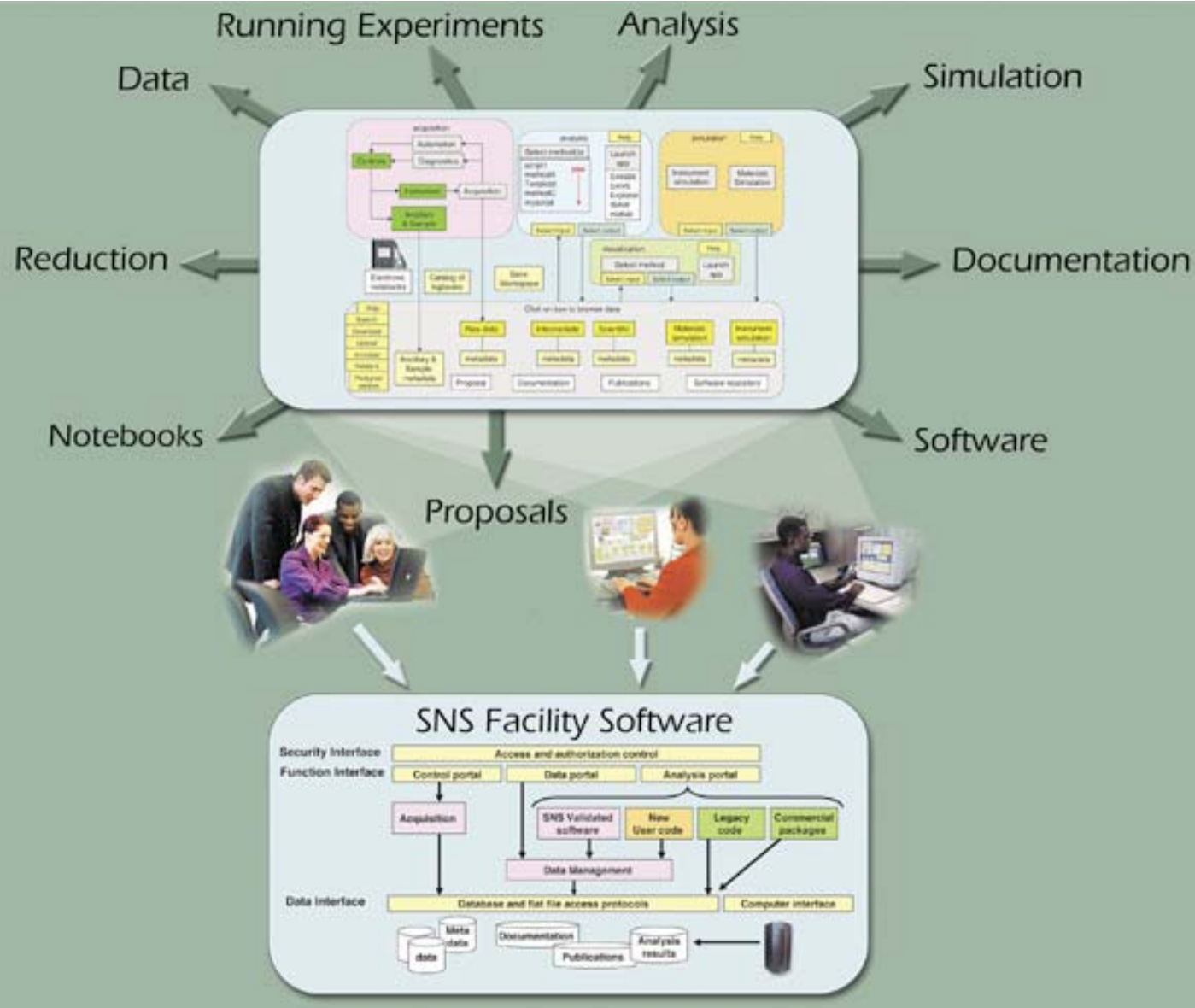
CCA Research at ORNL:

- Collective Ports – spec. and NxM
- Data Distribution in parallel components
- Computational Steering components
- Dynamic environments



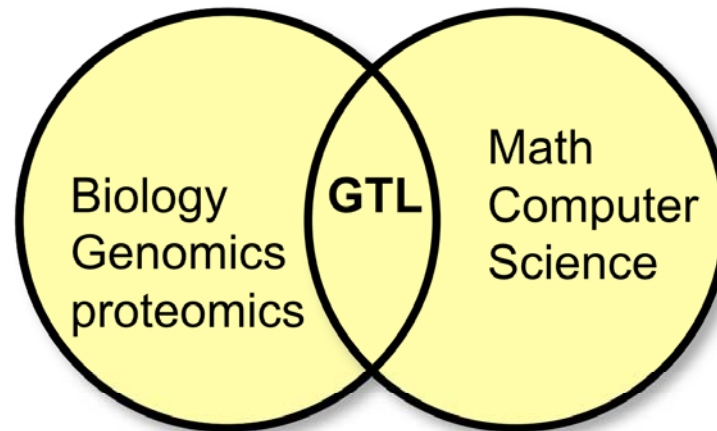
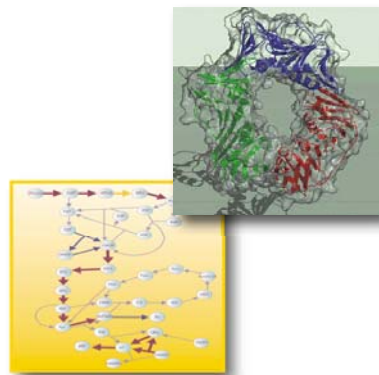
SNS Facility Software

Designed to enable much more than just data collection



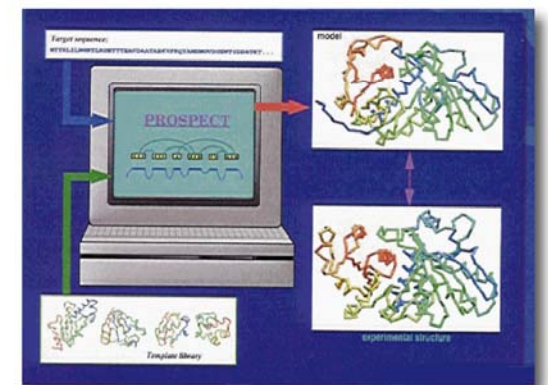
DOE Genomes to Life Program

Understand how genes, proteins, and cells work in intricate networks to form dynamic living systems exquisitely responsive to their environment.



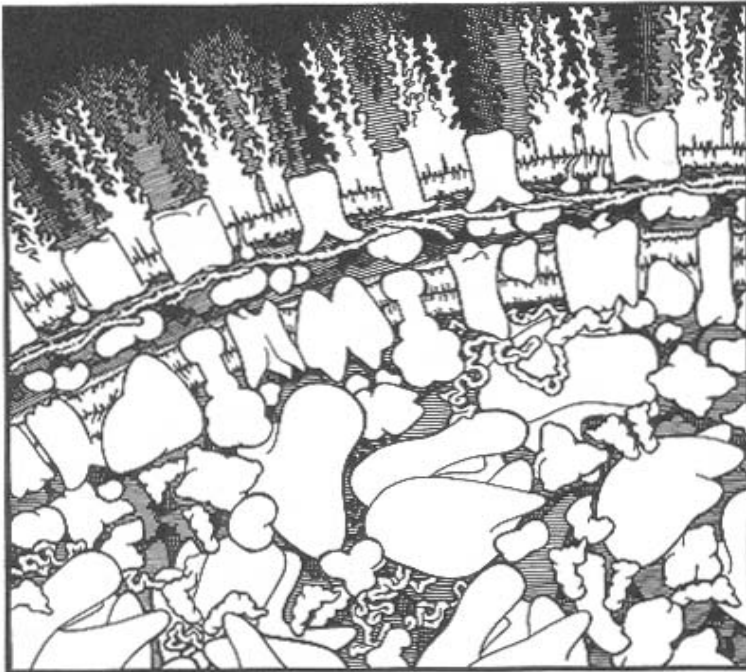
Start with simple life form – single cell organism

- Discover how microbial genes, proteins and cells work together
- Use supercomputers to analyze data, predict, model, and simulate how protein machines interact through complex interconnected pathways

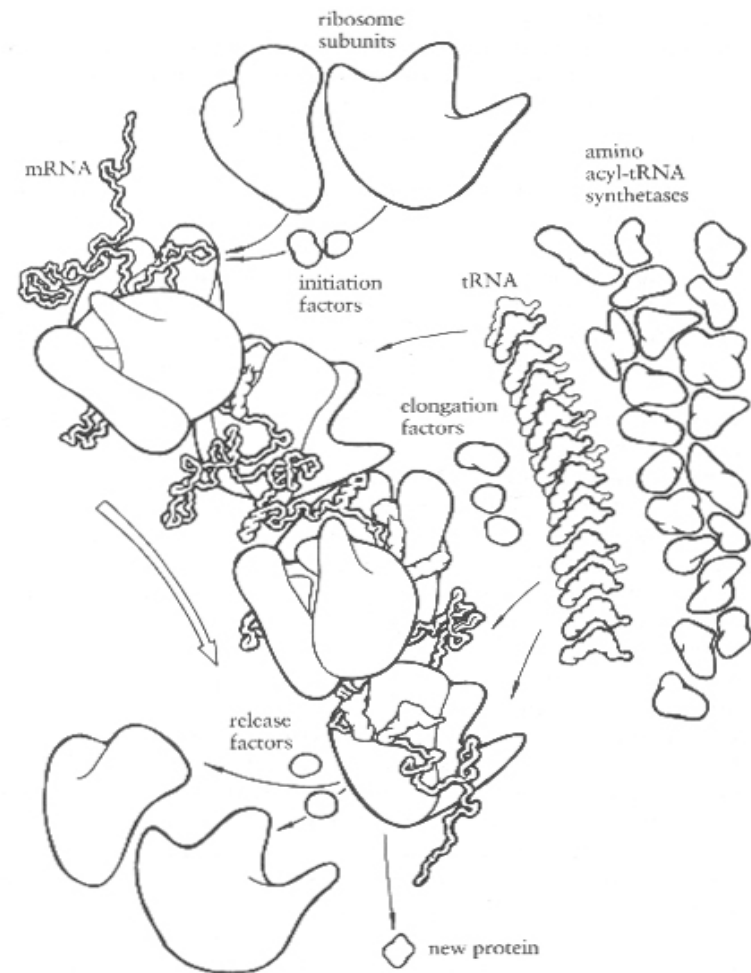


Molecular Machines Fill Cells

Many interlinked proteins form interacting machines

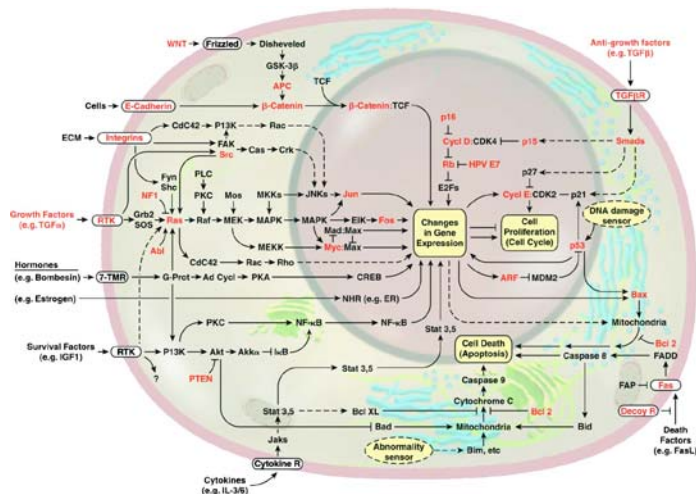


From The Machinery of Life, David S.
Goodsell,
Springer-Verlag, New York, 1993.

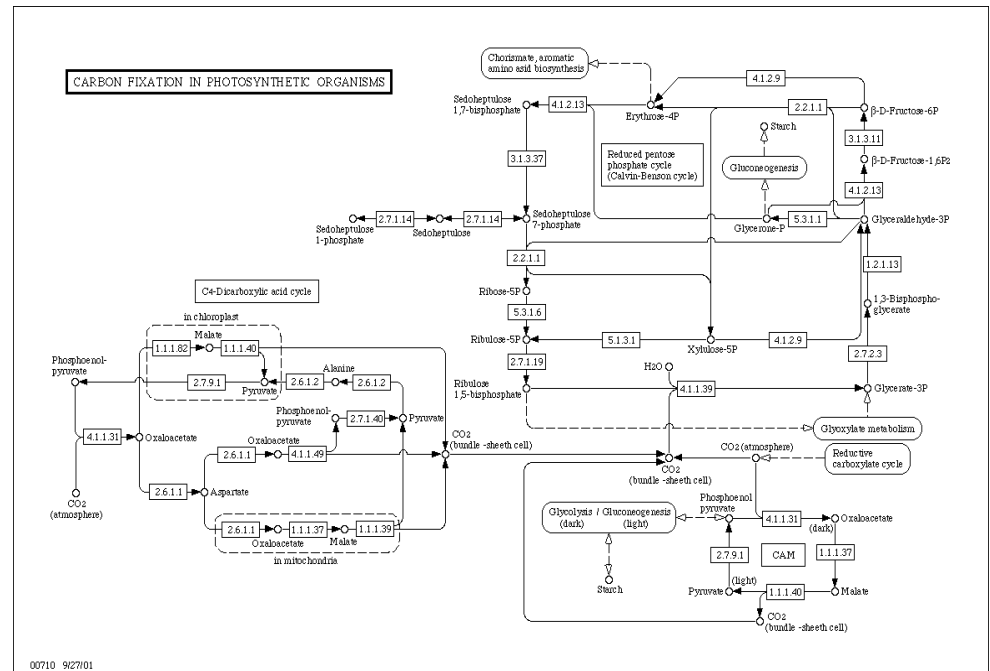
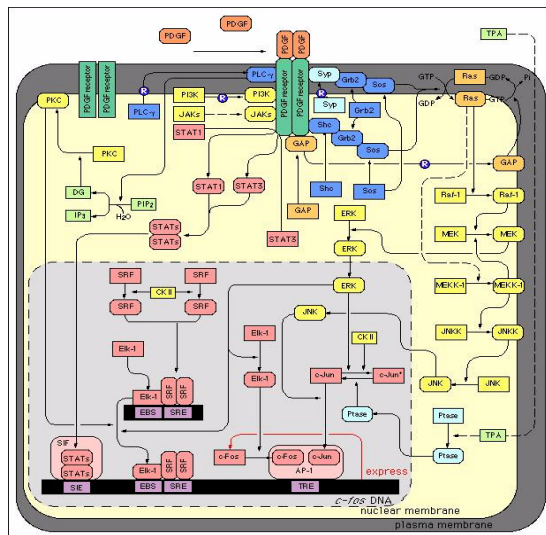


Regulatory Networks Control the Machines

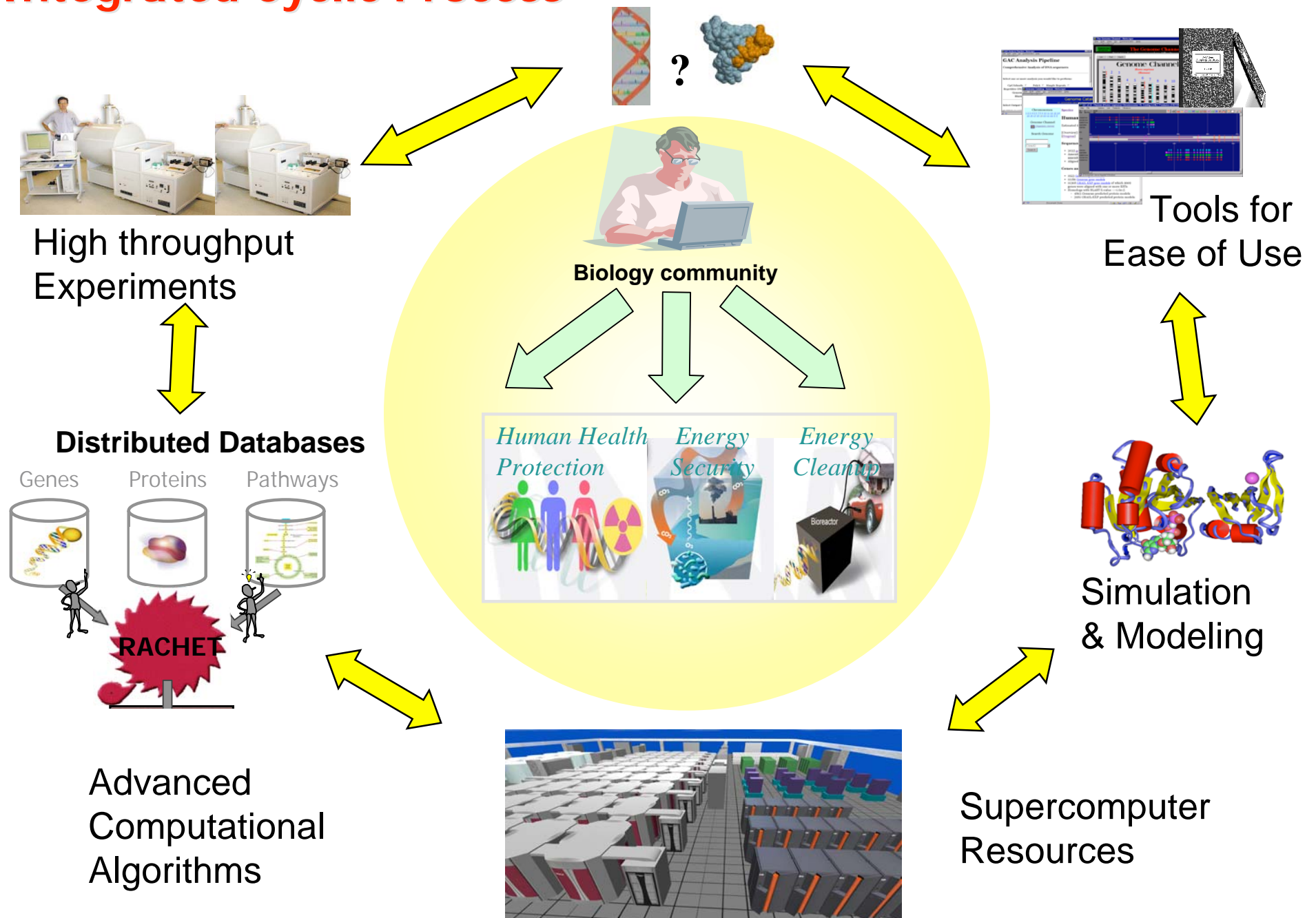
Gene regulation controls what genes are expressed



Proteome changes over time and due to environmental conditions



Integrated Cyclic Process



New Computational Biology Tools

- SVMMER protein functional characterization **web portal**
http://www.csm.ornl.gov/comp_biology/projects/SVMMER
- ROBETTA protein structure prediction **web portal**
<http://www.lanl.bakerlab.org/>
- PAT Pattern analysis tool for statistical comparative analysis of protein-protein interfaces, surface patches and binding sites **web portal**
http://www.csm.ornl.gov/comp_biology/projects/PAT/
- DEB data entry, sharing, and browsing, easy interface for entering, viewing, and sharing microarray data
<http://sdm.lbl.gov/~opm7/sdmdev/www/>
- Proteomic Toolshop (Matlab-like tool for biology) understands biological data types and allows easy analysis and viewing (www.vigyaan.cd.org)

