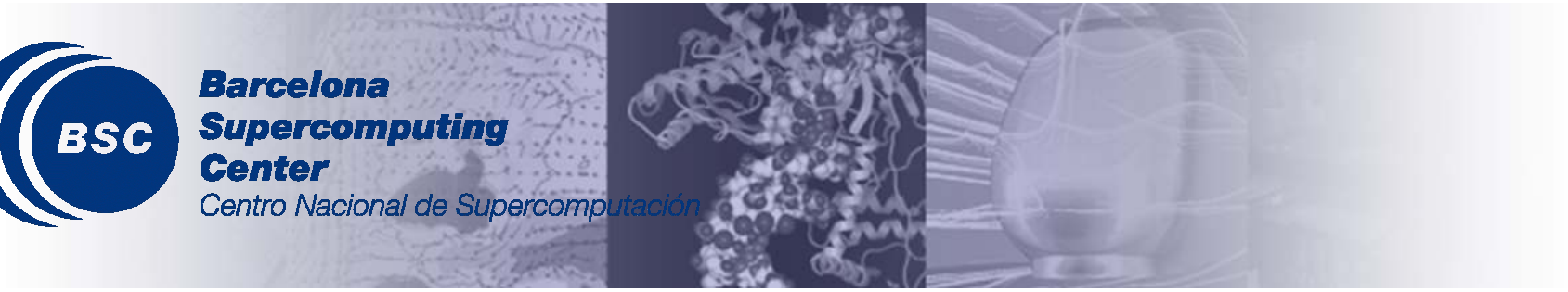
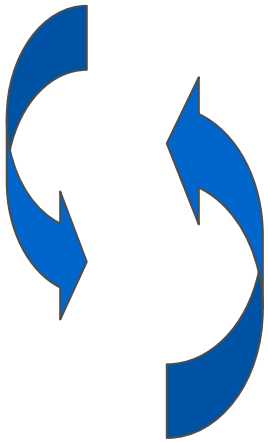
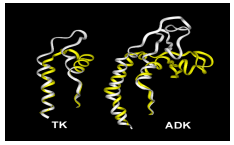
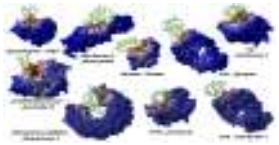
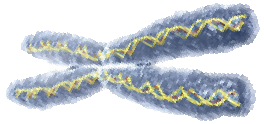
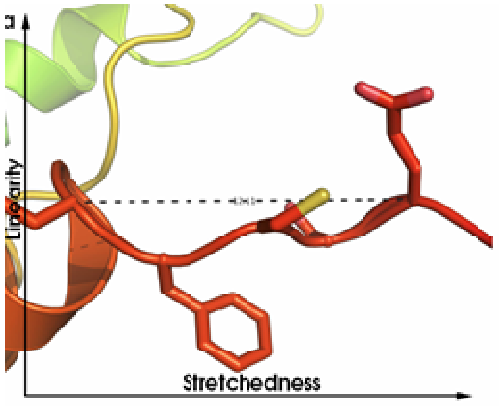


The Power of Computation in Life Sciences
Life Sciences - BSC

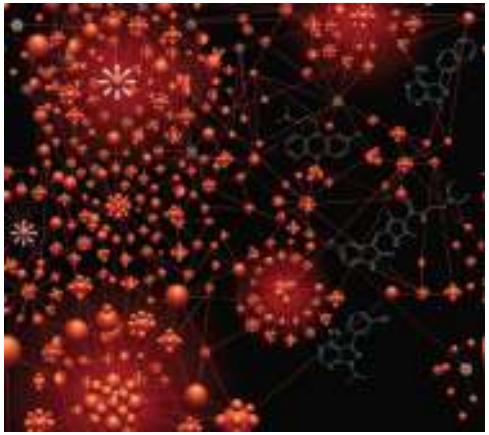


GROUP	HEAD		RESEARCH AREA
EAPM	Victor Guallar		Atomic (and electronic) modeling of protein biochemistry and biophysics
MMB	Modesto Orozco		Micro and mesoscopic modeling of macromolecules and Drug Design
PID	Juan Fernández		Identification of the structural bases of protein-protein interaction
CG	David Torrents		Analysis of genomes and networks to model diseases, systems and evolution
EBL	Montse Soler		Implementing advanced experimental approaches
AU	Josep Ll. Gelpí		Web services, applications, databases and software integration.

Research Lines

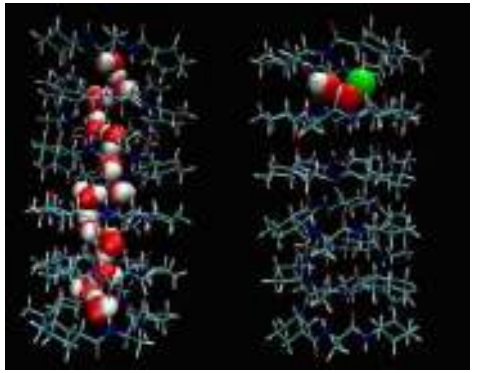


Target & Drug Discovery



Genomics & System Biology

BioSupercomputing



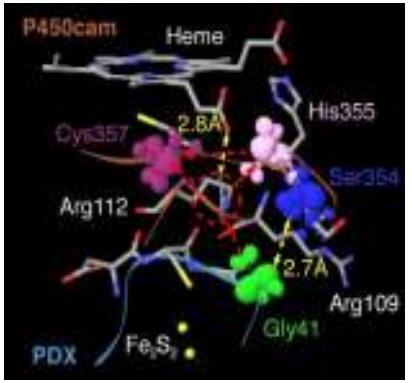
BioPhysiscs

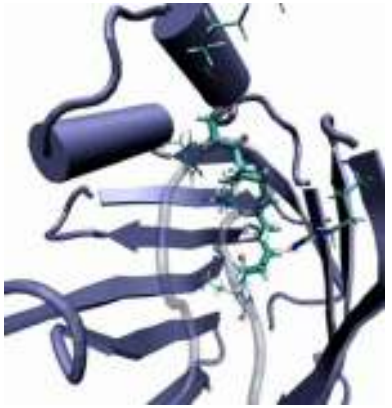
Electronic and Atomic Protein Modeling

Victor Guallar

icrea
Institució Catalana de Recerca i Estudis Avançats

- **Electron transfer processes and enzymatic catalysis.** Describe these processes using in-house developed mixed Quantum Mechanics/Molecular Mechanic (QM/MM) algorithms.
- **Protein-ligand and protein-protein docking.** Sample interactions using atomic and coarse grain force fields combined with Monte Carlo methods and protein structure prediction techniques.





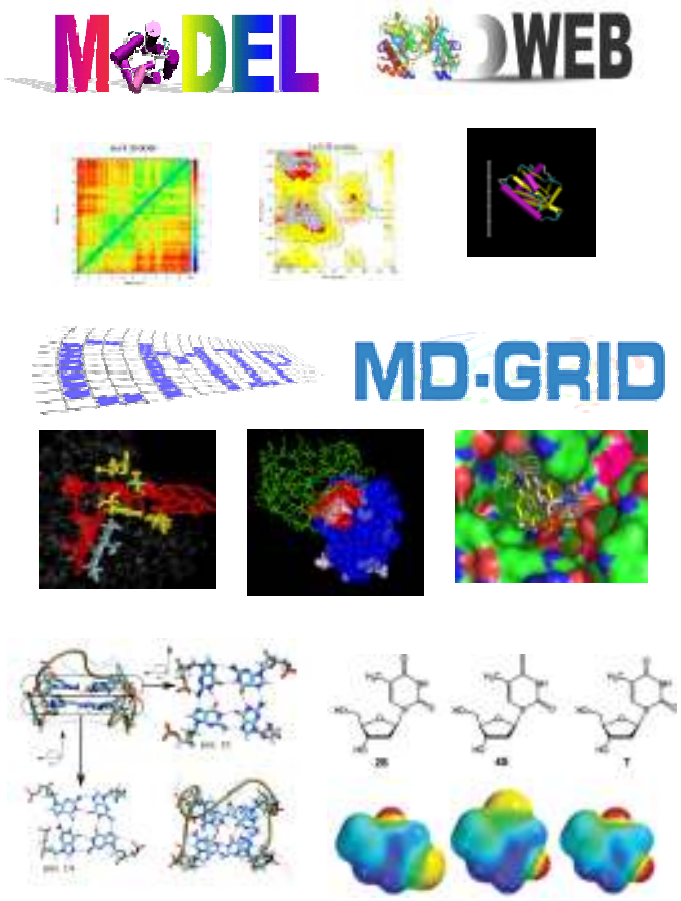
Molecular Modeling & Bioinformatics

Modesto Orozco

IRB BARCELONA

UNIVERSITAT DE BARCELONA

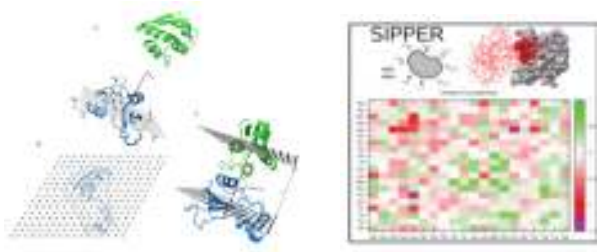
- **Study of protein flexibility**. MODEL, flexibility & binding, ed-MD, MDWEB and Coarse Grained Models.
- **Drug and target discovery**. SNPs and pathology, docking, binding site characterization, lead optimization and dirty drugs study.
- **Nucleic acids and chromatin**. Nucleosome characterization, promoter prediction and physical properties of unusual DNA.

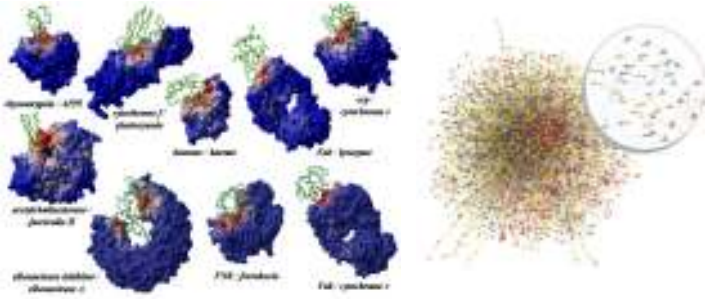


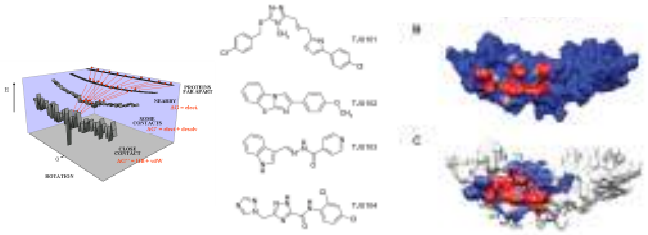
Protein Interactions and Docking

J.Fernández Recio

- **Development of new algorithms for protein docking.** Rigid-body sampling, Flexible & ensemble docking and Docking scoring
- **High-throughput application.** Porting to high-performance platforms, Binding site prediction for proteomics, Multi-protein complexes and Interaction networks modeling.
- **Study of association mechanism and biomedicine applications.** Mechanism and energetics of association and Drug design targeting protein interactions







Computational Genomics Group

David Torrents

icrea

- **New Generation DNA Sequencing.** Tomato genome assembly and Cancer genome project of Chronic Lymphocytic Leukemia
- **High-throughput and Omics data systems biology.** Study of complex diseases (Diabetes) and study of infectious diseases (Malaria)
- **MetaGENOMICS**. Human Intestinal Tract
- **Identification of Gene Regulatory Regions.**

LEUCEMIA

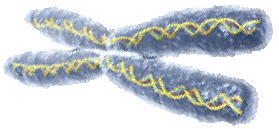
TOMATO

MiTiN

DIABETES

MALARIA

Meta HIT



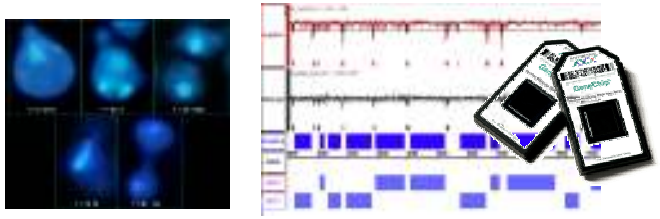
RELA

Experimental Bioinformatics Laboratory

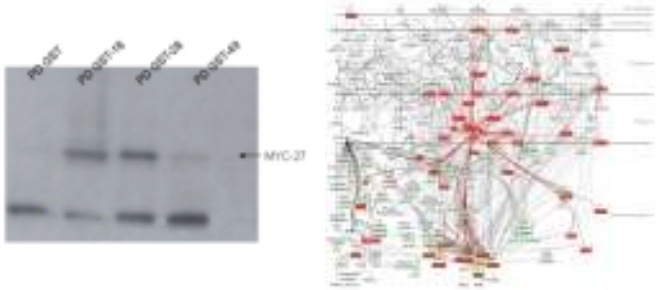
Montse Soler

IRB BARCELONA

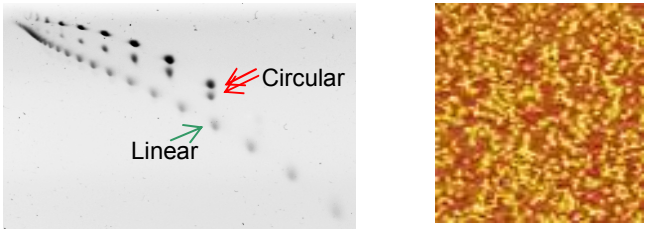
- **Nucleosome positioning.** Validation of physical predictive models and effect of methylation and cell cycle.



- **Pathological Pathways.** Systems biology/Network medicine, Colon Cancer, Alzheimer and Breast Cancer.



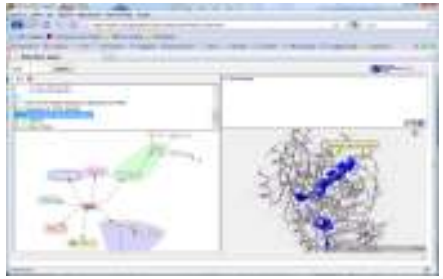
- **Epigenetics and DNA physics.** Towards a mesoscopic model for epigenetics.



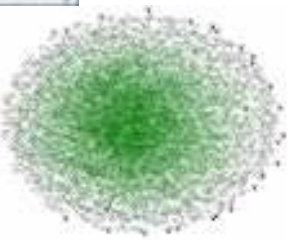
INB-BSC Algorithmic Unit

Josep Lluis Gelpí

- **Web & Database Development.** MobyMiner, INB workflows and Databases



- **Epidemiology.** Reutmathoid arthritis (IMIDKIT)

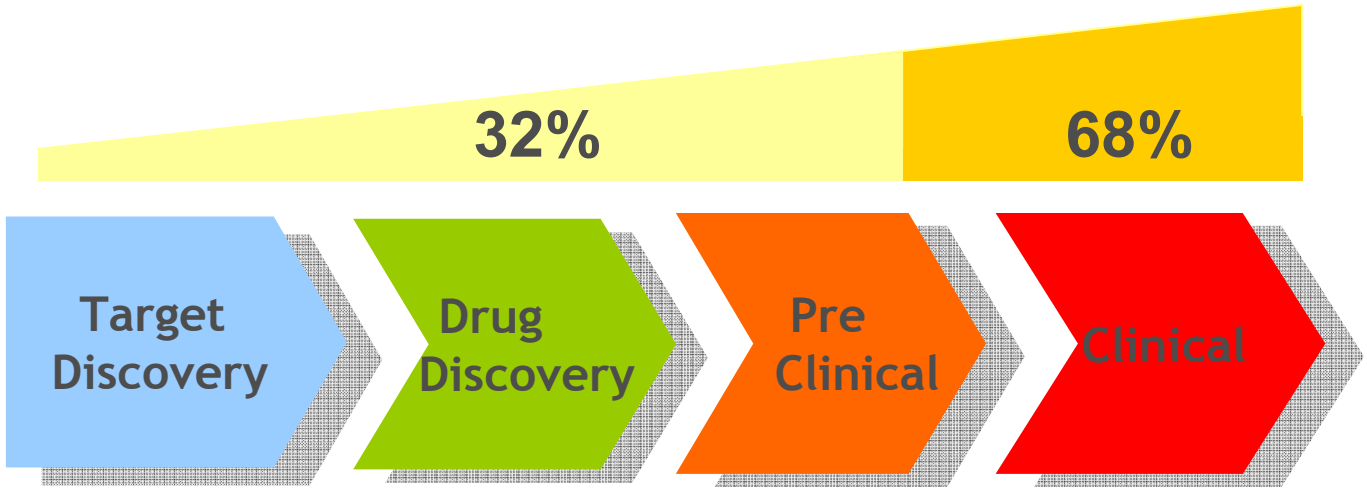


- **Elixir and ICGC projects.** Datamining, GSC and CSCm Mobylite.

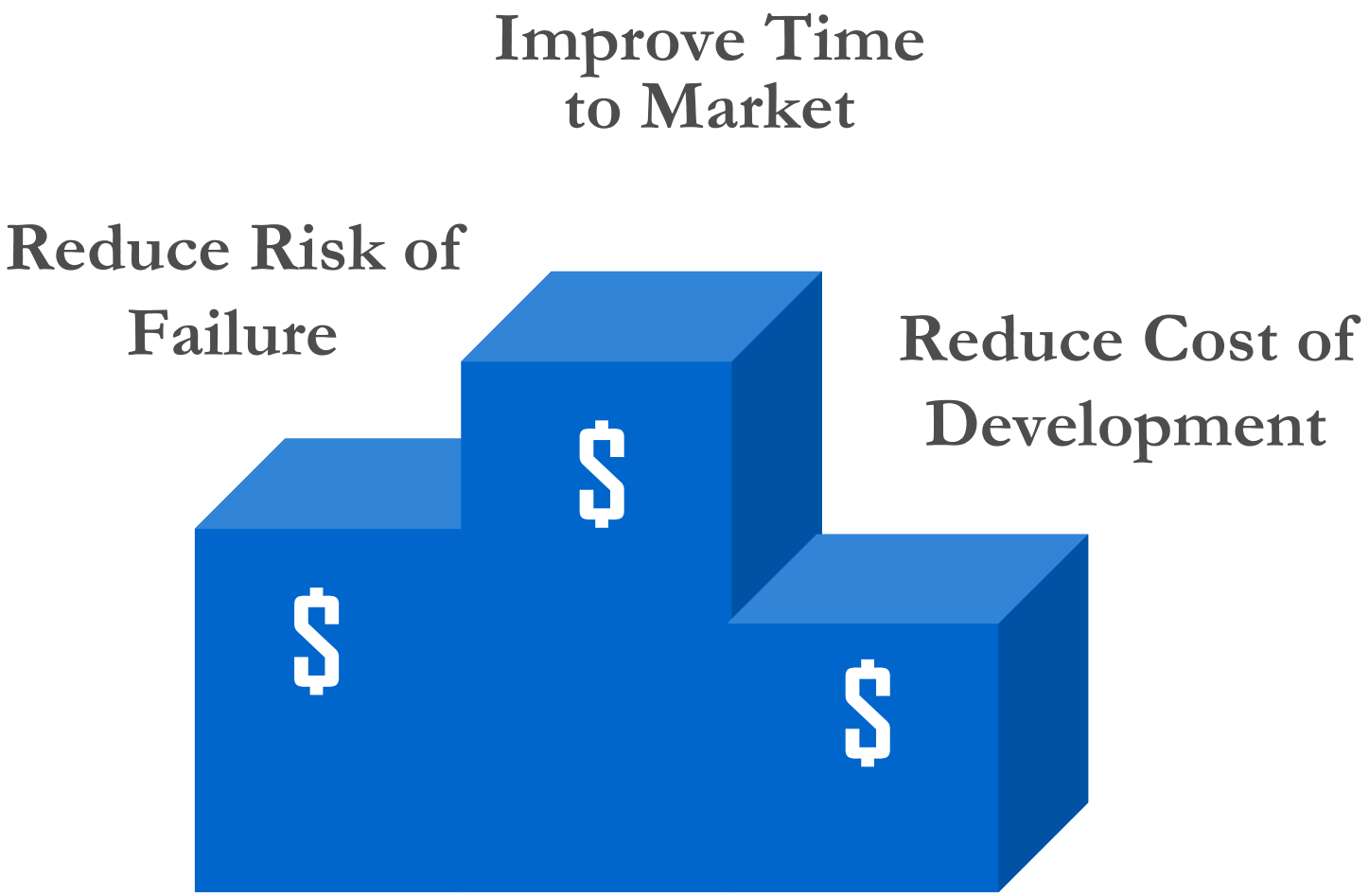
Rational Drug design

Drug Development

Cost: $1.2B / drug



Computer-Based Drug Design



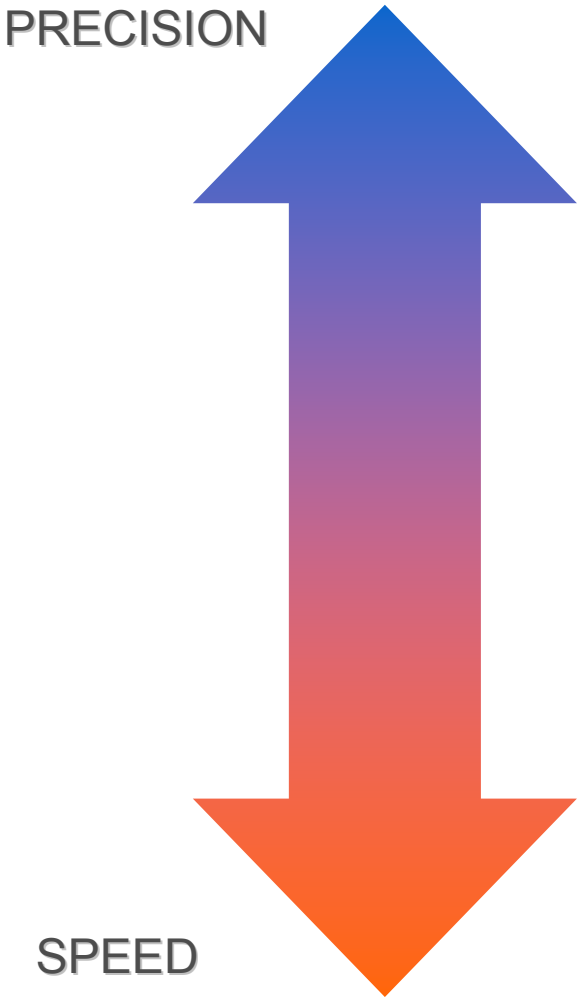
Computer Simulation

- Why and when we use it
  - **To validate a known model**
  - **As a cost-effective alternative**
  - **As the only realistic approach to solve a problem**

- The structure of bio-molecules are hardly modeled. The dynamics through experiments are only available for small molecules.

- There are different methods with different levels of complexity and realism

Molecular Simulation



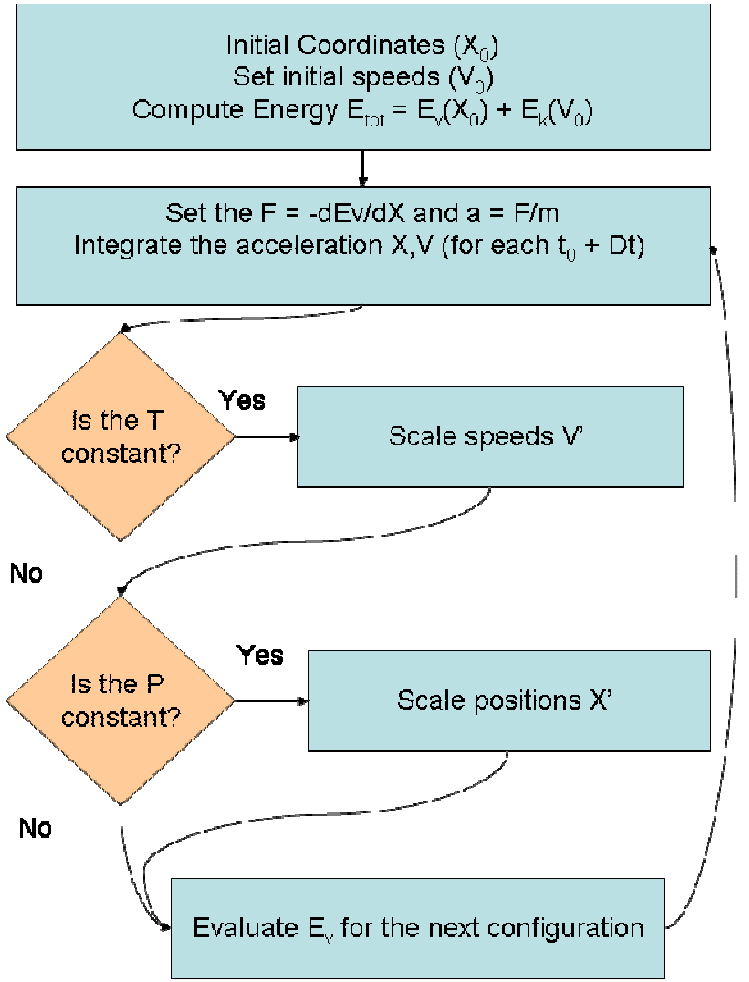
Quantum Mechanics

Molecular Dynamics (Newton motions)

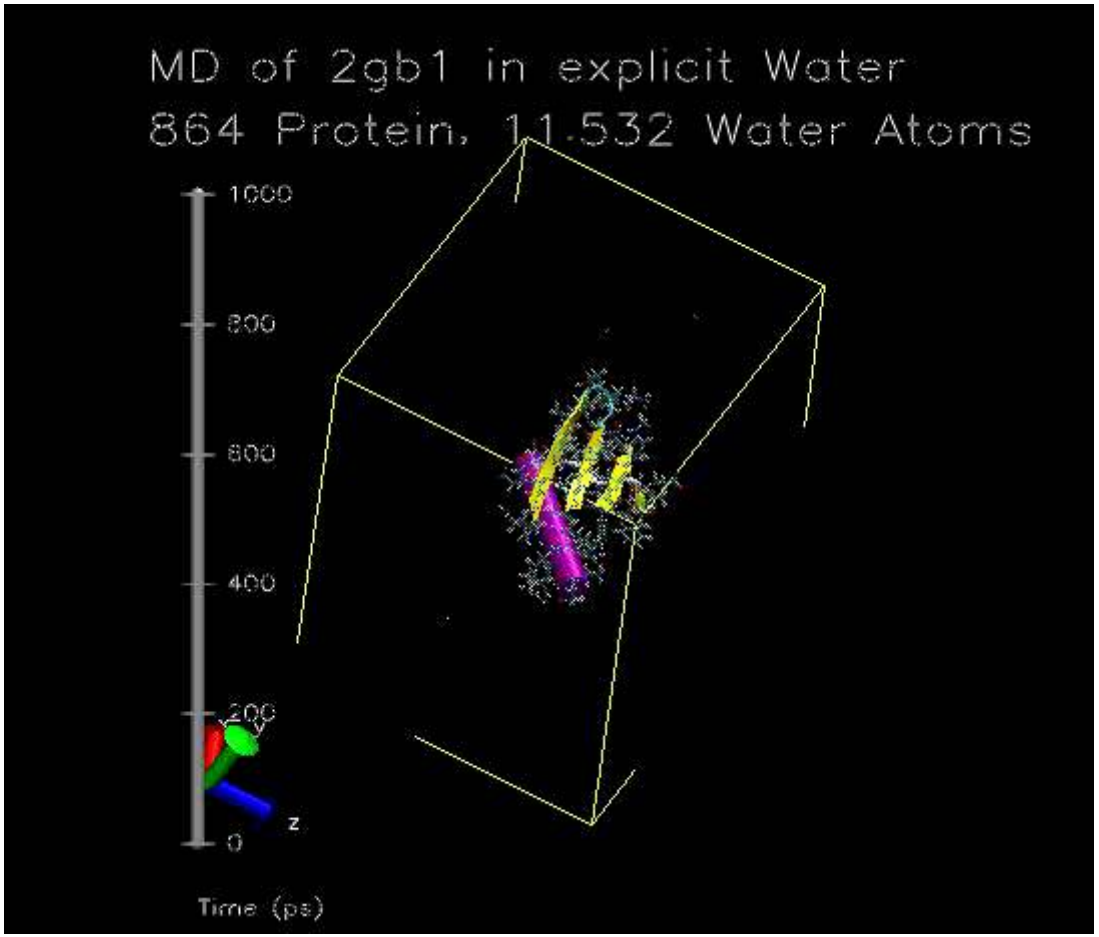
Corse Grained – MD (Pseudo-atoms)

Molecular Dynamics

- Atoms and molecules are allowed to interact for a period of time by approximations of known physics.

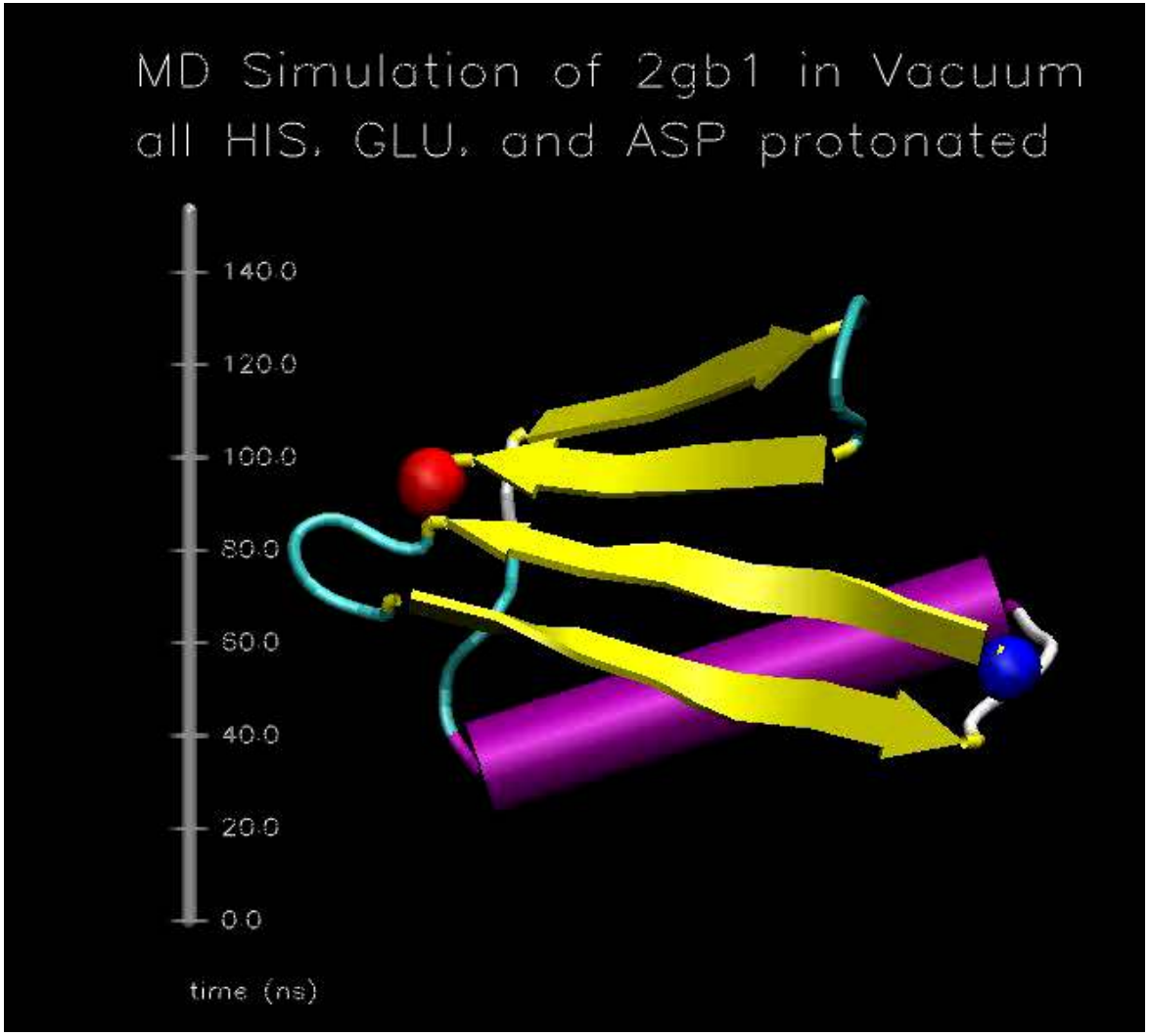


Molecular Dynamics of Solvated Protein



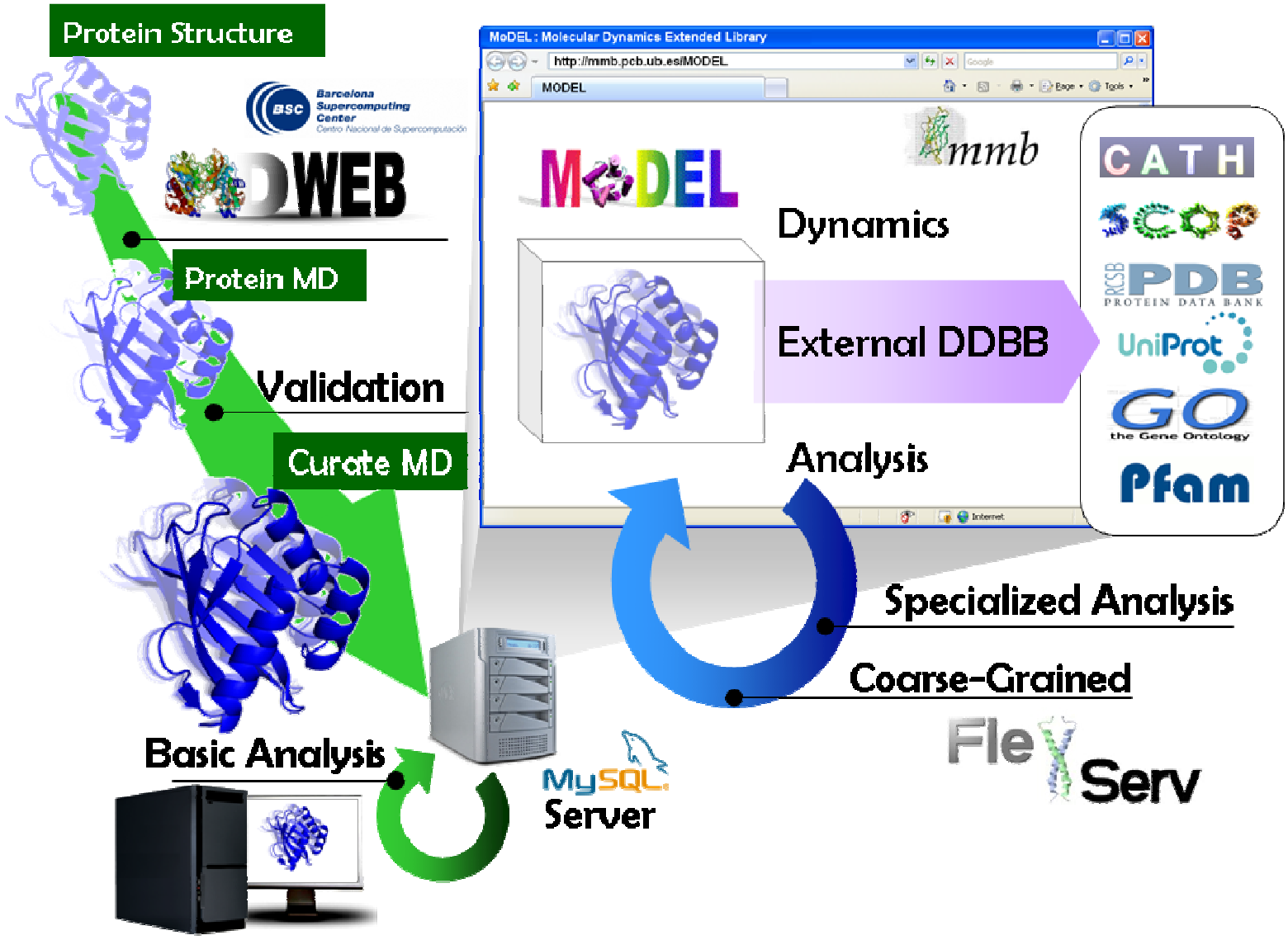
- Snapshot every femtosecond (10^{-15}s)
- System of 10^4 atoms
- 10 operations per atom pair-mate
- Using 16 processors we are able to simulate 10 nanoseconds (10^{-9}s) per day

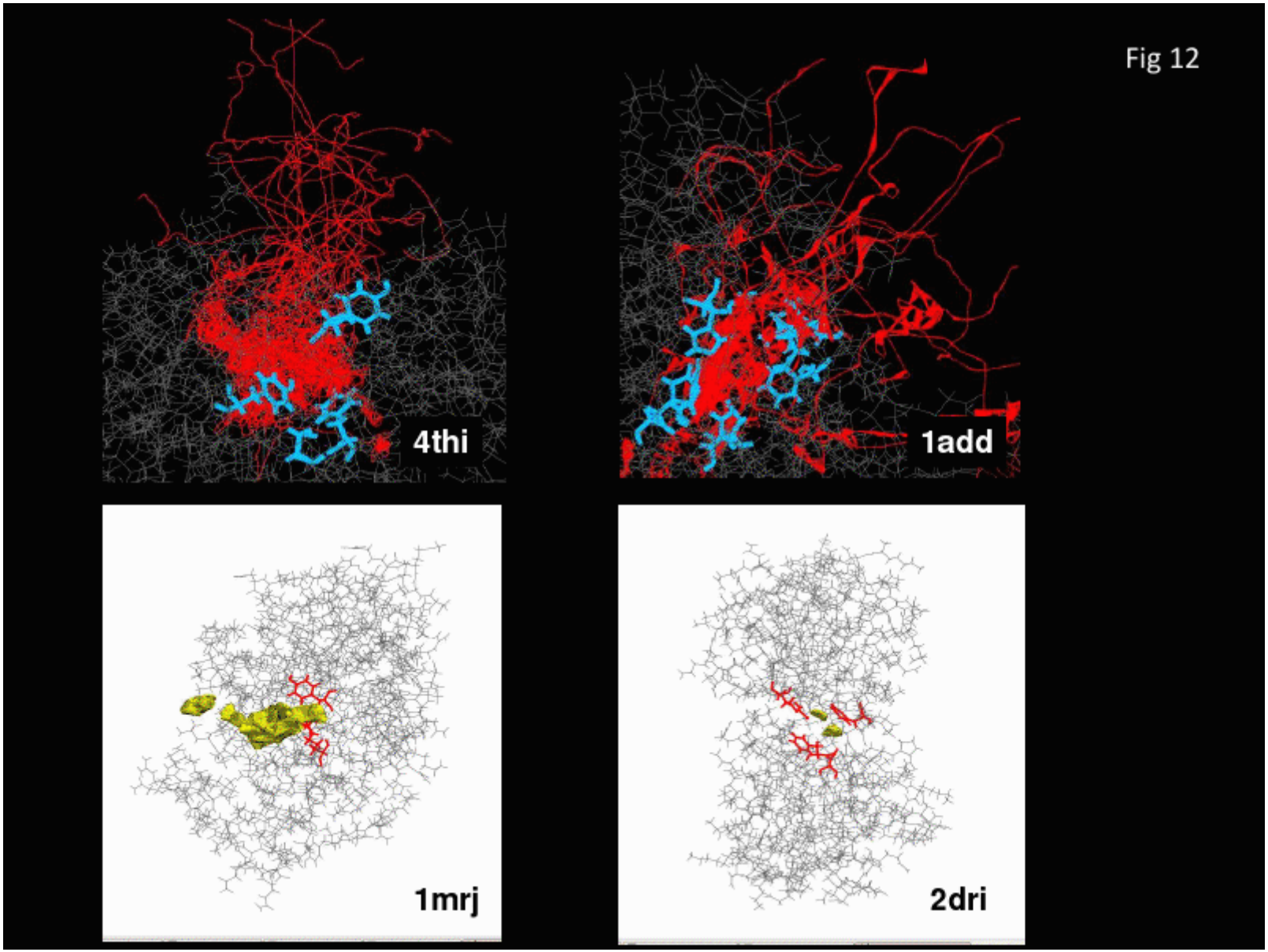
MD Scale



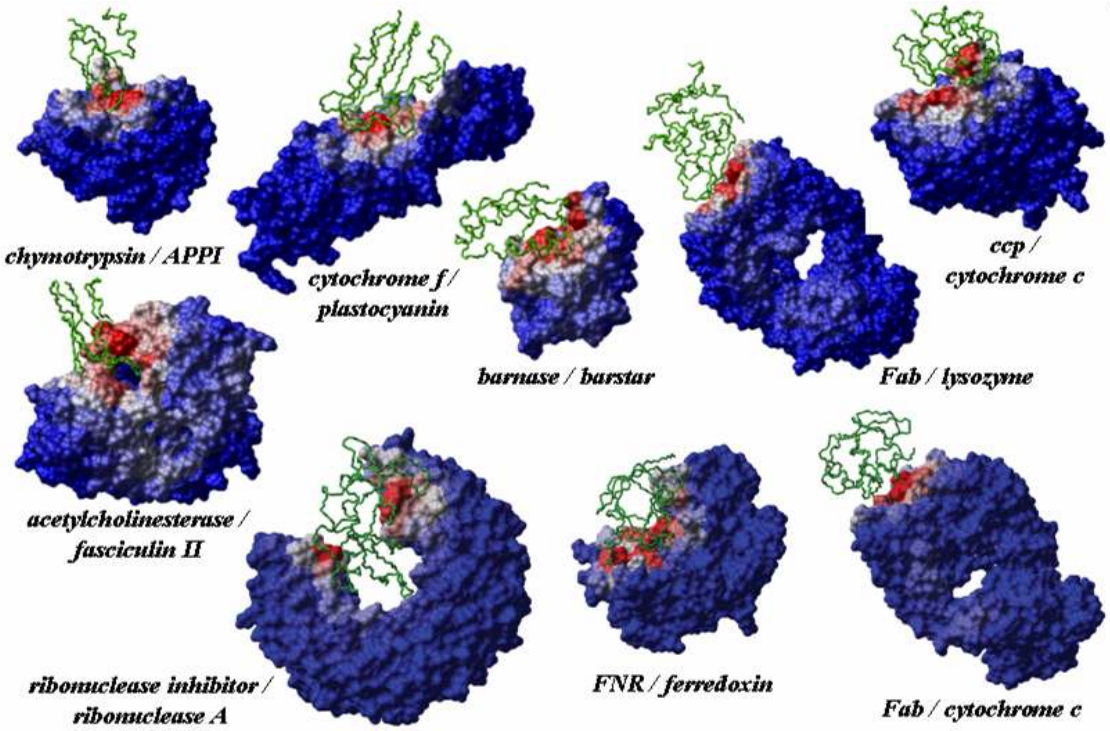
- Today the millisecond scale (10^{-3}s) is reached using specific hardware

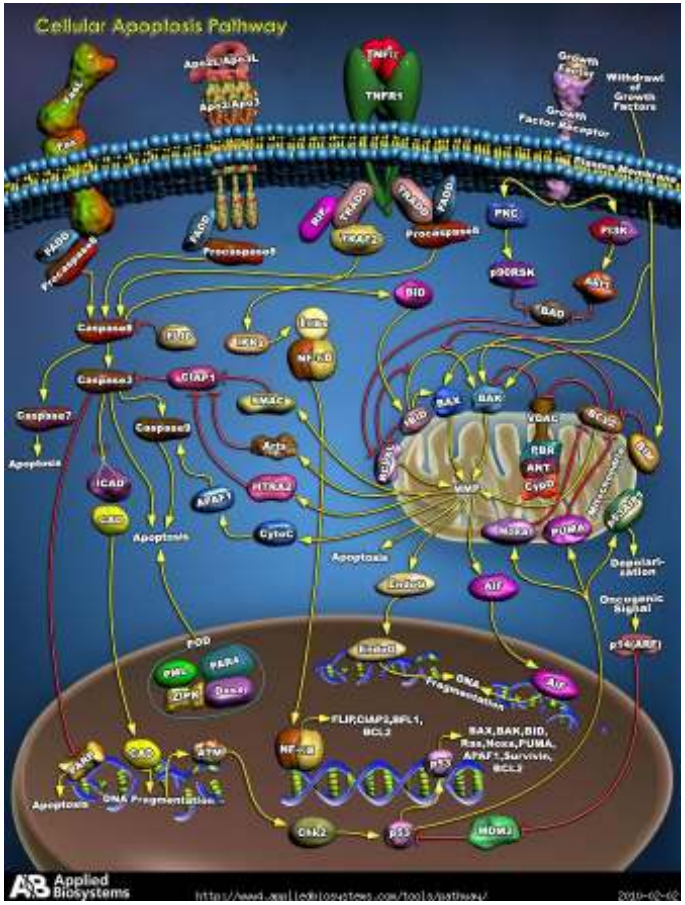
(512-processor)





Protein Interaction

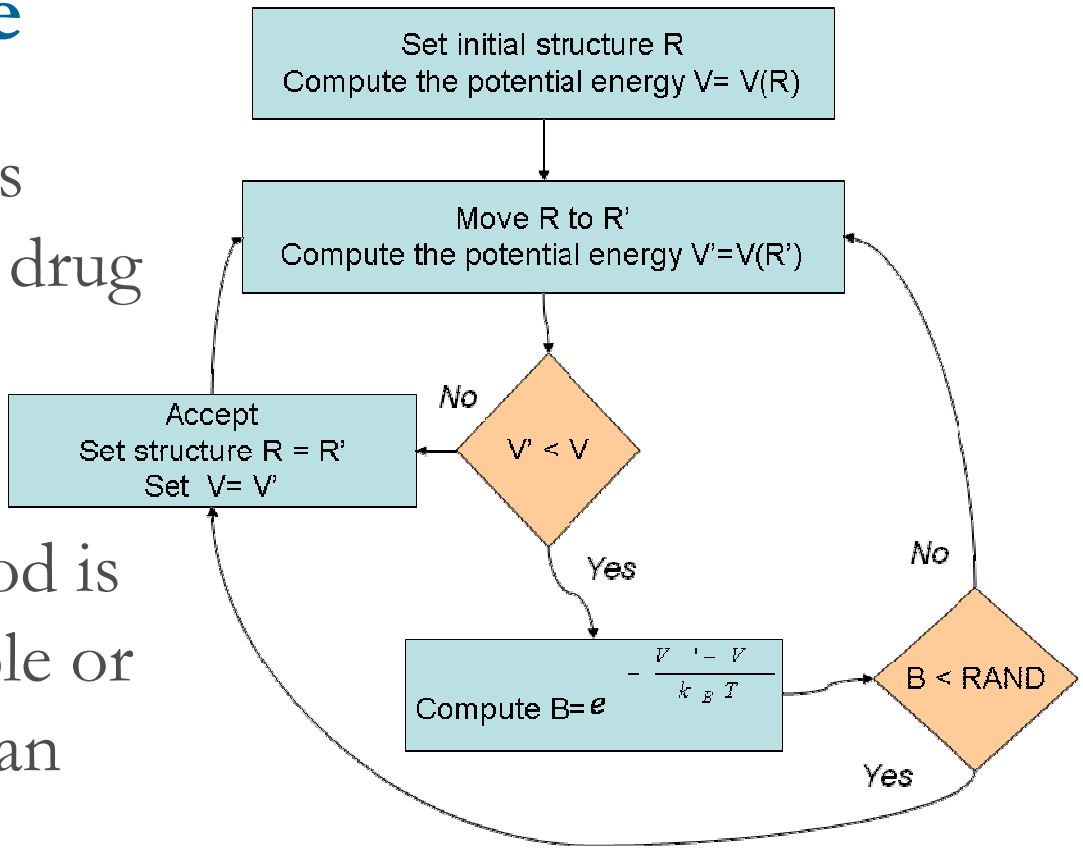




Simulate and Explore

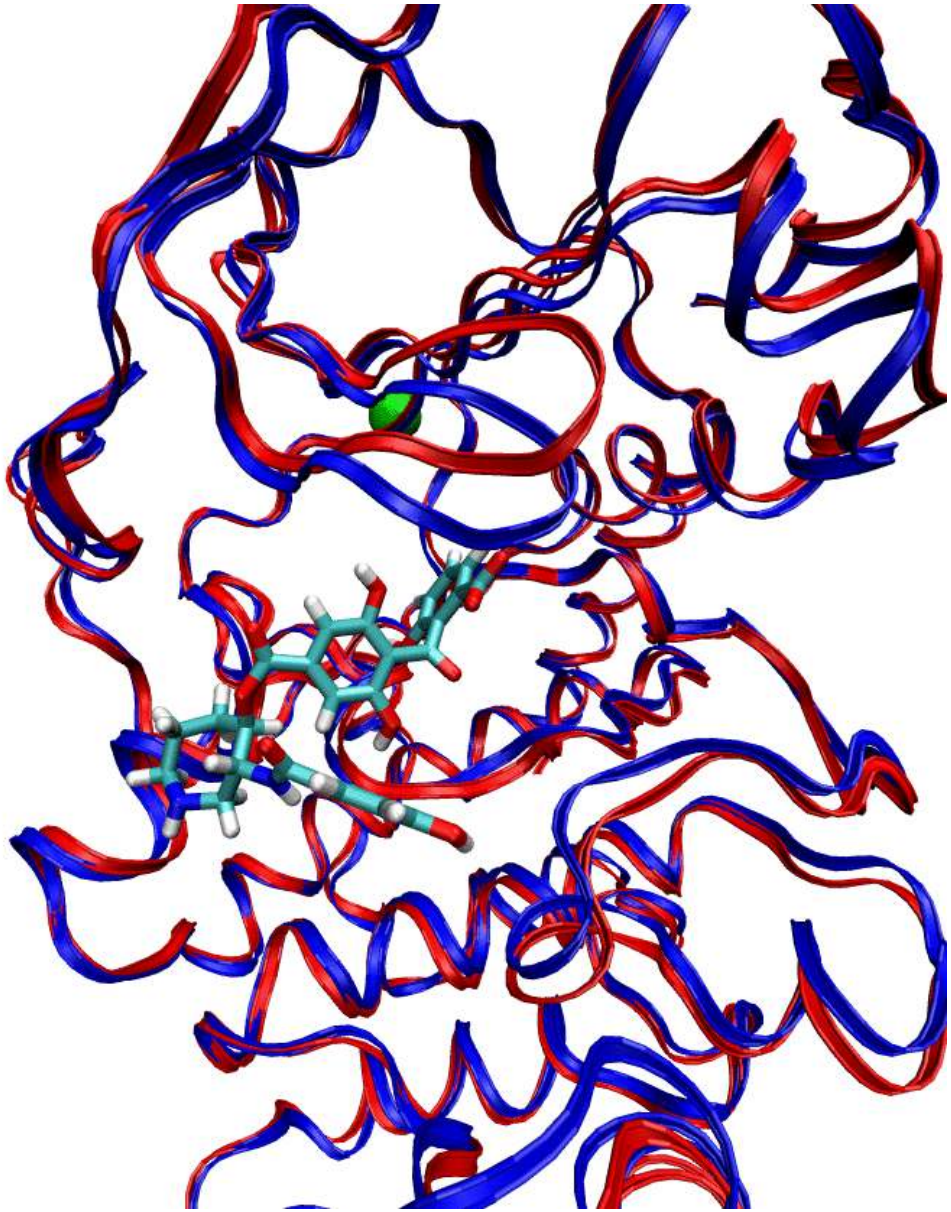
- Explore how to proteins interact. Explore how a drug binds its target.

- The Monte Carlo method is used when it is unfeasible or impossible to compute an exact result with a deterministic algorithm.



Drug Discovery

- Red: Cristal Protein
- Blue: Modeled Protein (Coarse Grained)
- Drug explores protein binding site using a Monte Carlo algorithm



Technology Follow up | Market opportunity | Product Plan | IP Protection | Financial Resources | Prototype | Market

David Torrents
BSC Group Leader

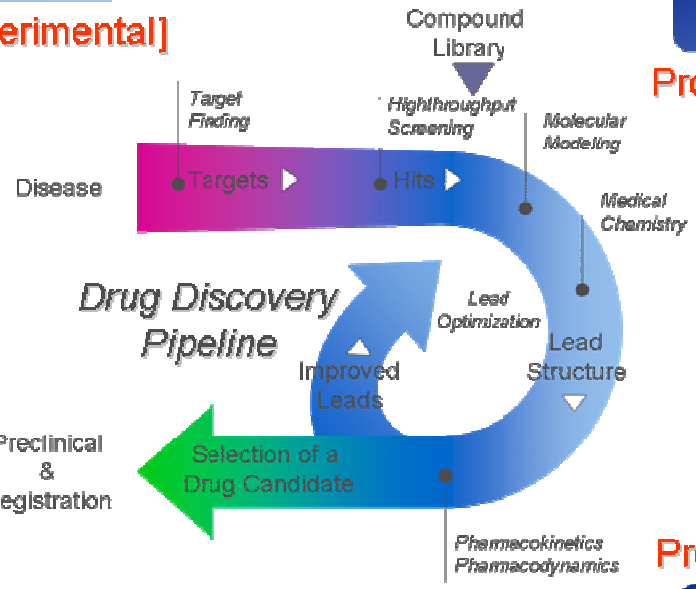
Sequencing [Analysis]

Montse Soler
Platform Leader

Sequencing [Experimental]

Juan Fernandez
BSC Group Leader

Protein-Protein Docking



Josep Ll. Gelpí
Platform Leader

SOAP Biology

Modesto Orozco
Director

Protein Dynamics

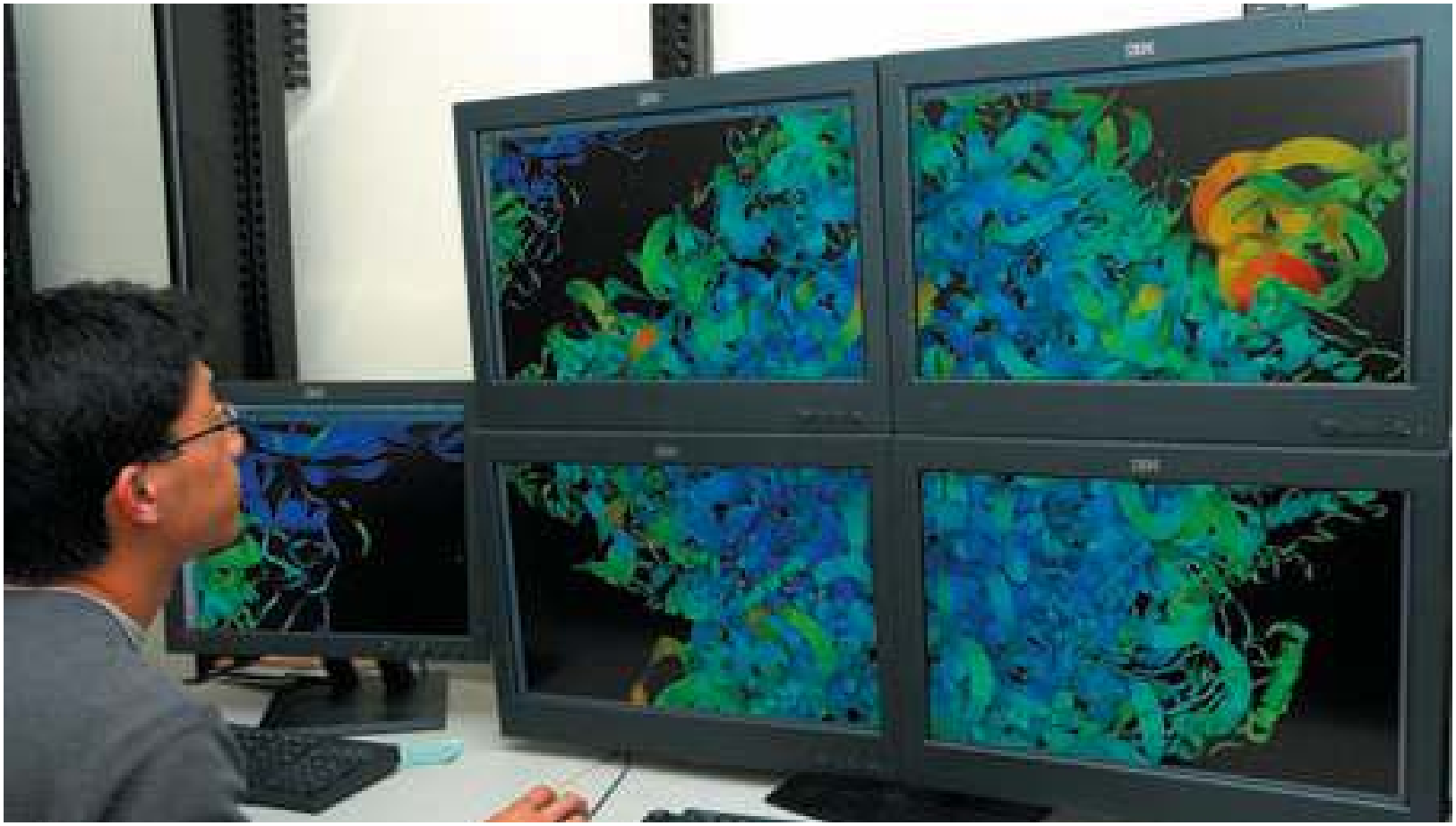
Protein-Ligand Docking

Victor Guallar
BSC Group Leader

Protein Networks

Patrick Aloy
BSC-IRB Joint Program

BioSupercomputing: The EXASCALE Challenge

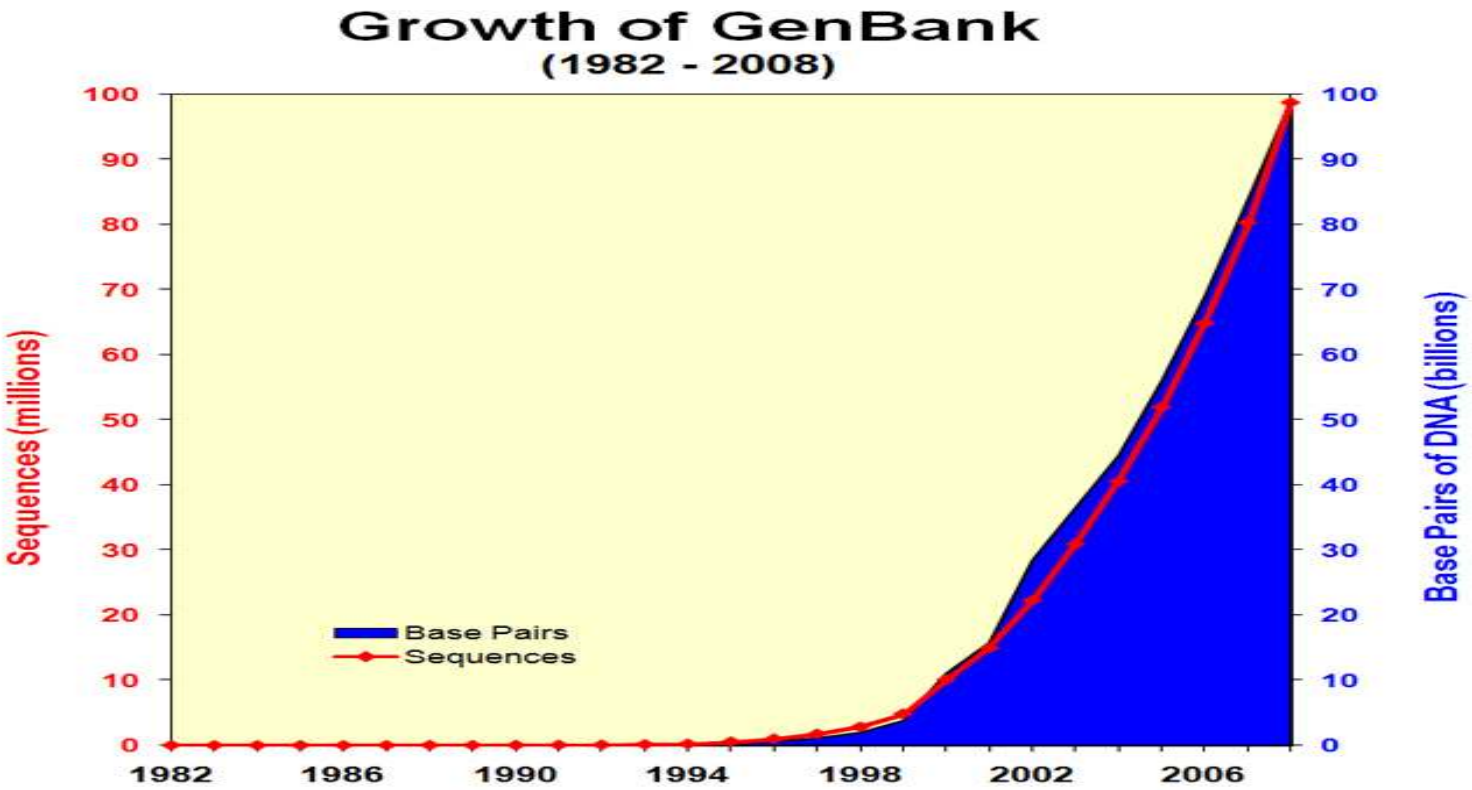


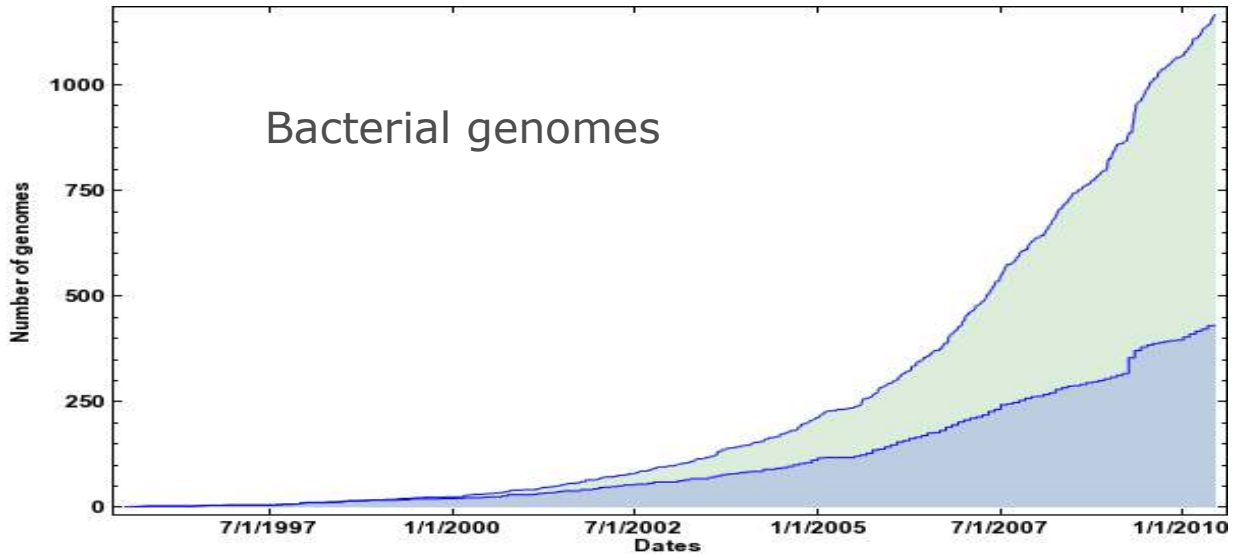
Life Sciences

- The biological problem is diverse in nature. Input of biologists should be important when selecting architectures
- Many biological problems are hard to scale
- Computer
  - **A) Manage data (Store)**
  - **B) Manipulate data (Calculator)**

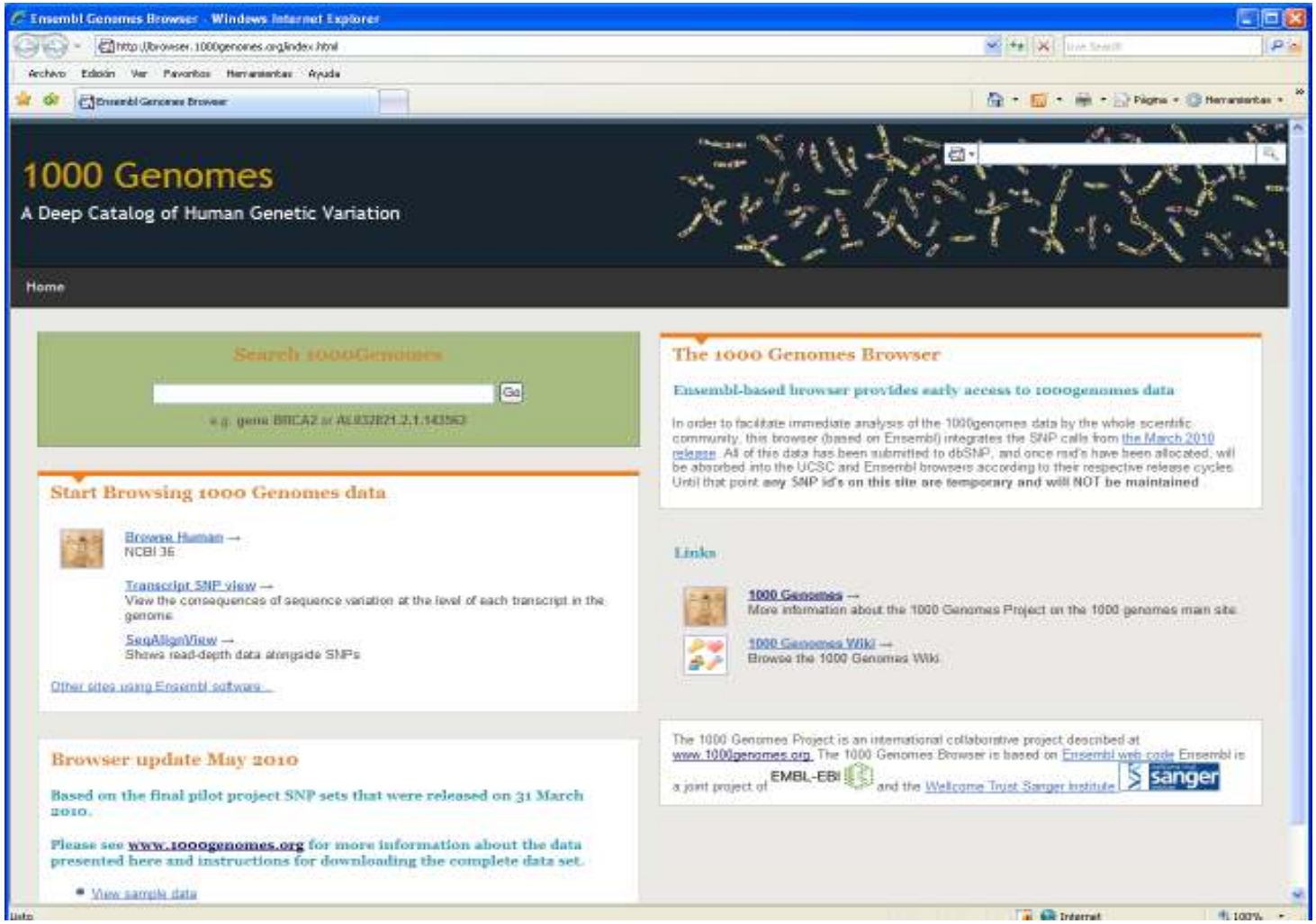
The computer as data manager

- Data growths exponentially
- Data management more complex
- Data integration a crucial issue
- Processing of data very costly

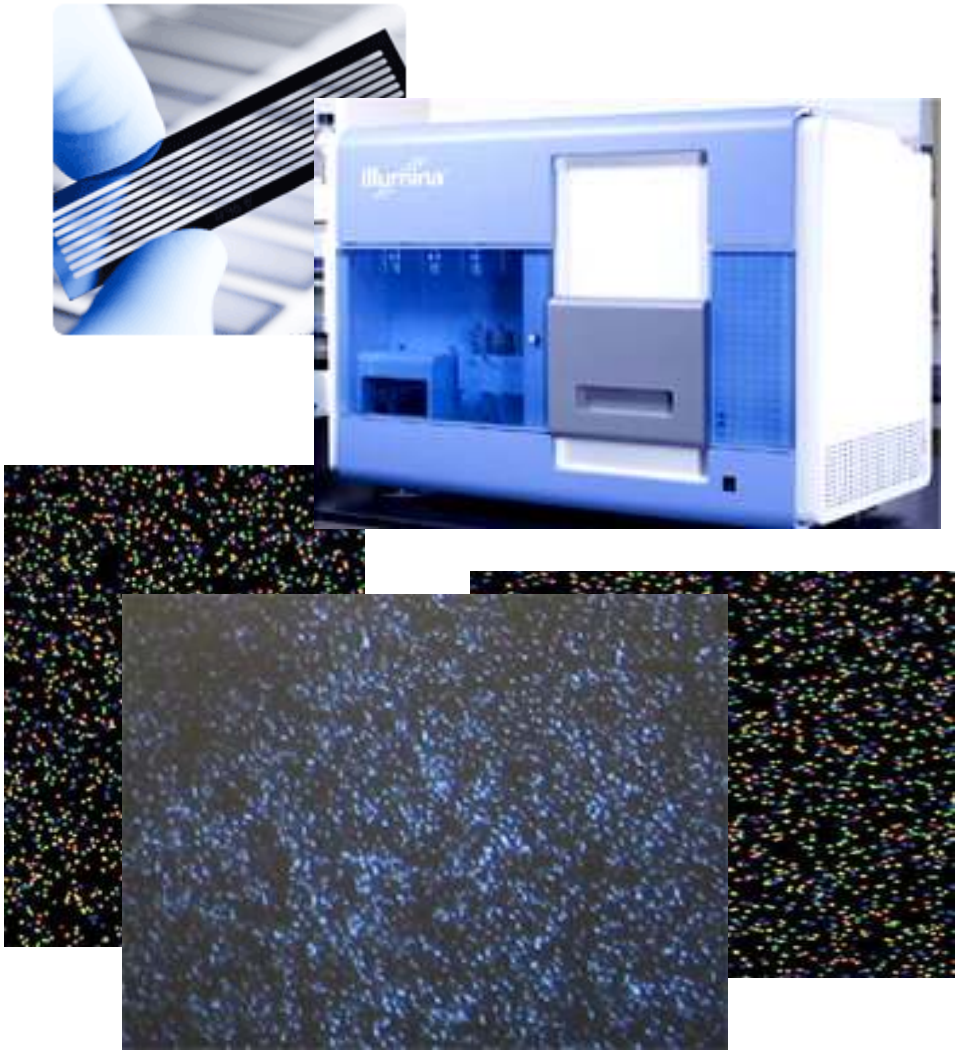

Growth of GenBank
(1982 - 2008)
Sequences (millions)
Base Pairs of DNA (billions)
Base Pairs
Sequences
0
10
20
30
40
50
60
70
80
90
100
1982
1986
1990
1994
1998
2002
2006


Bacterial genomes
Number of genomes
0
250
500
750
1000
7/1/1997
1/1/2000
7/1/2002
1/1/2005
7/1/2007
1/1/2010
Dates

Massive (human) genomic projects



Genomics: Next generation sequencing



- Every experiment 2 Tb of data
- Every machine 2 experiments a week
- A medium sized center 10 machines

A sequencing center generates around 2 Petabytes data a year!

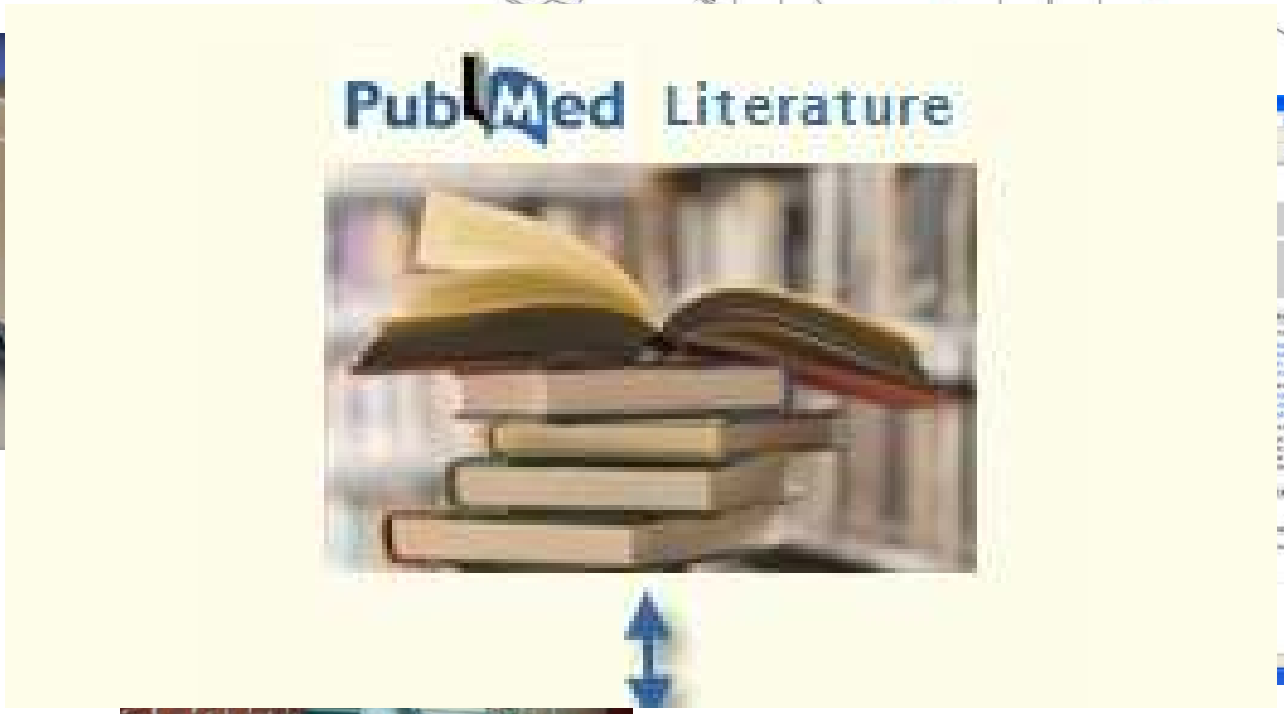
Massive (human) genomic projects

50 cancers

25000 cancer genomes

Data Link

Structures, chemical, literature,...



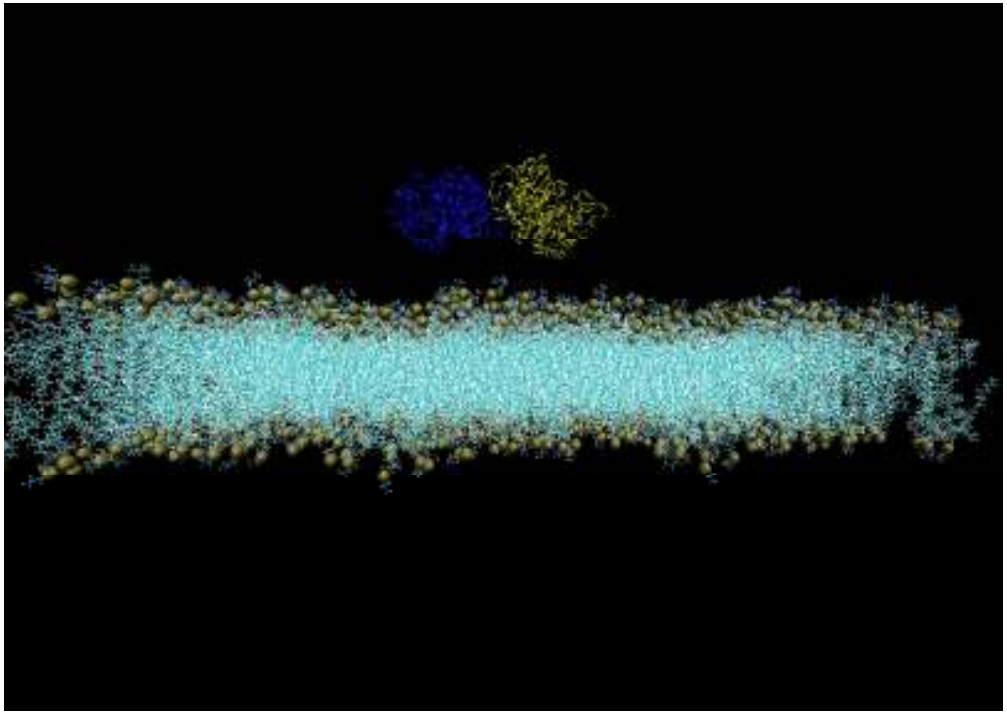
Chemical banks in the order of 10^7 compounds

The computer as a calculator

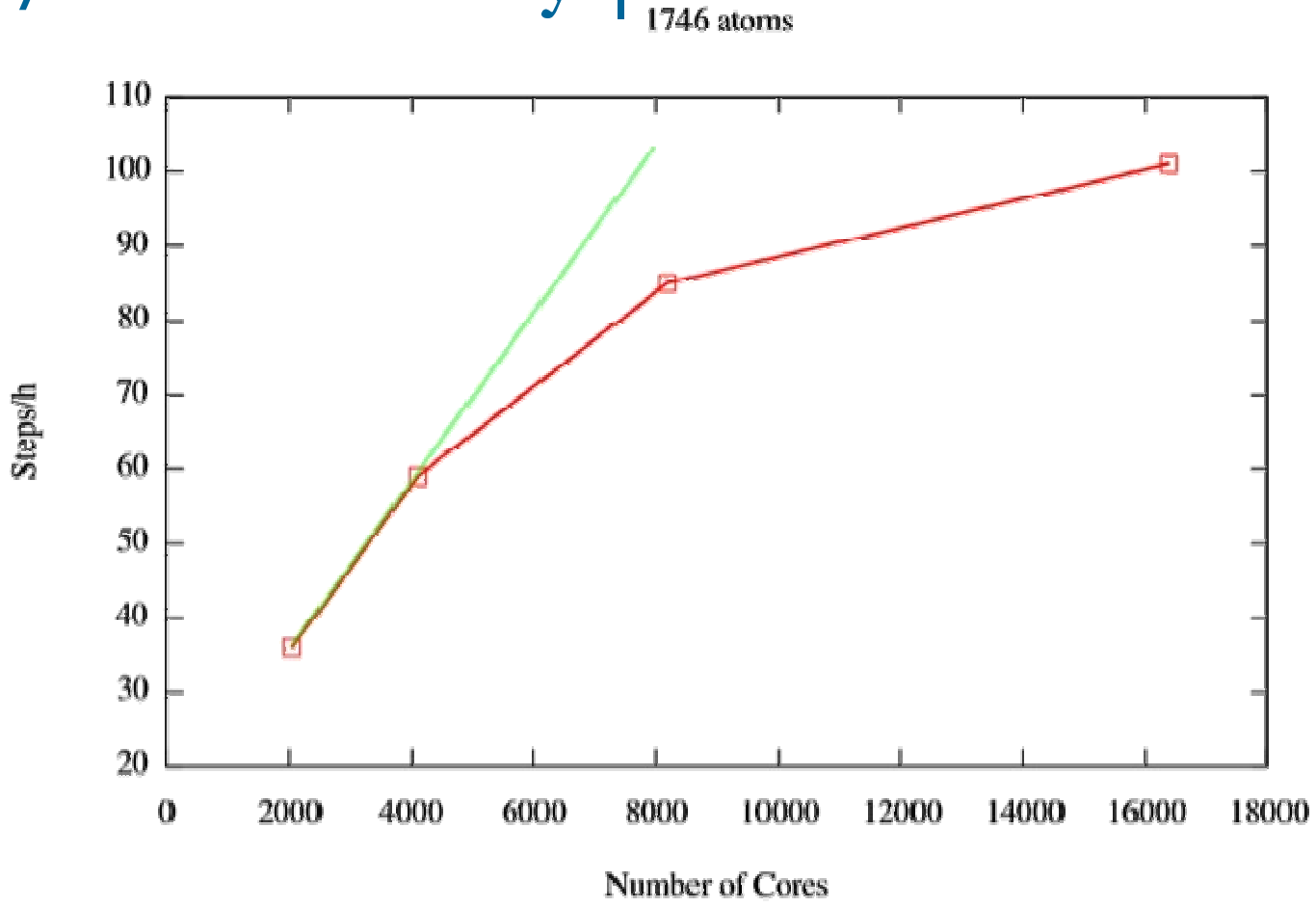
- Importance of simulation increases with:
  - **Increase in data on biological systems**
  - **Better definition of the problem**
- Different types of algorithms
- Often set-up conditions are unclear

Current limitations in MD

- Size of the system
  - **Typically: 10^4 - 10^5 particles**
  - **Flagship: 10^6**
- Simulation length (10^4 particles)
  - **Typically: 10^1 – 10^2 ns**
  - **Using HPC: μs**
  - **Using Anthon: ms**



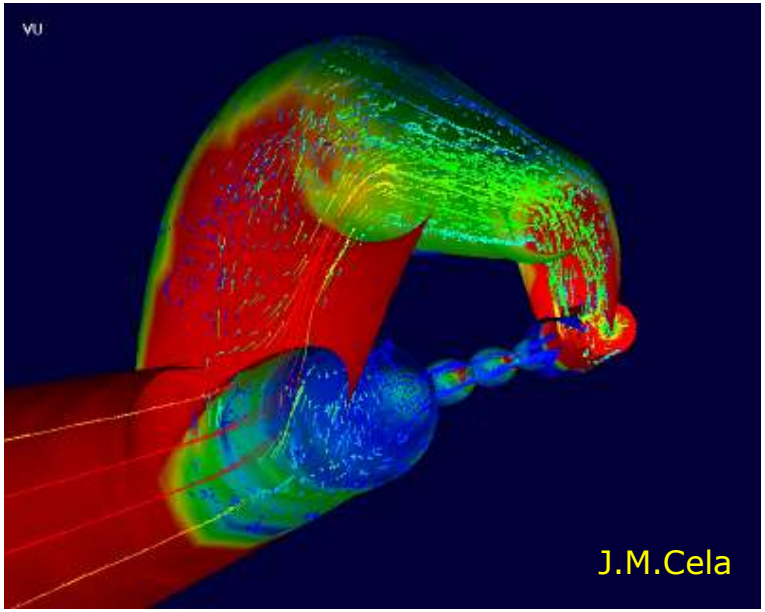
QM/MD Scalability | CPMD



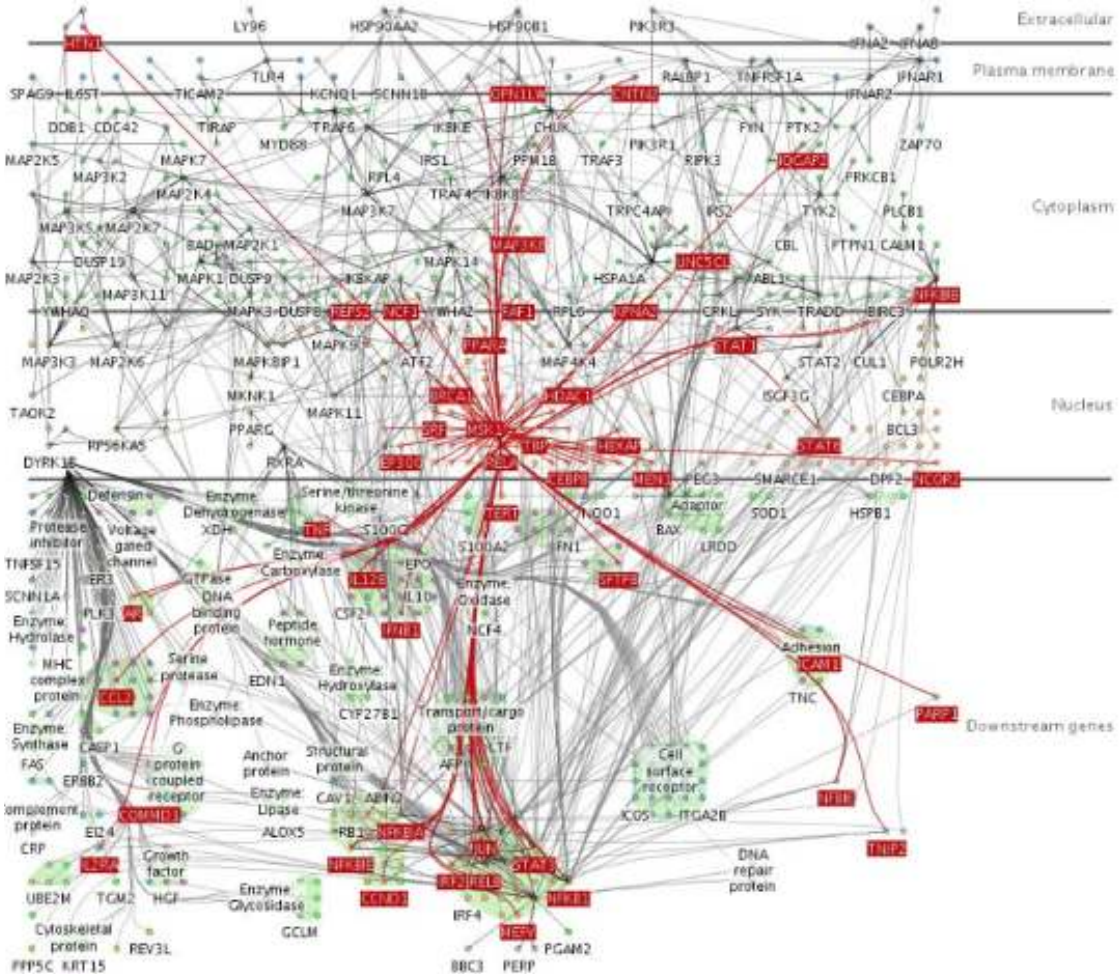
- www.cpmd.org implemented Jugene Curioni et al. (IBM Zurich)

Simulation scenario in Life Sciences

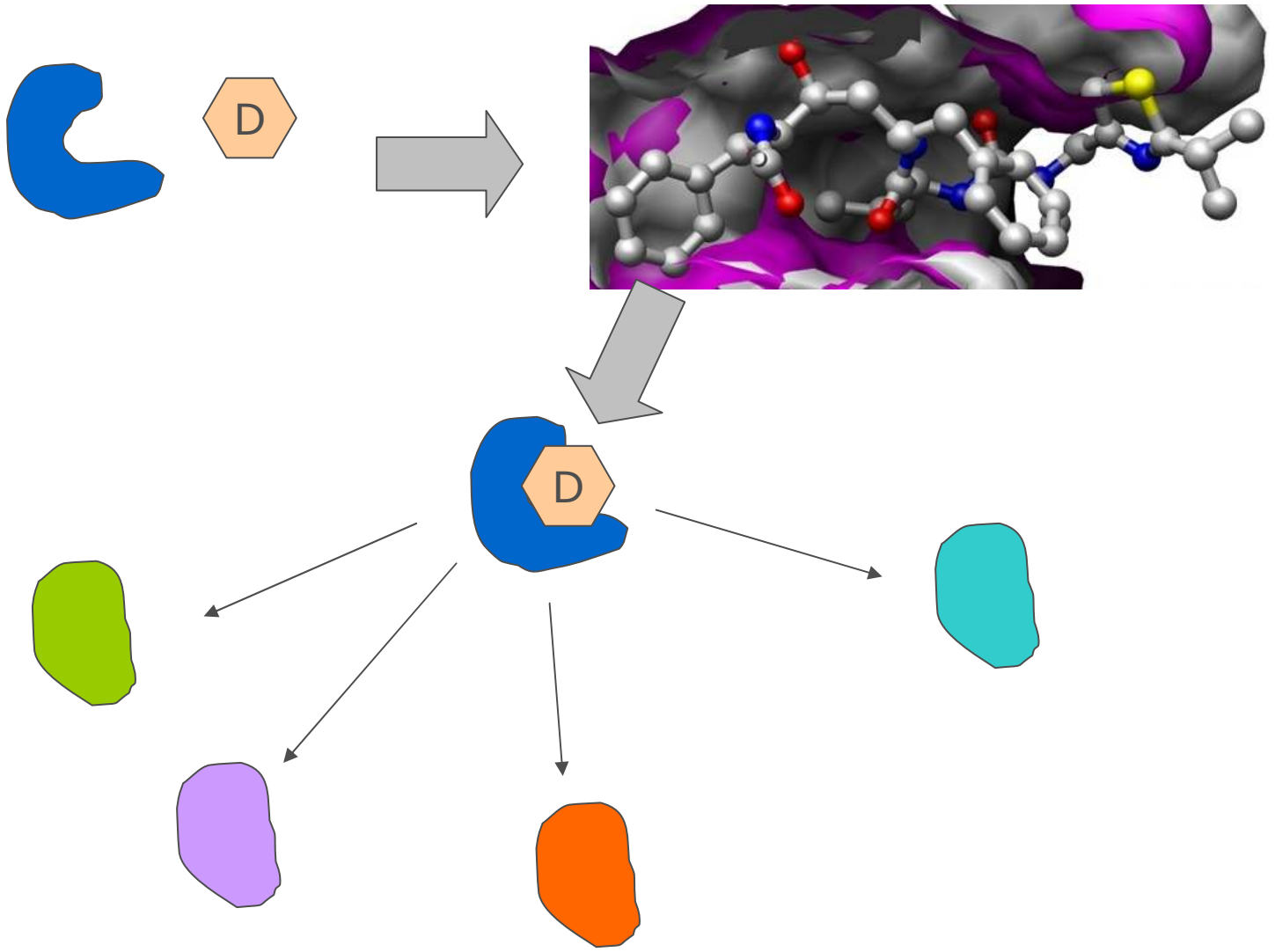
- Molecular simulations
  - **Structural prediction**
  - **Docking**
  - **Atomistic simulation**
  - **Cell-scale mesoscopic simulations**
- Gene inter-relations
- Cell simulation
- Organ simulation
- Ecosystem simulation



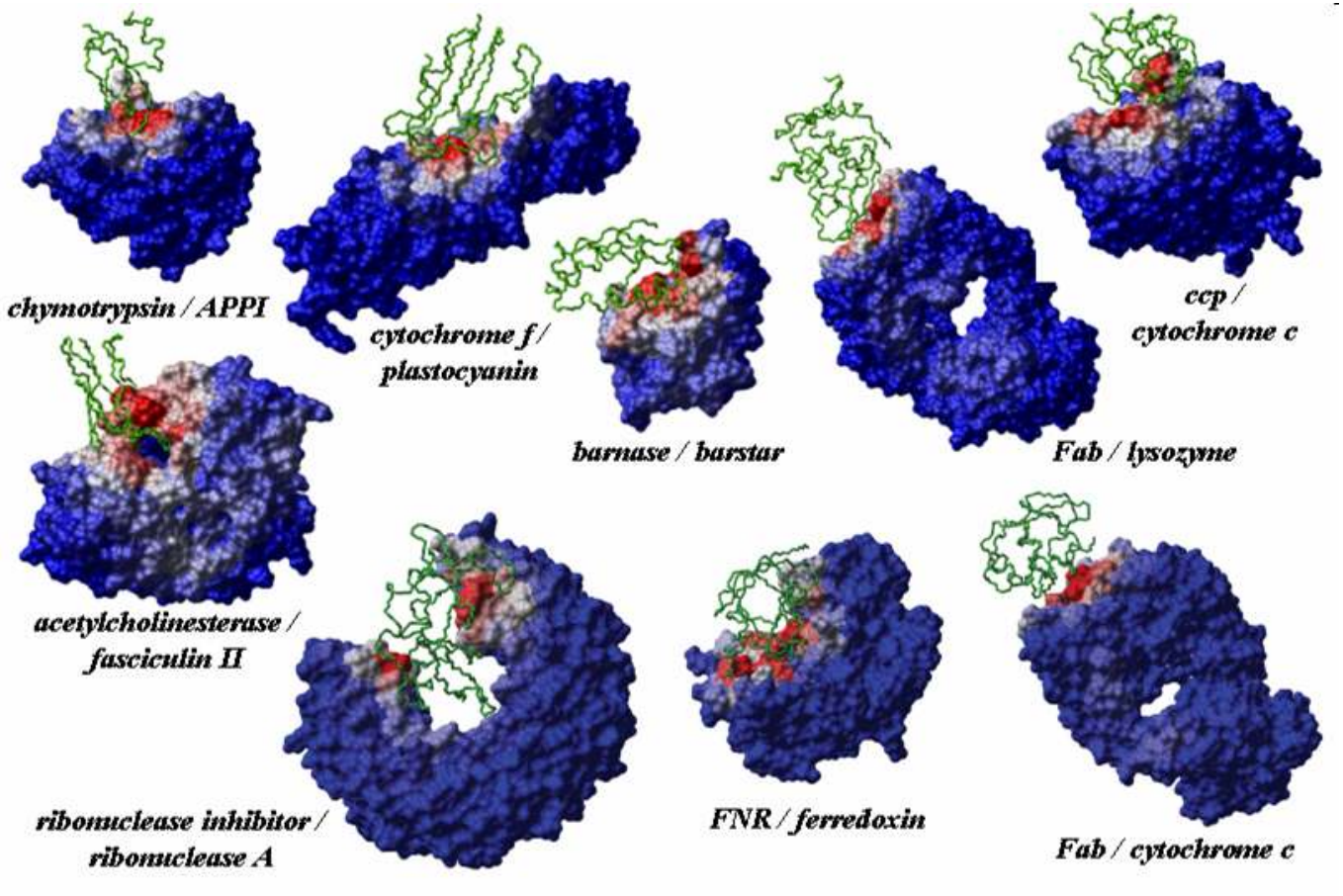
Breast cancer interactome



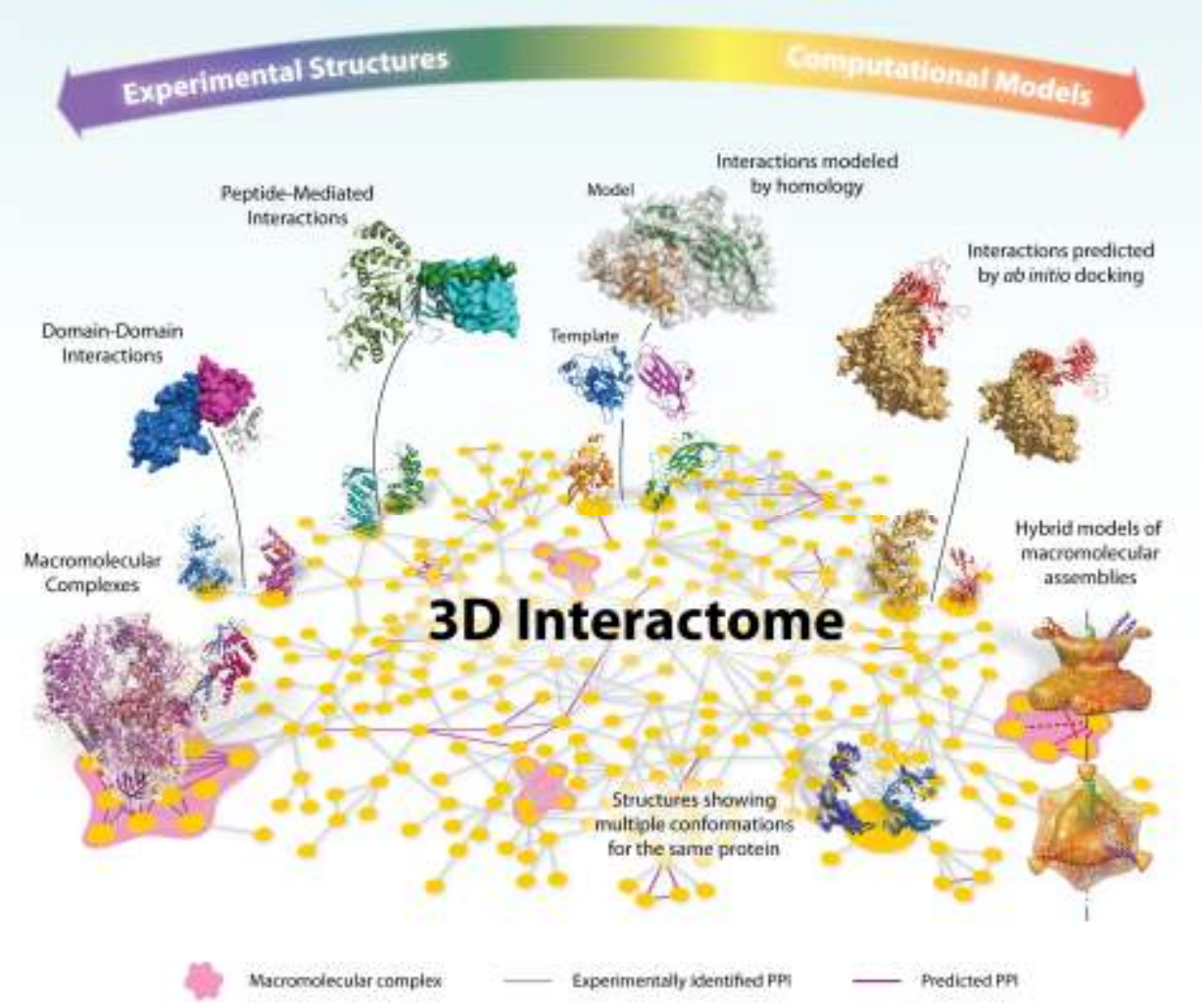
Cancer Interactome



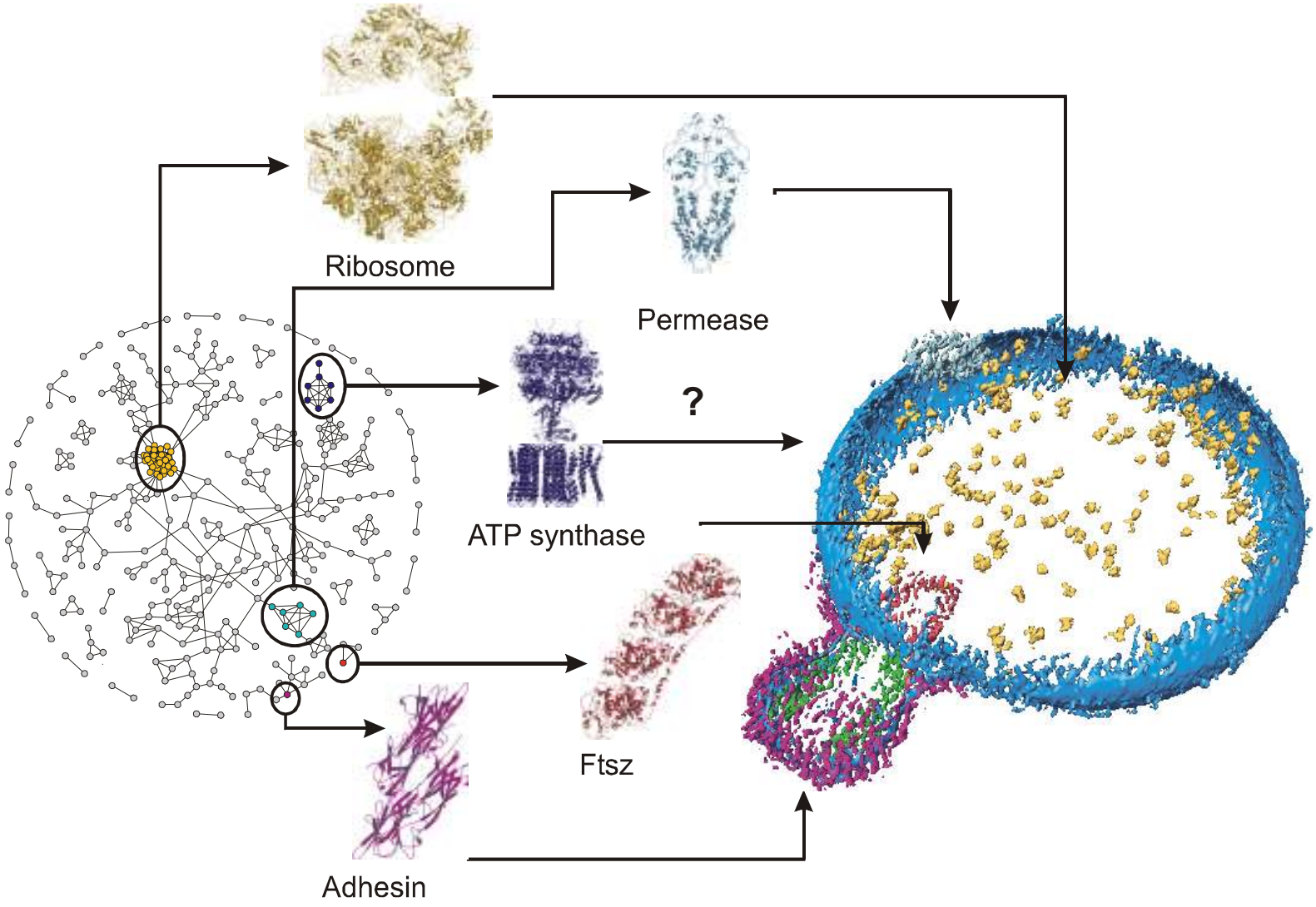
Protein-Protein Docking



YEAST 3D Interactome



From abstract networks to real cells



EXASCALE
Life Science
List of Experts

UF UNIVERSITY of FLORIDA
Jacksonville Healthcare, Inc.
Jacksonville Physicians, Inc.

Yale UNIVERSITY

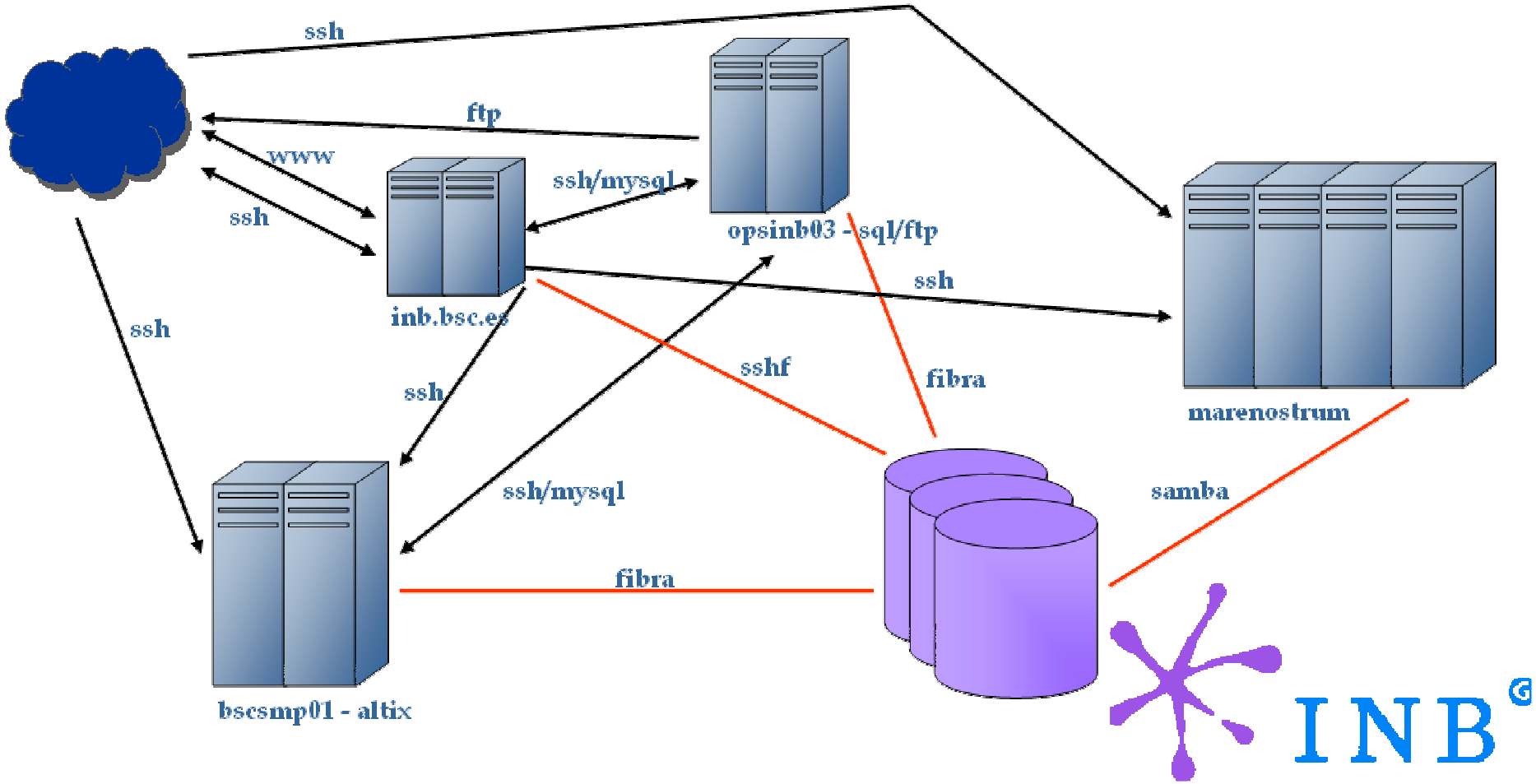
Biocomunity HPC Support

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

INB

ELIXIR

BioSupercomputing open to "BioCommunity"

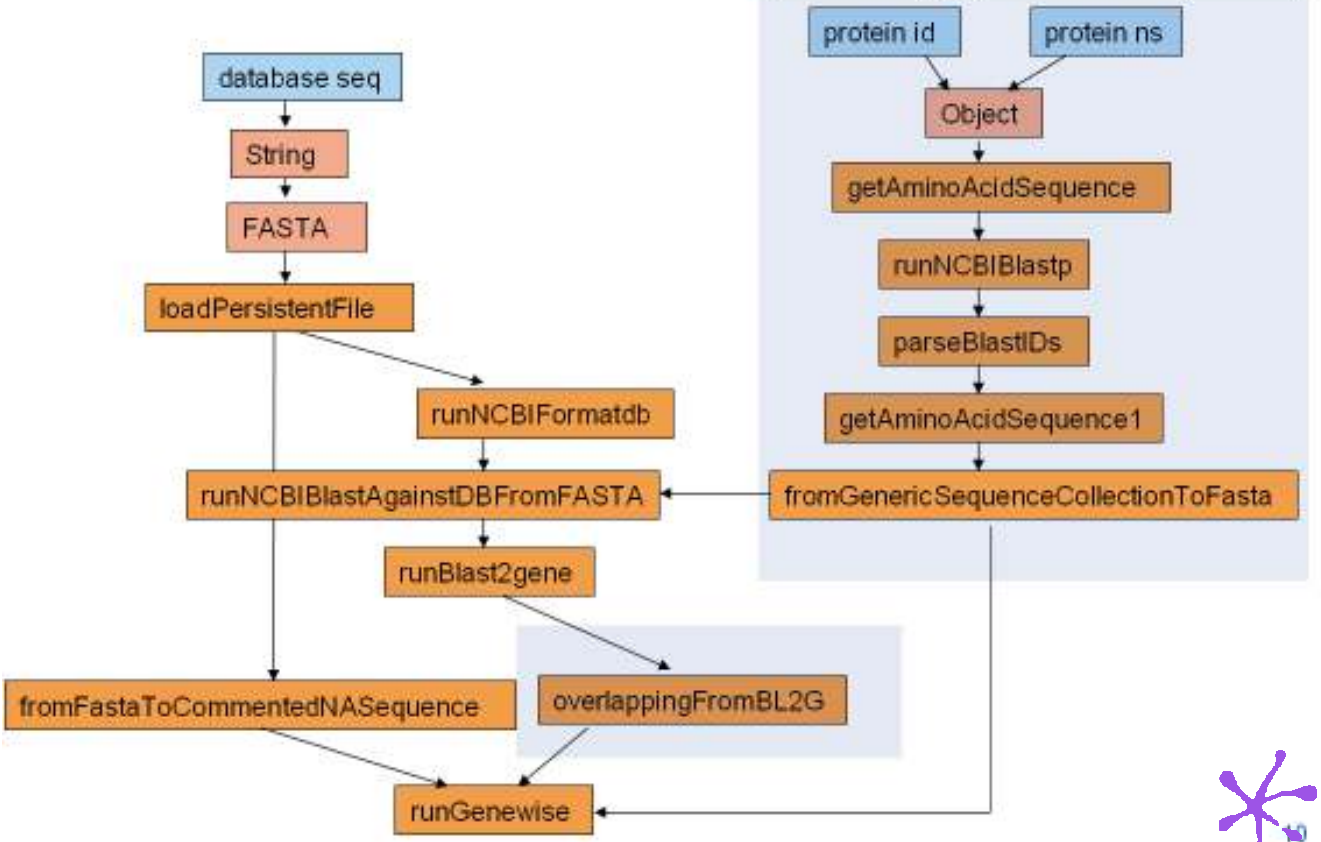


Pre-defined Workflows

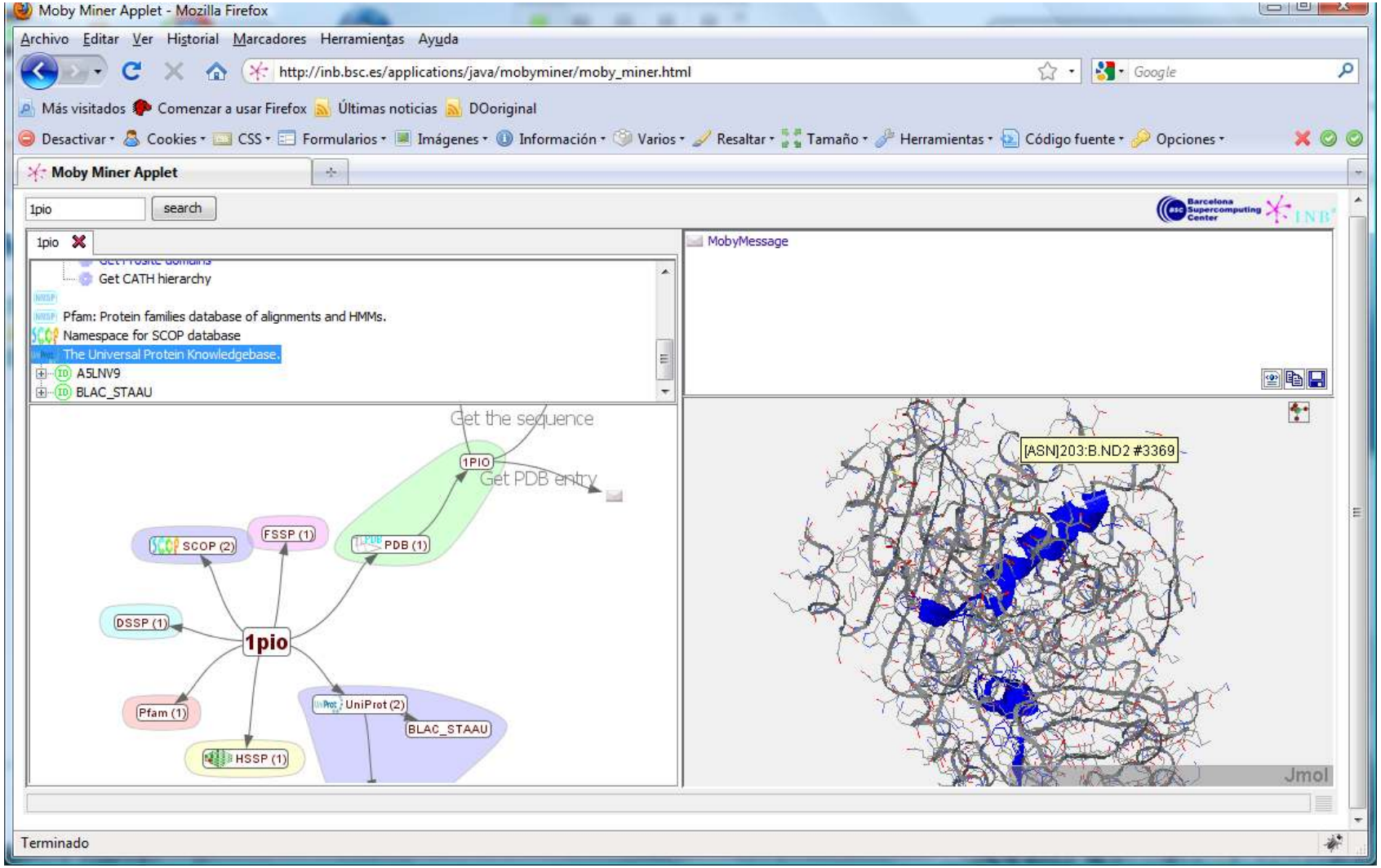
ISC 2008
A BioMoby-based workflow for gene detection using sequence homology. Royo, R., López, J., Torrents, D., Gelpi, J.L.

Workflow

Gene detection by homology (D. Torrents, BSC.)



User-friendly for non-expert



Molecular Modeling & Bioinformatics

Electronic and Atomic Protein Modeling

Protein Interactions and Docking

Computational Genomics

Experimental Bioinformatics Laboratory

INB-BSC Algorithmic Unit

Contact:

Ramon Goñi
Senior Researcher
BSC, Life Sciences
ramon.goni@bsc.es