



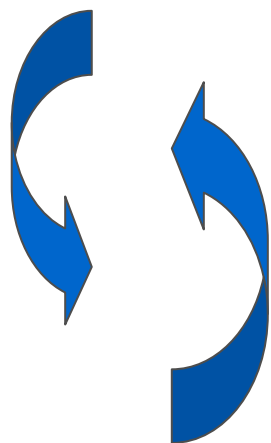
**Barcelona
Supercomputing
Center**

Centro Nacional de Supercomputación

The Power of Computation in Life Sciences

Life Sciences - BSC





GROUP

HEAD

RESEARCH AREA

EAPM

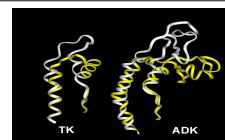
Victor
Guallar



Atomic (and electronic) modeling of protein biochemistry and biophysics

MMB

Modesto
Orozco



Micro and mesoscopic modeling of macromolecules and Drug Design

PID

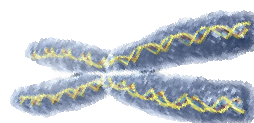
Juan
Fernández



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

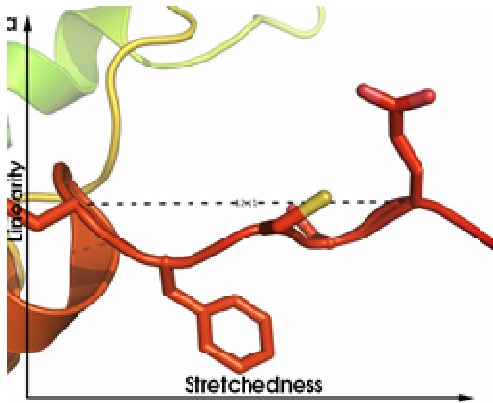
AU

Josep Ll.
Gelpí



Web services, applications, databases and software integration.

Research Lines



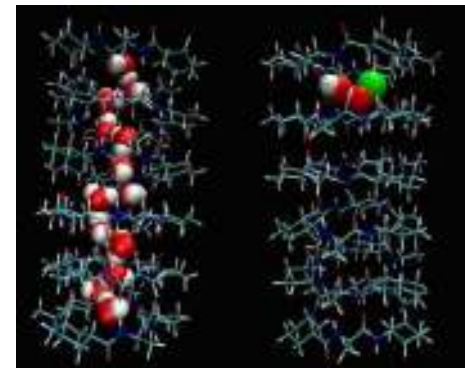
Target & Drug Discovery



Genomics & System Biology



BioSupercomputing



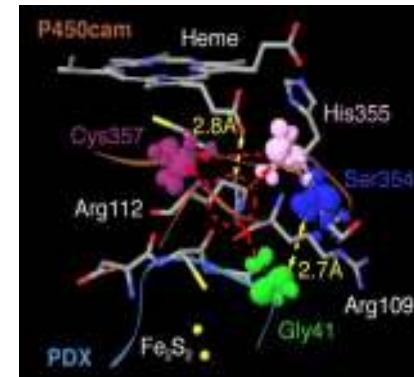
BioPhysics

Electronic and Atomic Protein Modeling

Victor Guallar



- **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.

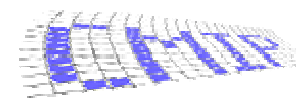
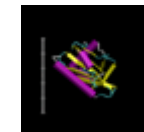
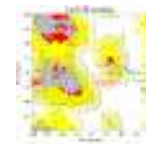
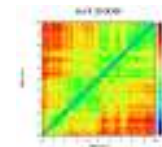


Molecular Modeling & Bioinformatics

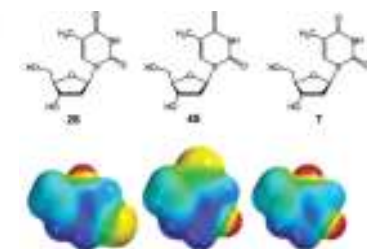
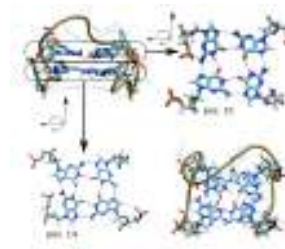
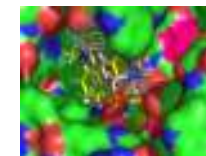
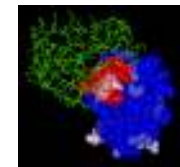
Modesto Orozco



- **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.



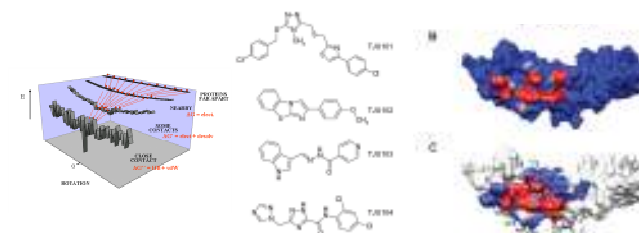
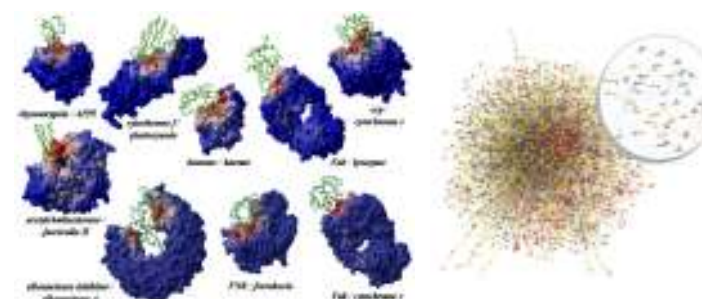
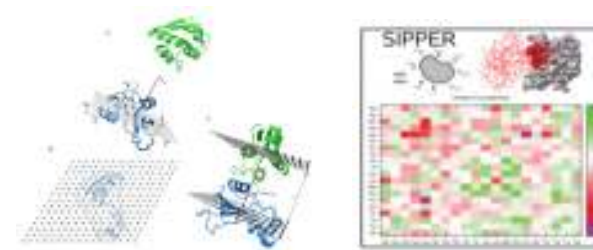
MD-GRID



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  Institució Catalana de Recerca i Estudis Avançats

- **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



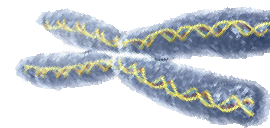
TOMATO



DIABETES



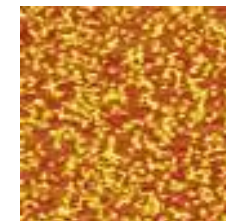
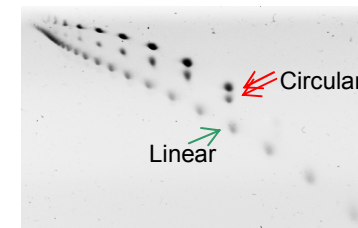
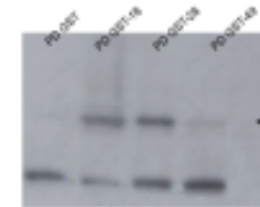
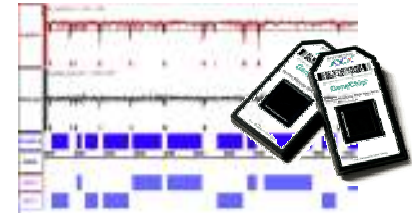
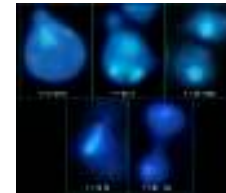
MALARIA



Experimental Bioinformatics Laboratory

Montse Soler 

- **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.

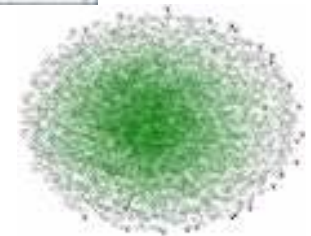


INB-BSC Algorithmic Unit

Josep Lluís Gelpí



- **Web & Database Development.**
MobyMiner, INB workflows and Databases
- **Epidemiology.** Reutmathoid arthritis (IMIDKIT)
- **Elixir and ICGC projects.**
Datamining, GSC and CSCm Mobylite.

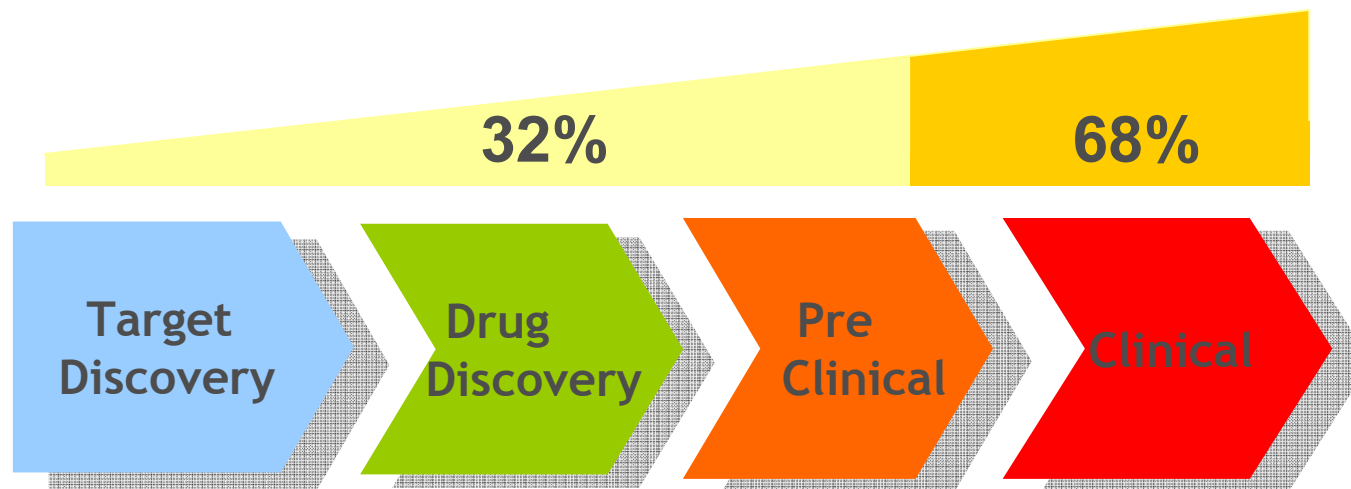


Rational Drug design



Drug Development

Cost: \$1.2B / drug

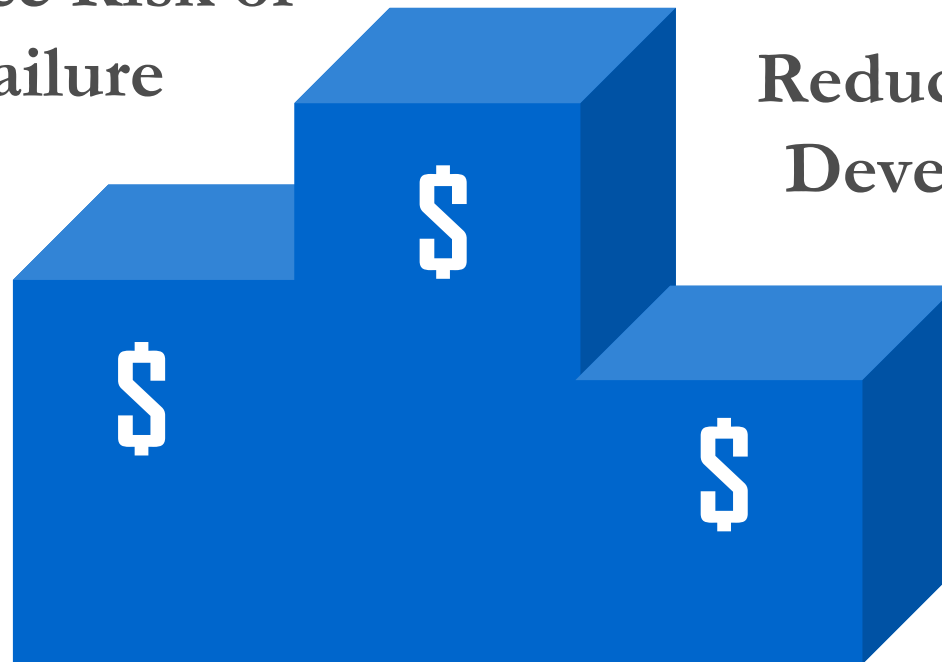


Computer-Based Drug Design

Improve Time
to Market

Reduce Risk of
Failure

Reduce Cost of
Development



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

PRECISION



Quantum Mechanics

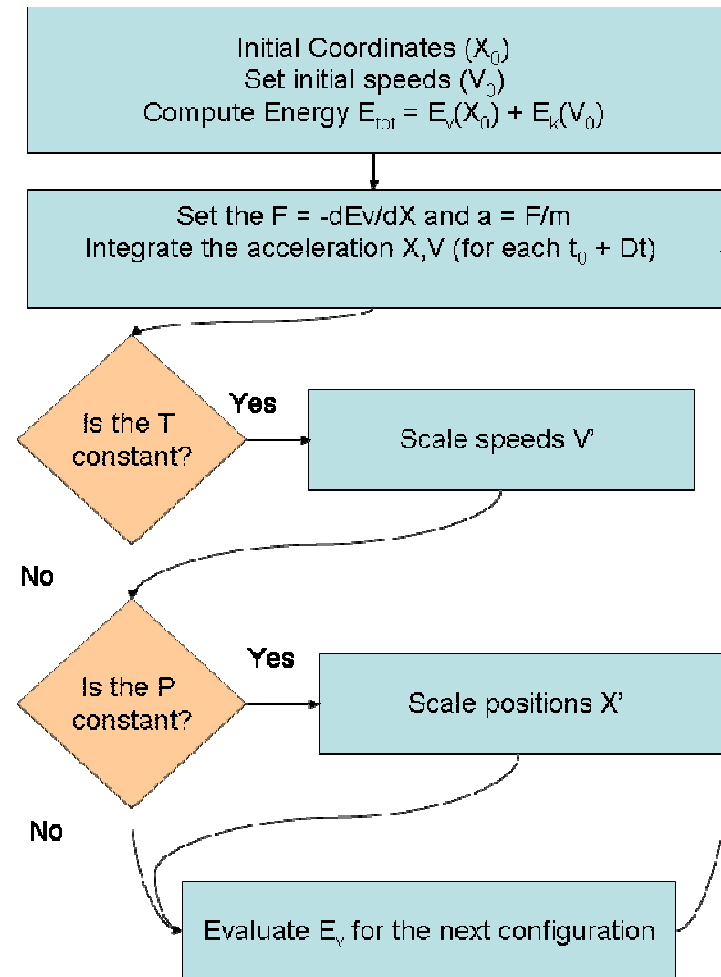
Molecular Dynamics (Newton motions)

Coarse Grained – MD (Pseudo-atoms)

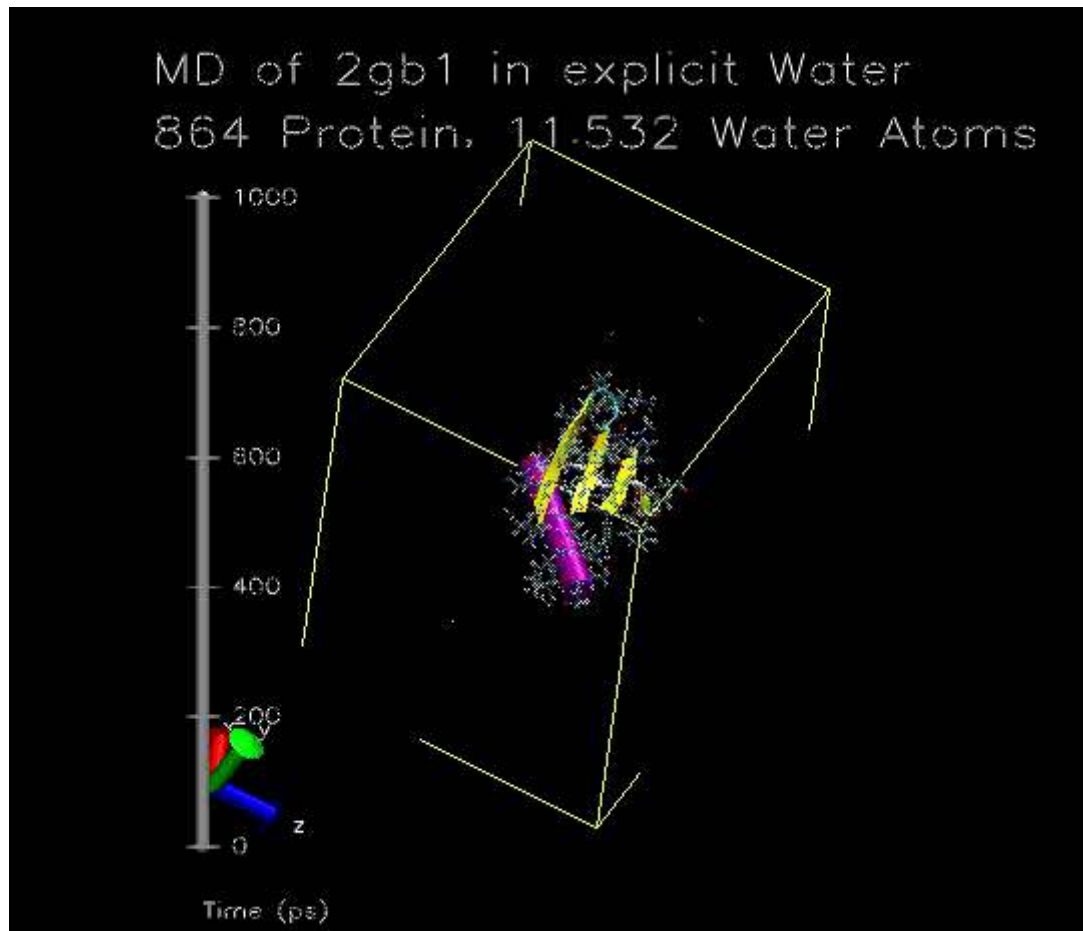
SPEED

Molecular Dynamics

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



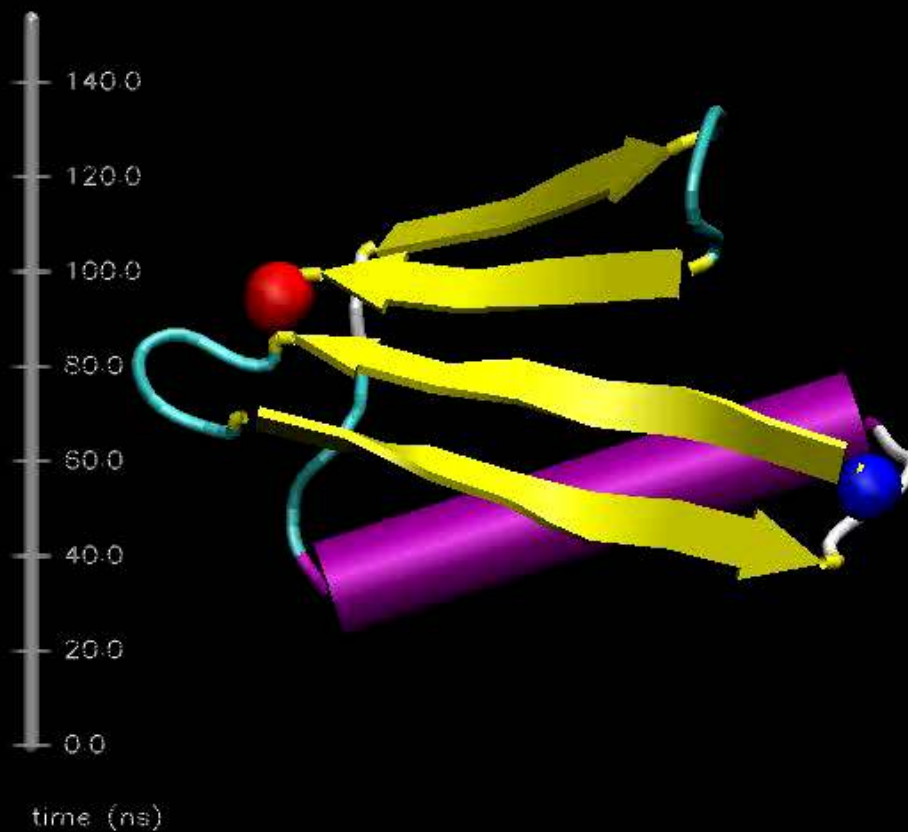
Molecular Dynamics of Solvated Protein



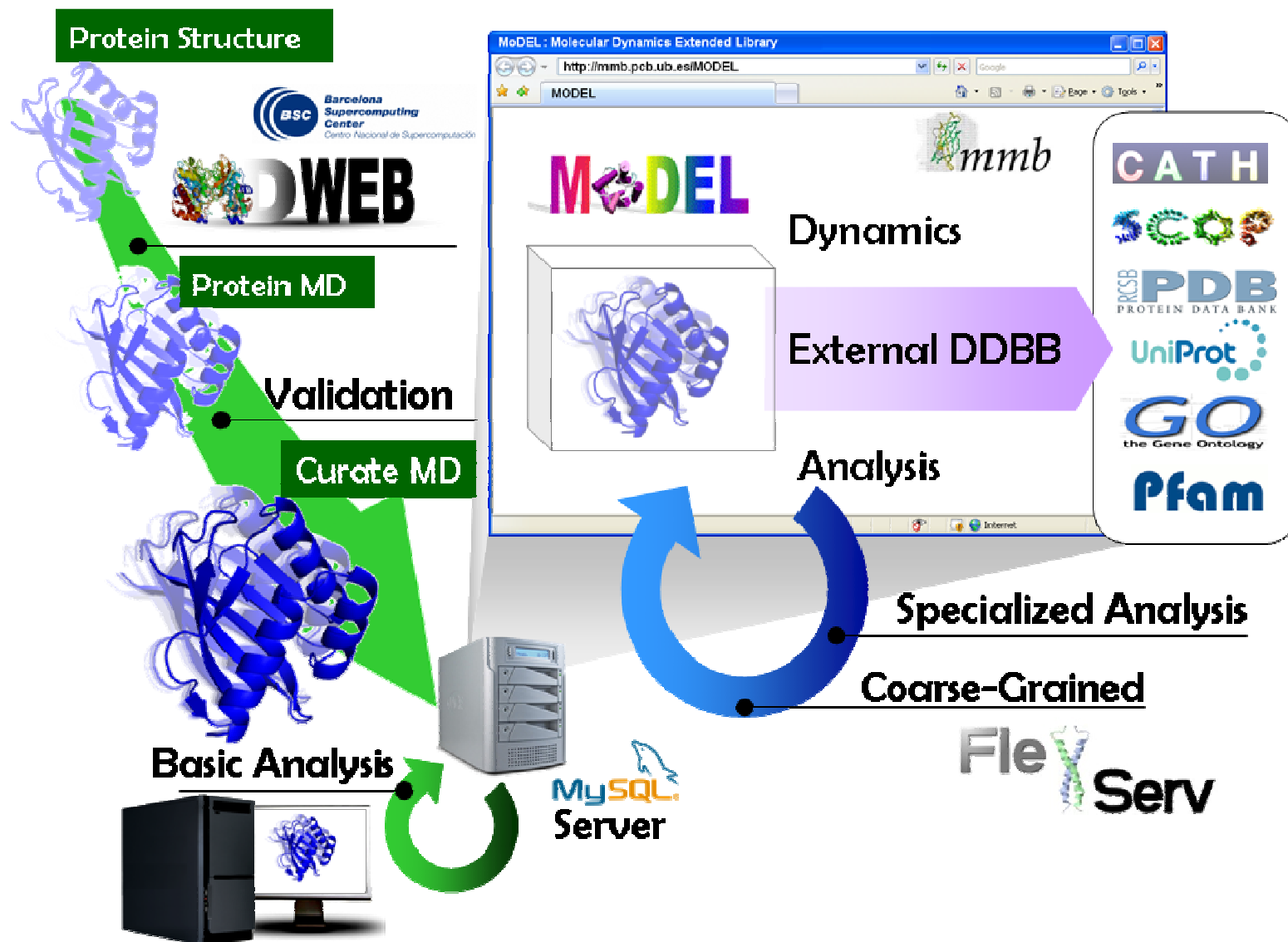
- Snapshot every femtosecond (10^{-15}s)
- System of 10^4 atoms
- 10 operations per atom pair-mate
- Using 16 processors we are able to simulate 10 nanoseconds (10^{-9}s) per day

MD Scale

MD Simulation of 2gb1 in Vacuum
all HIS, GLU, and ASP protonated



- Today the millisecond scale (10^{-3} s) is reached using specific hardware (512-processor)



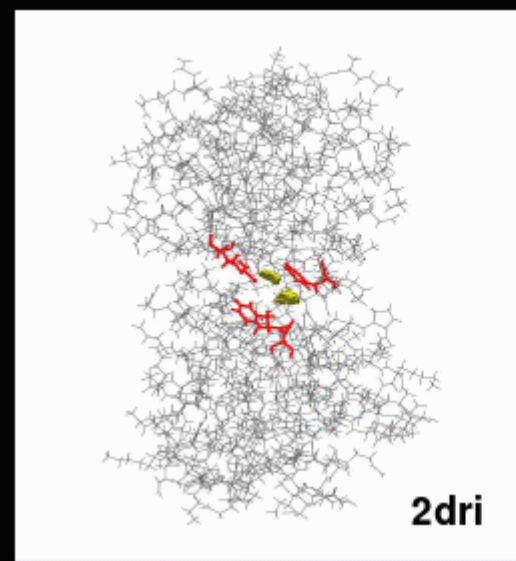
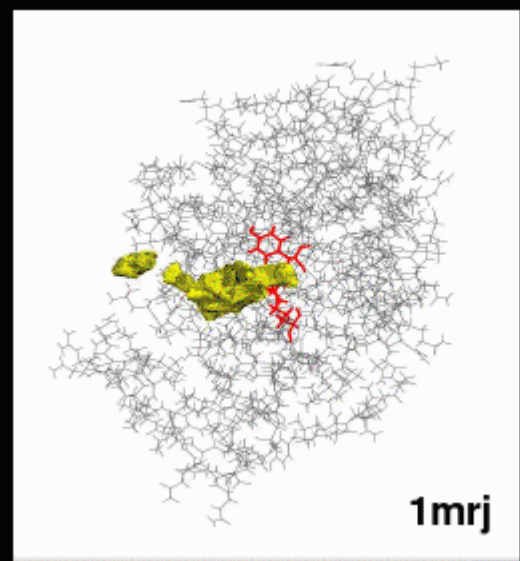
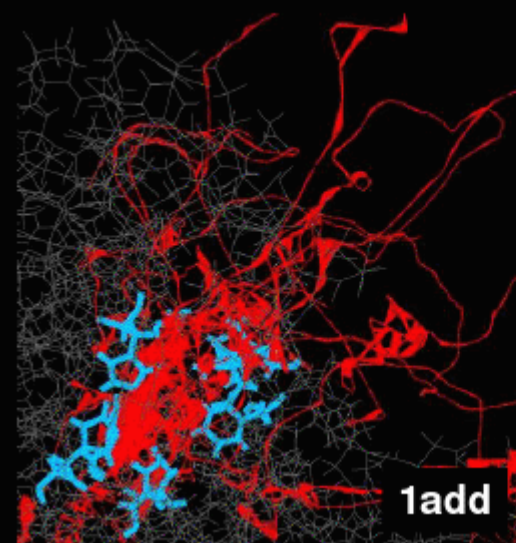
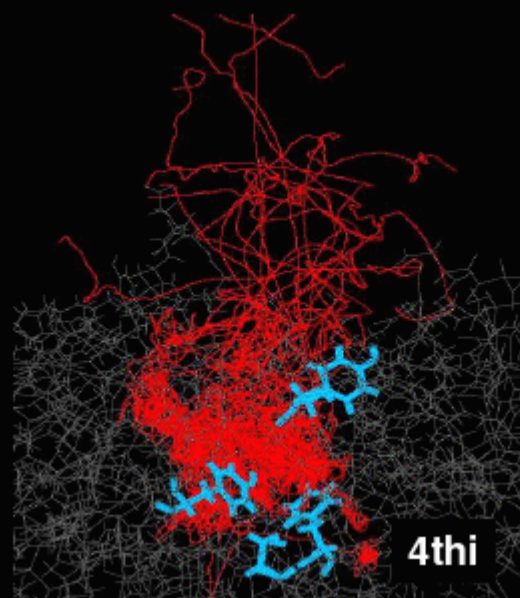
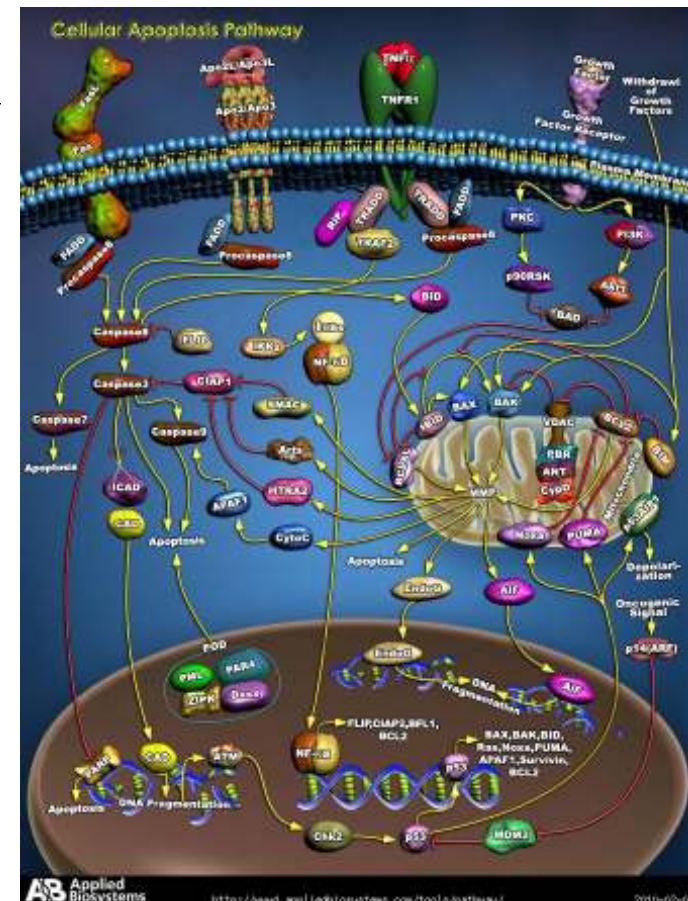
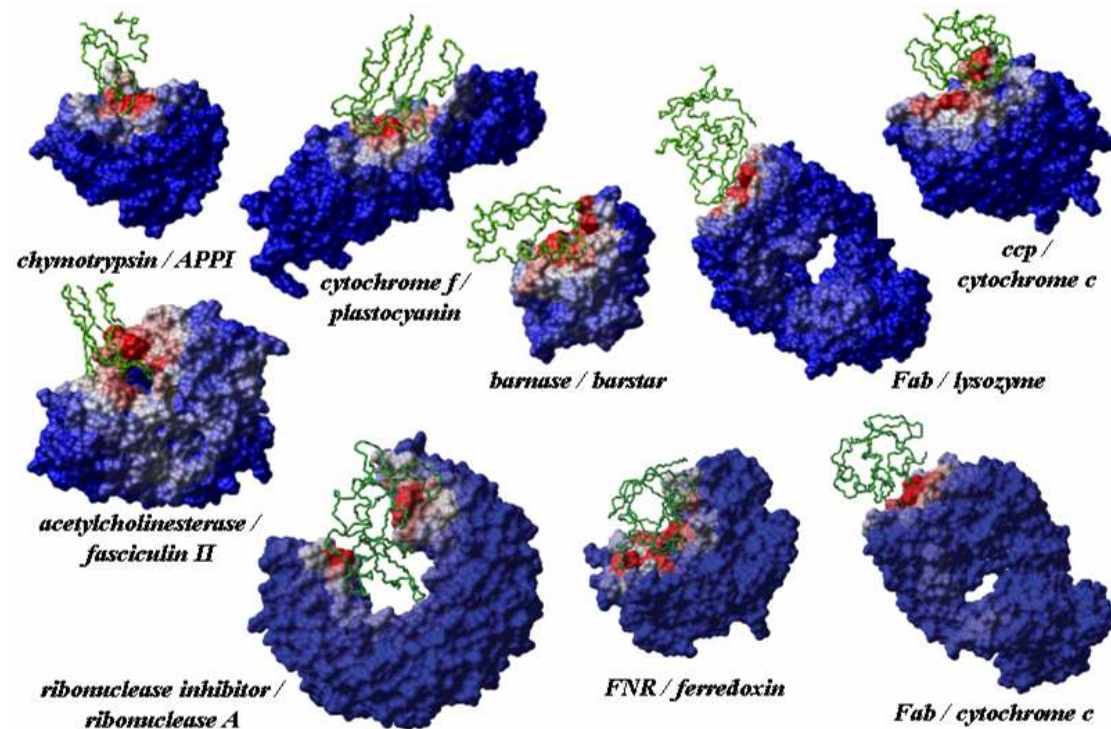


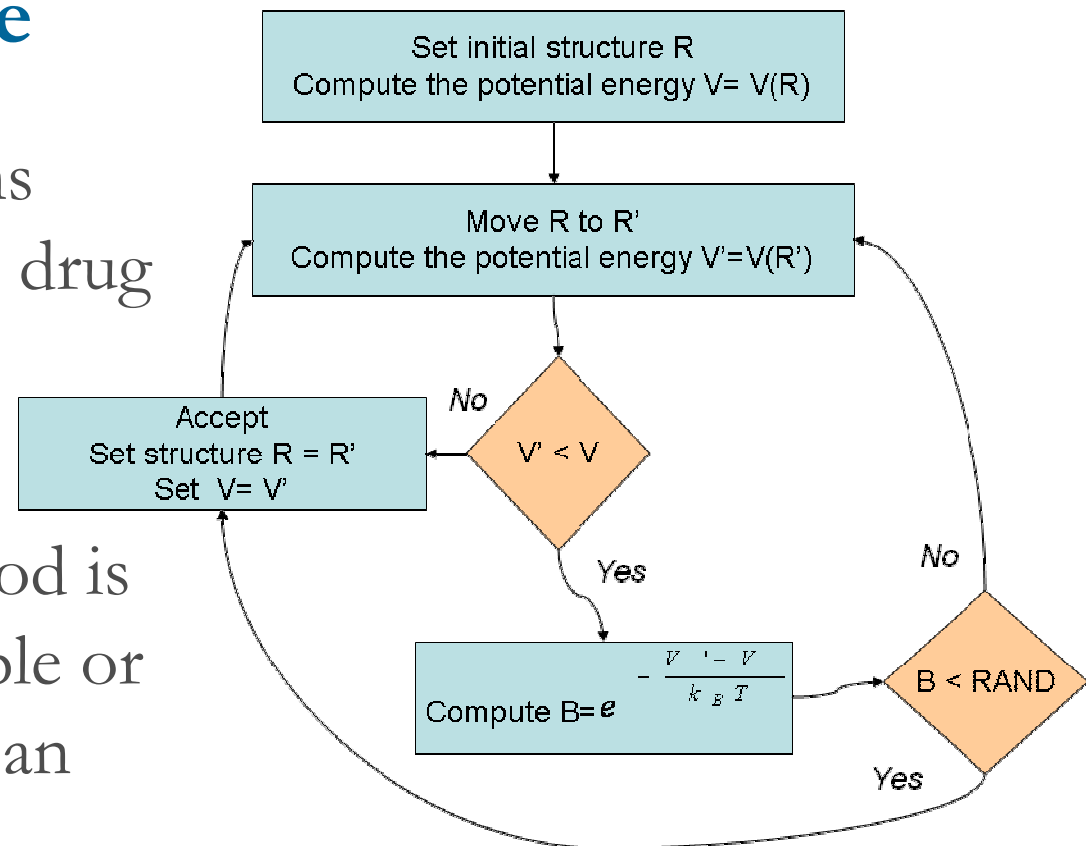
Fig 12

Protein Interaction



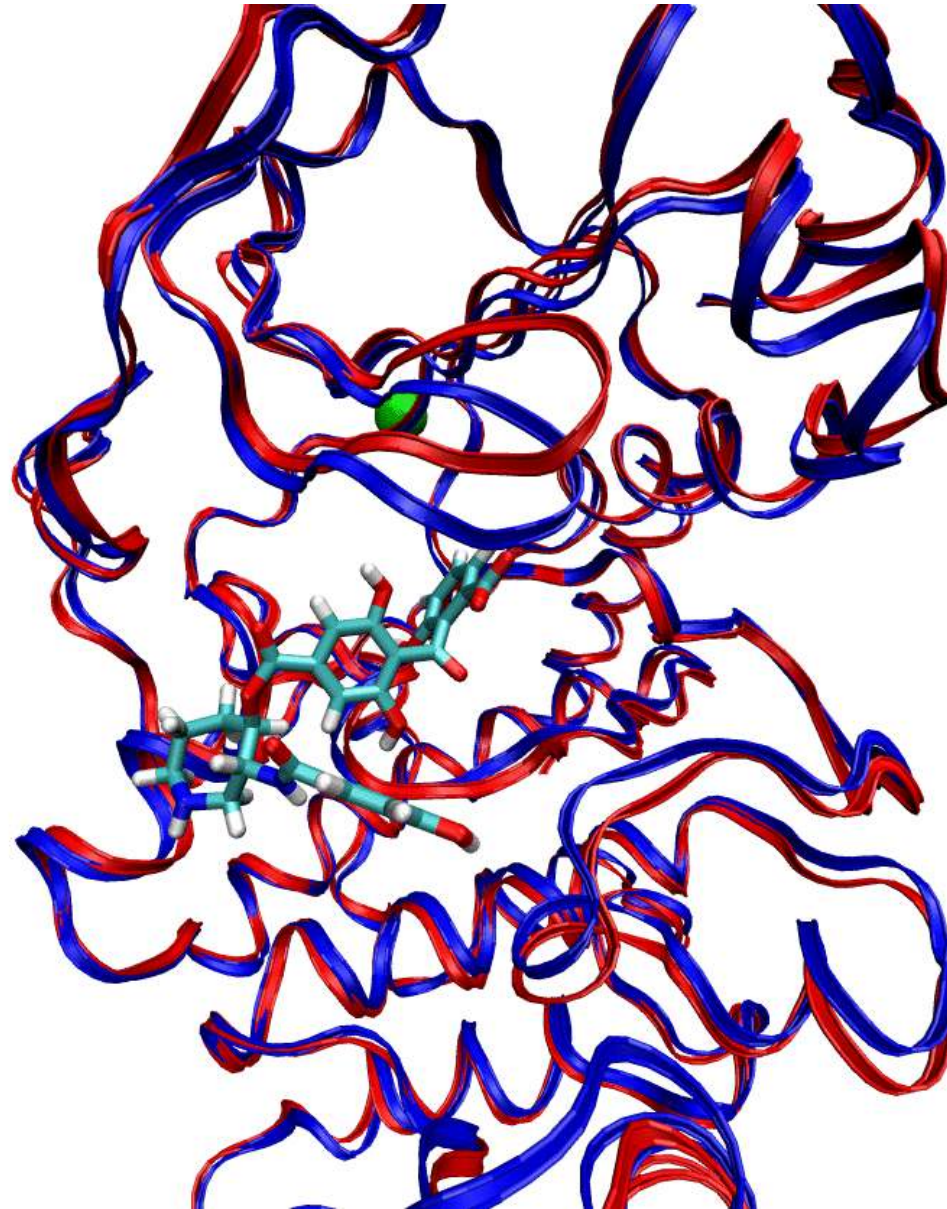
Simulate and Explore

- Explore how 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





RESEARCH

TEAM

Strong IP

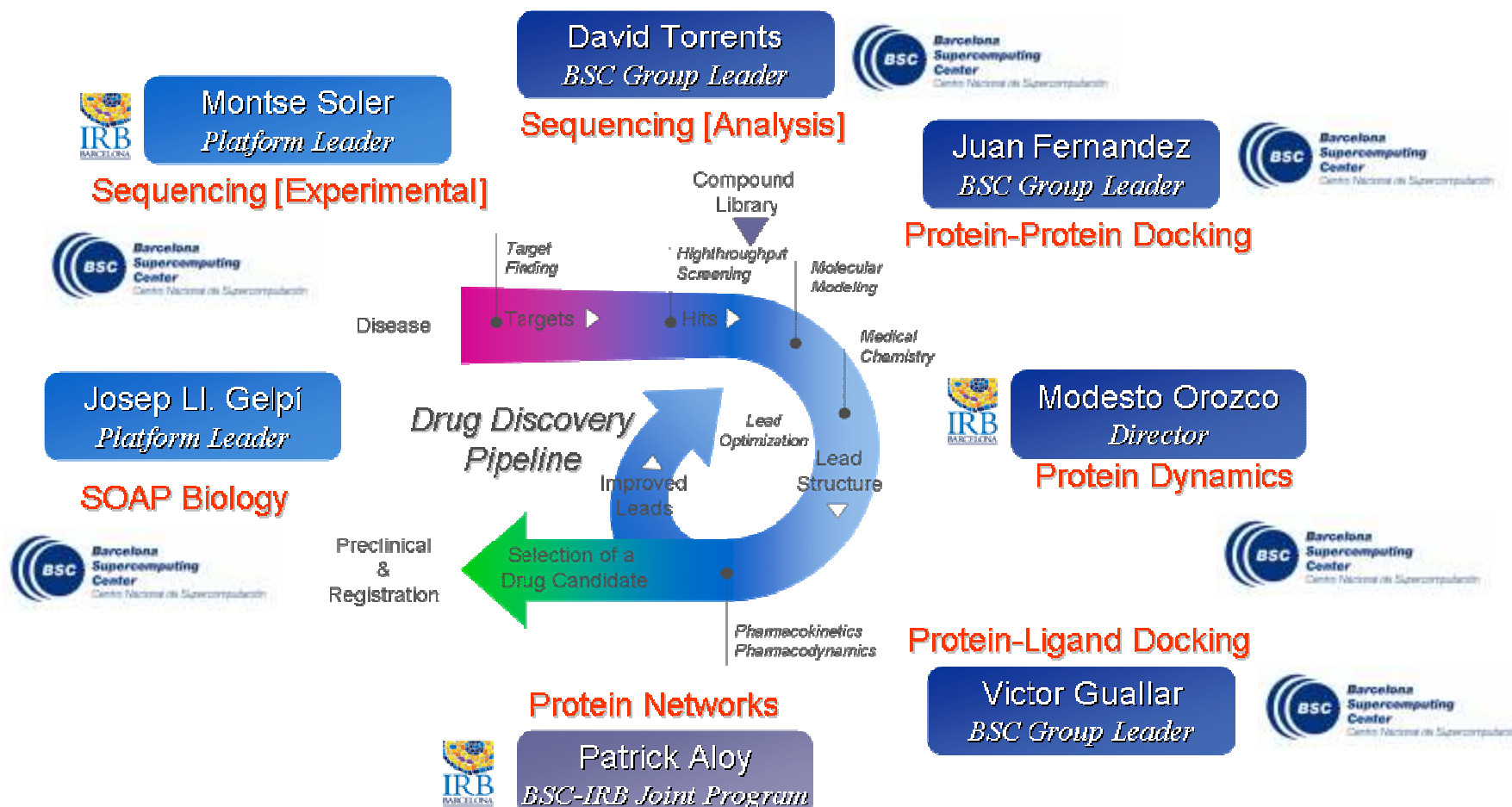
Product

Way to Market

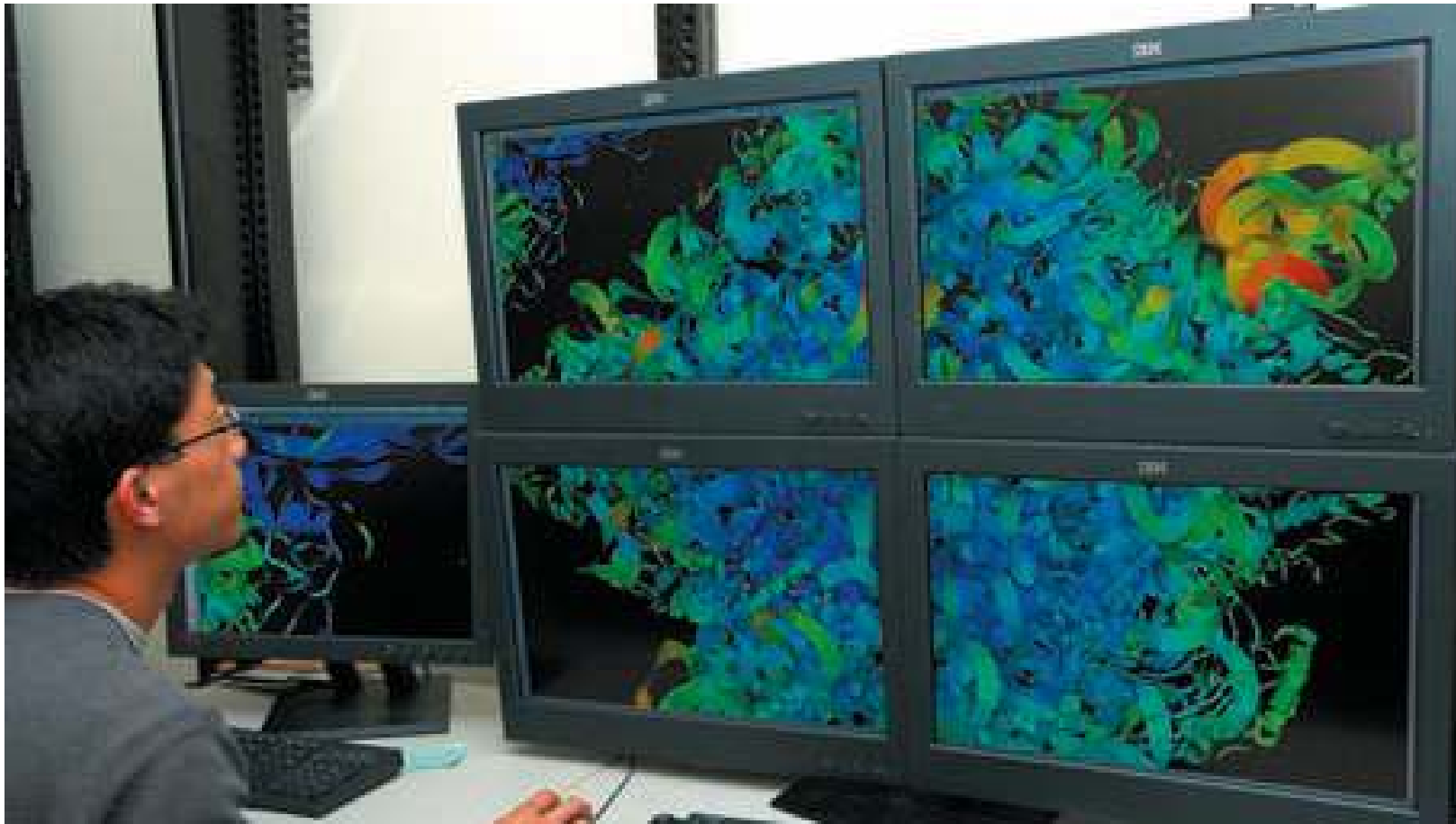
Big Market

Business Model

MARKET



BioSupercomputing: The EXASCALE Challenge



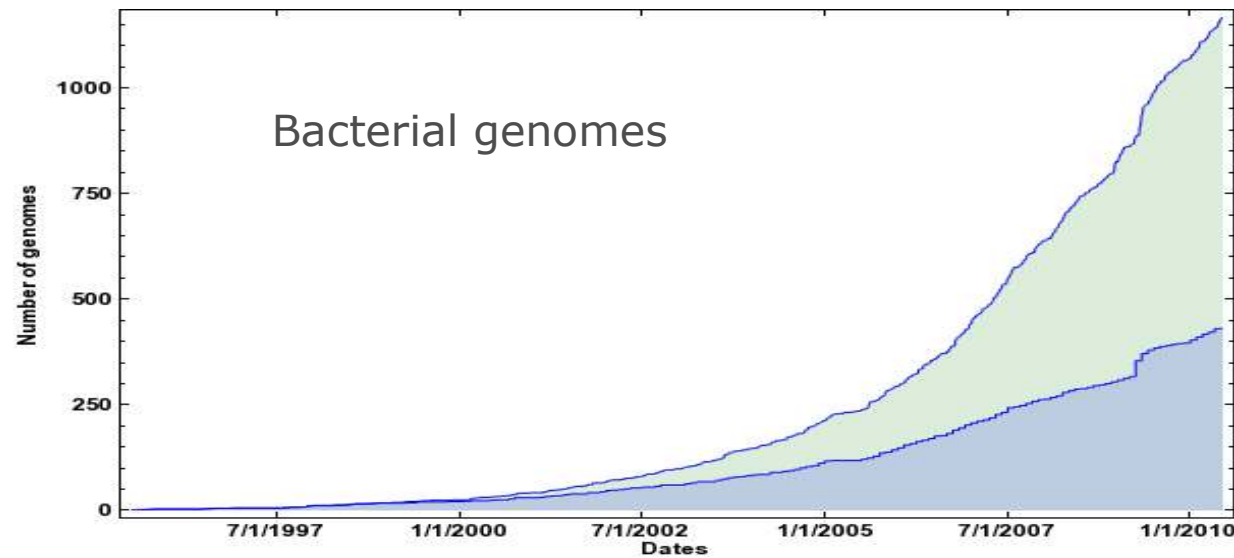
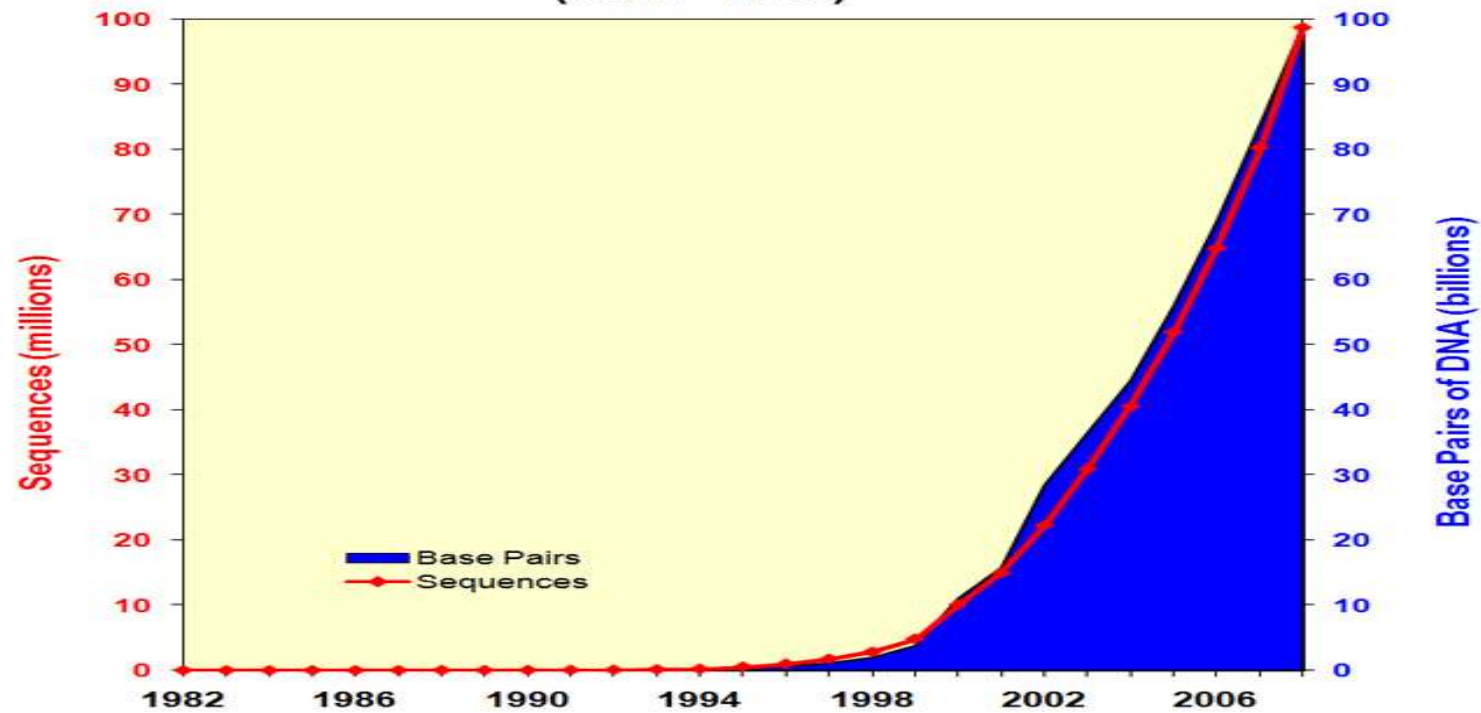
Life Sciences

- 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)

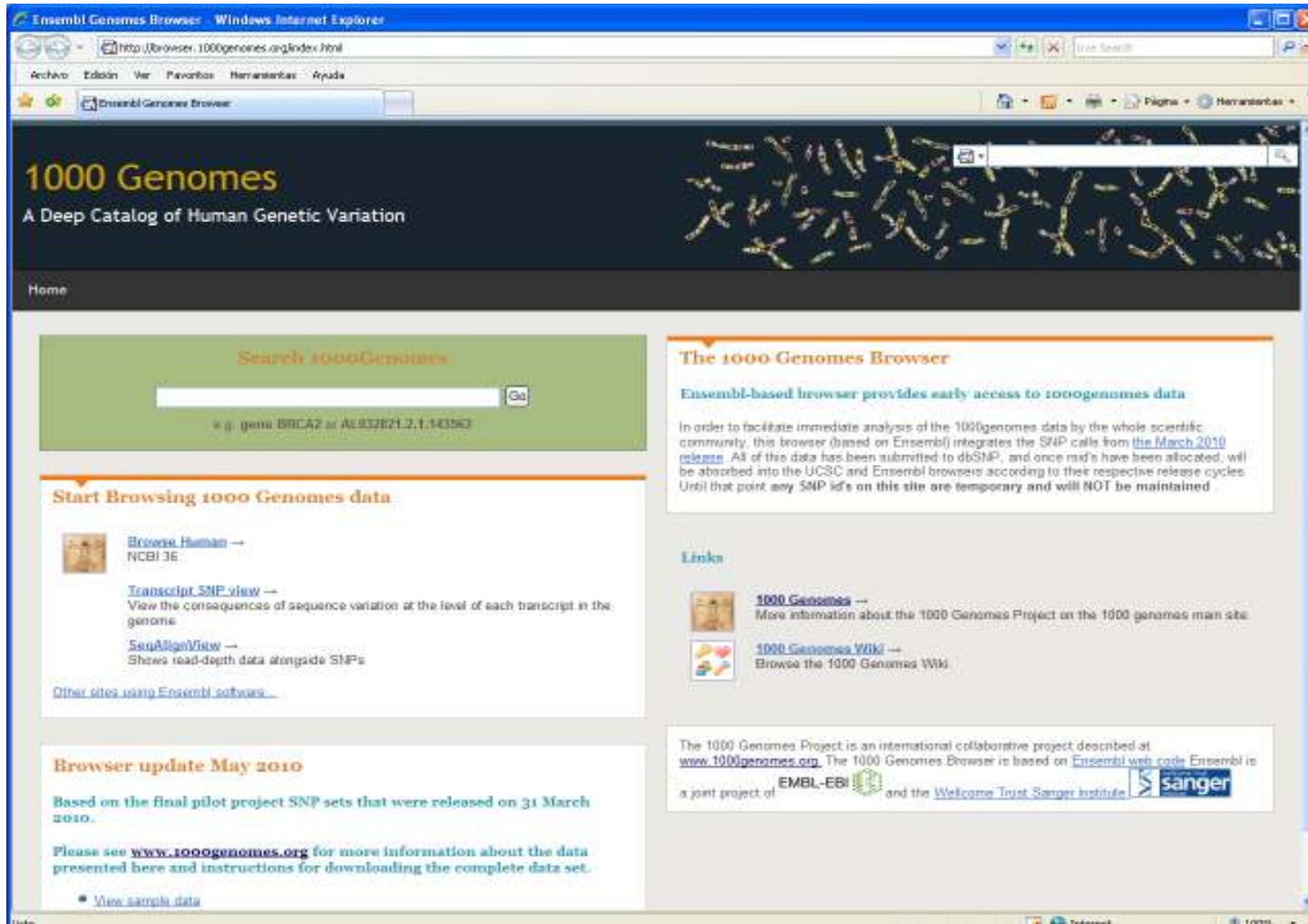
The computer as data manager

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

Growth of GenBank (1982 - 2008)

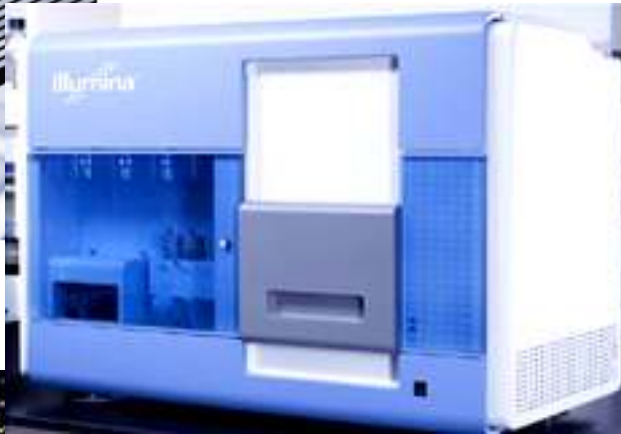
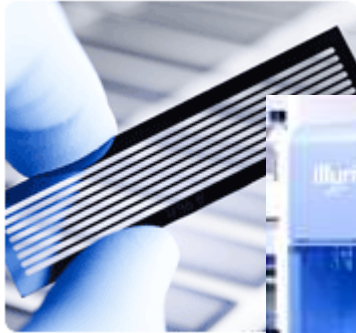


Massive (human) genomic projects

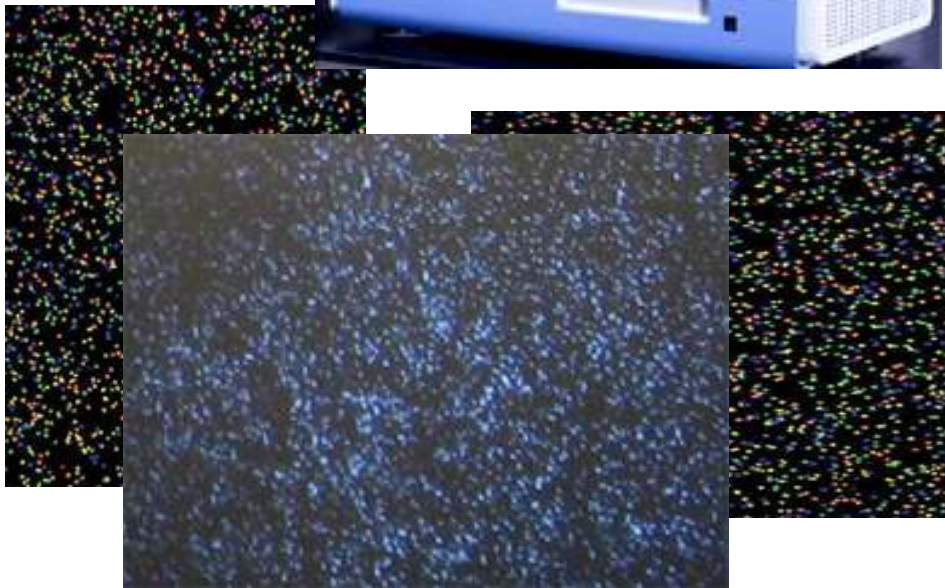


The screenshot shows the 1000 Genomes Browser website. The browser window title is "Ensembl Genomes Browser - Windows Internet Explorer". The address bar shows "http://browser.1000genomes.org/index.html". The website header features the "1000 Genomes" logo and the tagline "A Deep Catalog of Human Genetic Variation". Below the header, there is a search bar with the text "Search 1000Genomes" and a "Go" button. A sample search query is provided: "e.g. gene BRCA2 at AL032821.2,1:143563". The main content area is divided into several sections. On the left, under "Start Browsing 1000 Genomes data", there are links for "Browse Human" (with a NCBI 36 icon), "Transcript/SNP view" (described as "View the consequences of sequence variation at the level of each transcript in the genome"), and "SeqAlignView" (described as "Shows read-depth data alongside SNPs"). There is also a link for "Other sites using Ensembl software". On the right, under "The 1000 Genomes Browser", there is a paragraph explaining that the browser provides early access to 1000genomes data and that it integrates SNP calls from the March 2010 release. Below this, there is a "Links" section with icons and links for "1000 Genomes" (More information about the 1000 Genomes Project on the 1000 genomes main site) and "1000 Genomes Wiki" (Browse the 1000 Genomes Wiki). At the bottom, there is a "Browser update May 2010" section stating that the browser is based on the final pilot project SNP sets released on 31 March 2010. It also provides a link to "www.1000genomes.org" for more information and a link to "View sample data". The footer of the website mentions that the 1000 Genomes Project is an international collaborative project described at "www.1000genomes.org" and that the browser is based on Ensembl web code. It also states that Ensembl is a joint project of EMBL-EBI and the Wellcome Trust Sanger Institute, with logos for each organization.

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

The image shows a screenshot of the International Cancer Genome Consortium (ICGC) website. On the left, a list of countries participating in the project is visible, including Spain. On the right, a detailed view of the Chronic Lymphocytic Leukemia (CLL) project is shown. This view includes a list of funding organizations (Spain: Institute of Health Carlos III, Spanish Ministry of Science and Innovation) and a list of research organizations. The research organizations are listed in a grid, with the BSC (Barcelona Supercomputing Center) highlighted by a red circle. Other research organizations include the Catalan Institute of Oncology, Center for Cancer Research, University Hospital, Spain: National Genome Analysis Centre, CRG (Center for Genomic Regulation), CLÍNIC BARCELONA (Hospital Universitari), Hospital Clinic of Barcelona, UNIVERSITAT POMPEU FABRA (Pompeu Fabra University), CNIO (Centro Nacional de Investigaciones Oncológicas), Spanish National Cancer Research Centre, Deusto (Universidad de Deusto), University of Oviedo, USC (University of Santiago de Compostela), and the University of Barcelona.

Data Link



Structures, chemical, literature,...



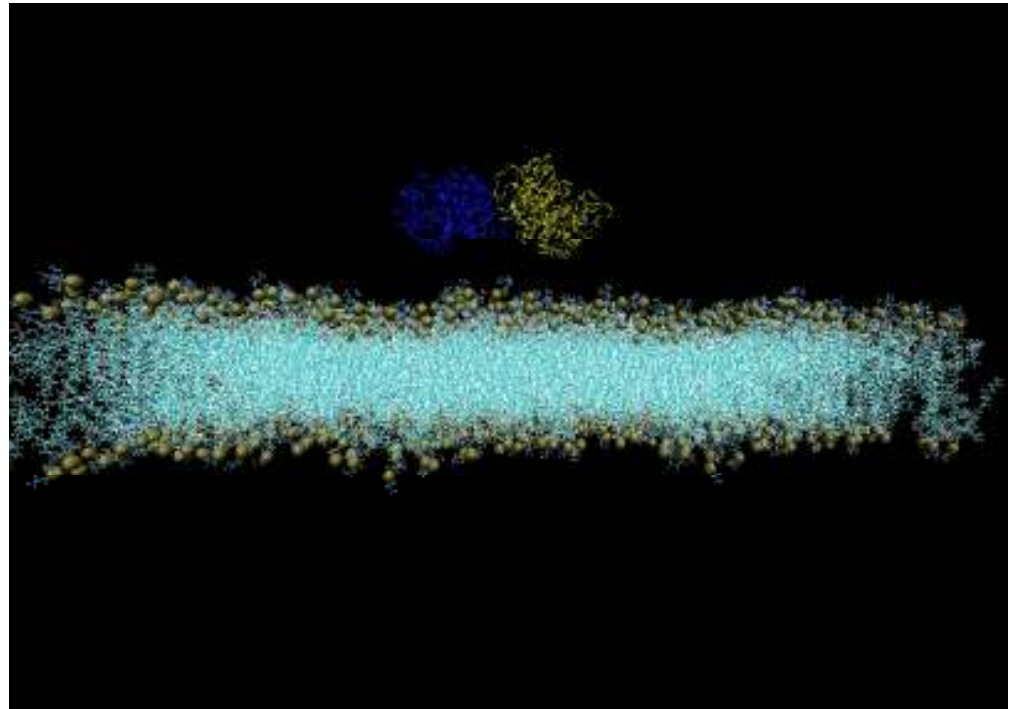
Chemical banks in the order
of 10^7 compounds

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

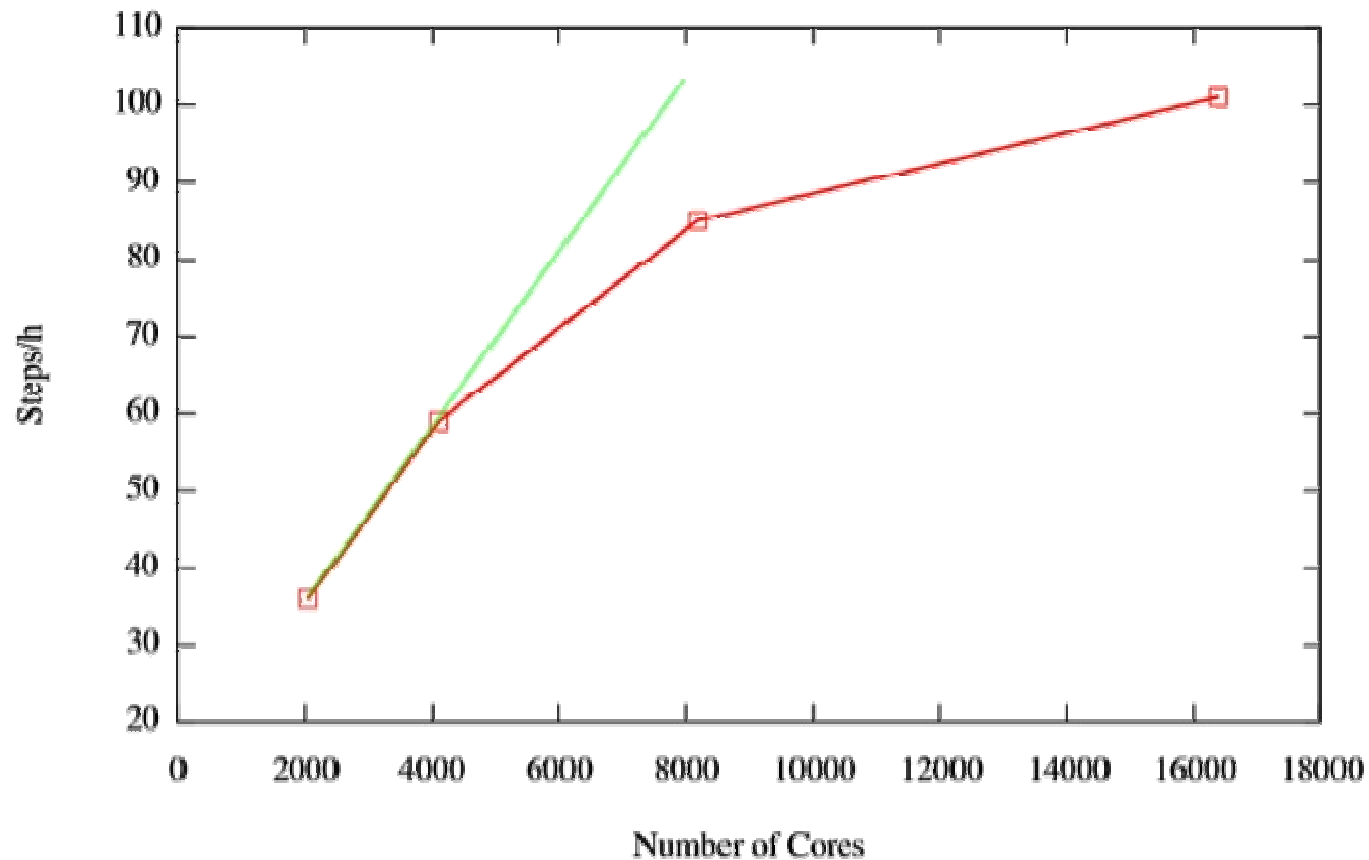
Current limitations in MD

- 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

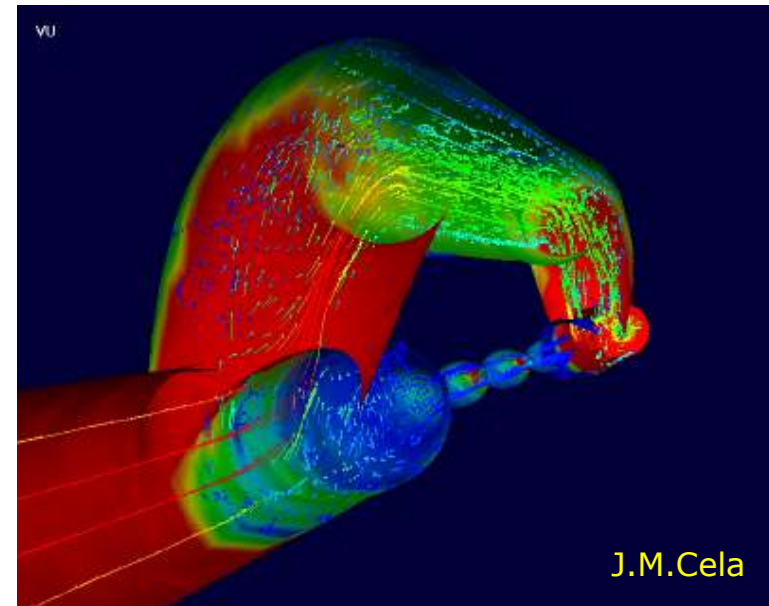
1746 atoms



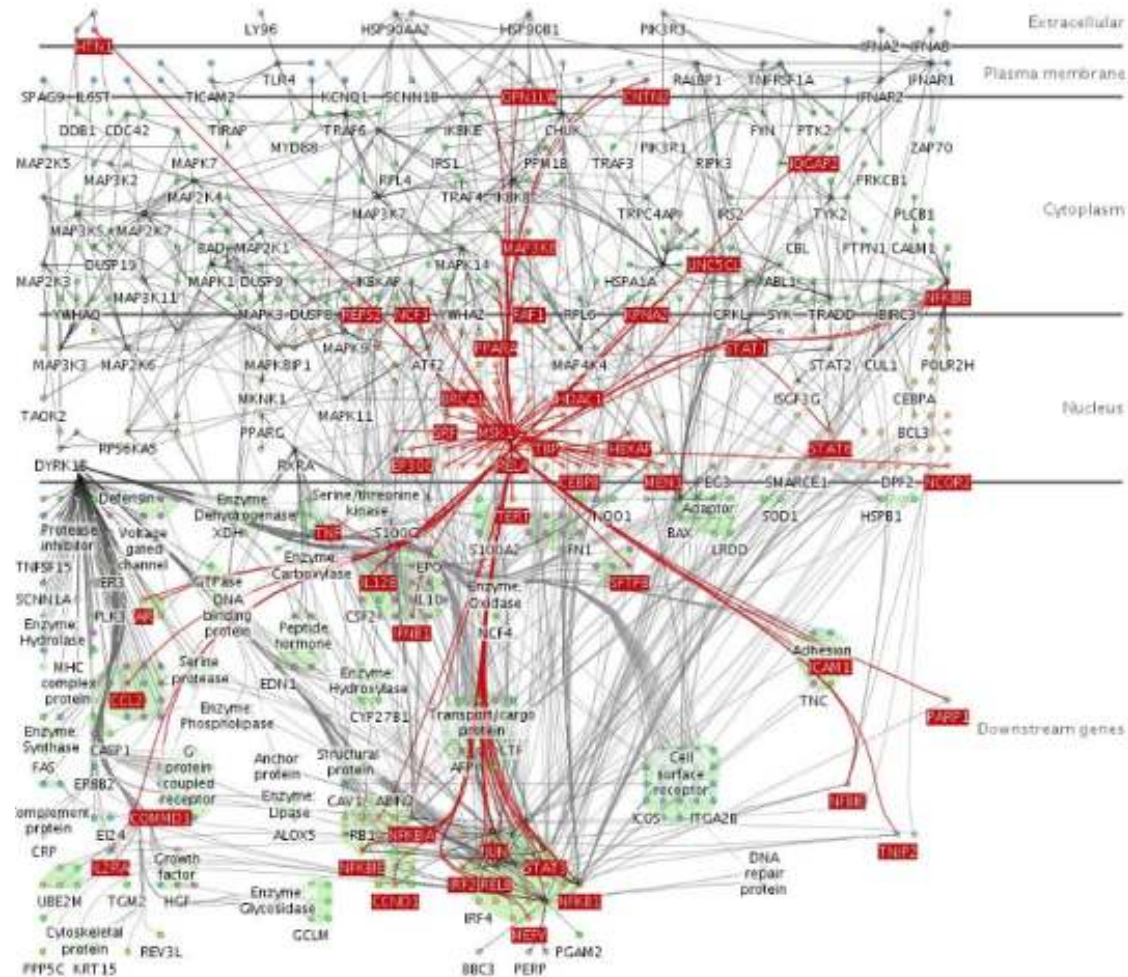
- www.cpmd.org implemented Jugene Curioni et al. (IBM Zurich)

Simulation scenario in Life Sciences

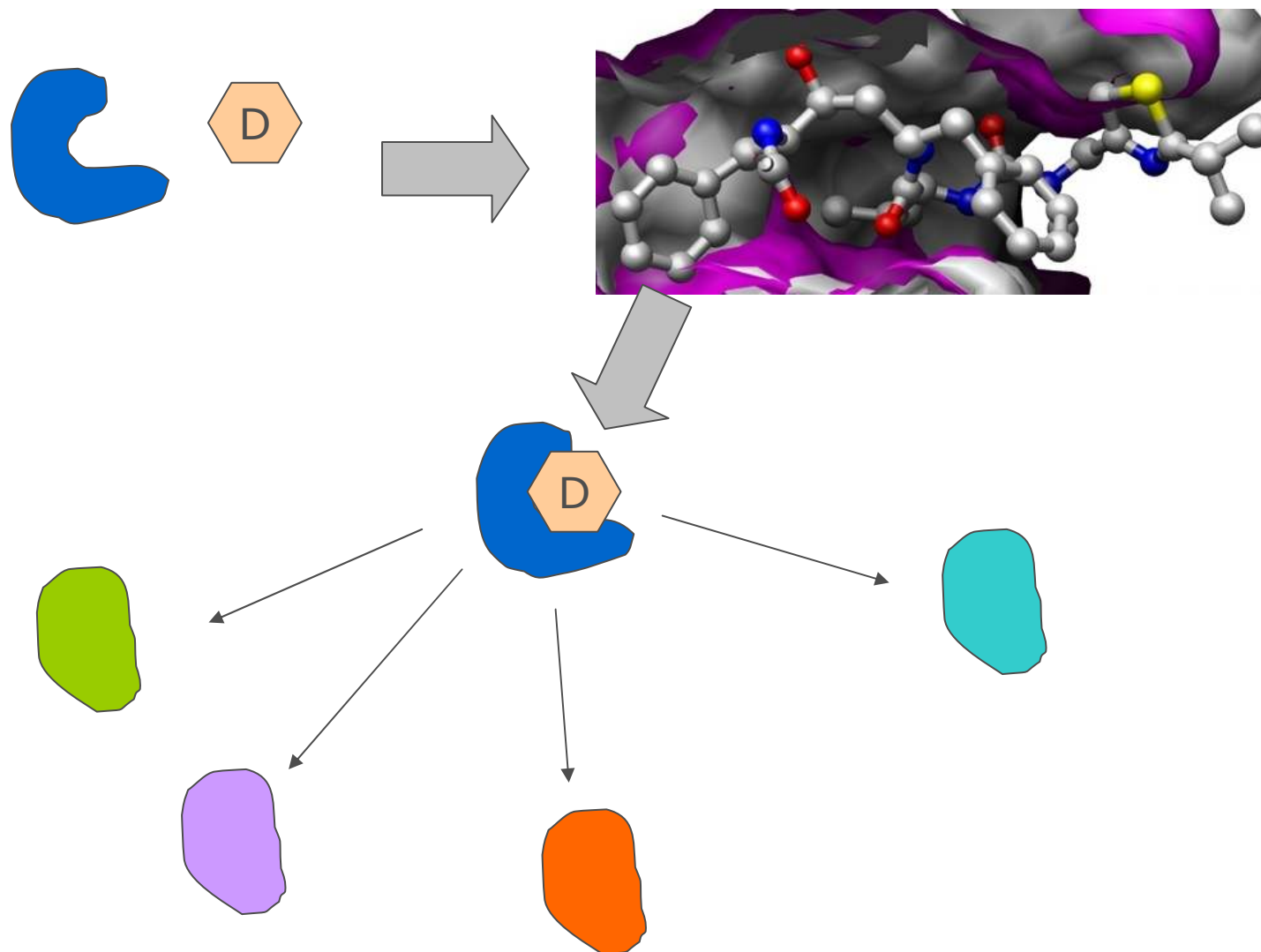
- Molecular simulations
 - Structural prediction
 - Docking
 - Atomistic simulation
 - Cell-scale mesoscopic simulations
- Gene inter-relations
- Cell simulation
- Organ simulation
- Ecosystem simulation



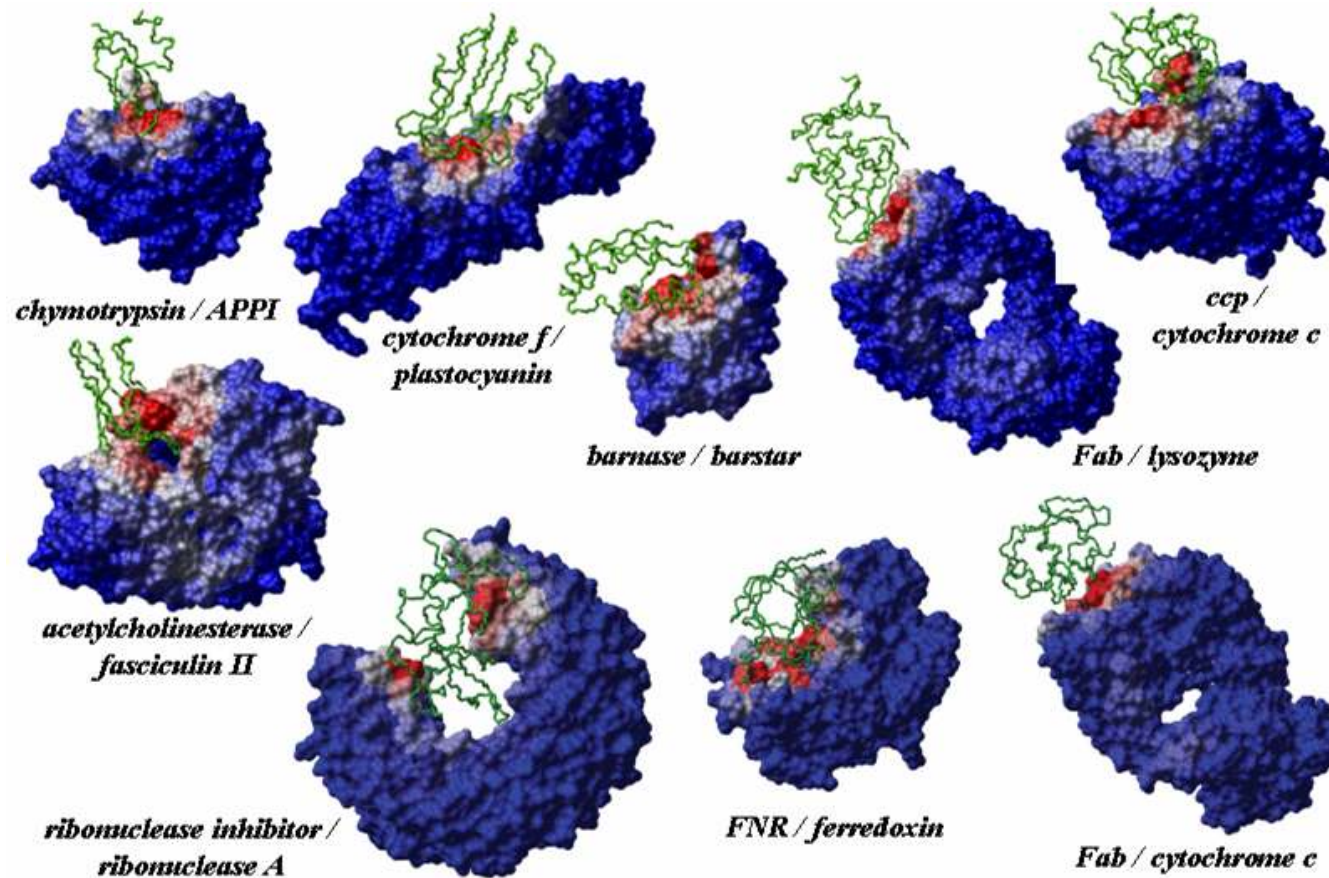
Breast cancer interactome



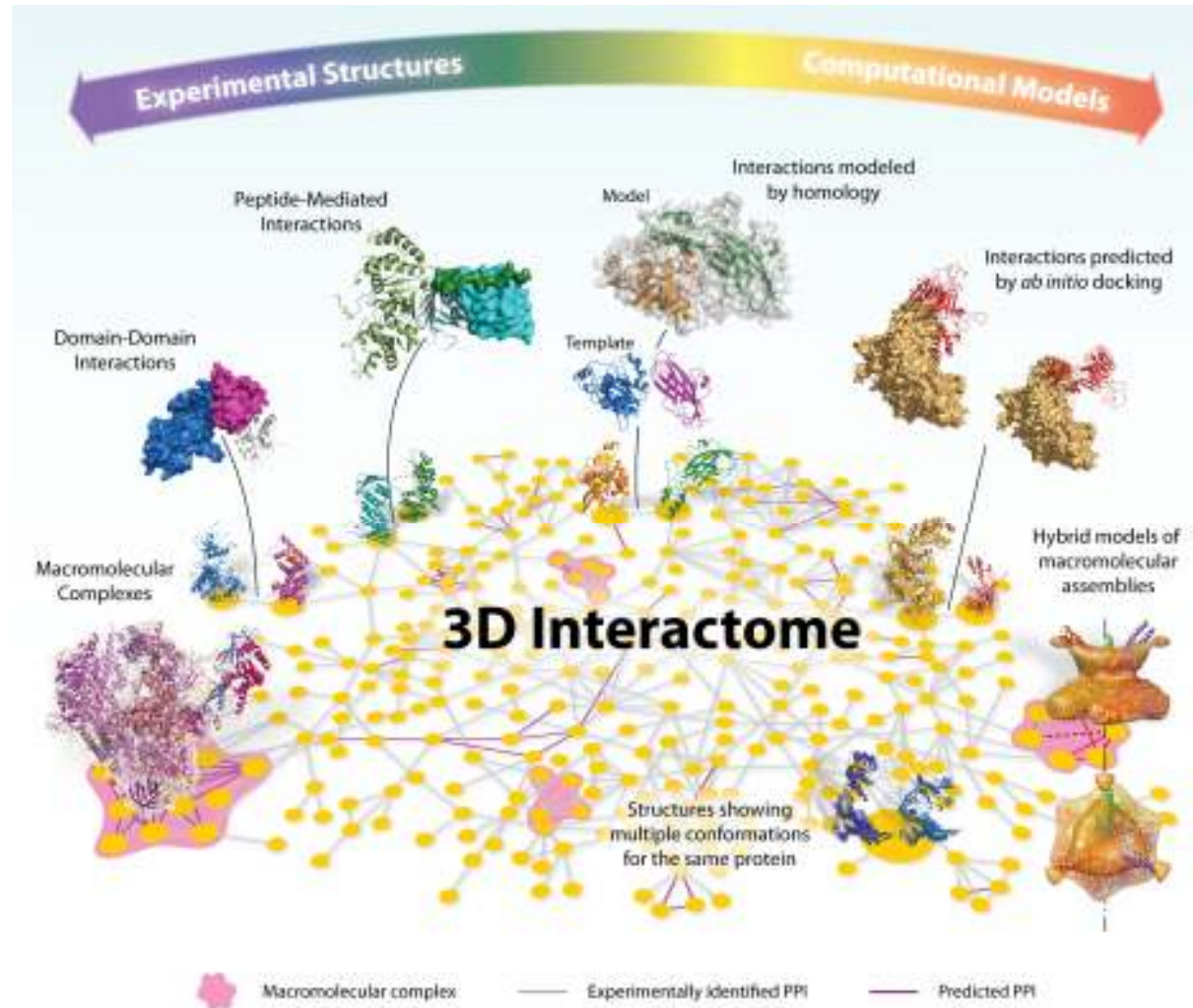
Cancer Interactome



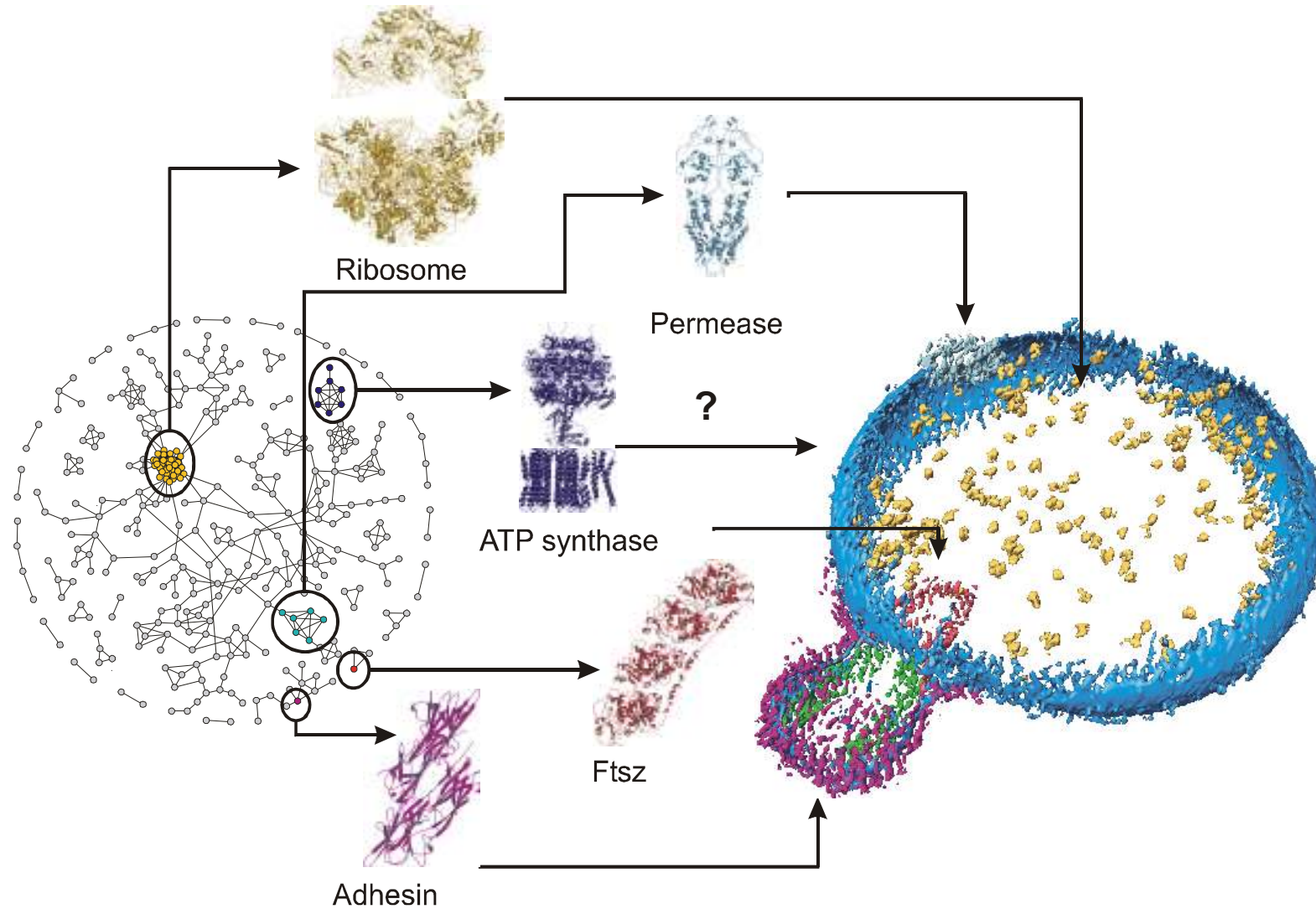
Protein-Protein Docking



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.

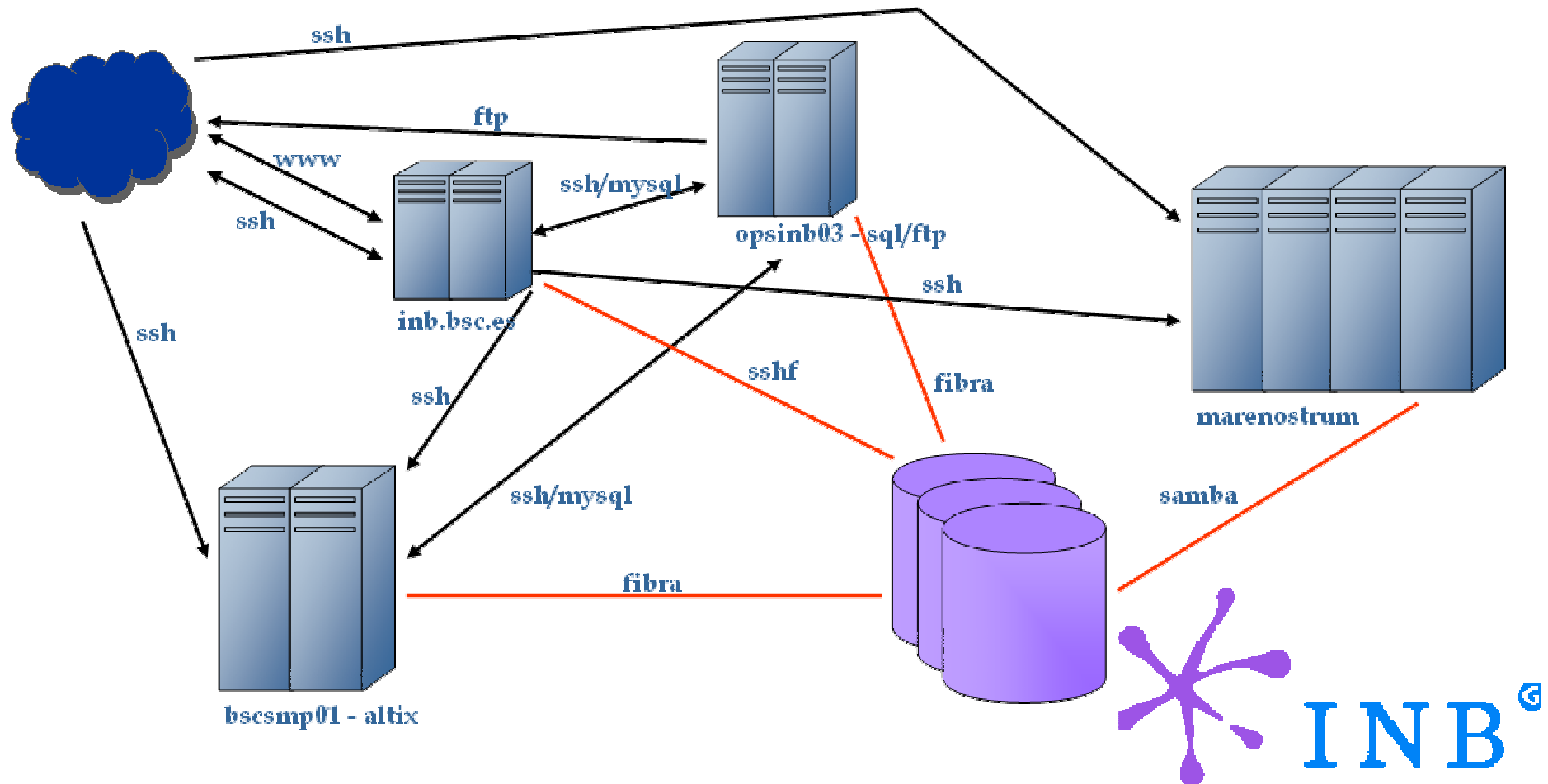
Yale
UNIVERSITY



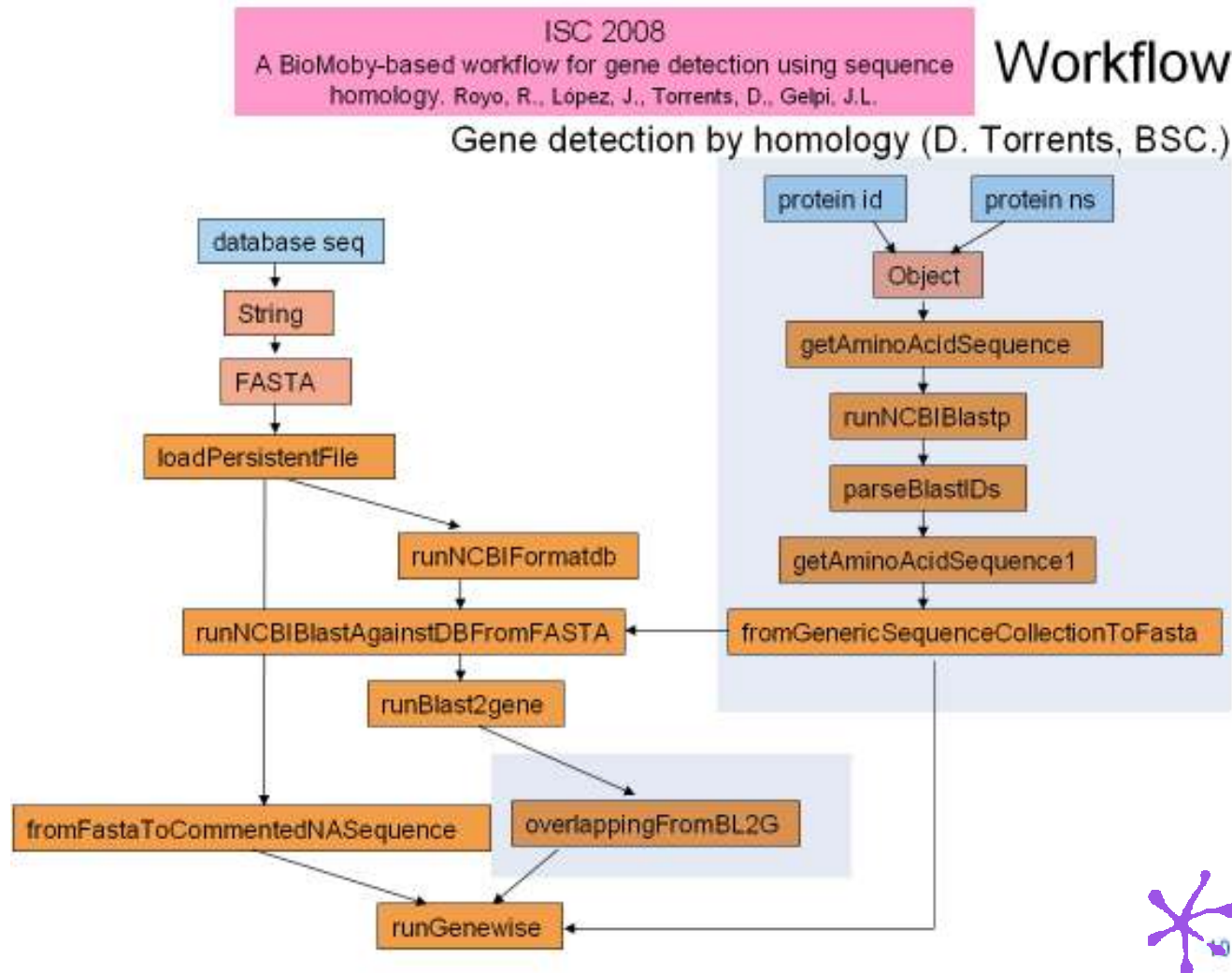
Biocomunity HPC Support



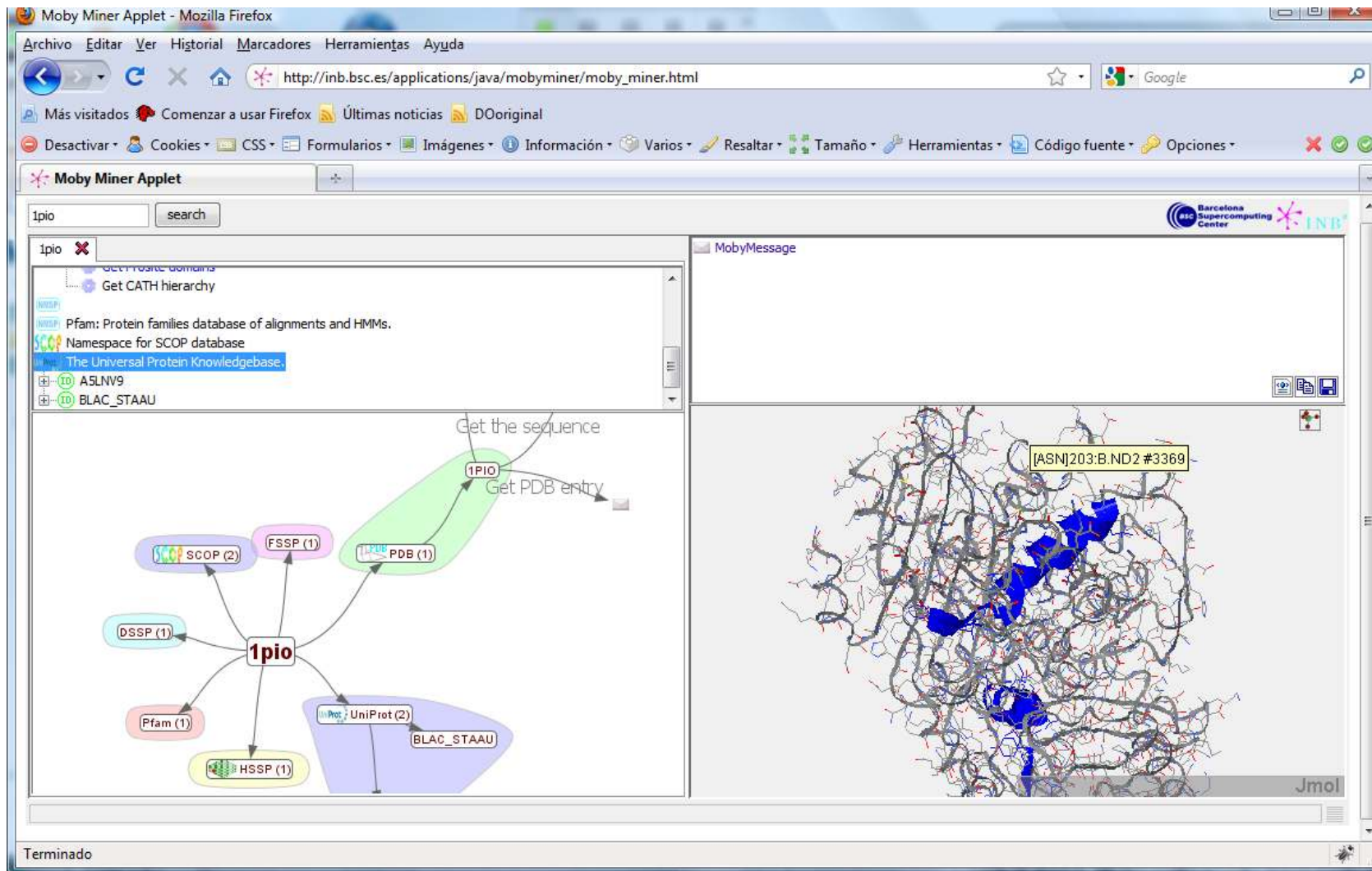
BioSupercomputing open to “BioCommunity”



Pre-defined Workflows



User-friendly for non-expert



Moby Miner Applet - Mozilla Firefox

Archivo Editar Ver Historial Marcadores Herramientas Ayuda

http://inb.bsc.es/applications/java/mobyminer/moby_miner.html

Más visitados Comenzar a usar Firefox Últimas noticias DOoriginal

Desactivar Cookies CSS Formularios Imágenes Información Varios Resaltar Tamaño Herramientas Código fuente Opciones

Moby Miner Applet

1pio search

1pio

- Get Protein domains
- Get CATH hierarchy
- Pfam: Protein families database of alignments and HMMs.
- Namespace for SCOP database
- The Universal Protein Knowledgebase.
- ASLNV9
- BLAC_STAAU

Get the sequence

1PIO

Get PDB entry

SCOP (2)

FSSP (1)

PDB (1)

DSSP (1)

Pfam (1)

HSSP (1)

UniProt (2)

BLAC_STAAU

1pio

[ASN]203:B.ND2 #3369

Jmol

Terminado



Molecular Modeling & Bioinformatics



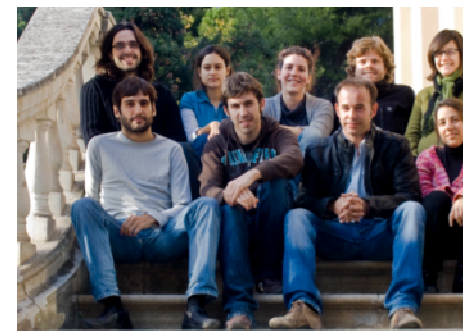
Protein Interactions and Docking



Experimental Bioinformatics Laboratory



Electronic and Atomic Protein Modeling



Computational Genomics



INB-BSC Algorithmic Unit

Contact:

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

