FP7 Support Action - European Exascale Software Initiative

DG Information Society and the unit e-Infrastructures

EESI WG: Life Sciences and Health

Modesto Orozco, Chair
Janet Thornton, Vice Chair
Ramon Goñi, Speaker

IESP April 6-7, San Francisco

EESI WG: List of Experts | Acknowledgements

Erik Lindahl — Stockholm University

Helmut Grubmuller — Max Planck Institute
Paolo Carloni — German Research School
Wofgang Wenzel — Karlsruhe Institute of Technology
Reinhard Schneider — EMBL

Janet Thorton/Andrew Lyall — EMBL-EBI
Charles Laughton — University of Nottingham

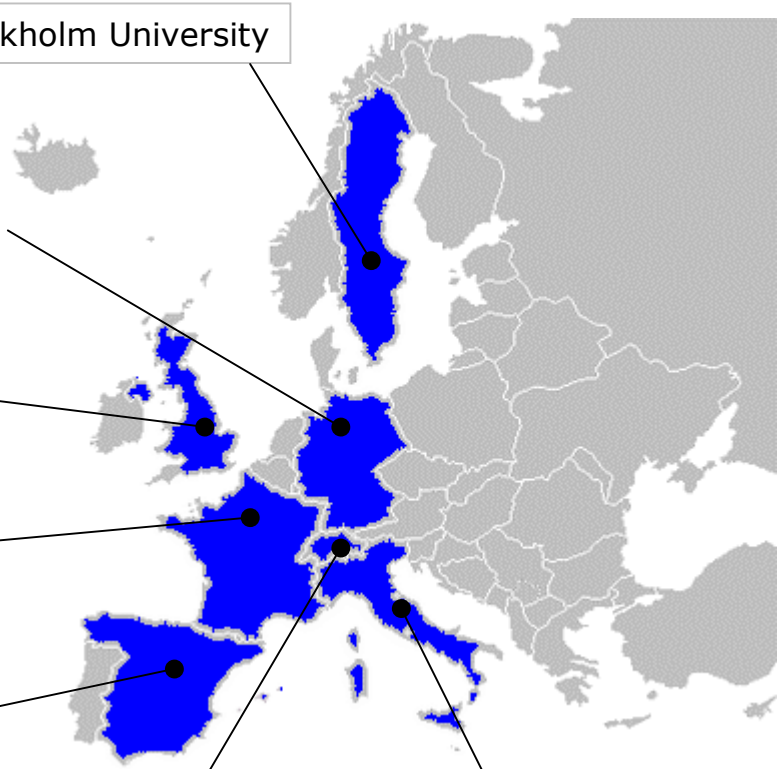
Richard Lavery — University of Nottingham
Nicolas Baurin — Sanofi-Aventis

Modesto Orozco — BSC
Alfonso Valencia — CNIO

Henry Markram/Felix Schuermann — EPFL
Manuel Peitsch — Swiss Institute of Bioinformatics

Anna Tramontano — University of Roma

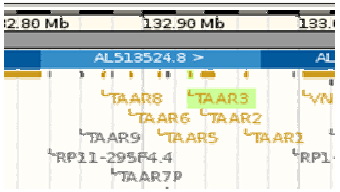
Julian.Tirado-Rives (US) — University of Yale
Adrian E. Roitberg (US) — University of Florida



Sub-pannels

- Molecular Simulation
- Genomics
- Systems Biology
- Medical Simulation

First experts meeting



Exascale Challenges in Computational Biology. 13-15 December, 2010 - Barcelona

As a result of the enormous potential of Molecular Biology to produce returns for human health, this discipline now figures as a main strategic research area. Molecular Biology has experienced a huge quantitative and qualitative jump thanks to high-throughput systems that have emerged in recent years, among these supercomputers. The challenge in the 90s was to build computers that reached the Teraflop scale. In 2007 the first computer to register 1,000 times this limit (1 Petaflop) was developed. This achievement has allowed us to exhaustively explore complex combinations of genetic variants, large scale Molecular Dynamics and high-throughput virtual screening.

In the coming years we expect to reach a computational power equivalent to 1,000-fold that of the current most powerful supercomputer (Exascale). There are three main issues in Excascale computing: i) to design computer architectures that can achieve this scale; ii) to design non-trivial software that can be scalable with a high number of processors; and iii) to choose the appropriate problems that can be tackled with this technology. With these challenges in mind, the conference "**Exascale Challenges in Computational Biology**" aims to bring together in Barcelona the top scientists working in bio-supercomputing in order to share an open discussion on the key questions of this field.

Agenda:

Systems biology (Monday 13th Morning)
Genomics (Monday 13th Afternoon
Molecular simulations (Tuesday 14th)
Simulation in medicine (Wednesday 15th Morning)

Confirmed speakers:

Alfonso Valencia (CNIO. Spain)
Henry Markram(EPF Lausanne. Switzerland)
Adrian E. Roitberg (U Florida, USA)
Ivo Gut (CNAG, Spain)
Oscar de Bustos Matin (BULL)
Manuel Peitsch (Swiss Institute of Bioinformatics. Switzerland)
Erik Lindahl (Stockholm University. Sweden)
Richard Lavery (IBCP. Lyon France)
Charles Laughton (Univ. of Nottingham. UK)
Nicolas Baurin (Sanofi-Aventis. France)
Paolo Carloni (German Research School Julich. Germany)
Julio Saez-Rodriguez (EMBL-EBI Hinxton. UK)
Johan Aqvist (University of Upsala. Sweden)
Manuel Doblare (Univ. Zaragoza. Spain)
Helmut GrubmAlfeuller (Max Planck Institute. Gottingen, Germany)
Reinhard Schneider (EMBL Heidelberg. Germany)
Julian Tirado-Rives (Univ. of Yale. USA)
Anna Tramontano (University of Rome "La Sapienza", Italy)
Andrew Lyall (EMBL-EBI Hinxton. UK)
Mariano Vazquez (BSC. Spain)
Wolfang Wenzel (Karlsruhe Institute of Technology. Germany)

The conference will be hosted by the **Casa de la Convalescència** in Barcelona (on the corner of Carrer St Quinti and Carrer Sant Antoni M. Claret). Talks will take place in the Aula Magna.

More information:
mmb.pcb.ub.es/exascale2010

Important Dates

20 October: Abstracts deadline and early registration

Organized by

IRB BARCELONA — INSTITUTE FOR RESEARCH IN BIOMEDICINE
BSC Barcelona Supercomputing Center — Centro Nacional de Supercomputación
INB

Sponsors

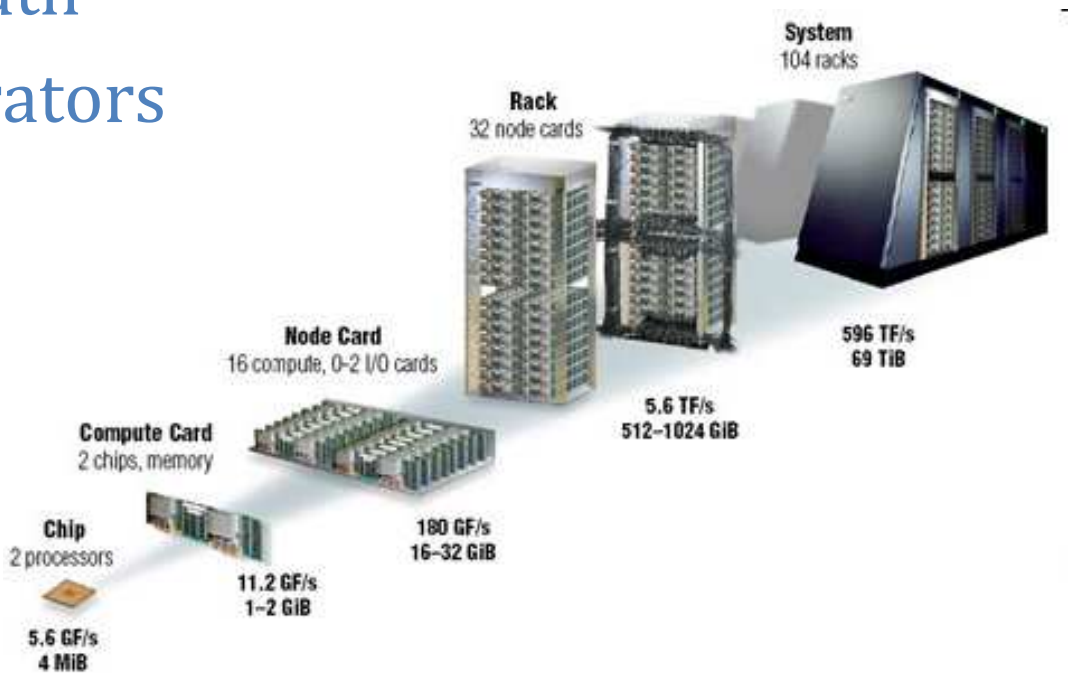
BULL
SyeC
scalalife
Genoma España
Generalitat de Catalunya

18 INTERNATIONAL EXASCALE SOFTWARE PROJECT
IESP April 6-7, San Francisco

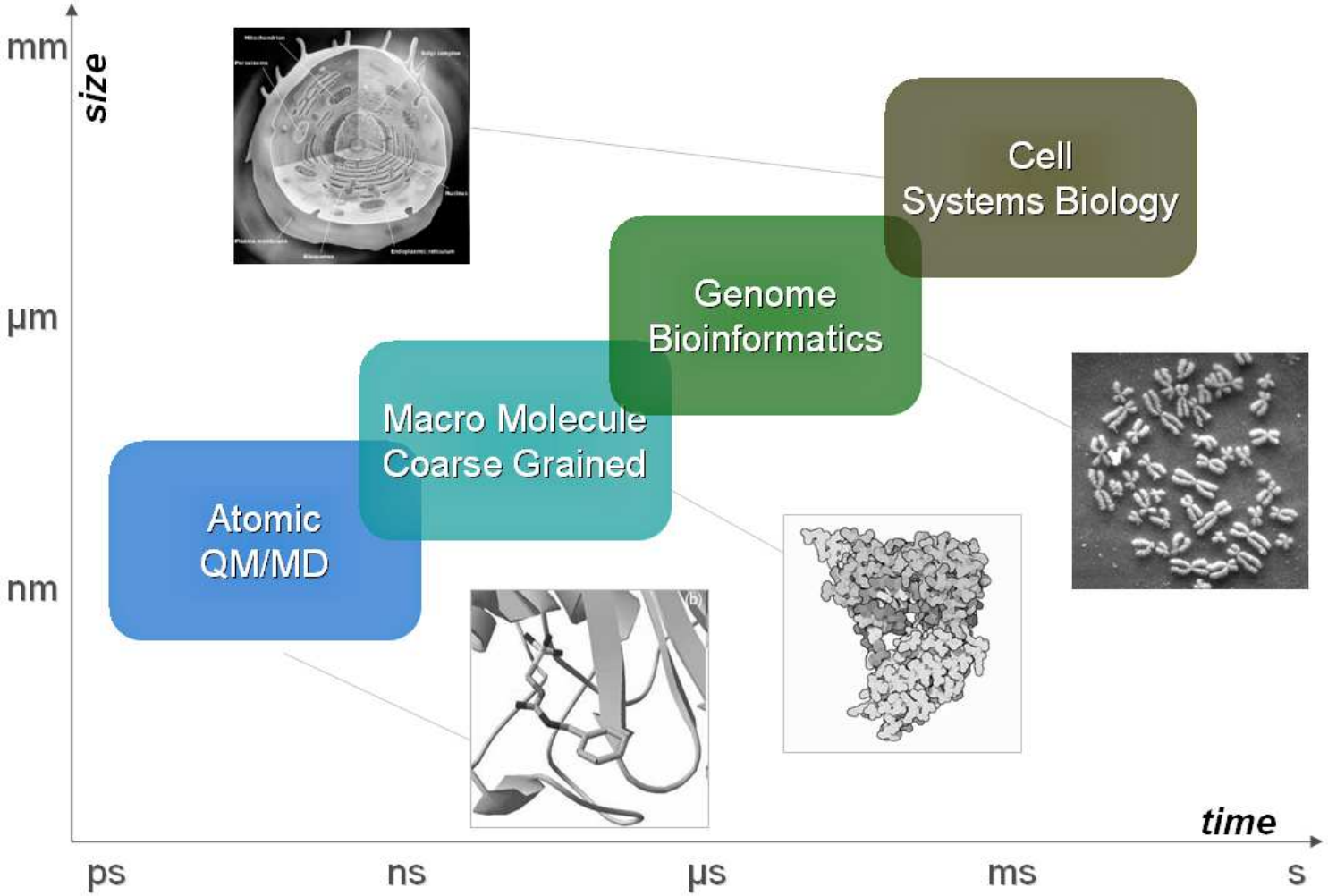
BSC Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Cross Issues (WG 4.1) Industry trends

- Data Storage
- Memory capacity
- Network bandwidth
- Hardware accelerators

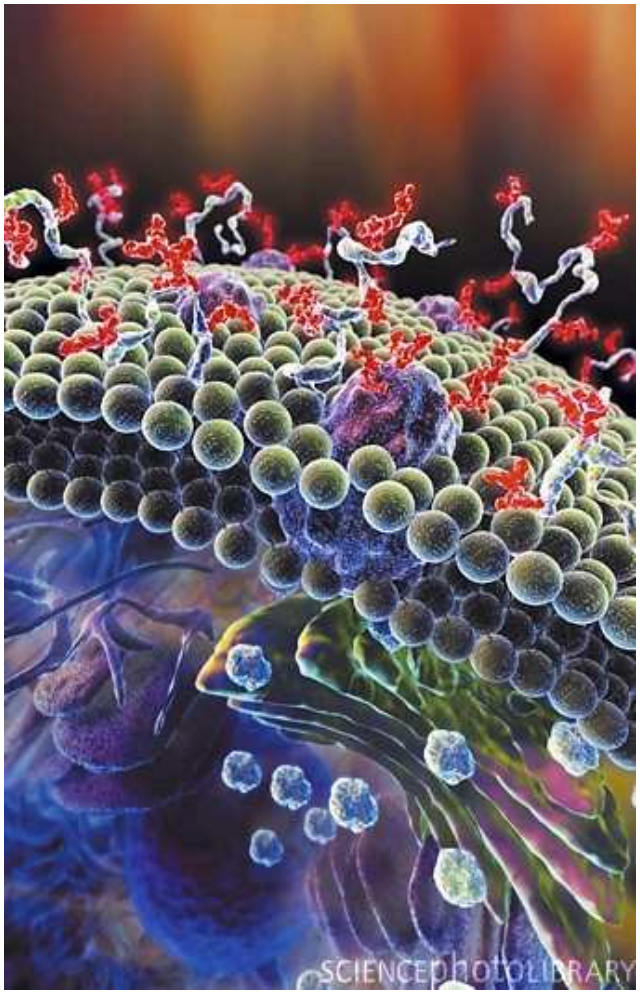


Cross-issues (WG 3.3 & 4.4)
Multi-scale coupled simulation



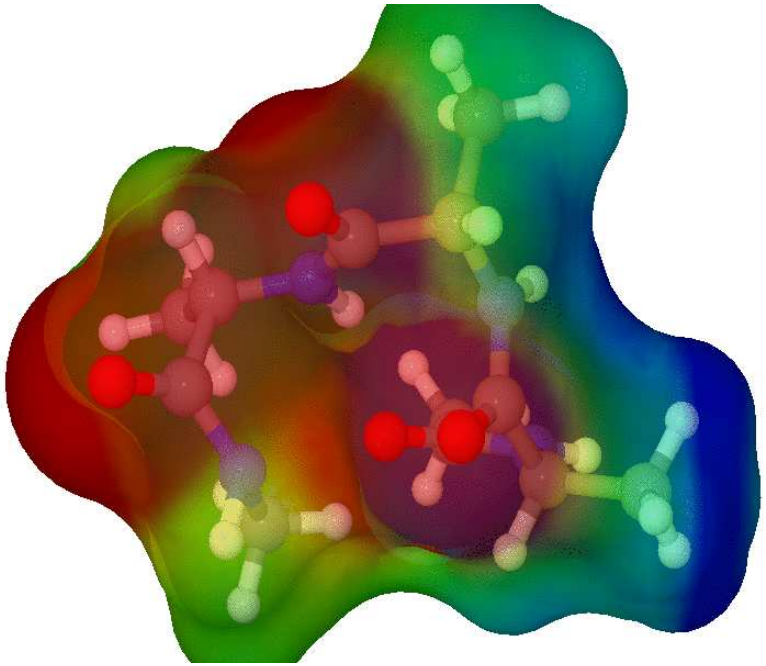
MD: Size

- Simulations of biological systems that are **1000 times bigger** than those possible today.
- An example would be realistic cell membrane models with drug permeation and binding.
- Because of the large data, distributed systems are not an option.



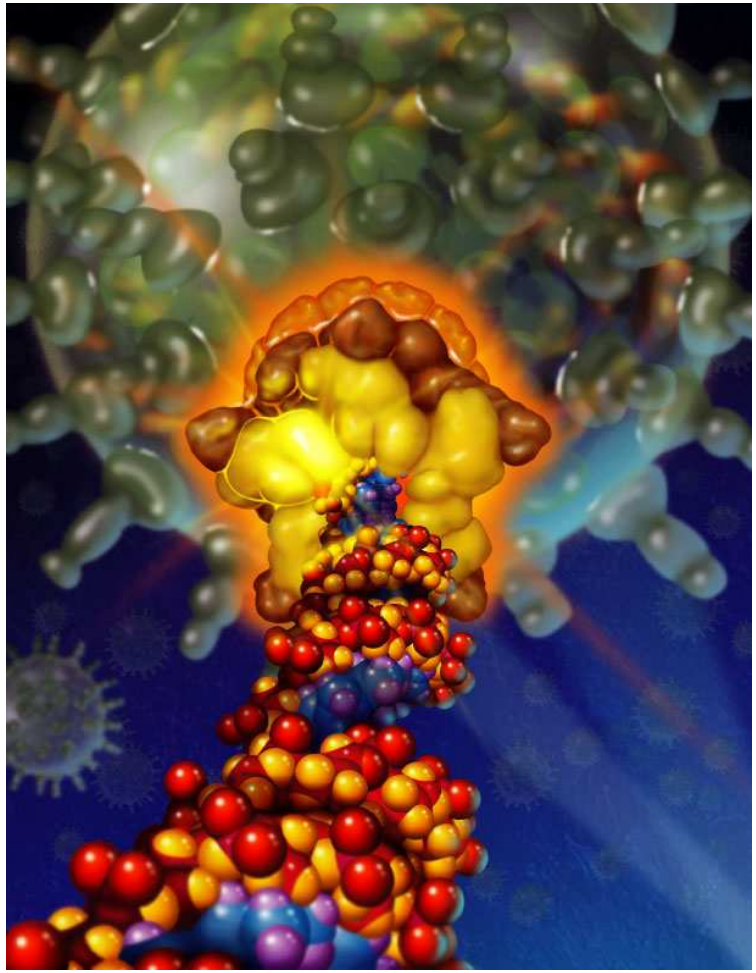
MD: Complexity

- Simulations that are **1000 times more computationally complex** than those possible today.
- An example would be simulations of biomolecules based purely on quantum mechanical principles.
- Exascale computing will be capable but will need high memory per core

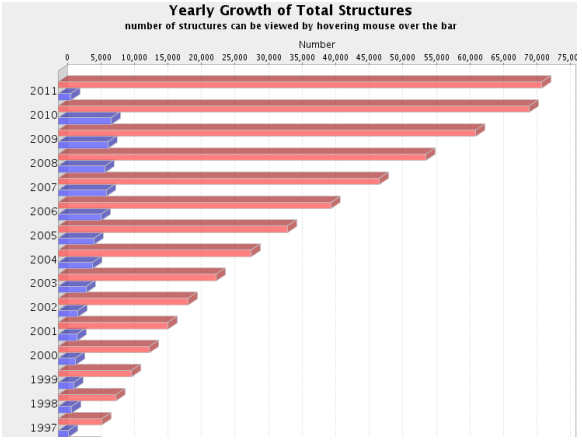


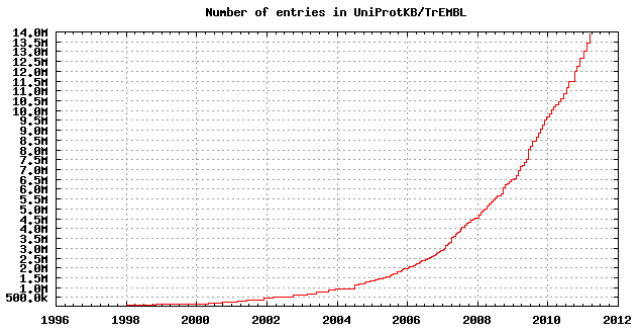
MD: Length

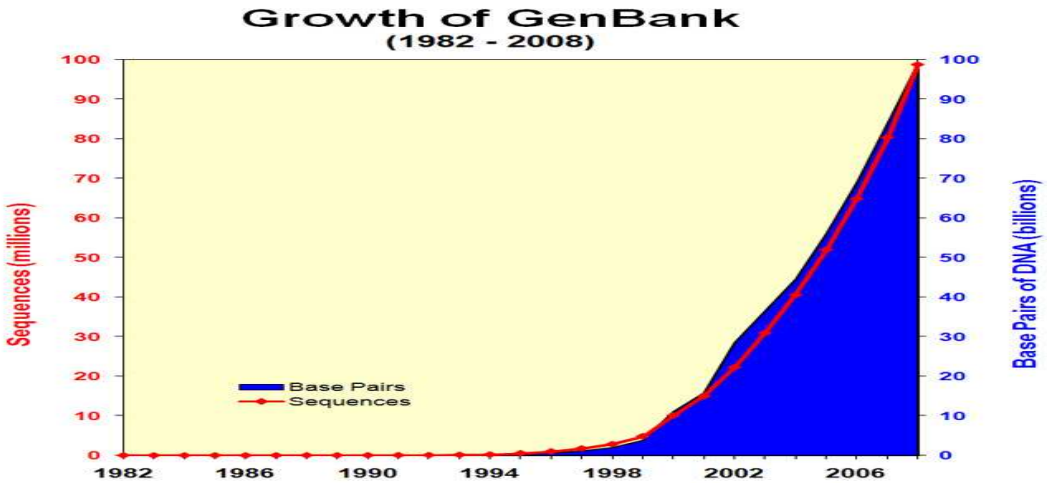
- Simulations that cover timescales **1000 times longer** than those possible today.
- Studies of the dynamics of protein folding and molecular motors that take up to seconds of real time.
- Time cannot be parallelized. Anton (a specific purpose architecture) is capable to run 100 longer (3~5 years ahead)



Genomics: Memory and Data issues



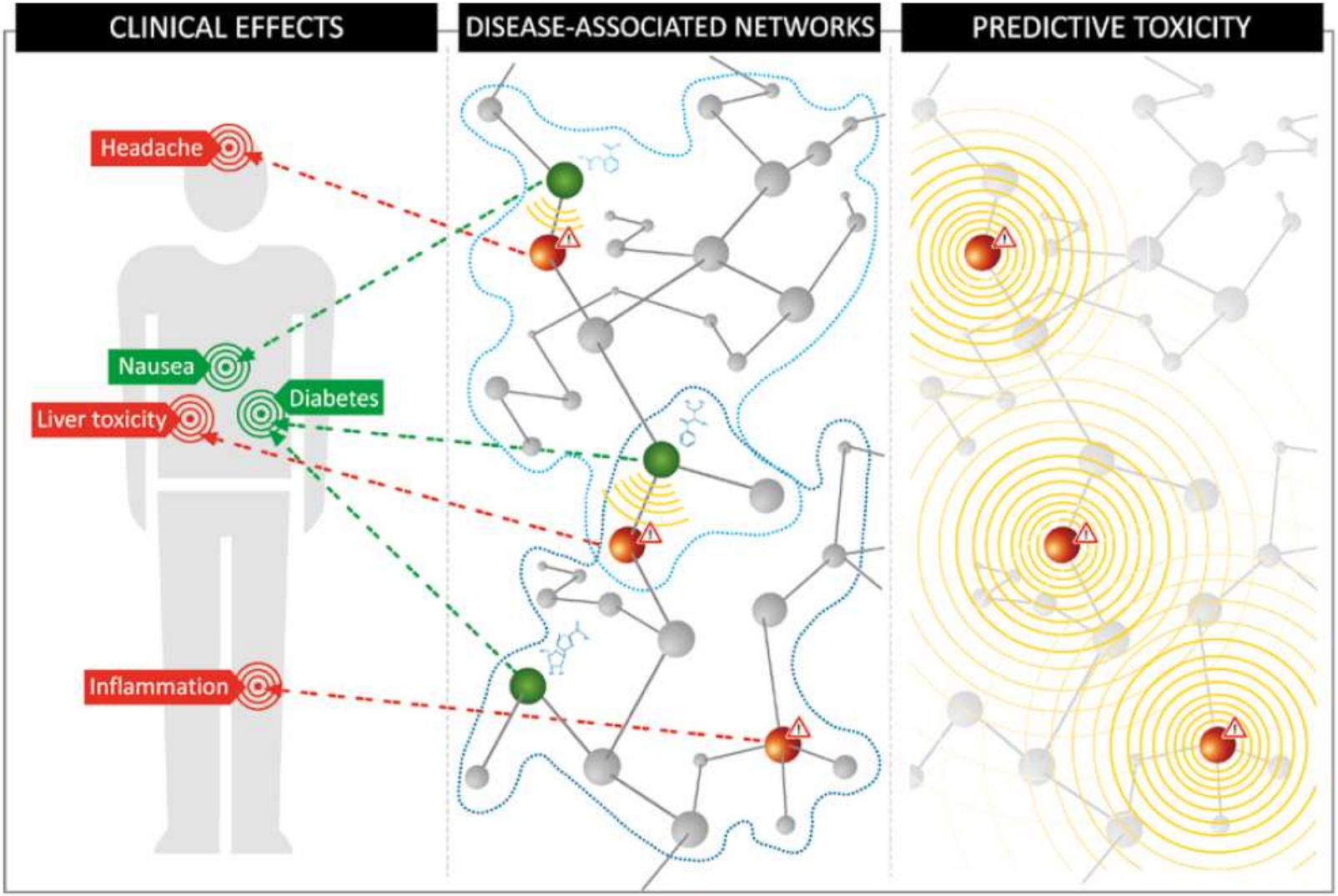




Data: Growing fast, Heterogeneous, Distributed, Partially confidential

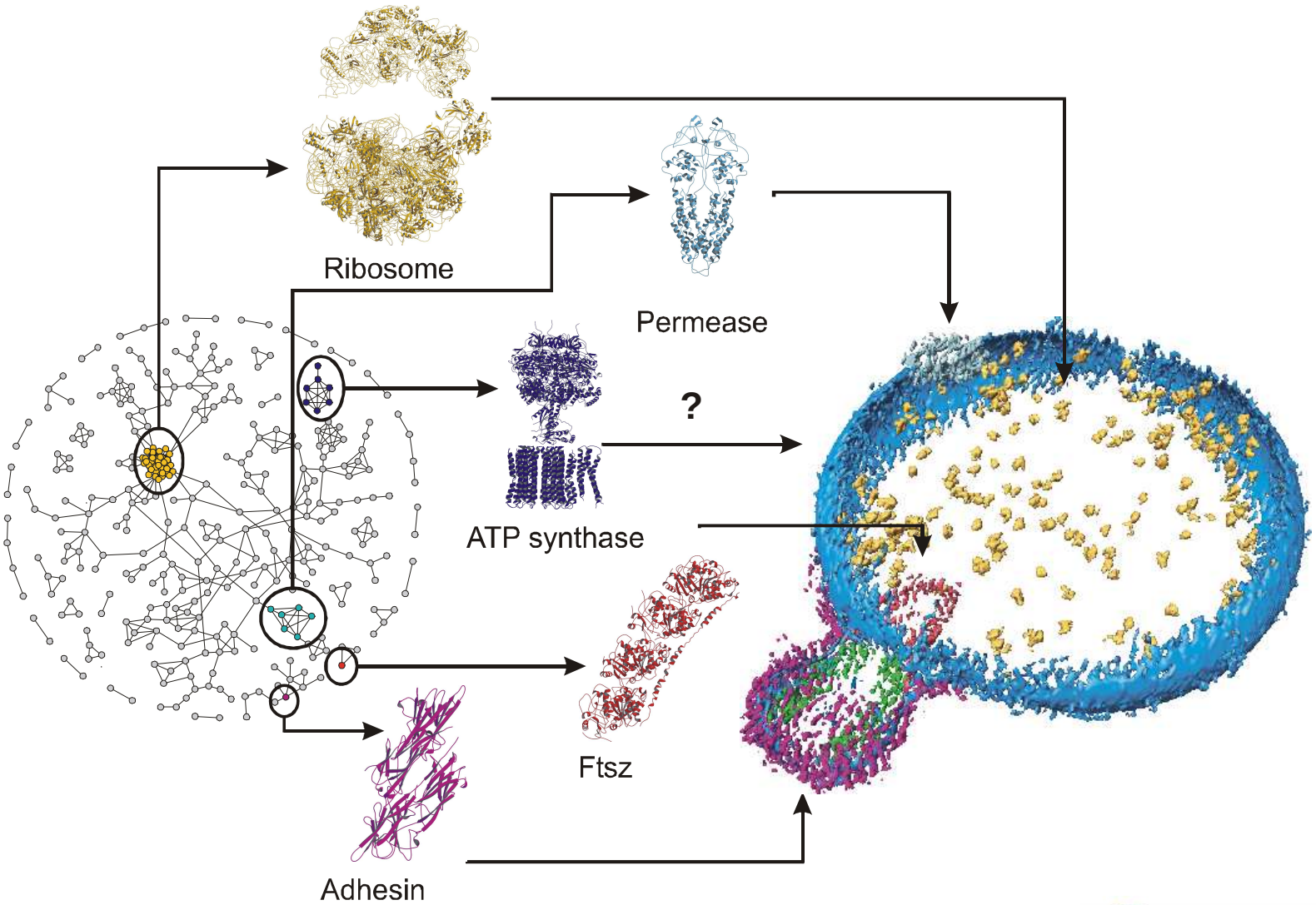
Genomes (euk) 10^2 (prok) 10^3
Sequences (bp) 10^{11}
Experiments (GEO-sampl) 10^5
Mutations 10^8
Literature 10^7
Proteins 10^7
Structures 10^4
Chemicals 10^7

Systems Biology : Complexity

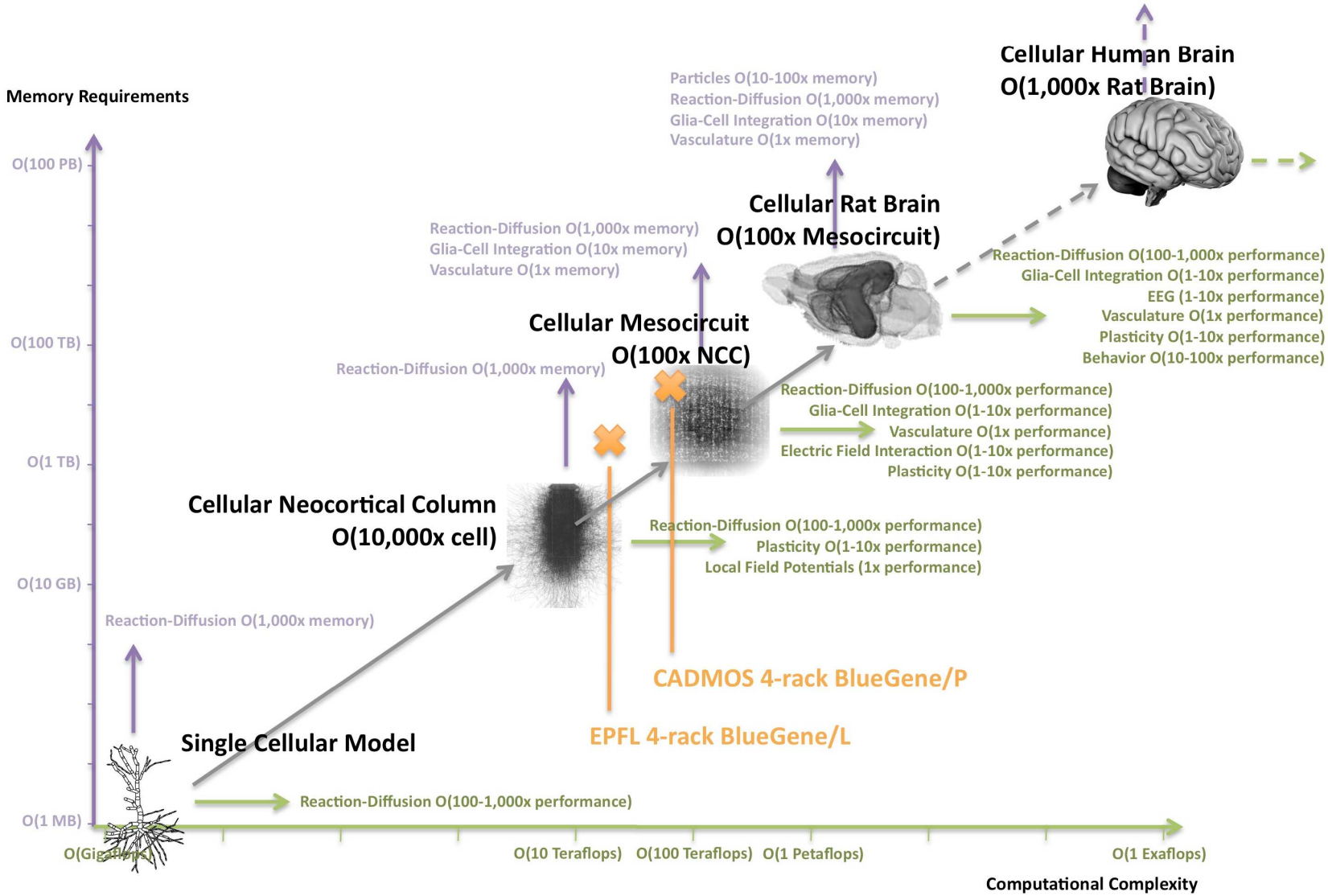


Patrick Aloy & Co, IRBB

Exascale models for Systems Biology



Medical simulation: Human Brain Project



Felix Schürmann, Herny Markram (Blue Brain Project, EPFL)

Software

- MD: Gromacs, AMBER, NAMD
- Quantum Chemistry: Dalton, GAMESS, Gaussian
- Sequence : BWA, BLAST/BLASTMPI, CLUSTALW, HMMER
- Docking: Autodock, Glide, Dock, Flexx, FTDock, LigandFit, PyDock

Summary

- Molecular Dynamics
  - Data expected to grow 10~100 times, competitiveness vs ASIC, Exascale capability for complex systems and biased-sampling techniques
- Genomics
  - Data management, high I/O requirement, constant application development, training
- Systems Biology
  - Multi-scale simulation, visualization tools, hardware flexibility for new software development, data handling
- Medical Simulation
  - Data Management, real-time monitoring, Exabyte memory, data transfer, Multi-scale