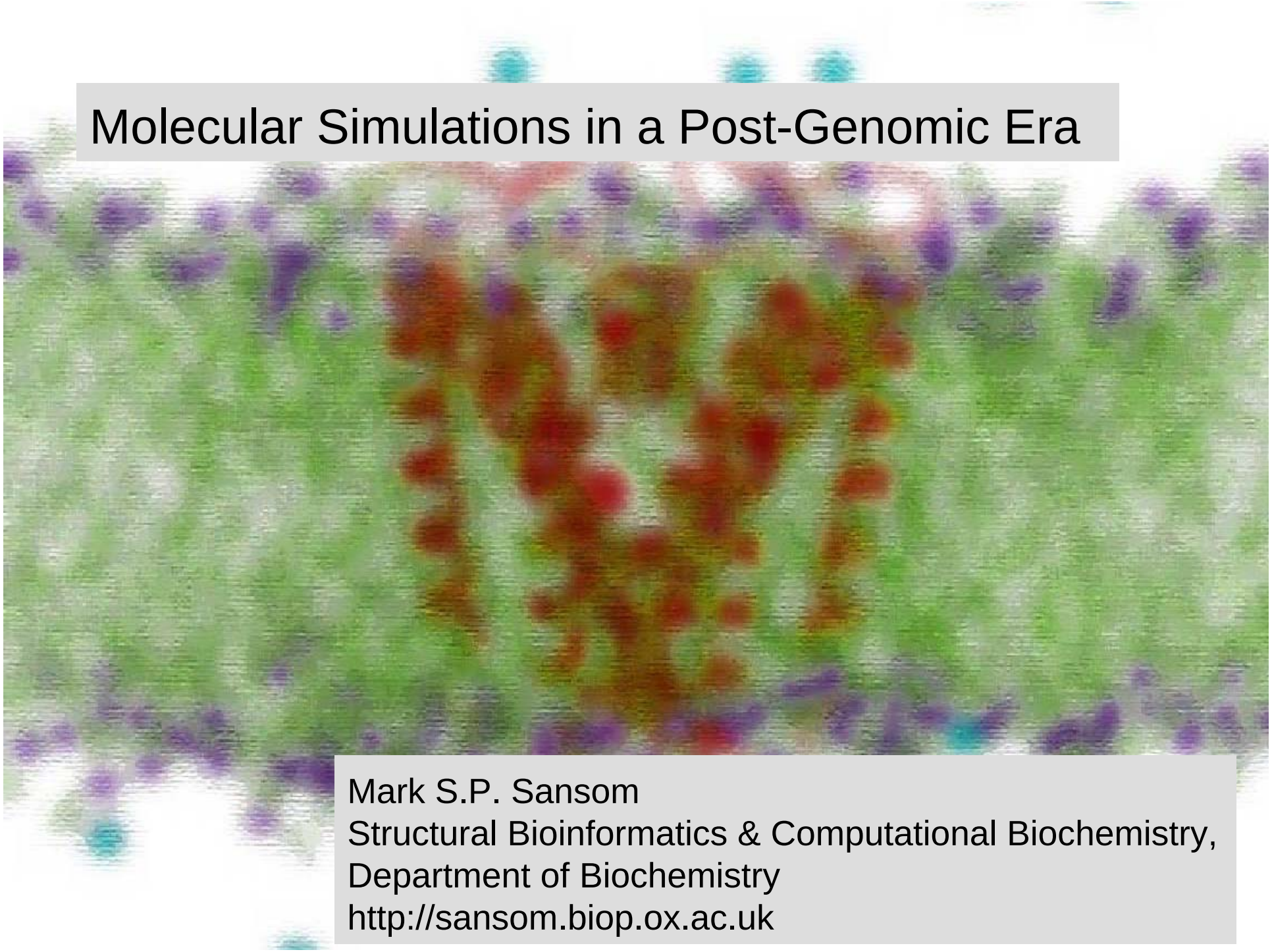
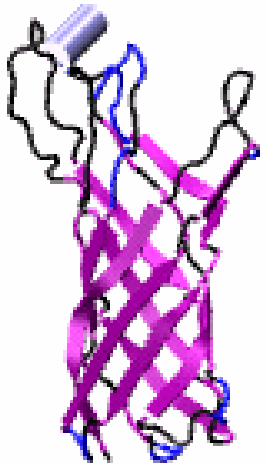
Molecular Simulations in a Post-Genomic Era



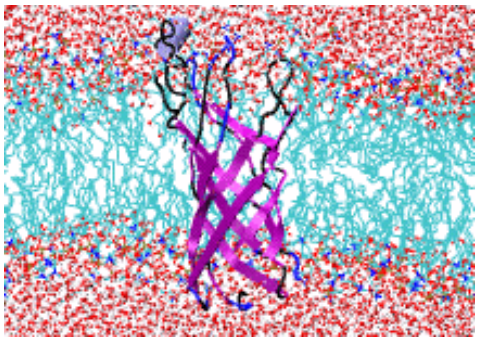
Mark S.P. Sansom
Structural Bioinformatics & Computational Biochemistry,
Department of Biochemistry
http://sansom.biop.ox.ac.uk

MD Simulations: from PDB to Dynamics

structure/model



simulation



analysis

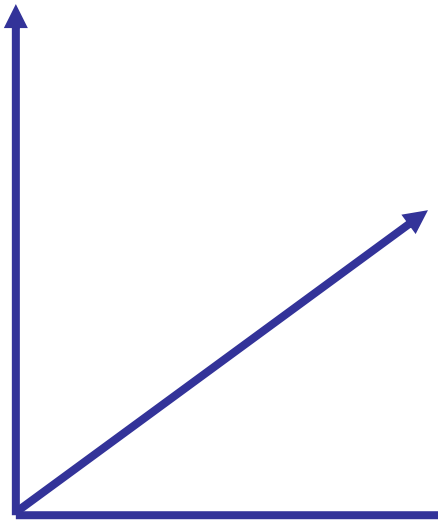
- Molecular simulations as a tool for protein structure analysis
- X-ray structure: average structure at 100 K in crystal
- MD simulations: dynamics at 300 K in ~cellular environment
- Challenge: to relate structural dynamics to biological function

Overview

- Challenges: from structural genomics & systems biology
- Advances in hardware: clusters, Grid computing & HPC
- Biology-driven simulations: structure-function relationships of proteins & more complex systems
- Simulation databases & pipelines
- BioSimGrid project (www.biosimgrid.org)
- IntBioSim: multi-level simulations
- National Grid Service (www.ngs.ac.uk) experiments

Three Key Directions

depth – improved models of interactions cf. physical chemistry

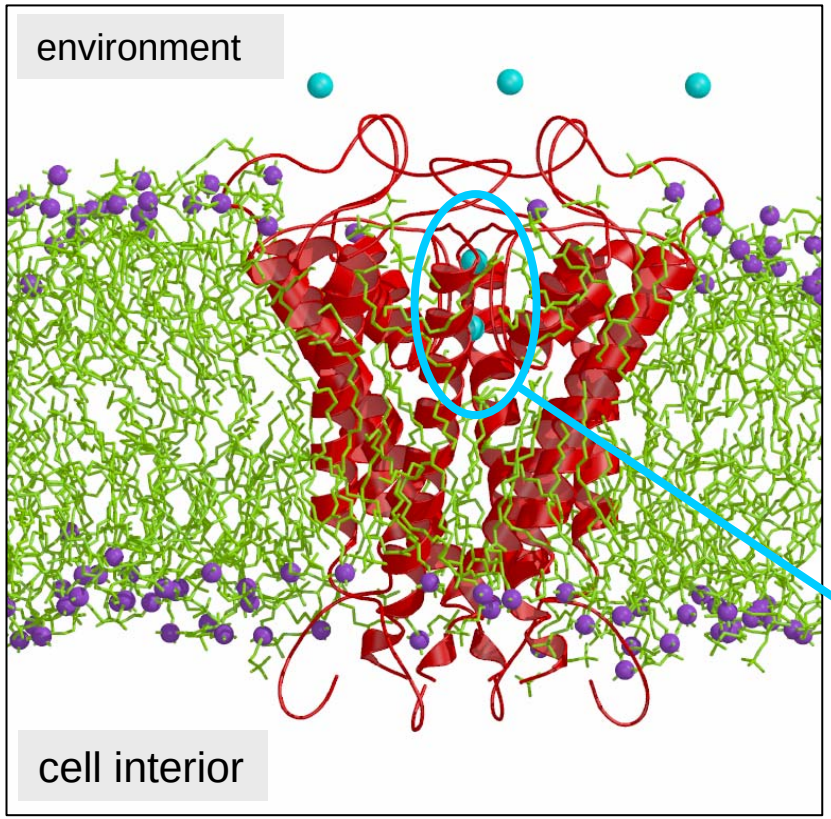


breadth – comparative simulations cf. structural genomics

complexity – multi-component systems cf. systems biology

- All three aspects – need HPC
- E-science issues – high-throughput & data integration

Complex Systems: MD Simulations of Biomolecules



water, ions

proteins & phospholipids

e.g. potassium channel

minimum system

50,000 to 100,000 atoms

biology

- Describe the forces on all atoms: $F = -dU(x)/dx$
- Integrate: $F = ma$ (a few million times…)
- Result: positions of all atoms for ~10 ns
- Experimental (static) structure → *in vivo* dynamics
- The challenge – to relate dynamics to function

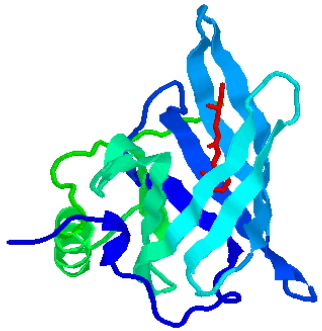
Molecular Dynamics Simulations of Biomolecules

- MD simulations of dynamics of biological molecules (proteins, DNA, membranes)
- Energy functions for MD
- Solvation & long range interactions
- Case studies:

  protein folding

  protein dynamics & large systems

X-ray (or NMR) experiments



static (average) structure

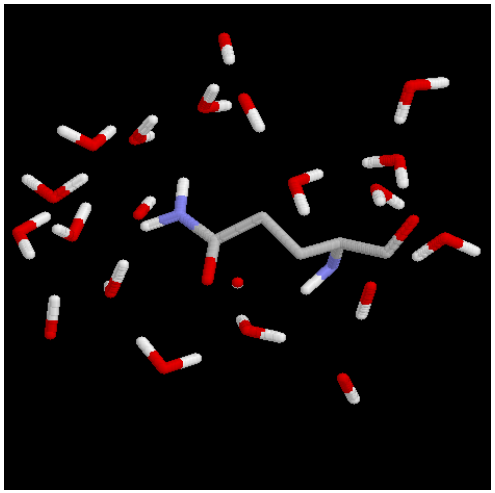
MD simulations

protein motions

Potential Energy Functions for MD

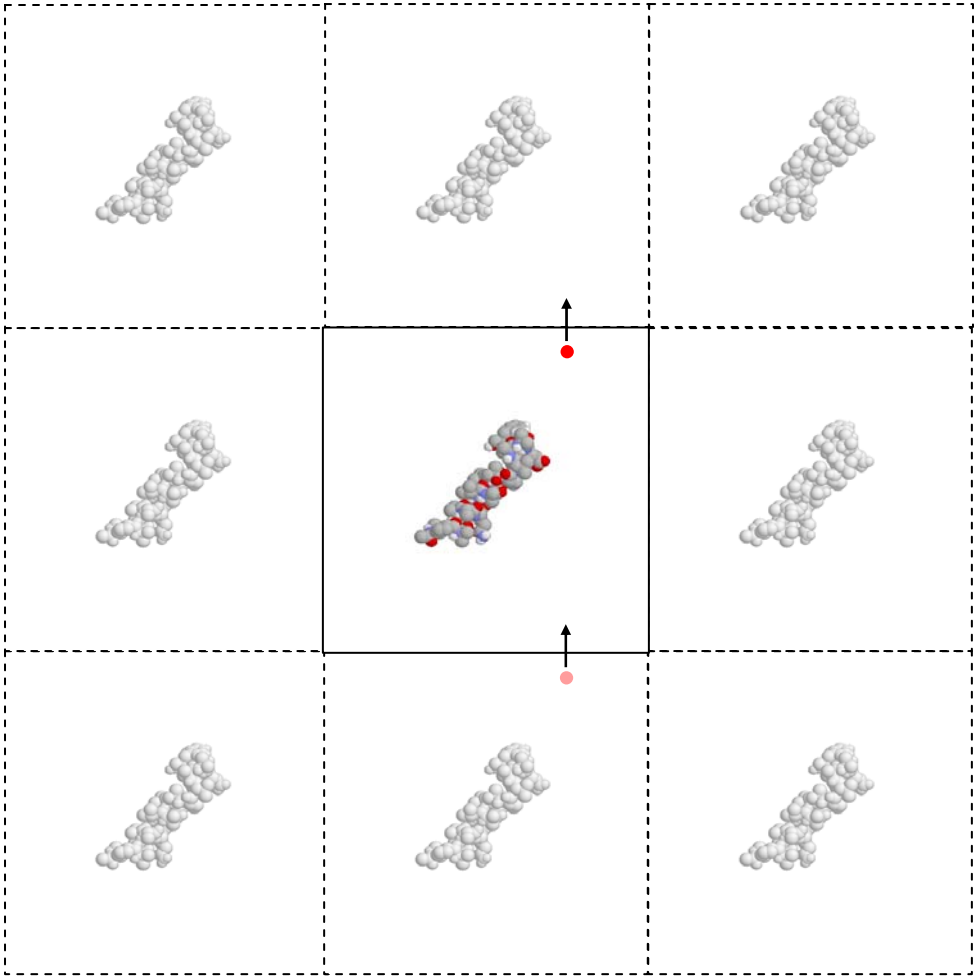
- Classical energy functions (no QM, i.e. “ball & springs” model)
- bonding interactions via simplified (harmonic) functions
- atoms treated as van der Waals spheres with single point charges
- Large systems – upto 10^6 atoms
- several packages - CHARMM, GROMACS, GROMOS, AMBER, NAMD

- $E = E_{BONDED} + E_{NON\text{-}BONDED}$
- $E_{BONDED} = E_{BONDS} + E_{ANGLES} + E_{TORSIONS}$
- $E_{NON\text{-}BONDED} = E_{VAN\ DER\ WAALS} + E_{ELECTROSTATICS}$
- Each term – simple function (e.g. quadratic)
- But … large number of pairwise interactions



Problems of a Finite System Size

- Restriction to relatively small systems e.g. 10x10x10 nm^3 = ca. $3x10^4$ atoms
- Finite system size introduces “boundary” problems
- Use of *periodic boundary* conditions to mimic an infinite system
- Long-range electrostatic effects are computationally expensive - use of either a cut-off or Ewald summation
- Need for solvation - large numbers of water molecules
- Timescales – 10 ns upwards (to ca. 1 μs)



Simulation & Analysis

Calculate forces on each atom

Update position of each atom after timestep δt e.g. δt = 1 fs (= 10^{-15}s)

Save atomic coordinates & energies every e.g. 1 ps (= 10^{-12} s) to give a *trajectory*

Off-line analysis

visualisation (movies)

sample structures (snapshots)

quantitative analyses
e.g. structural drift vs. time
e.g. fluctuations vs. residue

- New approaches – on line visualisation & interactive “steering” of simulations

Computational Resources

- Conventional supercomputers - research centres or universities
- e.g. HPCx (www.hpcx.ac.uk)
- Essential for very large scale simulations
- Fast communication between CPUs
- Code scaling issues

- PC (“beowulf”) clusters running Linux
- Individual research groups
- From small & hand-built (e.g. 16 nodes) to large & professional (e.g. 256 nodes)
- Communication between CPUs via conventional network technology

Parallelisation of Simulations

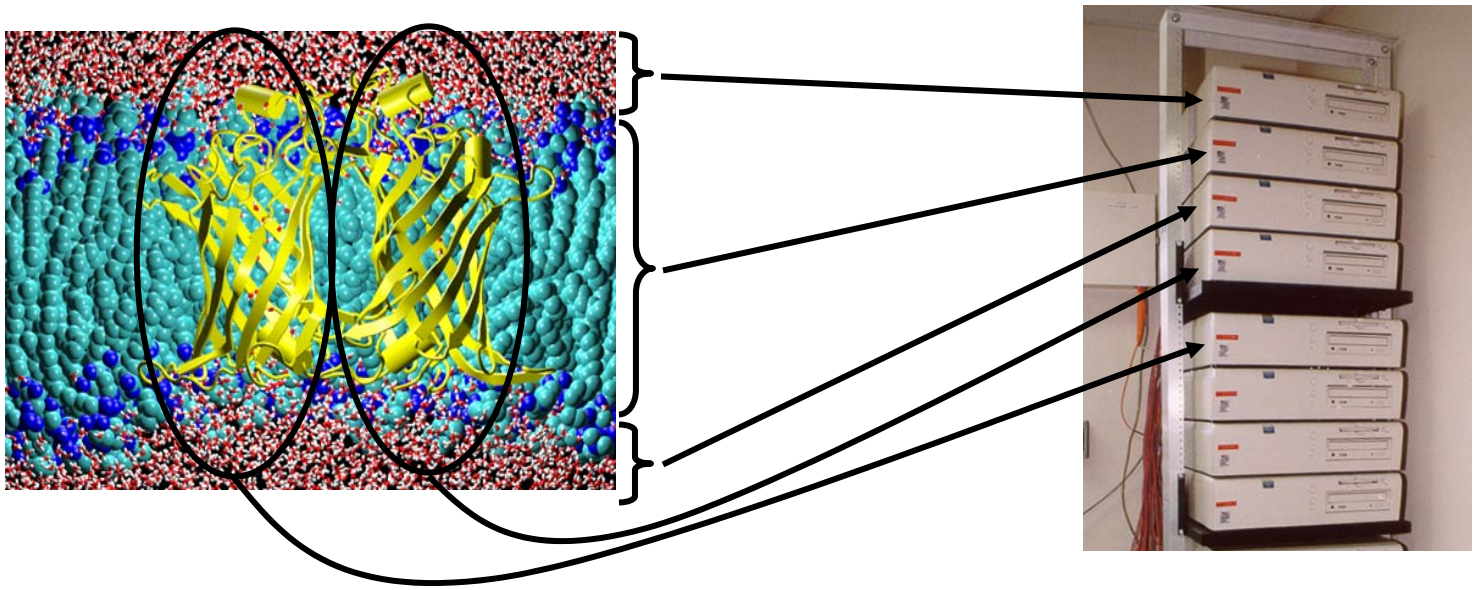
GRID
internet

Cluster
ethernet

HPC
shared memory

BlueGene
custom architecture

Inter-CPU communication speed



- Parallelization of codes – efficient partition of system onto CPUs
- Problem of long range interactions (FFT parallelization)

Reality Checks

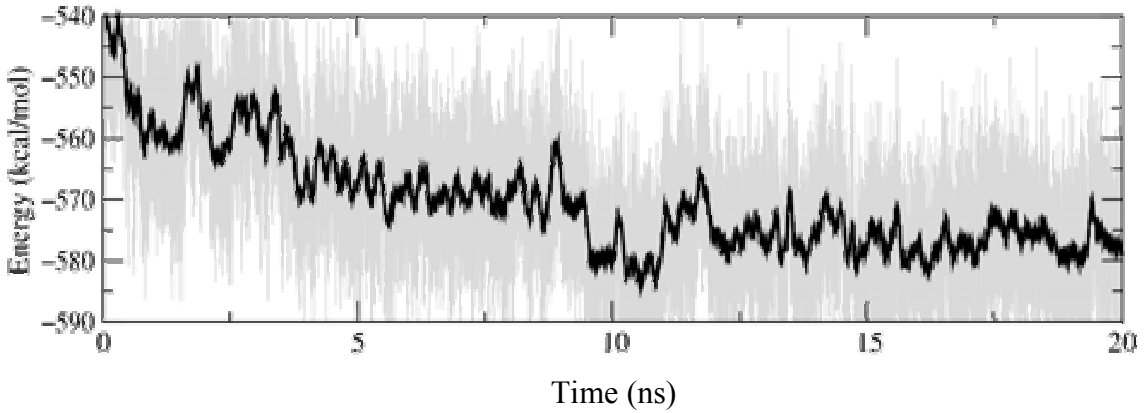
- How good are the forcefields?
- Sampling – need for longer simulation times
- Case studies:

  peptide folding – approaching NMR accuracy

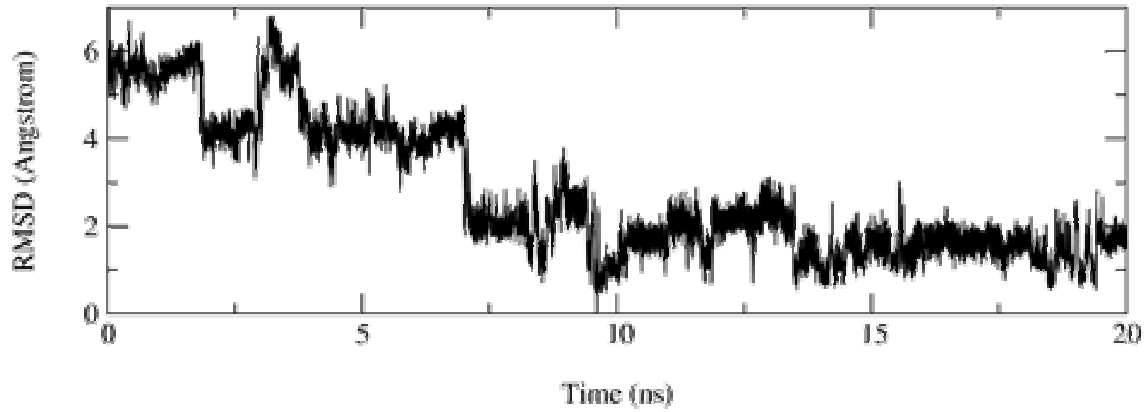
  K channels – improving on X-ray resolution

Trpcage Folding: A 20-mer Peptide

energy vs. time



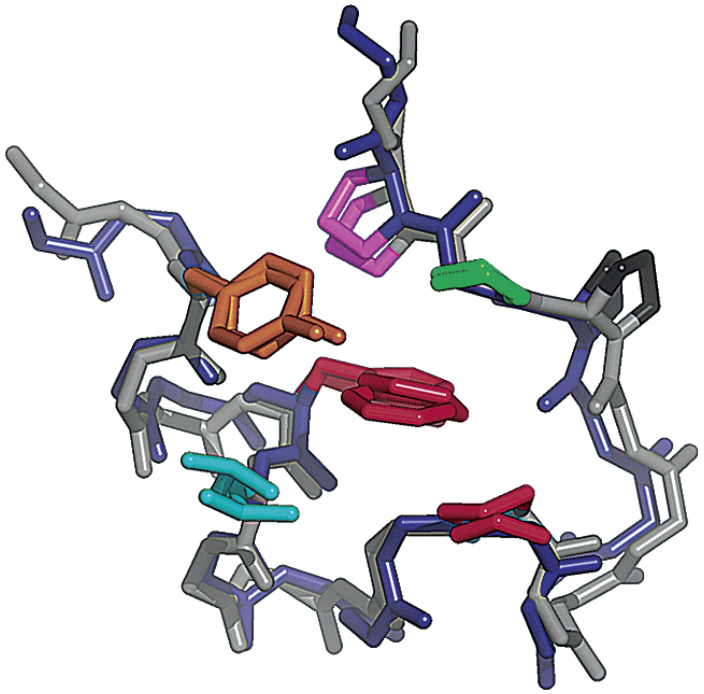
RMSD vs. time



B

All-Atom Structure Prediction and Folding Simulations of a Stable Protein
Simmerling et al. (2002) *JACS* 124:11258

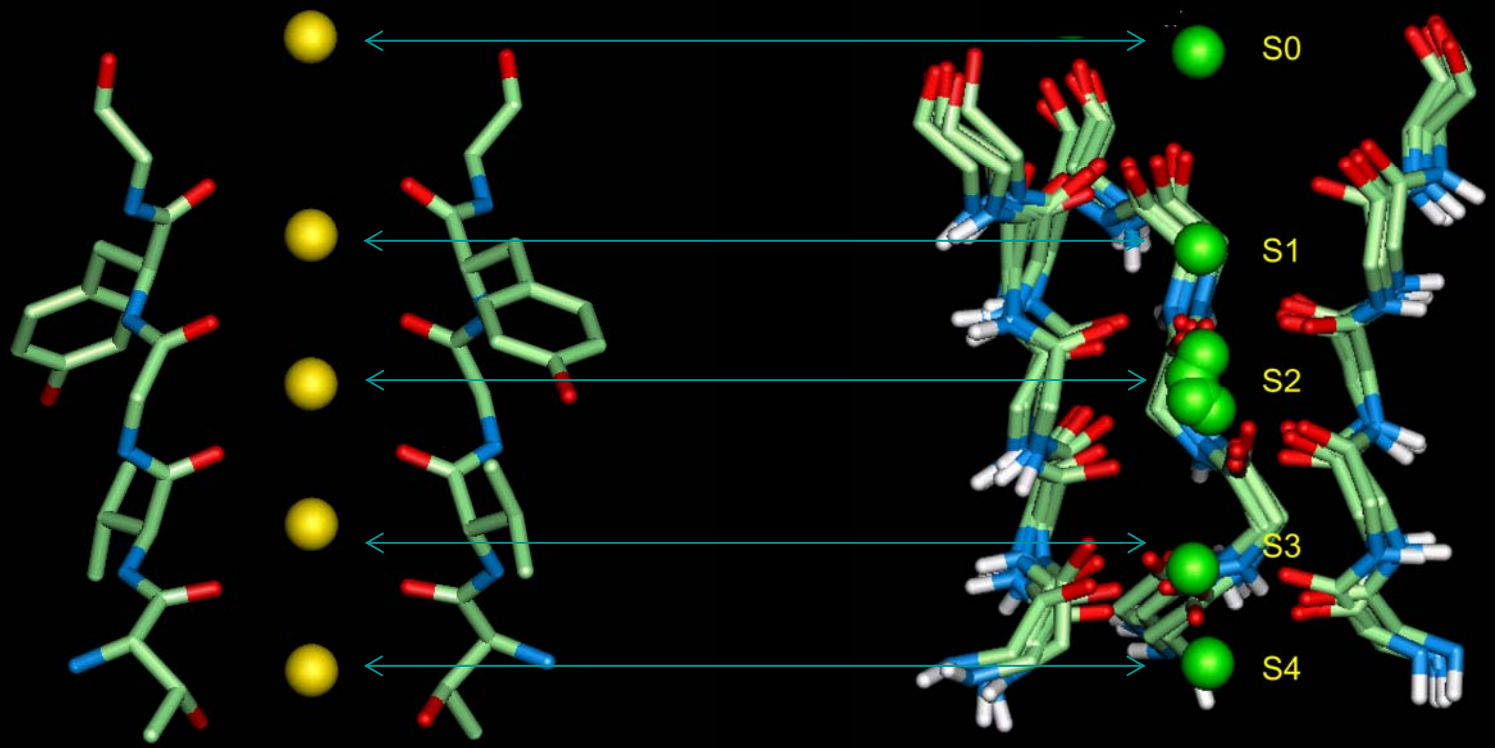
Trpcage: Experiment vs. Theory



- Simulated (blue) vs. experiment (grey)
- Good agreement (within experimental error)

- Only a very small “protein” – but successful
- Need to extend to larger & more complex systems

MD vs. X-Ray: A Reality Check

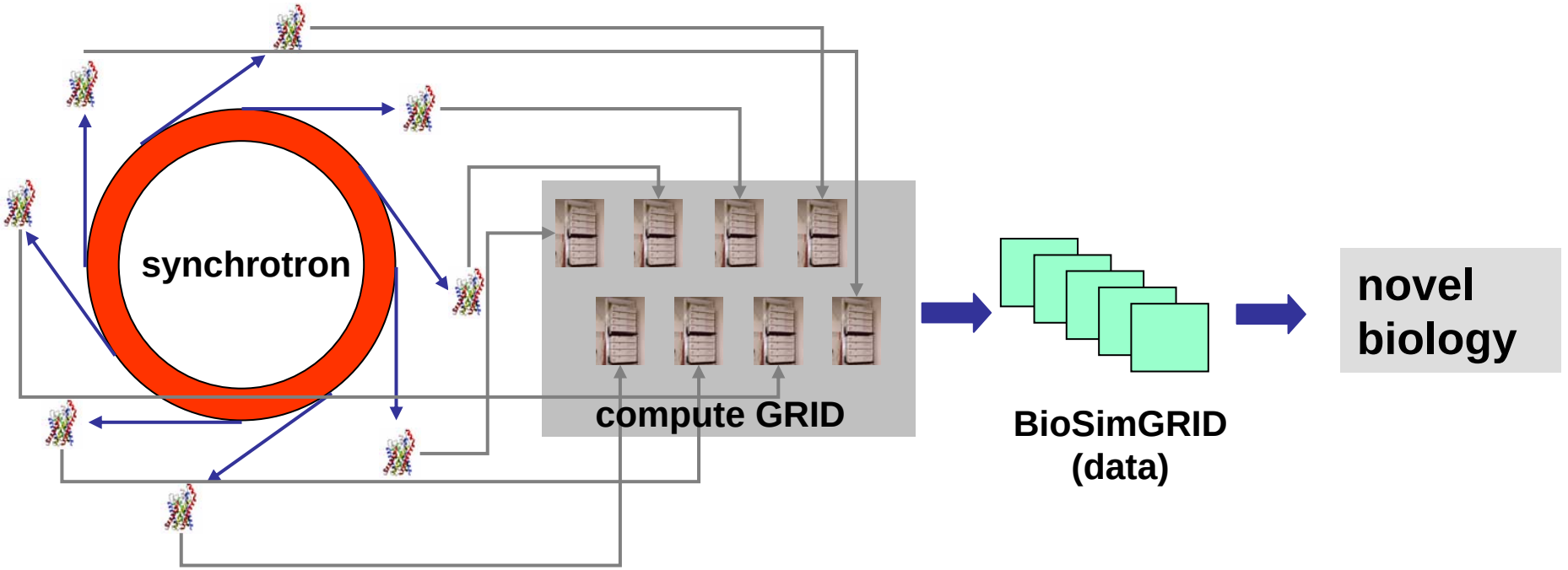


X-ray (2.0 Å)
MacKinnon et al. (2001)

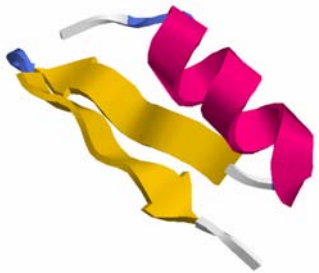
MD simulations
Shrivastava & Sansom (2000)

Towards HT MD

- Simulation pipeline
- QA tools & automated deposition

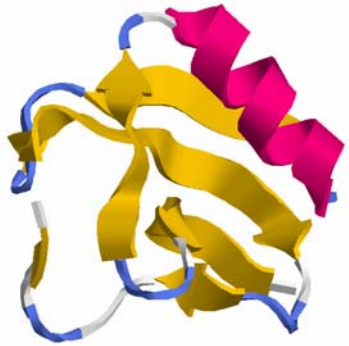


Comparative Simulations

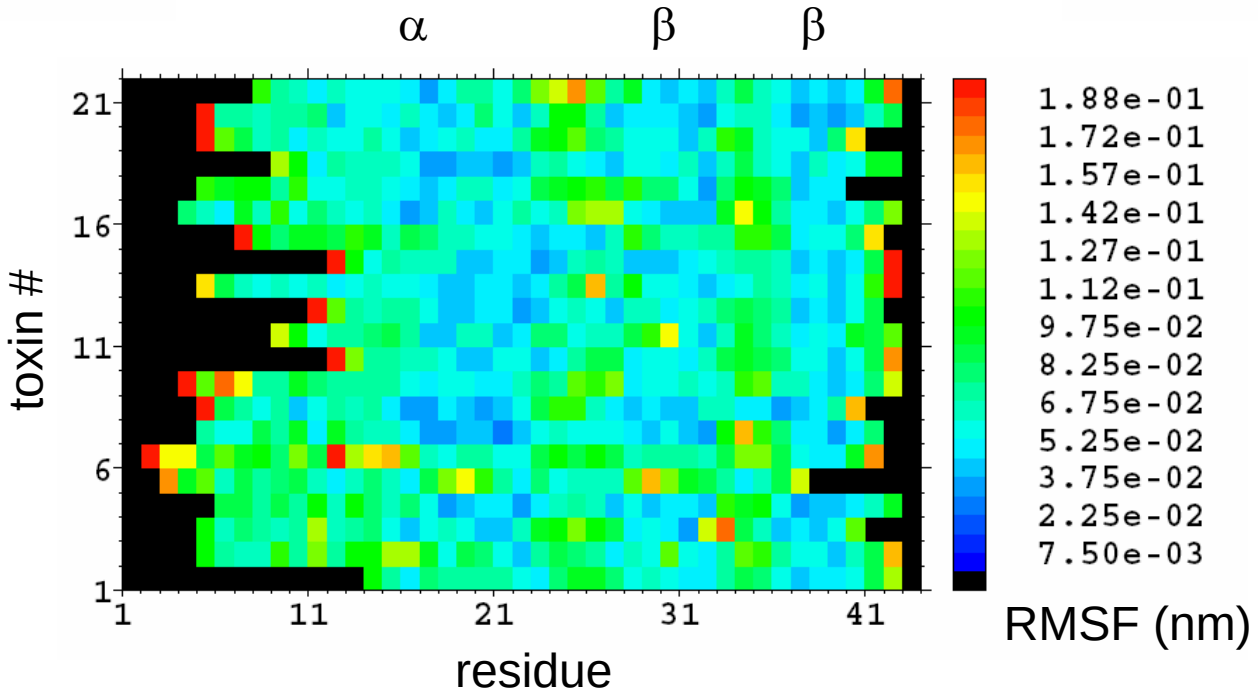


1acw

- 22 scorpion toxins (bind to K channels)
- Toxins have the same fold
- 10 ns MD run for each toxin
- Dynamic profile for a simple fold



1aho



Using the GRID

run simulation

(0.5 ns segments)

↓

check: energy, temperature & RMSD ‘on the fly’

↓

e-mail user re: simulation status

→

stop job if simulation unstable, or

if ok, resubmit for a further 0.5 ns

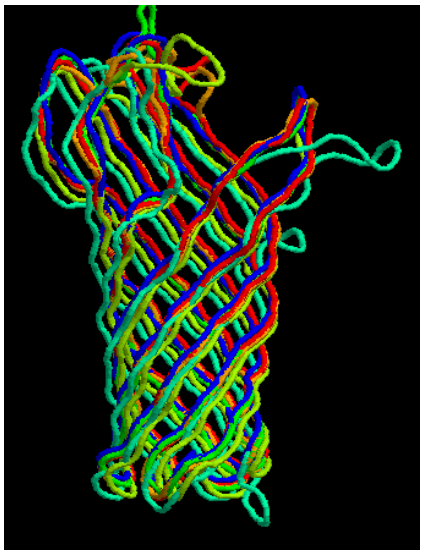
↑

iterate until simulation is complete

↑

integrate with biosimGRID for storage and analyses

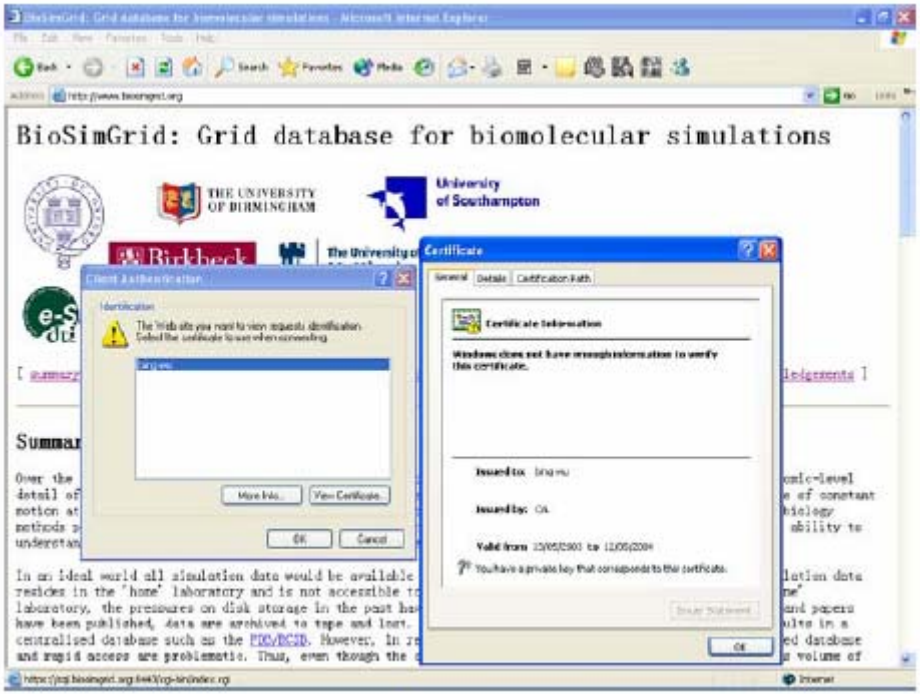
SHIDY- Shigella dysenteriae
SALTY- Salmonella typhimurium
ENTAE- Enterobacter aerogenes
SERMA- Serratia marcescens
HAEIN- Haemophilus influenzae



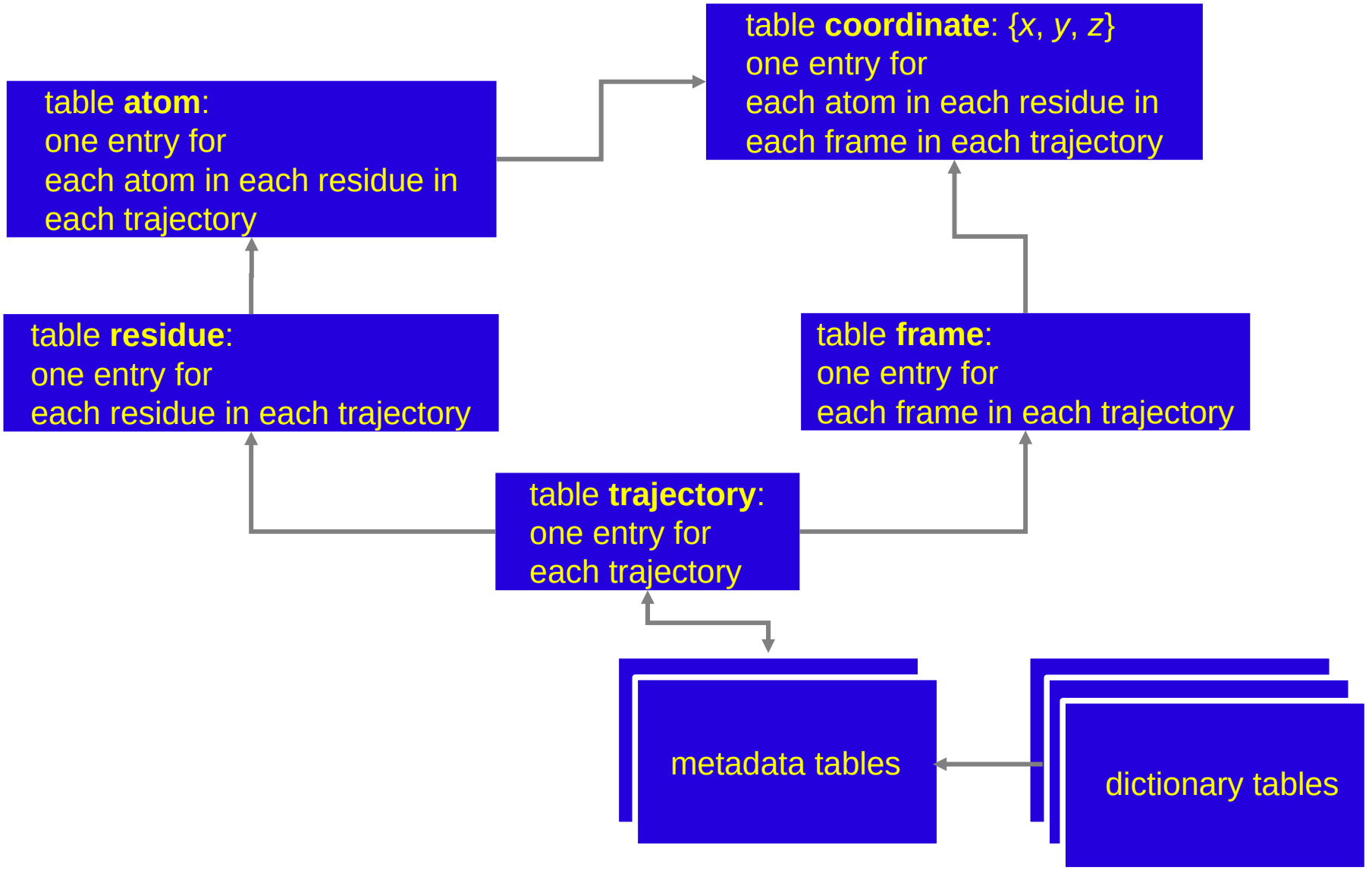
- Evaluation using the NGS (www.ngs.ac.uk)

Managing MD Data: BioSimGRID

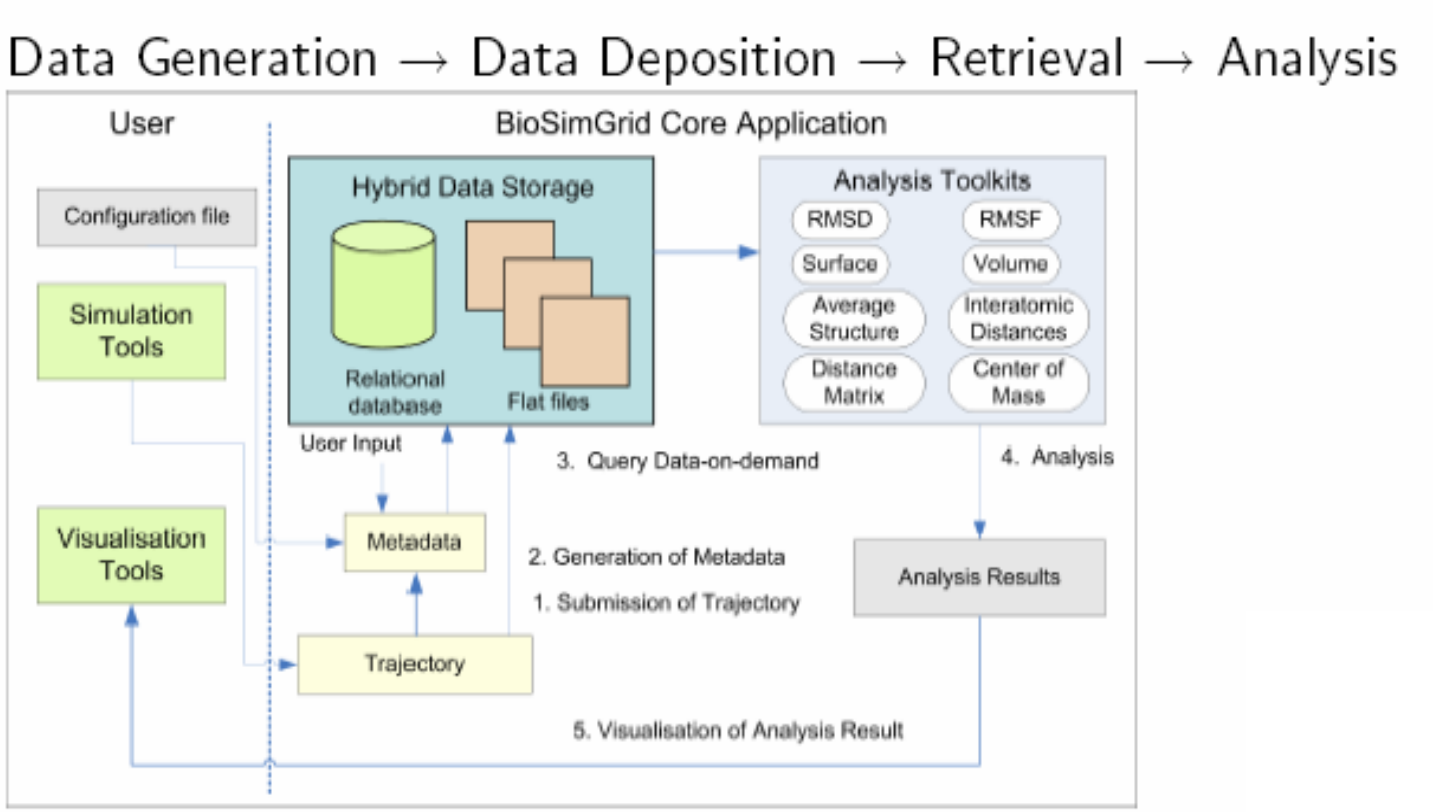
- www.biosimgrid.org
- Distributed database environment
- Software tools for interrogation and data-mining
- Generic analysis tools
- Annotation of simulation data
- Collaboration: Oxford, Southampton, Bristol, London, Nottingham, York

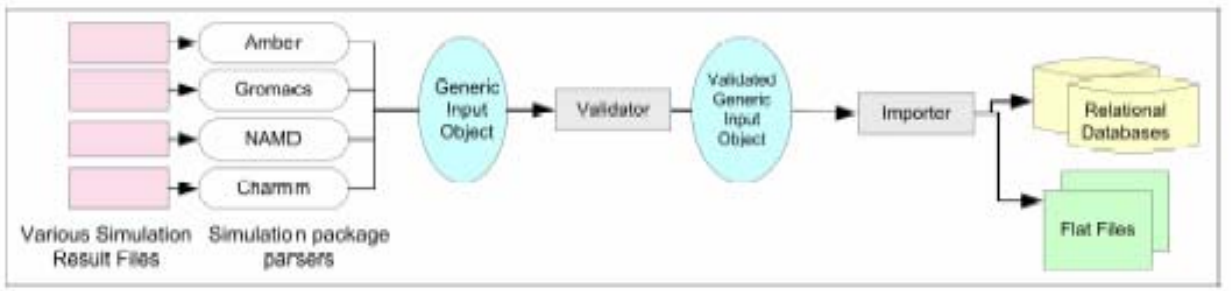


Database Design: Simplified

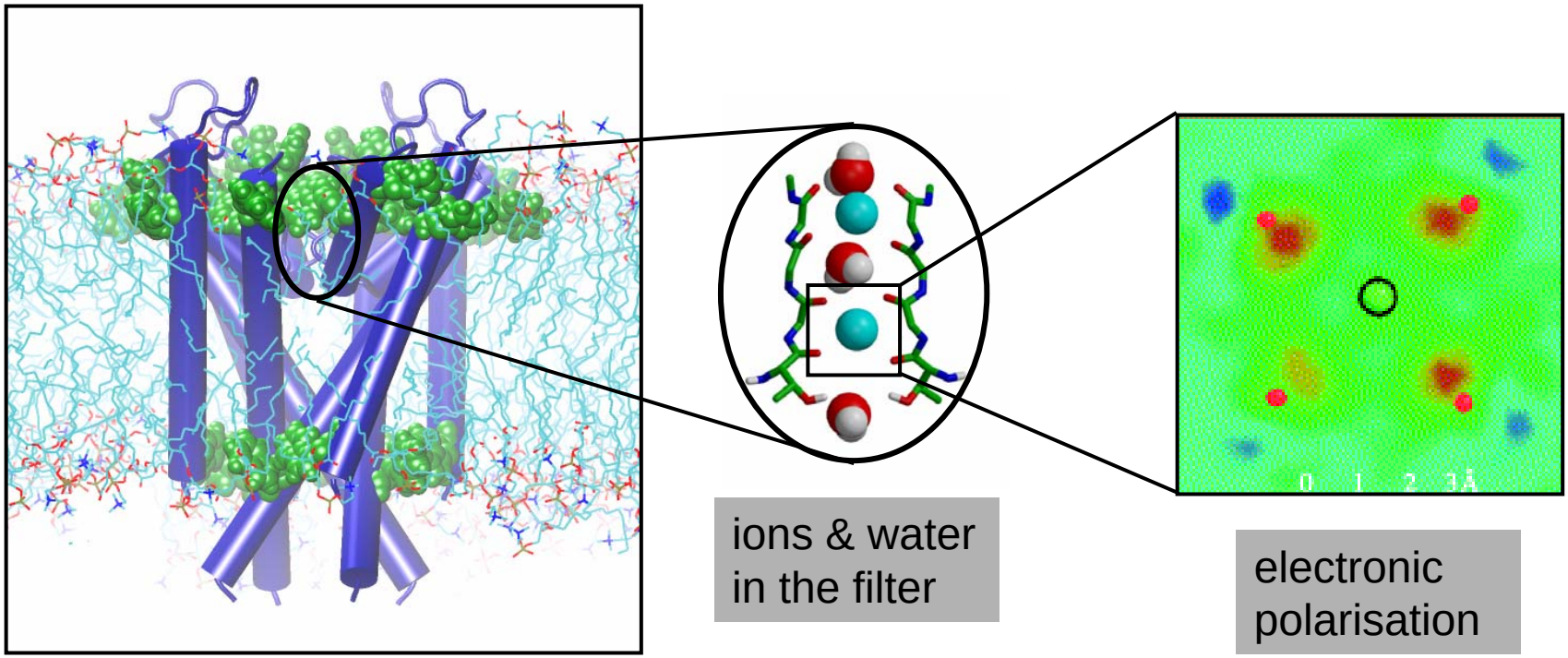


BioSimGrid Workflow





Back to Chemistry…



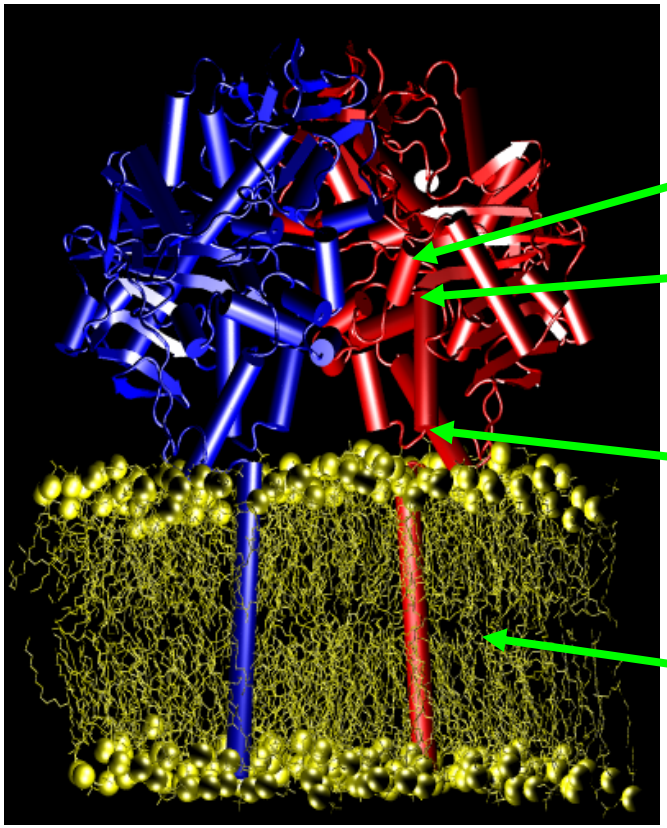
channel & membrane

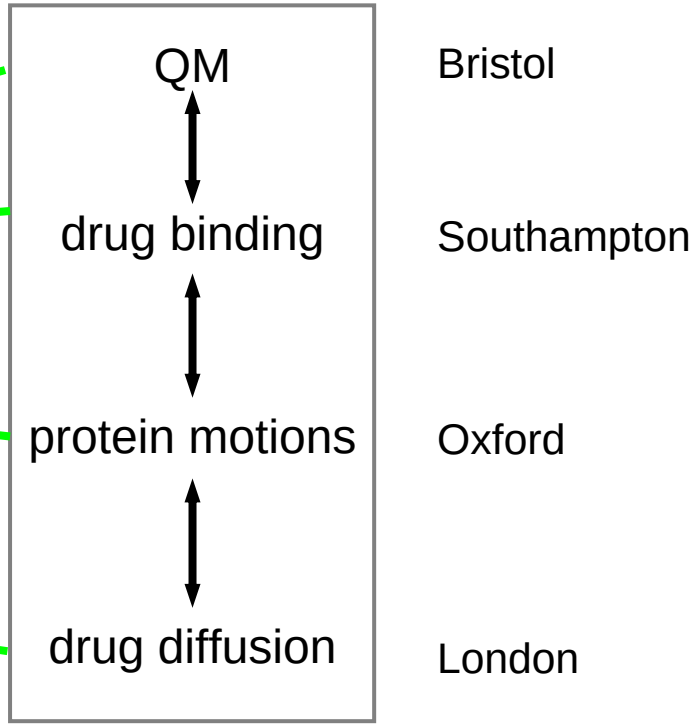
ions & water in the filter

electronic polarisation

- Ions and water in the channel – MD on >10 ns timescale reveals ion permeation Domene & Sansom (2003) *Biophys J* 85:2787
- *Ab initio* calculations for accurate energetics Guidoni & Carloni (2002) *BBA* 1563:1

Future Directions: Multiscale Biomolecular Simulations





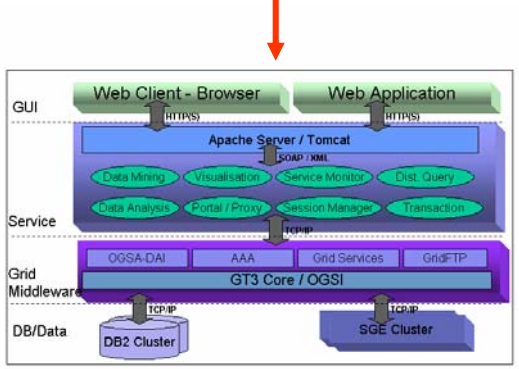
- Membrane bound enzymes – major drug targets (cf. ibruprofen, anti-depressants, endocannabinoids); gated access to active site coupled to membrane fluctuations
- Complex **multi-scale** problem: QM/MM; ligand binding; membrane/protein fluctuations; diffusive motion of substrates/drugs in multiple phases
- Need for integrated simulations on GRID-enabled HPC resources

Computational Challenges

HPCx

Linux cluster

IntBioSim

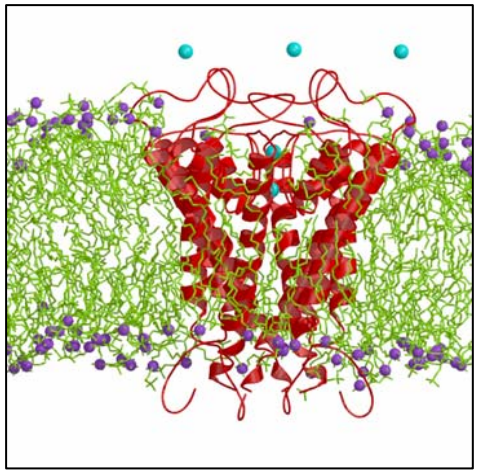


BioSimGRID database
www.biosimgrid.org

- Need to integrate HPC, cluster & database resources
- Collaboration: Oxford, Southampton, Bristol, London, Manchester

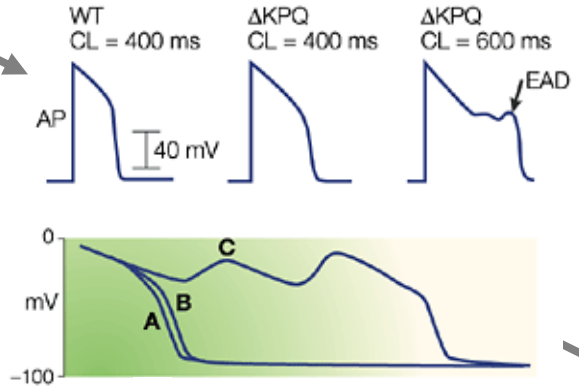
Towards Systems Biology

molecular

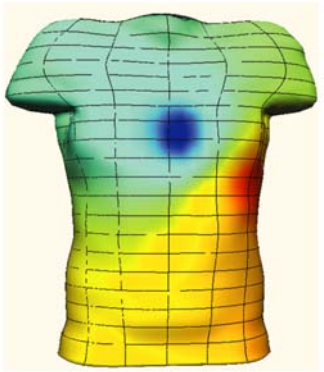


Noble (2002) *Nature Rev. Mol. Cell.Biol.* 3:460

cellular



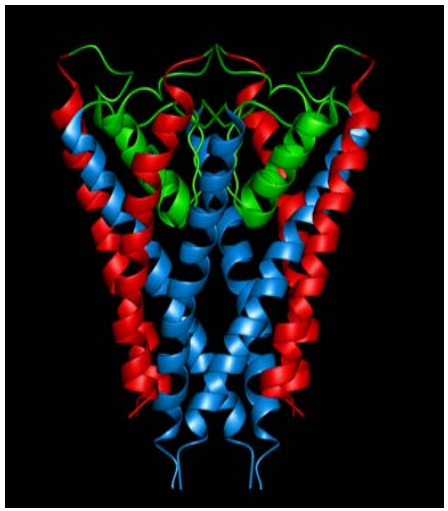
organism



- Aims: *in silico* knock outs and/or drug treatments
- Need to integrate *rigorously* if to deliver accurate & hence biomedically useful results
- Problem: how to link the different levels of description

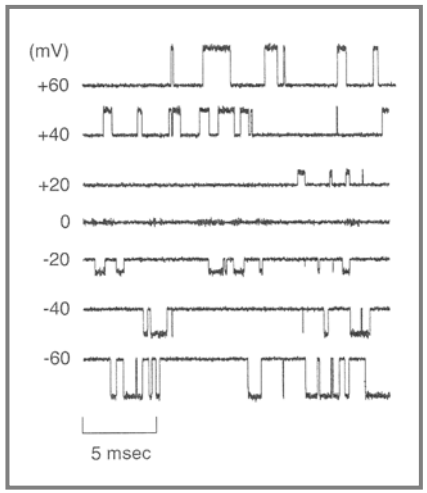
From Structure Towards Function

HT modelling & simulation
Channel & transporter model databases



Coarse-grained simulations

Physiological behaviour



Phil Biggin
Alessandro Grottesi
Andrew Hung
Daniele Bemporad
Shozeb Haider
Kaihsu Tai
Bing Wu
Jorge Pikunic
Syma Khalid
Zara Sands
Hyunji Kim
Oliver Beckstein
Yalini Pathy
Pete Bond
Jonathan Cuthbertson
Sundeep Deol
Jeff Campbell
Loredana Vaccaro
Jennifer Johnston
Katherine Cox
Robert D'Rozario
John Holyoake
Shiva Amiri
Samantha Kaye
Anthony Ivetac
Sylvanna Ho
Tim Carpenter
Victoria Caulfeild

Collaborating Teams
BioSimGRID
IntBioSim
Integrative Biology
MPSI
Fran Ashcroft & OXION
Heart Physiome

Funding
BBSRC
DTI
EPSRC
GSK
IBM & HPCx
MRC
NPL
OeSC (EPSRC & DTI)
OSC (JIF)
The Wellcome Trust
The Wolfson Foundation

<u>BioSimGRID & IntBioSim</u>
Leo Caves (York)
Simon Cox (Southampton)
Jon Essex (Southampton)
Paul Jeffreys (Oxford)
Charles Laughton (Nottingham)
David Moss (Birkbeck)
Peter Coveney (UCL)
Adrian Mulholland (Bristol)
Graham Riley (Manchester)
David Gavaghan (Oxford)
Steve McKeever (Oxford)

<u>Heart Physiome</u>
Peter Hunter
Denis Noble
David Paterson