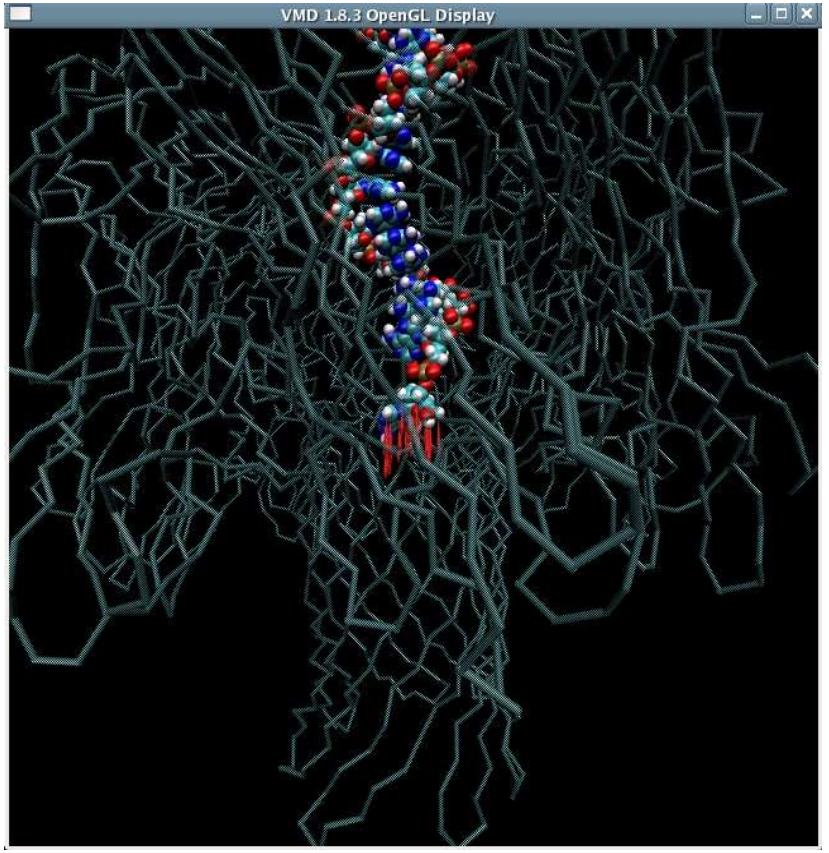
Simulated Pore Interactive Computing Environment: Using grid computing to understand DNA translocation across protein nanopores



RealityGrid

ESLEA
Exploitation of Switched Lightpaths for e-Science Applications

Jha, Harvey, Coveney, Pickles, Pinning, Pezzi & Clarke

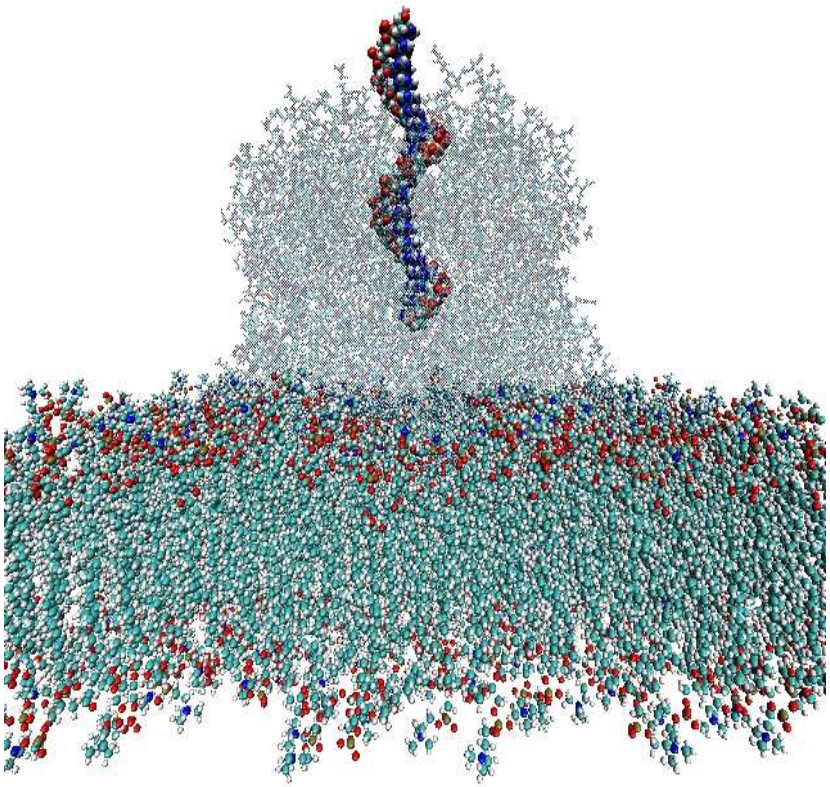
Simulated Pore Interactive Computing Environment

- Transport of biomolecules through protein pores not well understood. Semi-flexible polymer in confined geom.

- Fully atomistic simulations crucial to capture pore-protein interaction. Long timescale (~10s microseconds).

- Size, complexity & timescale make simulations computationally intensive.

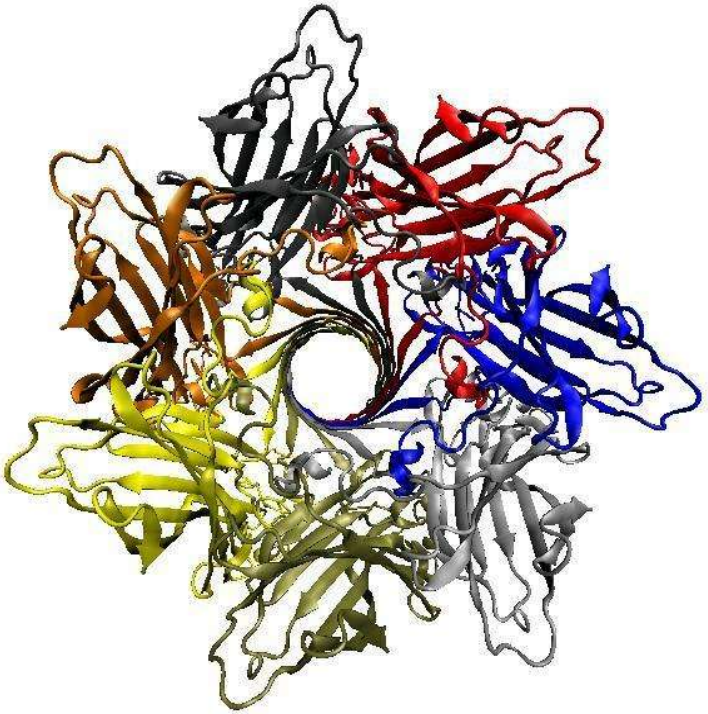
3X10^7 CPU (*tightly-coupled*) hours using “simple” MD. Need to do better...



DNA translocation through an alpha-hemolysin protein pore embedded in a lipid layer. ~275k atoms. Water not shown

SPICE: Grid Computing Using Novel Algorithms (1)

- Novel Algorithm: Steered Molecular Dynamics (SMD) to “pull DNA through the pore”. Jarzynksi's Equation (JE) to compute equilibrium free energy profile from the non-equilibrium pulling.
- Reduce computational cost by 50-100!
- Trivial – in theory, not in practise; Introduces two variable parameters. Determine “optimal” choice of parameters.
- Replace one long sim with multiple (~100) short(er) sims



A view into the pore from the top end. Note the seven fold symmetry of the system.

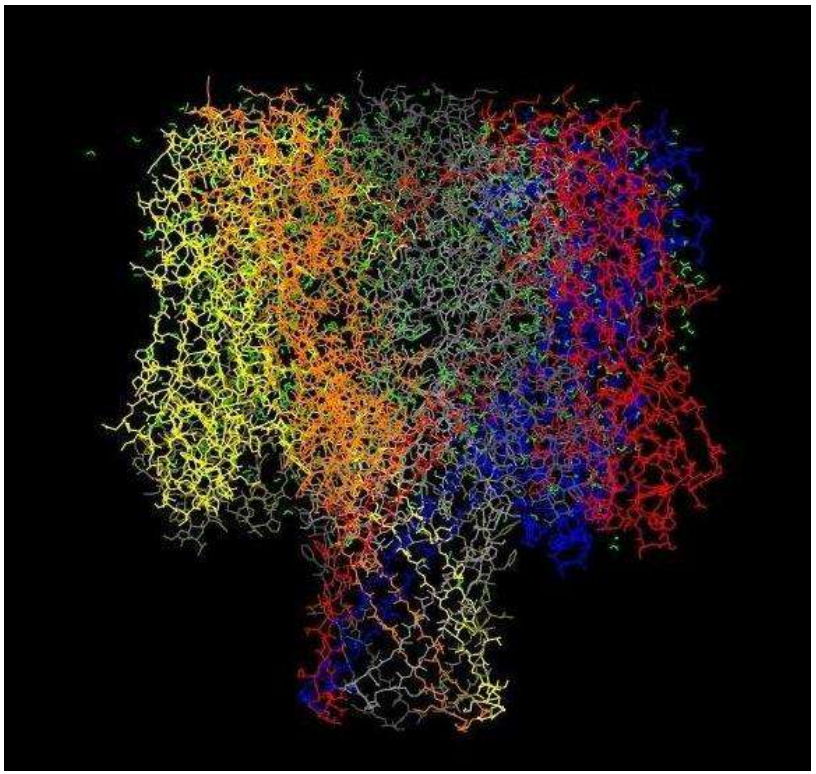
SPICE: Grid Computing Using Novel Algorithms (2)

- Distribute multiple large-scale sims
  Coupling delocalized sim and viz resources

  "Live coupling": Interactive simulations -- use viz to steer sim.

  Not just parameters...

- Grid facilitates greater resource utilization, novel analysis & uniform interface making ~100 simulations more manageable



Transform comp intractable problem using SMD+JE
Not just exploits grid infrastructure, but requires it...

SPICE: Computing the Free Energy Profile (FEP)

Step I: Understand structural features using static visualization

Step II: Interactive simulations for dynamic and energetic features
- Steered simulations: bidirectional communication.
  Qualitative + Quantitative (SMD+JE)
- Haptic interaction: Use haptic to feel feedback forces

Step III: Simulations to compute "optimal" parameters values
75 simulations on 128/256 processors each.

Step IV: Use computed "optimal" values to calculate full FEP along the cylindrical axis of the pore.

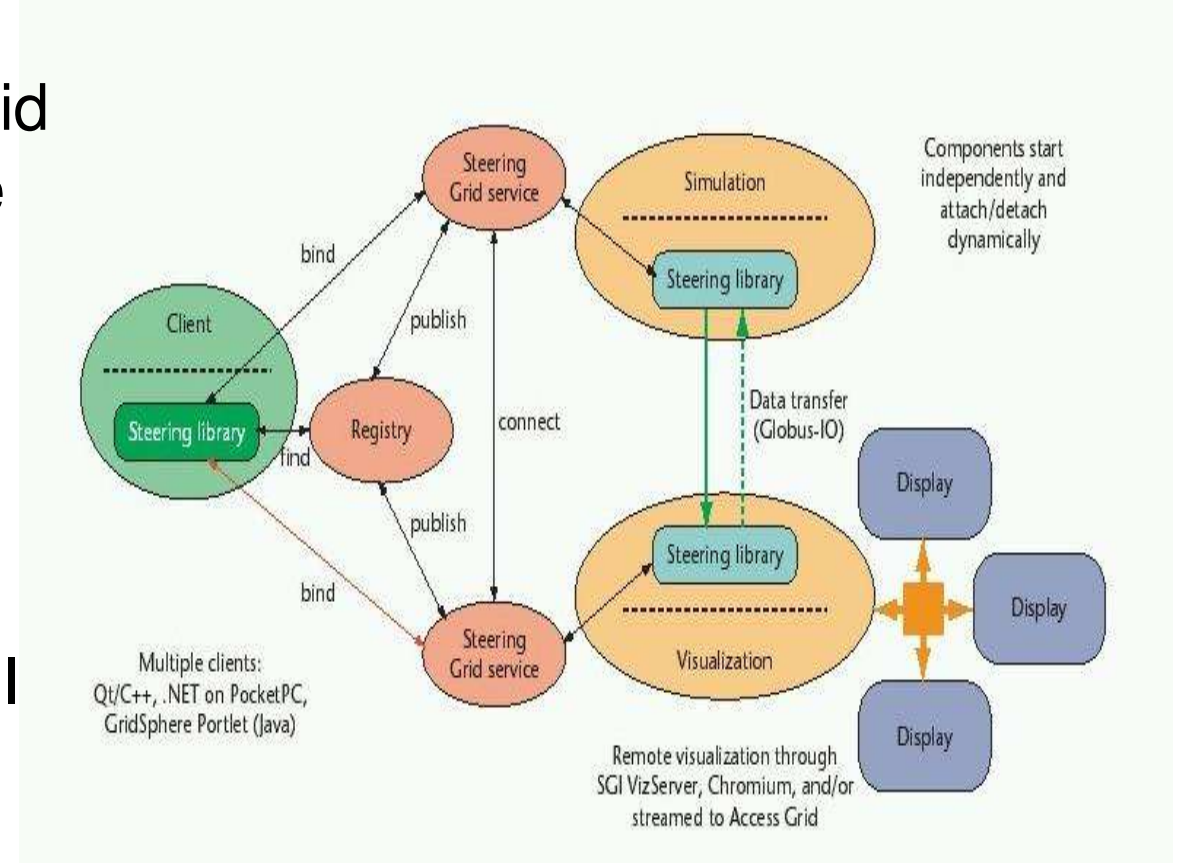
Requires ~250,000 CPU hours (100 simulations of 2500 CPU hours)!

SPICE: Grid Infrastructure (1)

RealityGrid Steering Infrastructure + NAMD (parallel MD engine)

App interfaces grid infrastructure through the ReG steering library

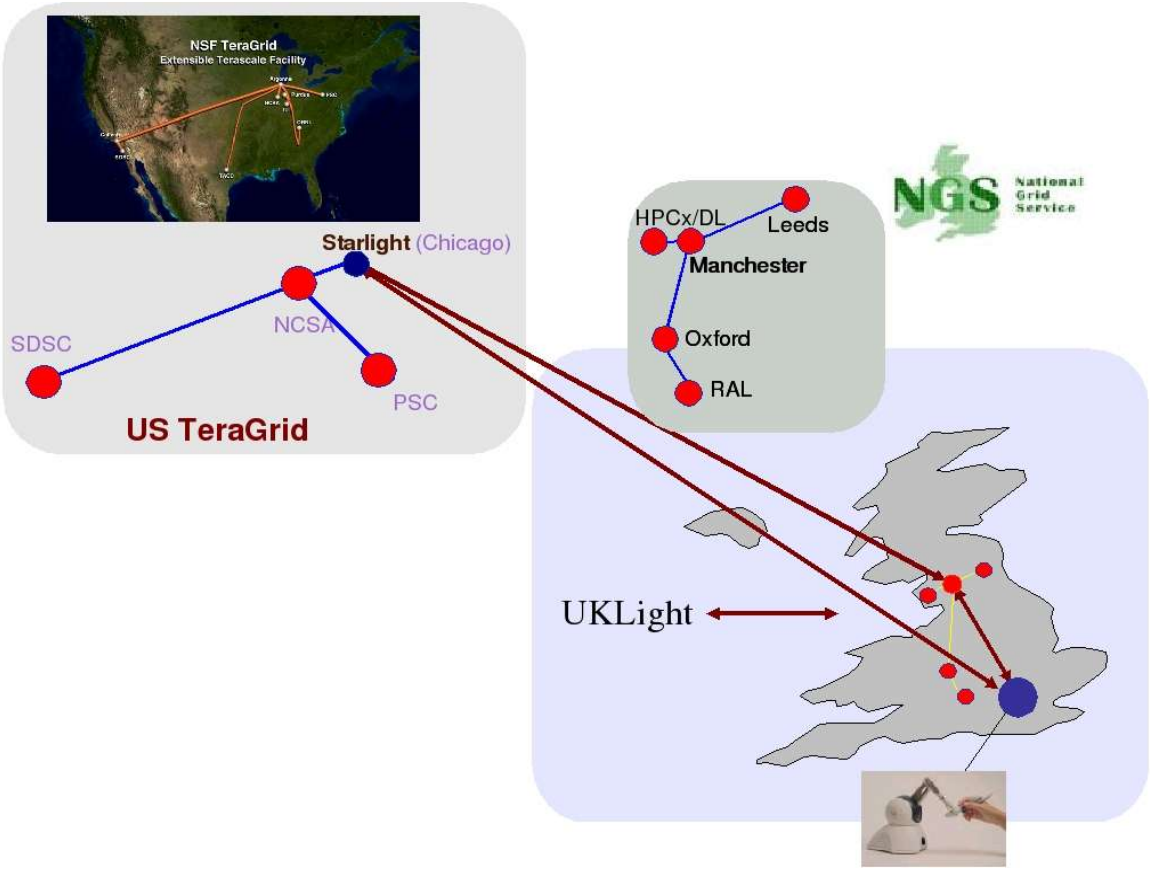
ReG Steering API:
-- launcher, steerer use server side API
-- App uses client side API



(WSRF::Lite based infrastructure for launching, steering, comm and notification soon)

SPICE: Grid Infrastructure (2)

Simulation: (II, III, IV)
NGS + TeraGrid
Visualization: (I,II)
Prism (UCL)
Sim–Viz connection:
UKLight/GLIF



Batch simulations post optimal parameter. N*ot batch sim to explore phase space*

Resources more *R-U-M* since TeraGyroid...
Middleware better? Or our trouble-shooting?

SPICE -- UKLight

+ Interactive simulations

  Qualitative (general) + Quantitative (specific)

  Require 128/256px of HPC to compute fast enough for interactivity

  Steady-state data stream (up & down) -- few MB/sec

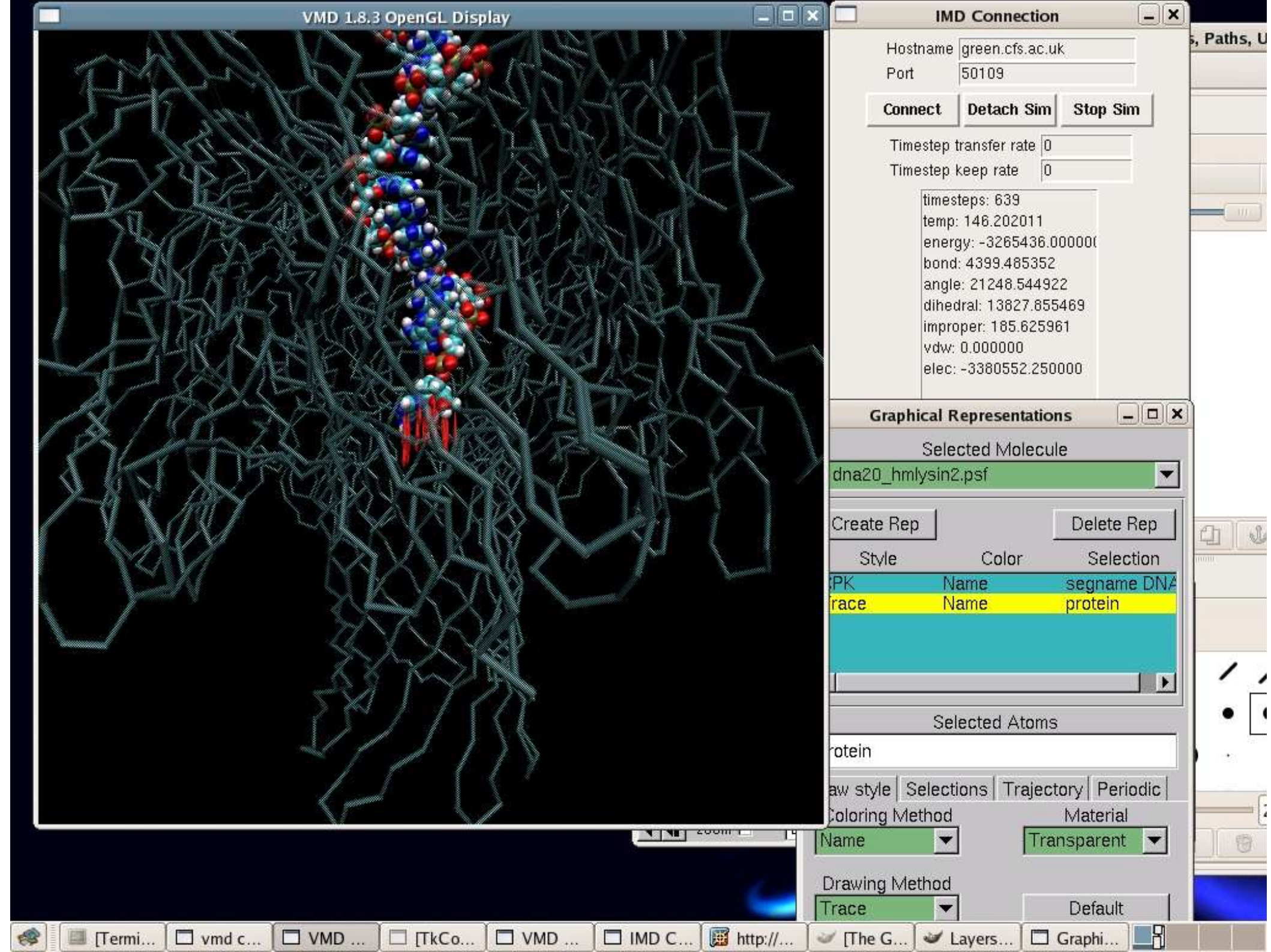
  Simple? "Unreliable" sim-vis commn simulation slowdown.?

  Interactive simulations perform better when using UKLight for communication between simulation and visualization!

  Advanced networks provide high QoS in terms of *packet loss, jitter* and *latency*.

  e.g ~20 vs 5 intermediate routers en route to NCSA from UCL-Prism

+ Total data generated in TeraBytes; replicated


VMD 1.8.3 OpenGL Display
IMD Connection
Hostname green.cfs.ac.uk
Port 50109
Connect
Detach Sim
Stop Sim
Timestep transfer rate 0
Timestep keep rate 0
timesteps: 639
temp: 146.202011
energy: -3265436.000000
bond: 4399.485352
angle: 21248.544922
dihedral: 13827.855469
improper: 185.625961
vdw: 0.000000
elec: -3380552.250000
Graphical Representations
Selected Molecule
dna20_hmlysin2.psf
Create Rep
Delete Rep
Style
Color
Selection
PK
Name
segname DNA
race
Name
protein
Selected Atoms
rotein
aw style
Selections
Trajectory
Periodic
Coloring Method
Name
Material
Transparent
Drawing Method
Trace
Default
[Termi...
vmd c...
VMD ...
[TkCo...
VMD ...
IMD C...
http://...
[The G...
Layers...
Graphi...

Thoughts...

- Getting infrastructure to work needs to be made easier (e.g. dual homing of systems -- adminstrative nightmare)
- Develop better theoretical underpinnings:
  Experiments to understand effect of network (latency, jitter) & RTT with simulation configuration (px's, IMD frequency)
- Testing protocols for low latency?
- User interaction with CPS? (e.g. dynamic app driven reservation of lightpaths)
  Co-scheduling with other resources? Via ReG API?
  GLIF - Global Plane & Grid Integration Middleware WG

SPICE @ AHM05

- Tue 13:00 - 14:00 UCL
- Wed 09:00 - 13:00 EPSRC
- Wed 14:30 - 15:00 ESNW
- Thu 12:00 - 15:00 EPSRC

SPICE a finalist at the SC05 HPC Analytics Challenge
(also at SC Global and numerous booths at SC05)

http://www.realitygrid.org/SPICE