

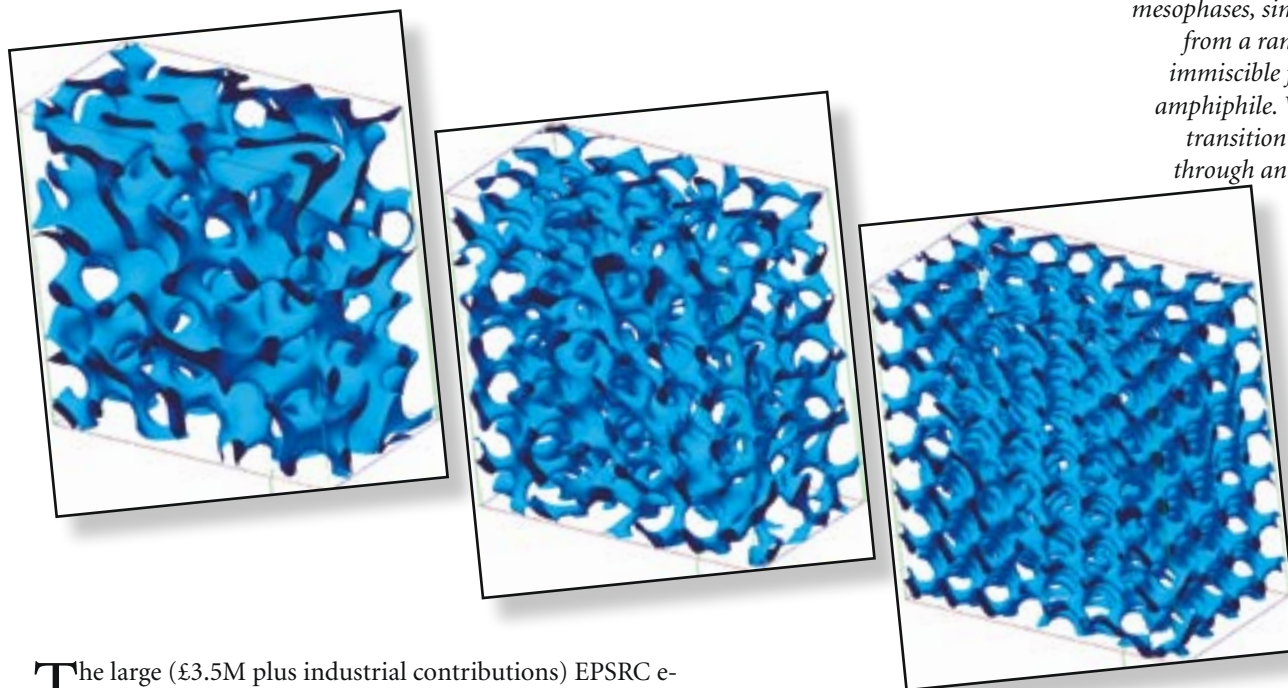
# RealityGrid: high-performance computing, visualization, computational steering & teragrids

Jens Harting,  
*RealityGrid Post Doctoral Research Fellow funded by EPSRC*

Shunzhou Wan,  
*RealityGrid Post Doctoral Research Fellow funded by the Edward Jenner Institute for Vaccine Research*

Peter V Coveney  
*Professor of Physical Chemistry and Director, Centre for Computational Science, University College London (Peter.V.Coveney@ucl.ac.uk)*

*Figure 1. Self-assembled equilibrium sponge (left) and liquid-crystalline gyroid (right) bicontinuous cubic mesophases, simulated using LB3D, from a random mixture of two immiscible fluids with dispersed amphiphile. We found a lyotropic transition from one to another through an equilibrium 'melted gyroid' mesophase (centre) (adapted from [2]).*



The large (£3.5M plus industrial contributions) EPSRC e-Science Pilot Project called RealityGrid, which commenced during the first half of 2002, is the main UK e-Science project dedicated to high performance grid computing research. A central objective of RealityGrid is to extend the concept of a Virtual Reality centre across the grid, linking it to massive computational resources at high performance computing (HPC) centres as well as to experimental facilities. A twin-track approach to research is being employed within RealityGrid: the 'fast track' uses currently available grid middleware to construct a working grid, while the 'deep track' involves computer science teams in harnessing leading-edge research to create a robust and flexible problem-solving environment in which to embed the RealityGrid.

RealityGrid is a collaboration between several teams of physical scientists, computer scientists and software engineers, see [1]. To meet its objectives, it is utilising a computing environment built around the UK's most advanced computing technology and infrastructure.

The scientific application that is being used as a test bed for the HPC, steering and grid technologies being developed within the RealityGrid project has mainly been a three-dimensional lattice-Boltzmann model for the hydrodynamically correct simulation of binary and ternary amphiphilic fluid mixtures ('LB3D' in the remainder). LB3D is fully computationally steerable using a steering library developed within RealityGrid; this allows users to interact with the simulations from local workstations and laptops while the compute jobs are running on a massively parallel platforms. LB3D was recently awarded a Gold Star rating for its excellent scaling characteristics for large problems (up to  $1024^3$  on 1024 processors of HPCx). The performance of the code has been optimised by EPCC who have also added parallel HDF5 support. Parallel HDF5 is a data format that allows every processor to write to the same file at the same time. It helps us to reduce memory requirements of the code and improves its scalability substantially.

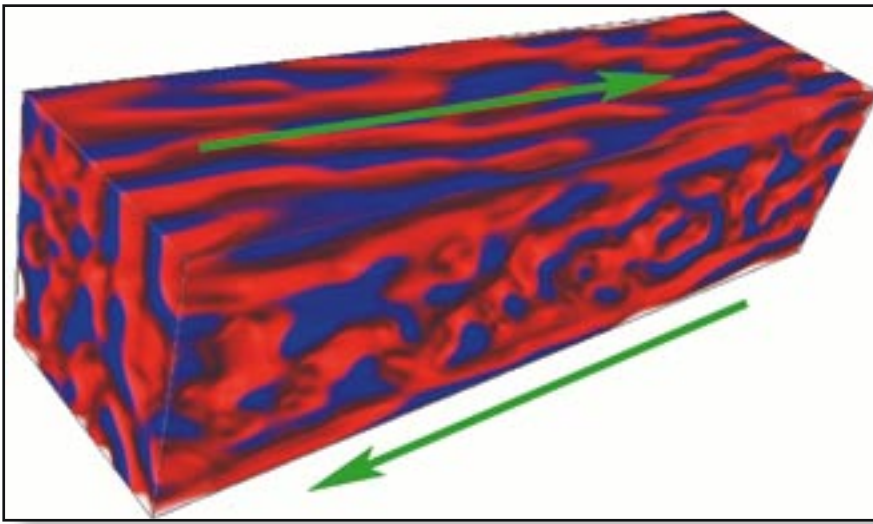
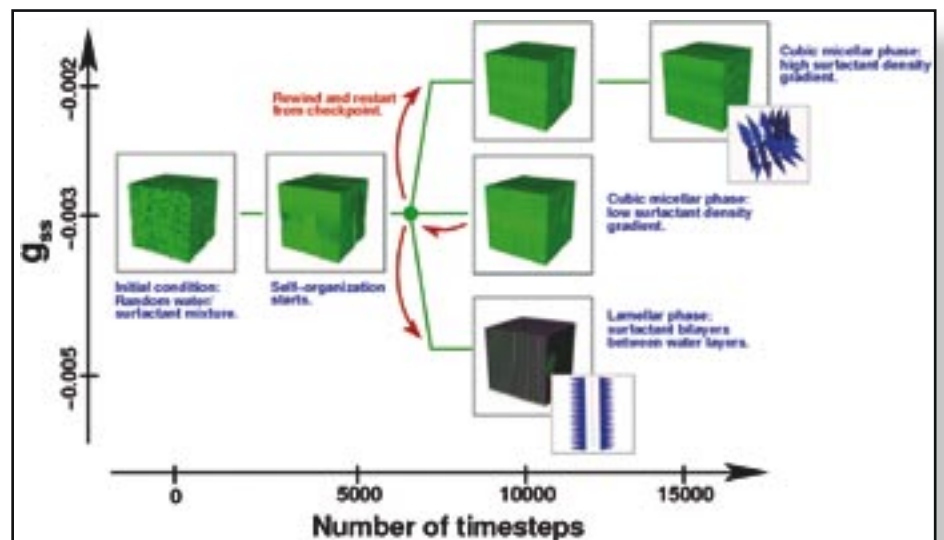


Figure 3. While studying fluid mixtures under shear (using Lees-Edwards boundary conditions, the arrows denoting the relative motion of upper and lower simulation cell boundaries), steering becomes very valuable as a tool to check for the occurrence of finite size effects. The figure shows a snapshot of a simulation of spinodal decomposition of a binary immiscible fluid mixture of size 1024 x 128 x 128.

Figure 2. Example of a parameter search using computational steering. From a random initial condition a binary water-surfactant mixture self-assembles in different final states while the user steers the parameter  $g_{ss}$ . The realisable phases are a cubic micellar phase and a lamellar phase where the surfactant molecules assemble in bilayers [3].



Using LB3D, we have made an extensive study of the equilibrium and non-equilibrium properties of 3D ternary amphiphilic fluids, and uncovered some very novel behaviour, including the self-assembly of the beautiful liquid crystalline gyroid phase (figure 1).

Working with Schlumberger, we are exploiting the scalability of the LB3D code, the existence of binary oil-water invasion data for flow within Bentheimer sandstone, x-ray microtomographic data on this rock sample collected at the ESRE, and magnetic resonance imaging (MRI) data of the flow (at much lower resolution) to perform some very large simulations of fluid invasion using LB3D. These simulations, on unprecedentedly large rock image samples, can now be run effectively using the UK's national supercomputing facilities. Executing our code needs anything upward of 256 processors. Even with the sizeable models we can now simulate, however, we still remain somewhat short of length scales on which direct comparisons can be made between MRI and simulation data.

RealityGrid partners from Manchester Computing's Supercomputing, Visualization and E-Science Group (SVE) have defined an API for computational steering, developed and released a library and some additional tools that implement the API, worked with scientists at UCL, Oxford, Edinburgh

and Loughborough to incorporate computational steering in several application codes, and conducted a number of public demonstrations of computational steering over national, European and trans-Atlantic Grids. SVE has also developed a graphical interface for steering any RealityGrid-compliant application.

These tools have already proved to be of value in working with LB3D, a fact now partly documented in our recent Contemporary Physics article [3]. As an example, the benefits steering offers for large-scale parameter searches are demonstrated in figure 2. In this example we are looking for the most suitable value of the parameter  $g_{ss}$  in the ternary amphiphilic fluid model, which is a coupling constant between two species of particle – in this case describing the strength of the surfactant-surfactant interaction. Starting from a random initial condition, the surfactant begins to self-assemble either into a micellar or lamellar phase. The scientist can change  $g_{ss}$  on the fly and eventually rewind the simulation in order to explore the parameter space for interesting effects. A more detailed description of computational steering in routine use can be found in [3].

Studying binary or ternary fluid mixtures under shear is another topic of current interest. In these simulations, finite size effects are

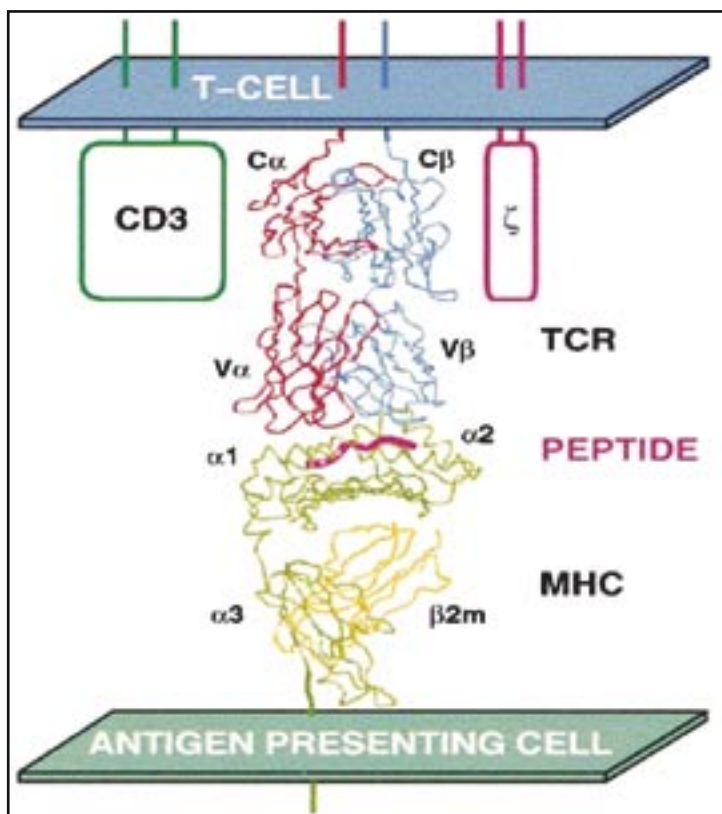
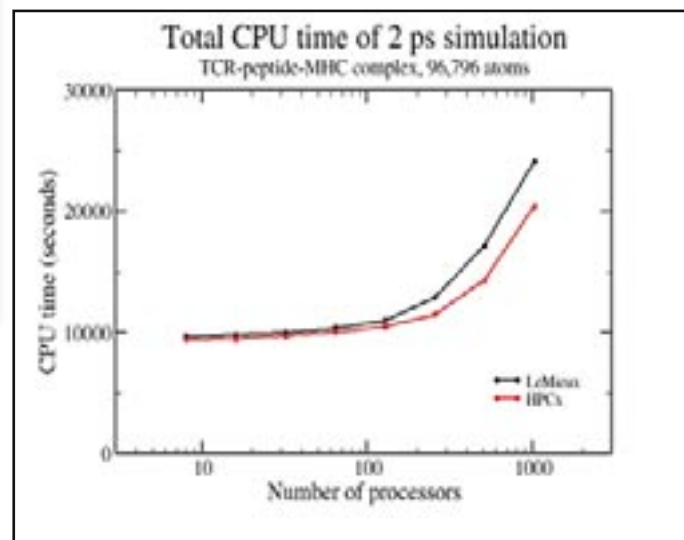


Figure 4. Peptide-MHC binding resembles the binding of drugs to other receptor. We can use molecular dynamics (MD) simulation method to examine/model TCR-p-MHC interaction. See [7].

a problem and force us to simulate large systems which require very substantial amounts of computing power. Steering allows us to interact with these simulations while they are running. By observing the output of an ongoing simulation, we can determine easily when finite size effects start to occur and then stop the simulation. This allows us to minimize the amount of CPU time used and/or disk space used for data storage. Figure 3 shows a snapshot from an LB3D simulation of spinodal decomposition under shear.

The UK has set up a group to explore and promote the linking of the two UK National HPC facilities at HPCx and CSAR with the Extended Terascale Facilities in the USA (comprising various Itanium systems and the Compaq Alpha system Lemieux at Pittsburgh Supercomputer Center) by combining the UK e-Science Level Two Grid and the USA's TeraGrid. The work has been promoted under the auspices of the UK High Performance Computing Strategy Committee and approved by the TeraGrid's Executive Committee. In its first stages, this project will attempt to deploy some RealityGrid applications across this trans-Atlantic grid. Our hope is to perform new science by exploiting the unprecedented access thereby afforded to these vast combined supercomputing resources – we plan to perform computational steering, migration and spawning of multiple LB3D simulations of complex fluid dynamics, utilising our capabilities to stream visualisation data over the Access Grid [4] to remote collaborating sites located within the USA and the UK, at all of which steering of simulations can be actively performed.

Figure 5. Performance of NAMD code when running on HPCx and Lemieux. Ideally, we would expect the ordinate to behave as a constant with respect to variations in the number of processes. The actual behaviour is not ideal, but it does not increase sharply.



We shall be demonstrating these capabilities live at Supercomputing 2003 in Phoenix, Arizona, USA, thanks to special funding from EPSRC and the National Science Foundation (USA). The UK supercomputing facilities at HPCx and CSAR will connect to the TeraGrid via the SuperJanet 4 development network, aka MB-NG (after the Managed Bandwidth - Next Generation Project) to London, then across a 1 Gbps link to Amsterdam (generously donated by BT), via Starlight to Chicago and thence into the TeraGrid backbone.

### Large scale molecular dynamics simulations of TCR-peptide-MHC

T cells are activated upon encounter with another cell that carries on its surface appropriate complexes of major histocompatibility complex (MHC) protein and peptide antigen. The interaction between TCRs (T-cell receptors) expressed by T cells and peptide-MHC (p-MHC) complexes displayed on various cell types is an active area of cellular immunology. Although our structural understanding of T-cell recognition has rapidly evolved due to recent crystallographic results, many of the most fundamental questions still remain elusive. Further molecular description of MHC restricted T-cell recognition will lift us from our current plateau of knowledge to the next level of understanding of the relationship of structure, function and dynamics. Molecular dynamics (MD) simulations can, in principle, provide the fundamental details of such phenomena. Our previous work [5] has shown that only a full atomistic model of p-MHC can accurately predict the binding of the peptide within the MHC

protein's groove; simplified models of only a part of the system, omitting the  $\alpha 3$  and  $\beta_2 m$  fragments, underestimate the binding energy. This work is being performed in collaboration with the Edward Jenner Institute for Vaccine Research.

To study the interaction between TCR and p-MHCs, we have completed a series of MD studies of TCR, p-MHCs and TCR-p-MHC complexes using the NAMD code [6], on HPCx and Lemieux. A full atomistic model, comprising around 100,000 atoms for TCR-p-MHCs (see figure 4), and 60,000 for TCR and p-MHCs, are used to predict the binding of the p-MHCs to the TCR. NAMD is one of the most scalable parallel molecular dynamics codes, with a large number of features which make it attractive to use for biomacromolecular simulations. It also includes non-grid based steering capabilities (a grid-enabled version is under development inside RealityGrid); a plot of its performance on HPCx and Lemieux can be seen in figure 5.

For the TCR-pMHC models, this code qualifies for a Bronze Star rating for capability computing on 256 processors on HPCx (much bigger models would likely receive Silver and Gold stars on 512 and 1024 processors). We are able to perform a 1 nanosecond run in 10 hrs of elapsed wallclock time on HPCx, where capability computing priorities mean that our jobs sit in the queue for only minutes before executing. Progress on tackling large scale MD problems in this area has been much faster than originally anticipated.

## Steering, high performance computing and the grid

We are able to use the UK e-Science Level Two Grid to access various compute and visualization resources, specifically HPCx and CSAR. At present, there are many additional overheads associated with running jobs on this grid, because each application requires its own scripts on each resource on which it may run, a situation which is very far from the ideal scenario, in which access to grid resources is made transparent to the computational scientist. Moreover, for effective utilization of computational steering in the context of high performance computing, it is very important that supercomputing centres evolve their current policies of job scheduling since co-allocation of compute and visualization resources becomes essential. Today, this is possible for small scale simulations which do not run on the grid when turn-around times are short, but for large-scale jobs one needs special arrangements with the resource owners. It is also important that one can make an advanced reservation of these resources for interactive steering at convenient times, ie during normal working hours rather than in the middle of the night.

With the major effort that is being invested worldwide in computational grids, we are confident that these problems will be solved, although our experience to date indicates that truly usable grid middleware will need to be substantially lighter than the leviathans currently being deployed. We believe that such grids will ultimately revolutionize the way high performance computational science is done, because they promise an easy and effective way to access distributed and powerful high-end resources in an optimal way for any given scientific problem.

## References

- [1] [www.realitygrid.org](http://www.realitygrid.org)
- [2] N. González-Segredo and P. V. Coveney, 'Self-assembly of the gyroid cubic mesophase: lattice-Boltzmann simulation' (e-print 2003, [www.arxiv.org/cond-mat/0310390](http://www.arxiv.org/cond-mat/0310390))
- [3] J Chin, J Harting, S Jha, P V Coveney, A R Porter, S Pickles, 'Steering in computational science: mesoscale modelling and simulation.' *Contemporary Physics*, 44, 417-434 (2003)
- [4] [www.accessgrid.org](http://www.accessgrid.org)
- [5] S. Wan, P. V. Coveney and D. Flower, 'Large scale molecular dynamics simulations of HLA-A\*0201 complexed with a tumour-specific antigenic peptide: Can the  $\alpha 3$  and  $\beta 2 m$  domains be neglected?' (preprint, 2003)
- [6] [www.ks.uiuc.edu/Research/namd](http://www.ks.uiuc.edu/Research/namd)
- [7] K C Garcia, et al, *Science* 279, 1166-1172 (1998)

## Further reading

J Chin, J Harting, S Jha, PV Coveney, A R Porter, S Pickles, 'Steering in computational science: mesoscale modelling and simulation.' *Contemporary Physics*, 44, 417-434 (2003)

JM Brooke, PV Coveney, J Harting, S Jha, SM Pickles, RL Pinning, AR Porter, 'Computational Steering in RealityGrid,' *Proceedings of the UK e-Science All Hands Meeting 2003*, pp 885-888; published online at [www.nesc.ac.uk/events/ahm2003/AHMCD](http://www.nesc.ac.uk/events/ahm2003/AHMCD)

N González-Segredo and PV Coveney, 'Coarsening dynamics of ternary amphiphilic fluids and the self-assembly of the gyroid and sponge mesophases: lattice-Boltzmann simulations.' (e-print 2003, [www.arXiv.org/cond-mat/0310768](http://www.arXiv.org/cond-mat/0310768))

PJ Love, M Nekovee, PV Coveney, J Chin, N González-Segredo, and J M R Martin, 'Simulations of Amphiphilic Fluids Using Mesoscale Lattice-Boltzmann and Lattice-Gas Methods', *Computer Physics Communications* 153, Issue 3, 340-358 (2003)