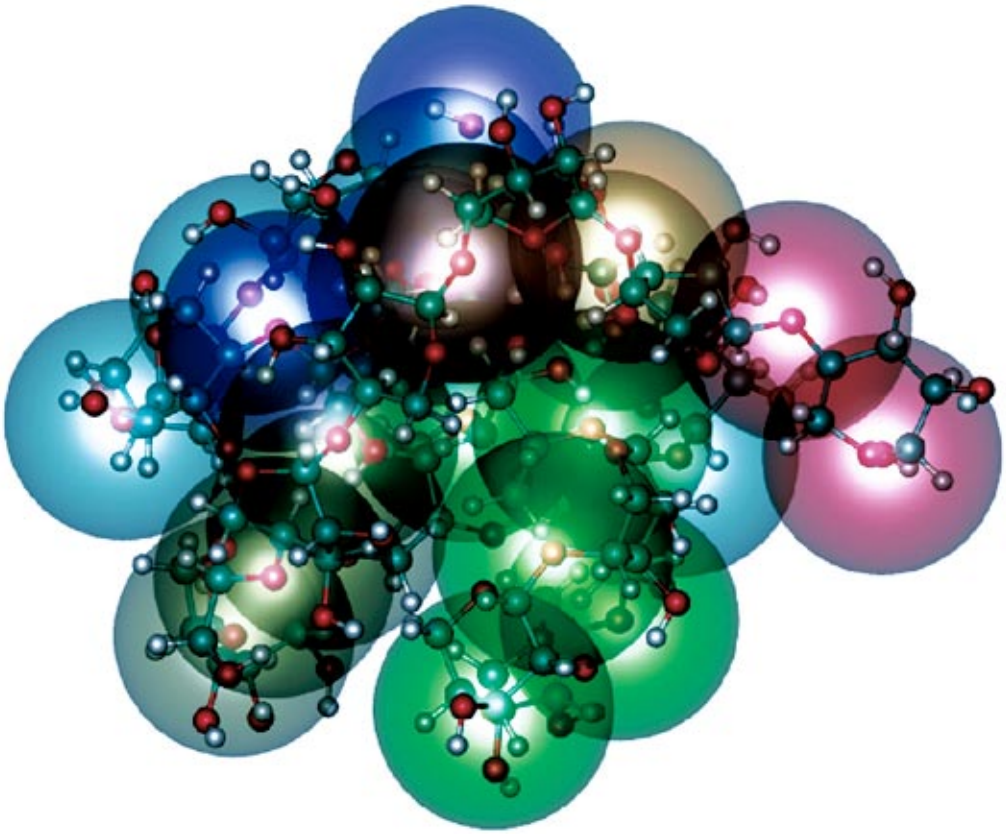
Capability Computing

The newsletter of the HPCx community

[ISSUE 9, SPRING 2007]



Simulations at an atomistic scale

CONTENTS

Editorial

Ilian T. Todorov, CSED, STFC Daresbury Laboratory

Hello and welcome to the 9th edition of Capability Computing. After four and a half years after its launch, HPCx is still the most reliable and most used HPC resource within the UK providing service to over 1000 active users distributed in over 60 projects. The HPCx consortium continues to build up expertise, port and scale up scientific software to meet the demands of scientists as well as increase the power of the service to give an edge to the most demanding simulations. With the arrival of HECToR, we hope that HPCx will also develop in new and distinctive ways, providing even more opportunities for existing consortia and reaching out to welcome new consortia for whom these developments are of the essence.

The theme of this issue reveals the variety of atomistic and molecular modelling carried out on HPCx. Four research papers address issues from molecular growth and self-assembly processes to biochemical insights of a medical condition and the nature of material damage due to irradiation. Although these well encompass modelling science in areas as diverse as physics, chemistry, biology and environment, this issue also presents software engineering contributions with an overview on cost of electronic structure calculations and scalability and performance of molecular dynamics software on processor counts beyond just the few thousands. Finally, we finish with a glance on the glamour of super IT with a review of SC06 from sunny Florida.

Announcing the Fifth DEISA Training Session

Gavin J. Pringle, EPCC, University of Edinburgh, UK

As regular readers of *Capability Computing* will know, HPCx is a member of the Distributed European Infrastructure for Supercomputing Applications, or DEISA.

DEISA seeks to deploy and operate a persistent, production quality, distributed supercomputing environment with continental scope and to enable scientific discovery across a broad spectrum of science and technology. Scientific impact (enabling new science) is the only criterion for success.

The 5th DEISA Training Session will be held at CINECA, Bologna, Italy, in October, 2007. The registration for this event will open early September. Scientists from all European countries and members of industrial organisations involved in high performance computing are all invited to attend.

The purpose of the training is to enable fast development of user skills and know-how needed for the efficient utilisation of the DEISA infrastructure.

The first part of the training will give a global description and introduction to the DEISA infrastructure and will describe the general middleware services, the use of the DEISA Common Production Environment and the detailed utilisation of UNICORE and the DESHL. The topic for the second part of the training will focus on a particular aspect of HPC programming of general interest to DEISA users. Past training sessions have included talks from scientists who have successfully exploited DEISA for their research purposes.

Most academic attendees, currently based outwith Italy, will have their travel expenses reimbursed.

For more information, please visit www.deisa.eu/training

Fig. 1 in-core RI-DFT
parallel performance (wall time/sec)

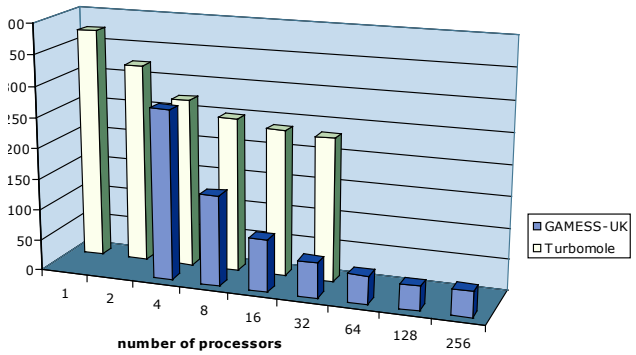


Fig. 2 sequential performance (33 atoms)
wall time/seconds

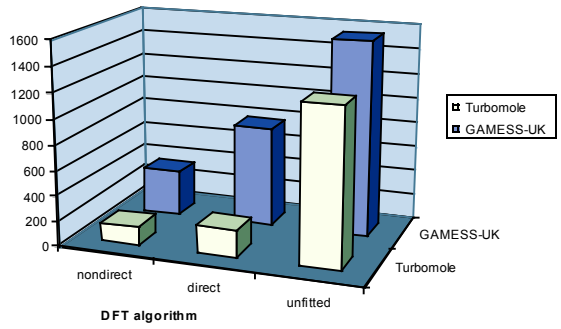
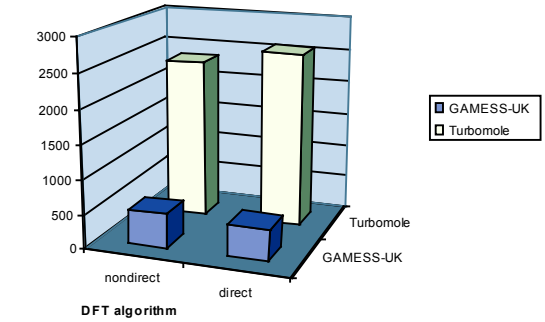


Fig. 3 125 atoms (64 processors)
wall time/seconds



Cost survey for electronic structure calculations

Graham D. Fletcher and M. F. Guest
Computational Chemistry Group, CSE Department,
STFC Daresbury Laboratory

The end-user of computational chemistry technology is presented with a bewildering array of methods and algorithms, implemented in a variety of software and hardware. Faced with this situation, most researchers adhere to a familiar code, remaining unaware of more effective alternatives. The need for an unambiguous comparison of methods is clear, and we begin here with a survey of those ab initio methods that might find first application to a given chemical system.

Hartree-Fock (HF) theory is often regarded as such a method for studying chemical systems at the electronic structure level. However, even for HF the computation of two-electron integrals, which nominally scales as the fourth power of the system size, poses a serious bottleneck for systems with hundreds of atoms. On the other hand, Density Functional Theory (DFT) has emerged as a popular method not only because it recovers electron correlation effects with effort similar to HF, but in certain formulations the quantum mechanical exchange terms can be neglected.

As a result the coulomb interactions can be approximated with three-centre integrals in a technique known as 'coulomb fitting'. This reduces the effort to third order, or less, and allows the possibility of storing the integrals in memory. The method is called RI-DFT1, since it is based on a resolution of the identity. HPCx is of particular interest here, owing to both its large processor count and large memory (2 gigabytes per CPU) in which to hold the integrals.

DFT methodology is now ubiquitous in the world of quantum chemistry, and several packages offer RI-DFT codes that are both scalable and capable of exploiting the available memory to store integrals. Among these, GAMESS-UK (developed by the Computational Chemistry Group (CCG) at Daresbury Laboratory, see http://www.cse.stfc.ac.uk/ccg/software/gamess-uk/) has focussed efforts in the area of parallel scalability, while Turbomole (see http://www.cosmologic.de), a leading commercial package, offers highly optimised methods for computing integrals.

So what is the overall cost? Both packages implement different strategies regarding each calculation step, so the focus here is on the overall 'time-to-solution'. Calculations on silicon oxide clusters (zeolites) were used to benchmark the codes. To assess the benefit of the fully in-core RI-DFT method, it is compared to the situation where the 3-centre integrals are re-computed when needed rather than stored (the so-called direct scheme), and also the conventional, or unfitted, DFT method with fourth-order scaling similar to HF. A sample of these results is given in figures 1-3.

Figure 1 illustrates the performance of the in-core RI-DFT codes with increasing processor counts (for a zeolite with 33 atoms), Figure 2 presents a 'snap-shot' of the code-method space for sequential execution, while Figure 3 presents the in-core and direct RI-DFT methods for a larger zeolite case on 64 processors. Figure 1 shows that the end-user can currently choose between software that minimizes either clock cycles or wall-time. When the former are expensive, Turbomole yields the shortest execution time on a single processor, whereas if computer time is less important than getting the answer as quickly as possible, GAMESS-UK ultimately provides the shortest wall times, but at the cost of greater multiprocessing. For smaller molecules on a single processor (Fig.2) the benefit of the in-core RI-DFT method is obvious, however, for larger molecules and processor partitions the issue of memory bandwidth with large storage arrays can have an impact (Fig. 3).

References

1. M. v. Arnim & R. Ahlrichs, *J. Comp. Chem.*, **19**, 1746 (1998).

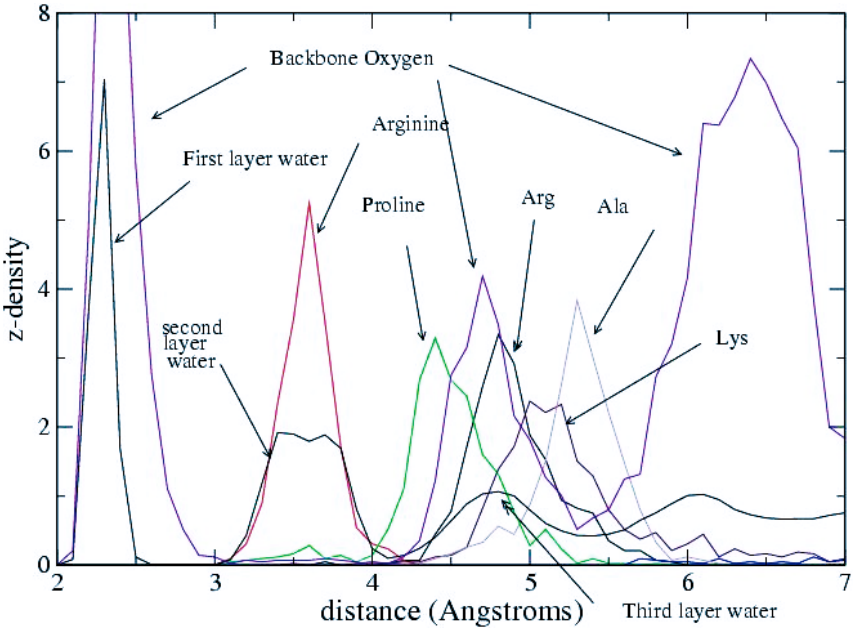


Fig 1: z-density profile for peptide binding to TiO_2 showing the various peptide units relative to the water layers. The TiO_2 surface is on the left.

The control of crystal growth by organic molecules

Colin Freeman, Mingjun Yang, John Harding (University of Sheffield, UK)

Dmytro Antypov, David Cooke, Yue Han, James Elliott, Alan Windle (University of Cambridge, UK)

Jenny Lardge, Maria Sushko, Dorothy Duffy, Mike Gillan, Alex Shluger, Marshall Stoneham (University College London, UK)

Taining Liang, Najib Ouja, Adam Skelton, David Quigley, Mike Allen, Mark Rodger, Tiffany Walsh (University of Warwick, UK)

There is a remarkable growth in our ability to exploit the interface between materials and biology. Bio-material *interfaces* often determine biomaterial *properties*. The development of biosensors and the full exploitation of probe microscopies for chemical resolution need detailed models of bio-material interfaces. Furthermore, the control of biomineralisation on organic scaffolds and templates is a key to the directed self-assembly that is essential if devices based on molecular-scale electronics are to be mass-produced. While recent results demonstrate clearly that proteins can control the assembly of nanoparticles[1], further progress requires an understanding of the specifics of interface recognition.

We are addressing these problems at several levels. First, we are investigating ways of obtaining new, consistent sets of force-fields to model organic-inorganic interfaces. Libraries of force-fields for organic and mineral systems exist; a systematic method of describing the interface between them does not. We have developed force-fields for the interaction of titania and calcium carbonate with water and organic systems and tested them with a range of experimental data and *ab initio* calculations[2].

Using these force-fields, we[3] have performed molecular dynamics simulations of a hexapeptide adsorbed onto the 110 surface of rutile TiO_2 in water. The water close to the mineral surface is highly structured. This structuring both influences, and is influenced by the peptide. The z-density profile (Fig 1) shows that some of the carbonyl oxygens are in the first layer, that the aspartic acid replaces some of the second layer of water removing some of the density and that the rest of the peptide is in the less structured second and third layer.

Calcite shields known as *coccoliths* are produced by unicellular algae. Complex organic molecules are known to be intimately associated with coccolith formation, although much about the mechanisms by which they function are unknown. We have therefore modelled the interactions between simple polysaccharides and several calcite surfaces (including steps and polar surfaces) using MD simulations. The polysaccharides have significantly different interaction energies depending on the type of calcite surface, which suggests that they selectively cover specific surfaces to control the crystal growth. The introduction of acid groups onto polysaccharides greatly changes the adhesion energy, resulting in the polar surfaces having the strongest interaction energies[4].

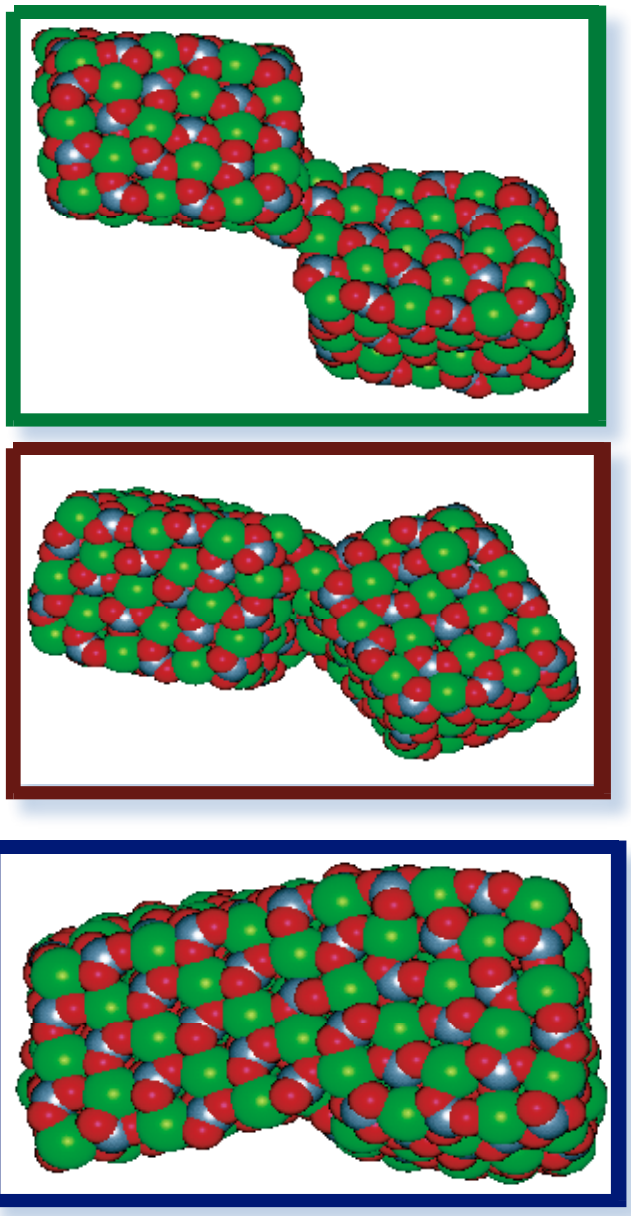


Fig 2: Amalgamation of calcite nanoparticles.

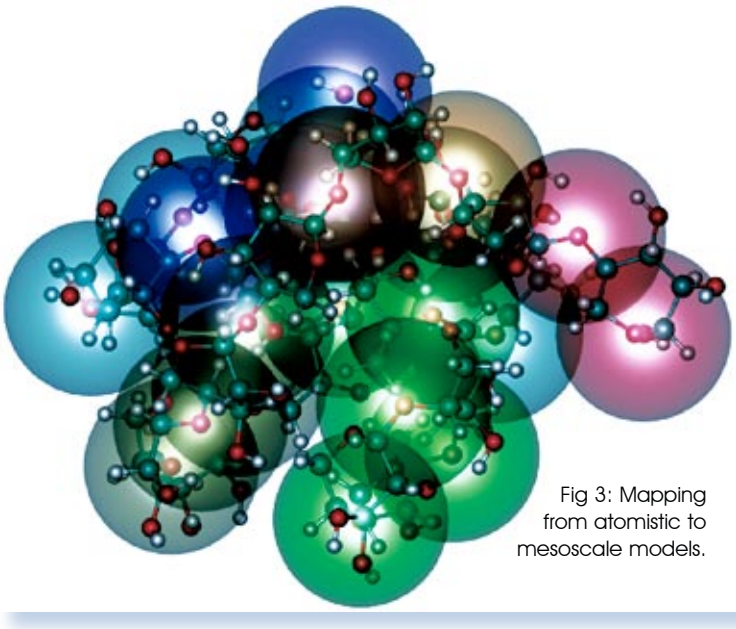


Fig 3: Mapping from atomistic to mesoscale models.

Other mechanisms of crystal growth and control are possible, based on the crystallisation of amorphous calcium carbonate. We have simulated $CaCO_3$ nano-particles of sizes ranging from 18 to 324 formula units, in vacuum and in water. They have shown that breakdown of structural order in the small particles is caused by the rotation of CO_3^{2-} groups on the surface when there is little bulk mineral to stabilise the structure. In the larger particles, the bulk units stabilise the system. We[5] have also begun to model the process of mineral growth via aggregating nano-particles. An example of this behaviour is shown in Figure 2.

The processes of crystal nucleation and growth cannot be modelled by standard molecular dynamics methods alone. We are developing: (i) a range of long timescale methods, based on meta-dynamics and hyper-dynamics approaches[6], (ii) mean field methods using simplified models to understand which are the dominant forces controlling the absorption of large molecules on surfaces[7] and (iii) coarse-graining methods whereby MD simulations are mapped onto a simpler model of the 300 systems that retains the essential physics[8]. Fig 3 shows the basic strategy. This enables us to construct phase diagrams of possible behaviours of the system. Applications include polymer-nanoparticles system and polymer adsorption on flat walls. We are developing coarse-graining models for polysaccharides in water and at mineral surfaces.

Acknowledgements

This work is funded under the EPSRC Materials Modelling Initiative as the consortium *Structural Modelling of the Biological Interface with Materials*. The work with Maria Sushko and Alex Shluger was initially funded by the Interdisciplinary Research Collaboration (IRC) for Nanotechnology.

References

1. S.R. Whaley, D.S. English, E.I. Hu, P.R. Barbara and A.M. Belcher, *Nature*, **405** (2000) 665

2. C.L. Freeman, J.H. Harding, D.J. Cooke, J.A. Elliott, J.S. Lardge, D.M. Duffy; submitted to *J. Phys. Chem. C.*

3. T. Walsh and A. Skelton, unpublished work

4. Collaboration with the experimental group of Susan Stipp, Copenhagen. New experiments are planned to test the predictions of the simulations.

5. D.J. Cooke and J.A. Elliott, unpublished work

6. D. Quigley and P.M. Rodgers, unpublished work.

7. M. Sushko and A.L. Shluger and C. Rivetti, *Langmuir*, **22** (2006). 7678-7688

8. D. Antypov and J.A. Elliott, unpublished work

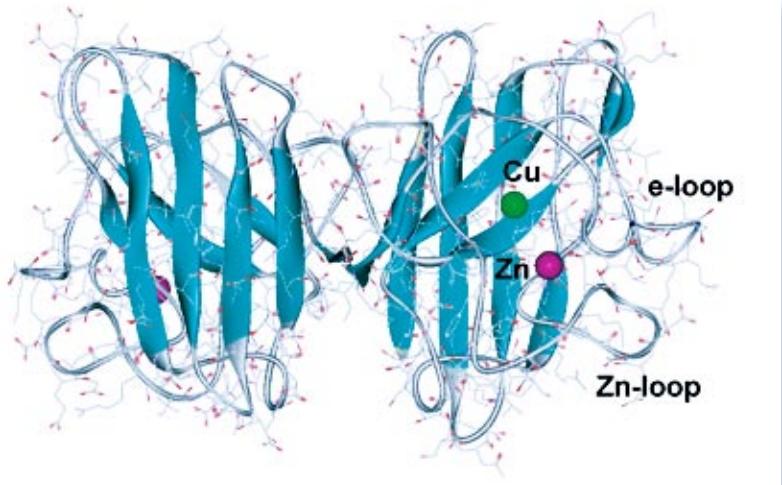


Figure 1.

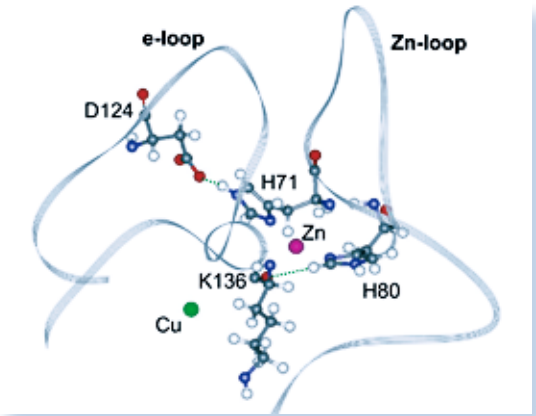


Figure 2a.

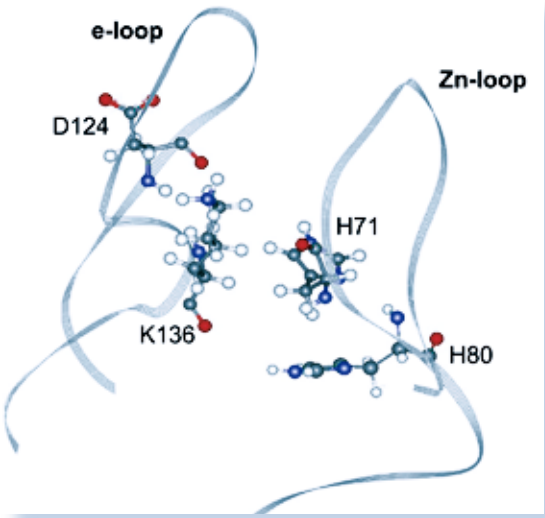


Figure 2b.

Probing a disease-causing protein at an atomistic scale

C. W. Yong, Computational Chemistry Group, STFC Daresbury Laboratory.

Amyotropic lateral sclerosis (ALS) is a neurodegenerative disorder characterised by the destruction of large motor neurons in the spinal cord and brain (lower and upper neurons, respectively). This results in progressive muscle weakness, causing paralysis. It was first discovered by the French neurologist Jean-Martin Charcot in 1869. The disease is known by several names: Lou Gehrig's in the United States, motor neuron disease (MND) in England and maladie de Charcot in France. Among the other neurodegenerative diseases such as Alzheimer's and Creutzfeld-Jakob, MND is considered to be most aggressive whereby a patient typically dies with two to five years of symptom onset. Worst still, the disease etiology is still not clearly known and there is no effective treatment to cure the disease.

However, we do know that a mutated gene called SOD1 that encodes a protein enzyme called copper-zinc superoxide dismutase can cause the genetic form of MND. The SOD1 is an essential metallic protein enzyme that is responsible for the removal of a toxic radical called the superoxide in the body. The enzymes are found in a wide range of aerobic organisms, both prokaryotic and eukaryotic, and are highly homologous across the phyla. In human, it exists as a dimer and orientated about 180° to each other and is often found in red blood cells and neurons. Fig. 1 shows the structure of the metallated protein (the *holo*-SOD1). Note that the amino acid sequences that traced out the two loops, the Zn loop and electrostatic loop, are collectively called the channel loops. This is where the superoxide ions enter in and subsequently react with the copper ion located at the bottom of the channel.

Despite decades of experimental research on the pathogenic nature of these mutated proteins, the aberrant biological roles of these mutations and their precise pathological contributions to the disease are still not well known. The mutation often involves only a single amino acid residue, of which more than a hundred distinct single-site mutations have been identified. An obvious explanation of the disease is dysfunctional enzymatic activity as a result of a mutation. However, it now becomes apparent that it is the change in the molecular structure in the absence of metals which leads to the gain of toxic behaviour via protein misfolding, unfolding and aggregation. To pinpoint these mechanisms, probing techniques at atomic scales would be needed. The protein X-ray crystallographic technique has been used successfully to investigate, down to atomistic levels, the structure of mutants and aggregates over many years, such as efforts carried out by the Molecular Biophysics Department under the direction of Prof. S. Hasnian here at the Daresbury Laboratory. However, the underlying atomic mechanisms leading to protein destabilisation have to be inferred from these static studies in crystalline forms and extrapolated to aqueous, physiological conditions.

To this end, we use the molecular simulation (MD) technique to aid the understanding of the molecular basis of protein destabilisation as a result of loss of metals[1]. With MD it is possible to follow how the SOD1 enzymes change over time and actually see how the mutant proteins are destabilised. The calculations were carried out using DL_POLY_3, a general purpose molecular dynamics (MD) software package developed here at the Daresbury

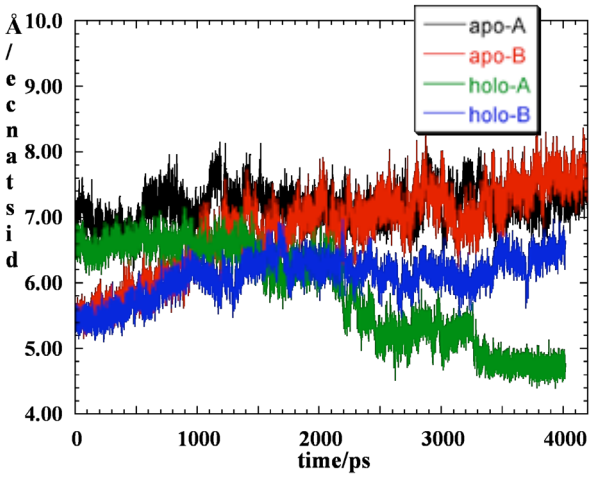


Figure 3.

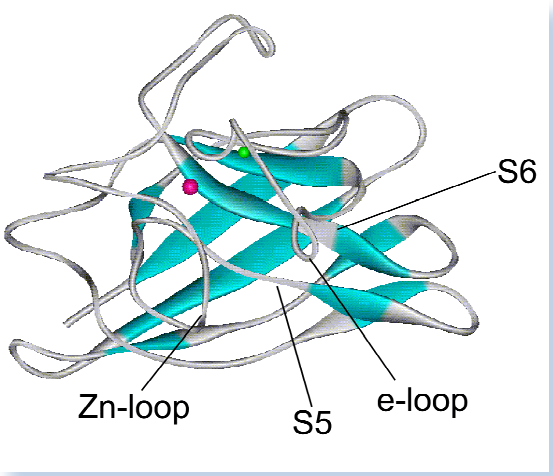


Figure 4a

Fig. 1 Molecular structure of a human SOD1, which consists of 2 identical monomer orientated in an approximate 2-fold symmetry.

Fig. 2(a) Atomic configuration (at 2.3 ns) of channel loops for the *holo*-SOD1 molecule in flat ribbon notation. The green dotted lines show the hydrogen bonds that stabilised the loops.

Fig. 2(b) Atomic configuration of channel loops of one of the monomer for the *apo*-SOD1 molecule in flat ribbon notation. The removal of metal ions result in the destruction of the hydrogen bonds.

Fig. 3 Distance between β-strands 5 and 6.

Fig. 4 Comparison of the one of the SOD1 monomer (a) *holo* molecule, (b) *apo* molecule. The *apo* structure shows widening of the cleft between β-strand 5 and 6, which may further enhance protein interaction to occur.

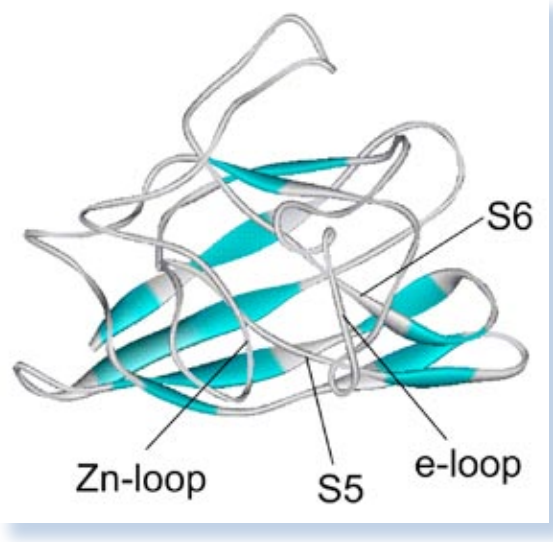


Figure 4b

Laboratory[2]. The software is highly scalable with an efficient parallel computing capability and had obtained a Gold Star rating on the HPCx system.

From experimental observation[3] we know that removal of metal ions (the apo-SOD1) can cause the protein to become destabilised and to undergo self-aggregation, via interaction of the unfolded channel loops with some amino acid residues located at the β-strands 5 and 6. These β-strands together form a deprotected depression and serve as the molecular platform for the non-native protein contacts to occur. However, the channel loop structures of a metal-void protein cannot be resolved by X-ray technique due to their disordered nature. Therefore, the use of MD technique together with the HPCx therefore is crucial to investigate the initial stages of loop disorders and the reason for the β-strands 5 and 6 to become deprotected.

Fig. 2a indicates the formation of some hydrogen bonds that help to stabilise and maintain the structural integrity of the channel loops. However, when the metals were removed, Fig. 2b, these hydrogen bonds were disrupted and subsequently the e-loops and the Zn-loops become more flexible and elongated. MD analysis showed that the channel loops were highly uncorrelated with each other and the loops were moved wide apart, before approaching each other again during the course of the simulation. The distance between β-strands 5 and 6 of both monomers of a SOD1 molecule was also measured. The results are shown in Fig. 3. In the case of *holo*-SOD1, one of the monomers has a smaller gap (at 5 Å) than the other (at 6 Å). However, in the case of *apo*-SOD1, the gaps between the strands were widened for both monomers. This shows that demetallation results in elongation of the loops, and the unintended interaction of these loops may also be further encouraged by the widening of the gaps between the strands. This is illustrated in Fig. 4 where the molecules are shown in ribbon notations.

This work demonstrates the feasibility of using MD simulations to elucidate the underlying atomic mechanisms that may give rise to the toxic nature of the SOD1 that leads to MND. With improved parallelisation techniques (as employed in the DL_POLY_3) and advances in computational capabilities (such as HPCx), long time-scale biomolecular simulations can be achieved in order to provide a better understanding of the structural biology, by providing a more complete picture of the structural evolution.

References

1. C. W. Yong, W. Smith, R. W. Strange and S. S. Hasnain, *Mol. Sim.* **32**, 963 (2006).

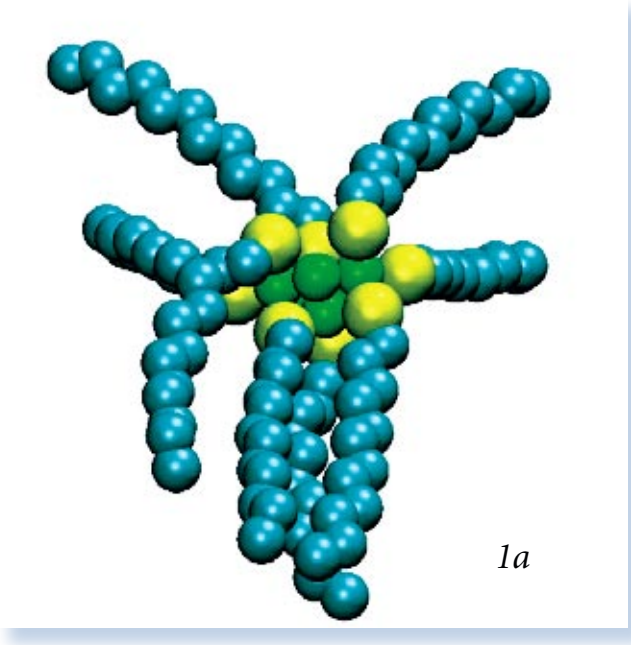
2. I. T. Todorov and W. Smith, Philos. *Trans. R. Soc. London Ser. A* **362**, 1835 (2004).

3. R. W. Strange, et. al. *J. Mol. Biol.* **328**, 877 (2003).

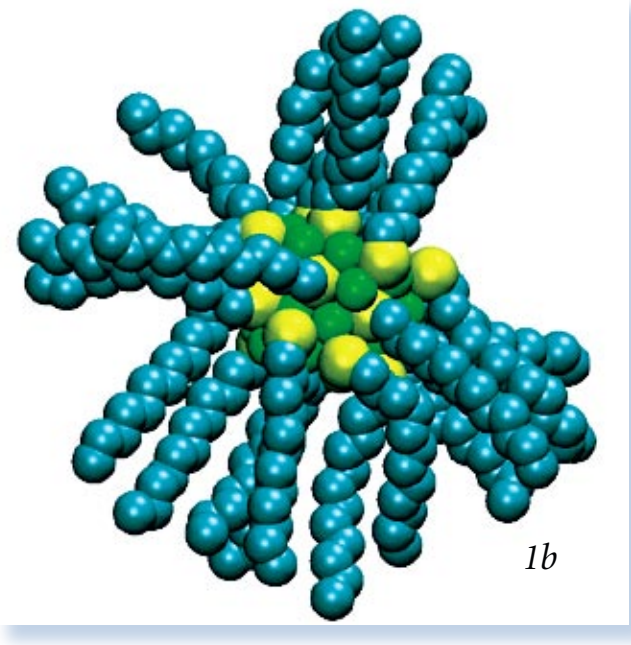
Nano-objects in supercritical media: solvation behaviour of metal nanoparticles in supercritical fluids

M Lal, Centre for Nanoscale Science, The Donnan and Robert Robinson Laboratories, University of Liverpool, UK

M Plummer and W Smith, Computational Science and Engineering, STFC Daresbury Laboratory, UK



1a



1b

The phenomenal surge of interest in nano-objects (metal-atom clusters, fullerenes, carbon nanotubes, nanowires, nanotapes, nanocomposites, mesoporous nanofilms, protein-nanoparticle conjugates etc) is due to their extraordinary optical, electronic, magnetic and conducting properties that differ significantly, and often qualitatively, from those of the corresponding bulk materials. These properties hold the promise of a technological revolution in areas such as optoelectronics, magnetic storage, sensing technologies, catalysis and fuel storage. In addition to structural and chemical characteristics, factors that feature crucially in the determination of the various properties of nano-objects are their size and shape. Therefore, a key requirement for pursuing systematic and accurate studies of these objects, with both scientific and technological objectives in mind, is the availability of samples of narrow size distribution. Recently, *supercritical fluids* have been investigated for their efficacy as fractionating media for nanoparticles and other nano-objects: the tunability of solvent characteristics of supercritical fluids allows the development of promising approaches for the production of monodisperse objects.

Above a certain temperature known as the *critical temperature*, a substance exists in one phase only, no matter how great the pressure applied. Supercritical fluids were found to possess extraordinary solvent properties well over a century ago, but their potential as effective media in processing and extraction technologies was first realized in the late nineteen-seventies. Today, supercritical fluids feature prominently in this role in technologies as diverse as the manufacture of decaffeinated coffee and nicotine-free tobacco, catalysis, synthesis and purification of polymers, extraction of perfumes and flavours, recovery of low-boiling-point chemicals from crude oil residue, reaction engineering, biotechnology and cleaning of electronic components.

The superiority of supercritical fluids over liquids as solvents and suspending media in these applications and others is attributed to their high compressibility, low surface tension, low viscosity and,

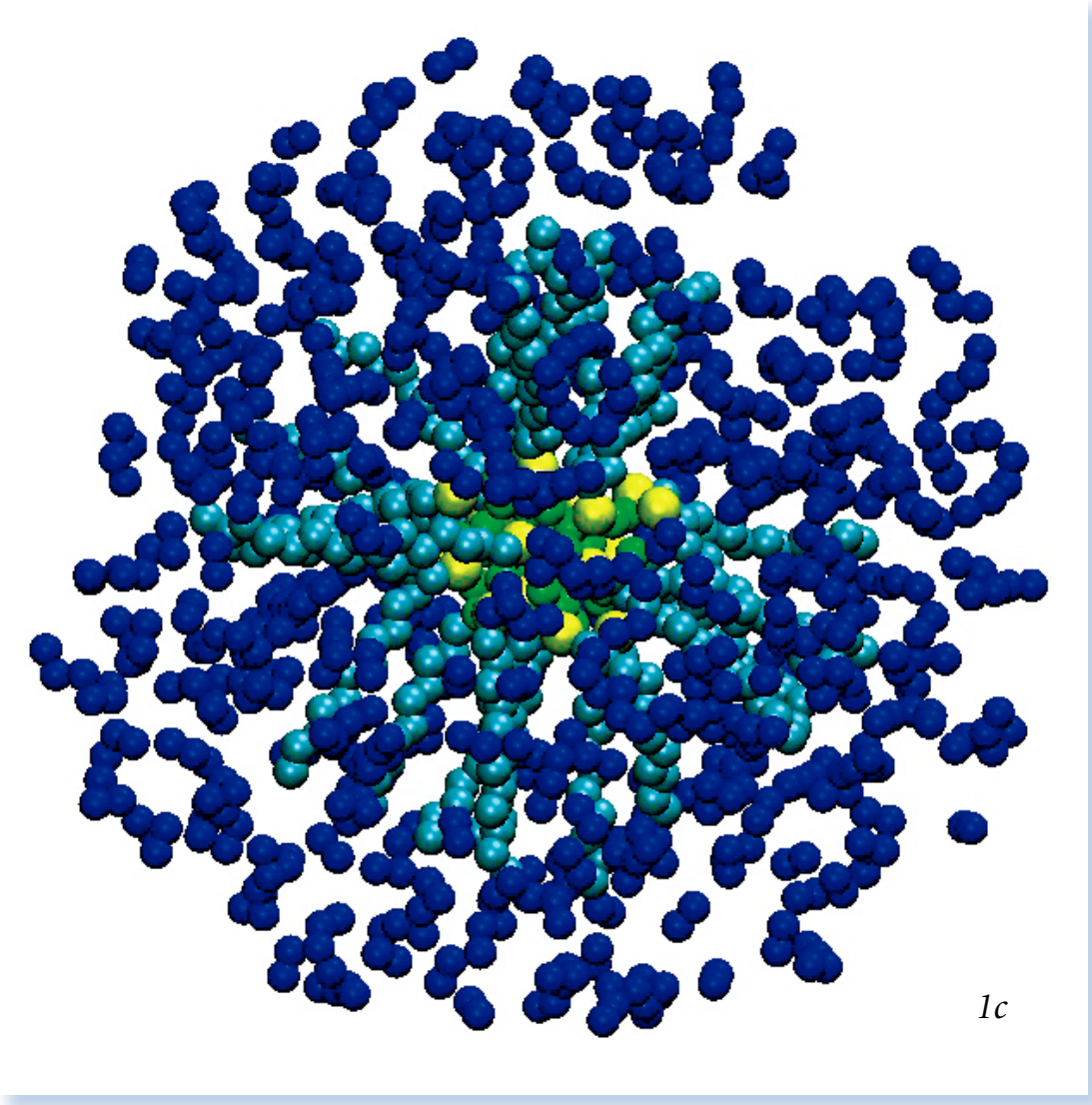


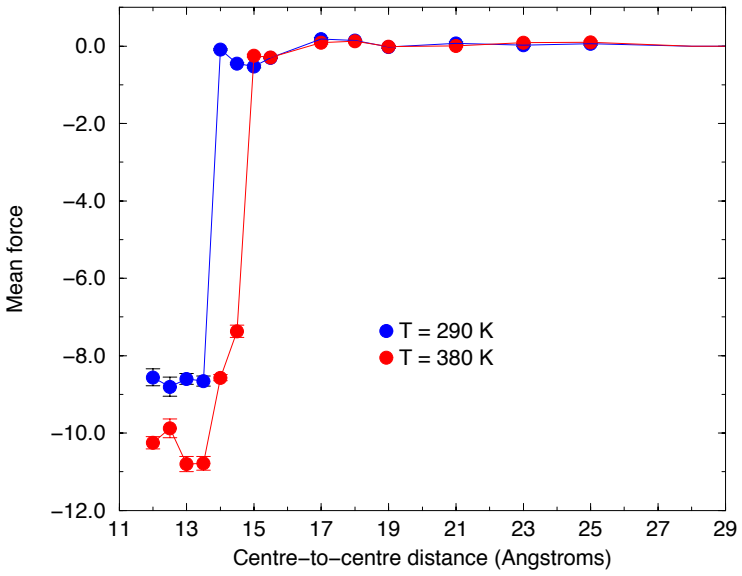
Figure 1:
(a) 8-gold-atom passivated nanoparticle.

(b) 38-gold-atom passivated nanoparticle.

(c) the 38-atom passivated nanoparticle in ethane. The ligands are alkane-thiolates ($C_{12}H_{25}S$).

perhaps most significantly, to their special mode of solvating the particles and solute molecules immersed in them. The focus of our studies in the field resides in this last aspect of the behaviour of supercritical fluids. The supercritical regime is characterised by appreciable inhomogeneity and fluctuations in density, particularly close to the critical point, which for simple systems have been shown theoretically to lead to the *clustering* or *depletion* of the solvent around the solute molecule, depending on the strength of the interaction between a pair of the solvent and the solute molecule relative to that between a pair of the solvent molecules. Enrichment or depletion of the solvent in the immediate surroundings of the solute particles would significantly affect the solvent-mediated interactions between them and hence their propensity to aggregate or remain dispersed in the solvent. Thus solvation has an important part to play in controlling the state of material dissolved/dispersed in supercritical media. This makes the development of molecular-based understanding of particle solvation a highly desirable objective.

Chemical methods developed in recent years yield *passivated* nanoparticles covered with ligands, in particular alkyl thiolates terminally attached to the particle surface through Au-S bonding. We have been engaged on the study of bare and passivated gold nanoparticles in ethane aimed at the elucidation of the mode and the extent of solvation of the particles in the supercritical regime of the solvent. Our published HPCx simulations[1,2] using the DL_POLY_2 code[3] have elucidated detailed solvation behaviour as a function of (sub- to supercritical) temperature at the critical density and as a function of density over a range of supercritical isotherms. The modes of solvation for bare and passivated particles, deduced from the radial distribution of the solvent about the metal-core centre of mass, are found to be qualitatively different from each other: while the molecules solvating the bare particle form a well-defined, two-region layer around it, those solvating the passivated particle are loosely dispersed in the passivating layer.



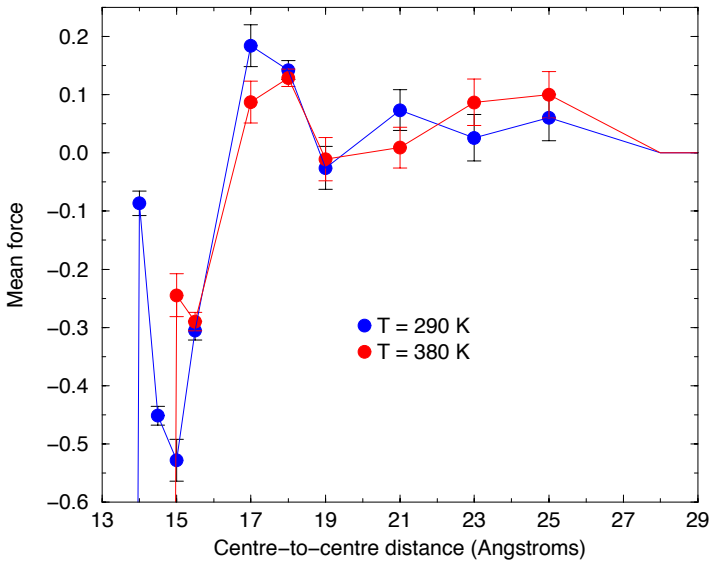
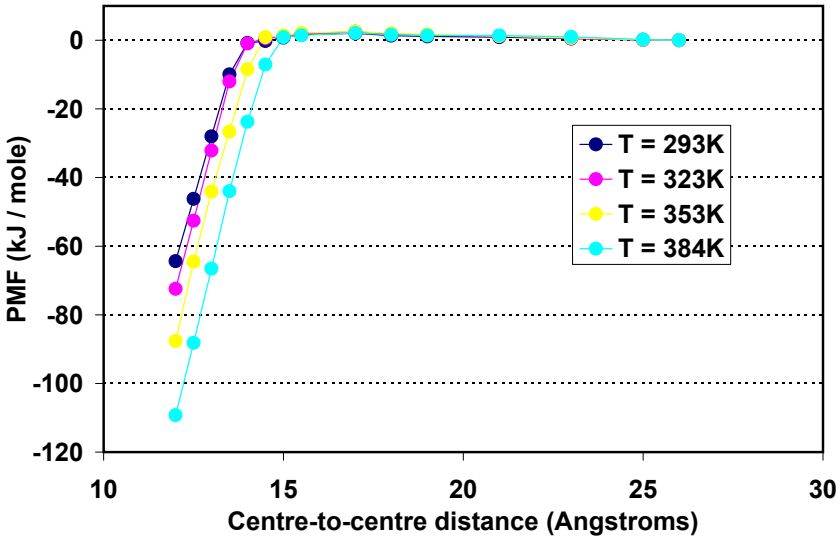


Figure 2 (above): relative mean force calculation for two 38-atom nanoparticles and solvent density ~8 molecules per nm^3 at T = 290K (liquid) and T = 380K (supercritical).

Figure 3 (right): potential of mean force calculation for two 38-atom nanoparticles and solvent density ~8 molecules per nm^3 at various temperatures.



We have recently embarked on the computation of partial molar volumes and potentials of the mean force (pmf) for the nanoparticles in the solvent over similar ranges of temperature and density: some preliminary results obtained for 38-atom bare gold particles of size just under 1nm in supercritical ethane are presented here. Our aim is to provide a tangible link between our simulations and experimental thermodynamics of nanoparticle solutions in supercritical solvents. The pmf may be regarded as the effective potential between nanoparticles in the solvent controlling their propensity to disperse or aggregate. High compressibility of supercritical fluids would provide a convenient means to change the pmf by systematically changing the density of the system through application of moderate pressures at constant temperature. Our objective is to explore how the pmf would depend on the solvent density above the critical temperature and in this way to identify those regions of the supercritical regime in which the solvent would serve as a dispersing medium and those in which it would serve as an aggregating medium.

At large interparticle separations the two particles retain their individual identity with a very weak repulsion between them. As the particles come closer, the surface atoms of the particles jump across giving rise to sintering, a process that drives the particles to merge with each other. With the progressive reduction in separation the fused doublet assumes a well defined dumbbell shape with a shrinking neck, eventually transforming into a cylinder. The onset of sintering is marked by almost an order of magnitude increase in the attractive force between the particles. Eventually the force will become repulsive as the particles become very close. The integration of the mean force with respect to the distance will immediately yield the free energy of sintering.

Limiting partial molar volumes of the nanoparticle have been calculated as a function of the solvent density in a range of isotherms in the supercritical domain. These are found to be *negative*, their magnitude decreasing with the increase in the solvent density in the regime $\rho > 1.5\rho_c$. This means that in this regime the degree of solvation of the particle is *positive* and is a

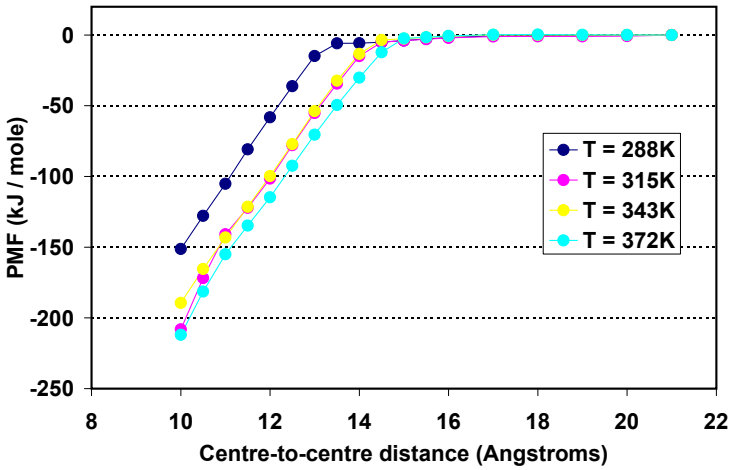
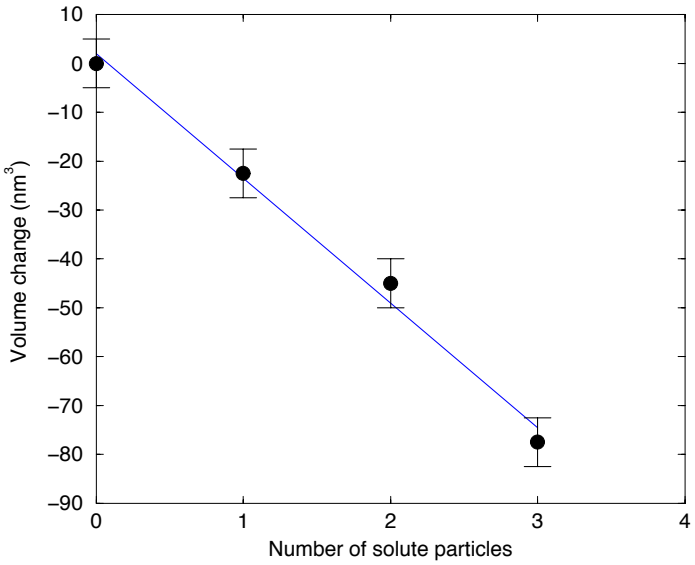


Figure 4: potential of mean force calculation for two 38-atom nanoparticles and solvent density ~4 molecules per nm^3 at various temperatures.



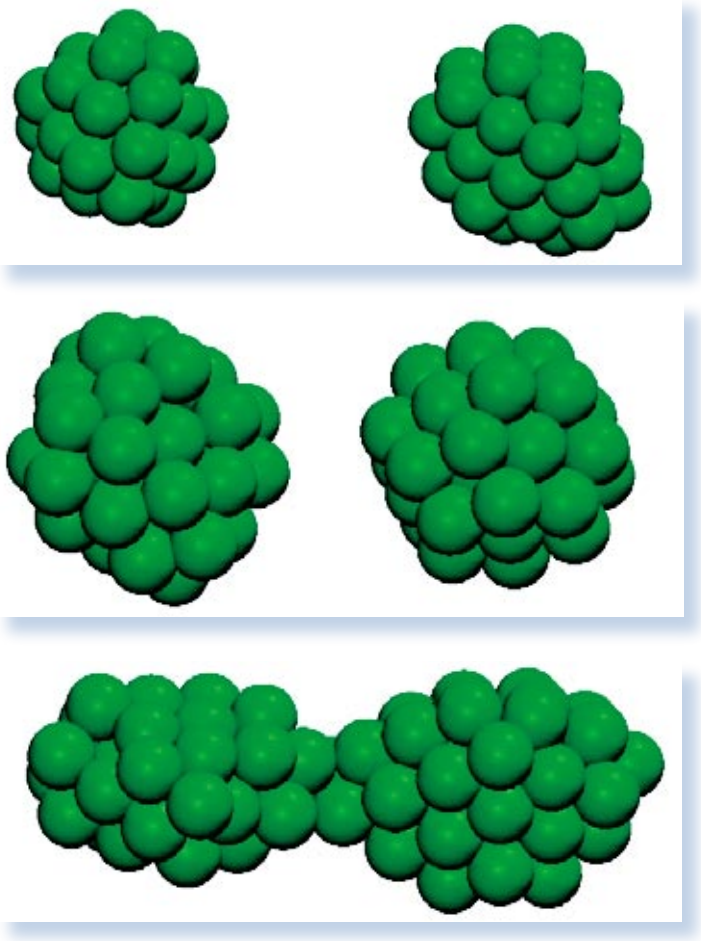


Figure 5 (left): sintering behaviour of the nanoparticles at sample separations 21, 15 and 13 Angstrom and T=290K, solvent density as in figure 2 (solvent molecules not shown).

Figure 6 (above): partial molar volume calculations at T = 310K and p = 0.054 katm. The partial molar volume $(\partial V/\partial n) = -22 \pm 3$ nm^3 per particle.

decreasing function of the solvent density, in excellent agreement with our results on solvation excess[1,2]. The computed values of the change in the volume as a function of the number of particles at constant temperature and pressure enable the determination of the partial molar volume with accuracy of ±3 nm^3 per nanoparticle.

Passivation of nanoparticles enables chemical modification of their surface to prevent sintering, to control the interparticle interactions and to render them adaptable to given solvent conditions. Our main current focus is the extension of the pmf and the partial molar volume computations to nanoparticles passivated with alkyl thiolate molecules anchored at their surface through the sulphur-gold bond. These results will be compared with recent pmf work in which the gold nanoparticle cores are represented as rigid bodies[4], and in which pmfs are calculated in vacuum[5]. We also plan to determine the free energy of dissolution of the nanoparticles in both supercritical and liquid solvents, initially for particles of size up to 2 nm.

References

1. M Lal, M Plummer, N J Richmond and W Smith, *J. Phys. Chem. B* 2004, **108**, 6052-6061; **108**, 10196.

2. M Lal, M Plummer and W Smith, *J. Phys. Chem. B* 2006, **110**, 20879-20888

3. Details of the DL_POLY codes may be found at: http://www.ccp5.ac.uk/DL_POLY/

4. N Patel and S A Ergorov, *J. Chem. Phys.* **126**, 054706, 2007

5. K Tay and F Bresme, *Molecular Simulation* **31**, 515-526, 2005

Computer resources on the HPCx service are provided via our membership of the UK's HPC Materials Chemistry Consortium and funded by EPSRC (portfolio grant EP/D504872).

Figure 1: radiation cascade in rutile TiO_2, created by 100 keV U recoil. The recoil travels from the upper left corner to the bottom right. The box size is 43 nm. Only atoms that have been displaced more than 1 Angstrom are shown.

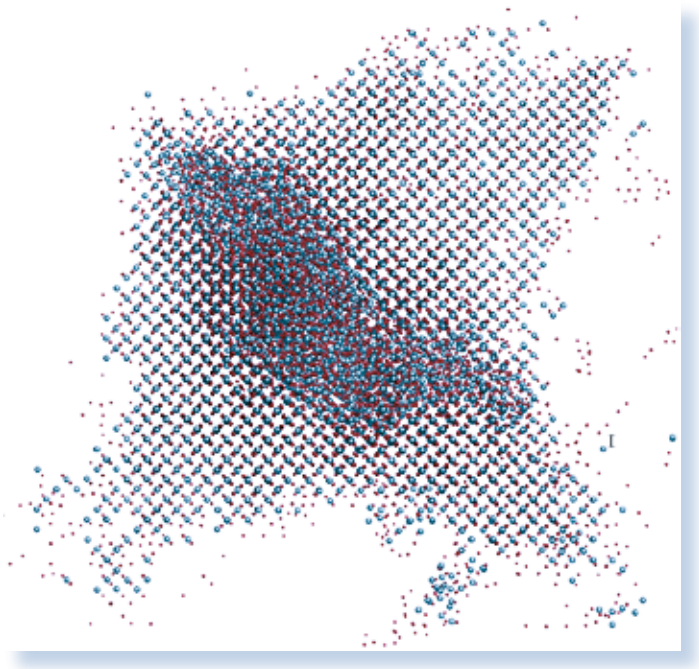


Figure 2. Atoms displaced more than 1 Angstrom in the 40 keV event at 1 ps. We observe large elastic deformation of the lattice around the cascade, which recovers at longer time.

Molecular dynamics simulations of radiation damage events

K. Trachenko[1], I. T. Todorov[1,2], M. T. Dove[1] and W. Smith[2]

1. Department of Earth Sciences, University of Cambridge
2. Computational Science and Engineering Department, STFC Daresbury Laboratory

To safely encapsulate highly radioactive nuclear waste, a number of materials have been proposed which will serve as immobilisation matrices (waste forms) that will prevent radioactive isotopes from leaking into the environment. If a waste becomes amorphised under irradiation by heavy recoils in alpha decay process, it can show increase of transport, which may lead to radioactive isotopes diffusing into the environment. A material resistant to amorphisation, on the other hand, would be free from this negative effect. This leads to the question of what defines the ability of a material to recover from the damage induced by a radioactive decay by recrystallisation. Two related interesting questions are the important time scales of damage recovery in these materials, and how damage recovery during the cascade lifetime of several ps compares to slow diffusive-type processes at the experimental time scale. These questions extend far beyond the area of waste forms, and are important in a wider branch of materials science and technology: modification of materials by ion bombardment.

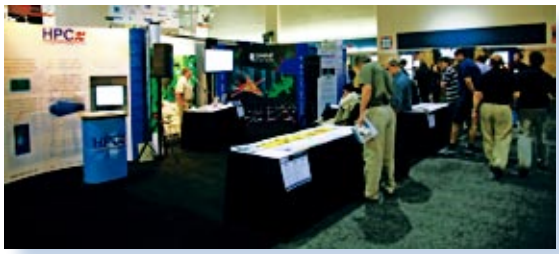
A suitable tool to address these questions is molecular dynamics (MD) simulation. Our approach is to simulate the behaviour that follows the sudden insertion of an energetic particle into the sample, either representing the radioactive decay of an atom or a beam of atoms incident on a sample surface. These simulations model the processes of creation and evolution of cascades of displaced atoms on the scales of nanometers and picoseconds, scales that are not accessible in experiments. Unfortunately, small system size has long been the limiting factor in these simulations. One needs to be able to simulate cascades produced by recoils of realistic experimental energies of about 100 keV. Depending on the material and ion type, the number of atoms in the MD box needed to contain the damage often exceeds the one million atoms benchmark, and when the force field includes electrostatics, a major factor behind the simulation slow-down, standard tools become impractical.

To tackle the issues outlined above, we have developed the DL_POLY_3 MD package. Unlike its predecessor, its parallelisation is based on the domain decomposition (DD) strategy coupled with the linked cells (LC) algorithm for a fast build of the short-ranged interaction lists[1]. The coulombic long-ranged interactions are calculated in a smoothed particle mesh Ewald manner by the use of an advanced three-dimensional fast

SC2006 – the premier international conference on high performance computing

Lorna Smith
EPCC, University of Edinburgh, UK

Pictures by Tim Franks, Computational Science and Engineering Department, STFC Daresbury Laboratory



SC2006 was held in the sunshine state of Florida in Tampa. Held in November, the sun made a welcome change from Britain's rather dismal weather at the latter part of last year.

This year's conference took its inspiration from Albert Einstein: *"Computers are incredibly fast, accurate, and stupid; humans are incredibly slow, inaccurate and brilliant; together they are powerful beyond imagination."* There were all sorts of new and novel systems and applications on offer.

The HPCx team was again well represented at the conference, with both EPCC and Daresbury Laboratory hosting their own booths on the exhibit floor. As usual, our booths were co-located, HPCx given pride of place between the two. It was good to see the UK so well represented: there were also booths from Manchester, The FPGA High Performance Computing Alliance and the UK e-Science programme to name but a few.

The Phase 3 system entered the Top500 for the first time, at a respectable number 43. The various technology refreshes throughout HPCx were designed to ensure UK science had a world leading system throughout the entire project – not just for the first couple of years. While the relationship between Top500 position and the performance of real applications on the system is crude, the fact that HPCx has remained in the top 50 throughout the duration of the project (which started in 2002) reflects the success of this strategy. Had we not had regular technology upgrades the original system (which entered at number 9 in 2002) would now be at around number 360.

In addition to the booths, HPCx staff gave presentations, taught tutorials and participated in birds-of-a-feather sessions. All in all it was a busy conference. In terms of interesting or new technologies, there was quite a lot of talk about multicore systems and FPGAs. The Blue Gene system at LLNL held on to the number 1 spot, with the Cray XT3 system at Sandia slipping in at number 2.

Fourier transform algorithm[2]. The latter fits conveniently within the DD/LC scheme, thus allowing for fast and efficient simulation of systems with millions of ions[1] on any large processor counts.

We have used this code to simulate high-energy radiation events in a number of materials using HPCx[3]. Shown in Figure 1 is the result of simulation of 100 keV U recoil in rutile TiO_2, a potential waste form material. Using a variable time step algorithm[1], the system of 8,014,200 atoms was simulated for 12 hours on 512 and 1024 processors to obtain the trajectory of the damage over 16,000 and 25,000 MD steps, respectively, or over 6 and 25 ps of real time.

We find that resistance to amorphisation by radiation damage is defined by relaxation processes that take place on a very short time scale of several picoseconds, during which most of the damage recovers if it is able to do so[3]. On this time scale, we observe two distinct relaxation processes: reversible elastic deformation around the radiation cascade (see Figure 2), and recovery of the in-cascade damage of high topological disorder. Our results show that the degree of structural damage can profoundly affect activation barriers for damage recovery: the barriers increase as damage increases. Finally, our results allow us to discuss how resistance to amorphisation is related to the details of interatomic potentials and to the nature of the chemical bond[3].

References

1. I. T. Todorov, W. Smith, K. Trachenko and M. T. Dove, *J. Mater. Chem.* **16**, 1911 (2006).

2. I.J. Bush, I.T. Todorov and W. Smith, *Comp. Phys. Commun.*, **175**, 323-329 (2006).

3. K. Trachenko, M. T. Dove, E. Artacho, I. T. Todorov and W. Smith, *Phys. Rev. B* **73**, 174207 (2006).

Report on the Blue Gene Scaling Workshop

Jülich 5th-8th Dec 2006

Ian Bush and Ilian Todorov
Computational Science and Engineering, STFC Daresbury Laboratory, UK

Number of Processors	MD	Link	VDW	Ewald -Short Range	Ewald - Long Range	Other
2048	2.70	0.383	0.333	0.256	1.333	0.389
4096	1.50	0.220	0.163	0.136	0.751	0.230
8192	0.72	0.124	0.083	0.067	0.327	0.123
16384	0.49	0.067	0.043	0.037	0.270	0.069

Table 1: execution times of DL_POLY_3.

Introduction

The portability and performance of DL_POLY_3 on Blue Gene (BG/L) was examined at the Scaling Workshop held at the Forschungszentrum Jülich from 5th-8th December 2006. The aim of the workshop was to provide selected applicants with an opportunity to test the scalability of codes across an 8-rack Blue Gene system (8192 nodes, 16384 processors). DL_POLY_3 is a classical molecular dynamics (MD) package developed at STFC Daresbury Laboratory[1]. It is a very widely used application, with a few thousand licenses being held worldwide, and may be used to study a very wide range of systems due to the flexibility of the force field that it supports. However, the code has never been run before on systems with appreciably more than 1000 processors.

The system chosen for the workshop was a model of radiation damage in a fluoritized zirconium pyrochlore. One of the native gadolinium ions in the system was replaced by uranium, which was then given a velocity consistent with a 100 keV recoil after an alpha decay. Due to the very high velocity of the Uranium ion it is necessary to study very large supercells, and the total system size we use is approximately 14.6 million particles. It should be noted that this is the first attempt to model this system with a realistic recoil, previous work having been carried out at appreciably lower energies[2,3,4]. Since both the required size of the system and the number of timesteps increase with the recoil energy it is only on machines with power comparable to the Jülich BG/L that these calculations may be performed.

DL_POLY_3

DL_POLY_3 is a totally distributed memory code. The scaling of the time to solution depends slightly on the force field employed, but it is always approximately O(N). To achieve both the time and memory scaling a link-cell algorithm[5] is used, which is essentially a domain decomposition method.

The force field used for the simulations is relatively simple, but is not trivial. The various terms can be generalised as:

1. Short range repulsion
2. Van Der Waal's (VDW) attraction
3. Coulomb forces

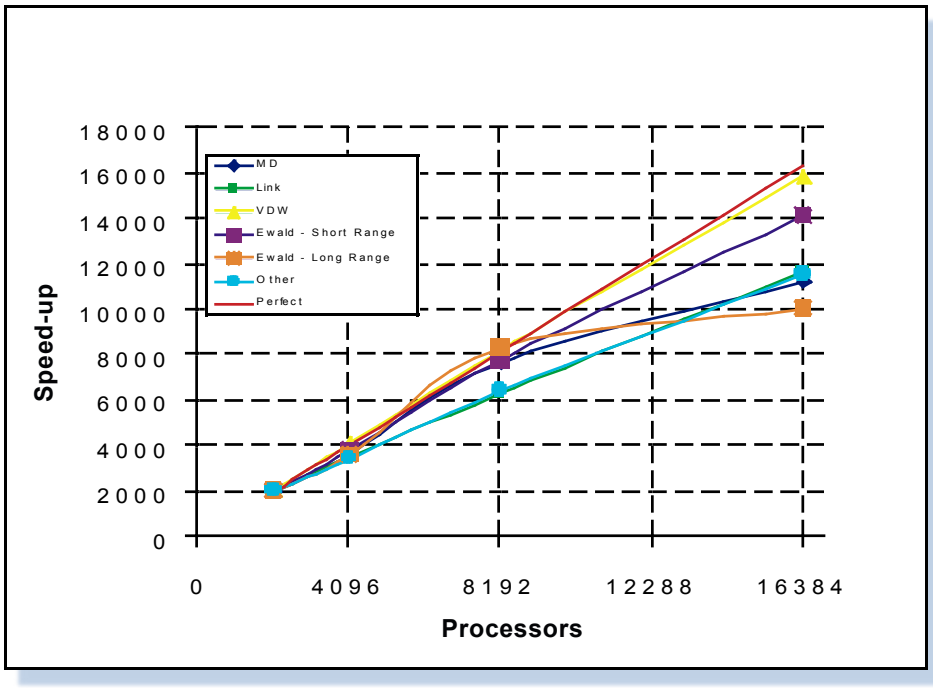
1 and 2 are both short range terms, and are handled together in DL_POLY_3. As such subsequent references to VDW terms should be understood to include both these terms. Due to the short range of these forces they should scale very well with processor count due to their spatial locality (compare halo exchange algorithms). Previous experience suggested that for the test case the scaling of the VDW terms should begin to fall away from perfect only when using the full size of the machine (16384 processors).

On the other hand, Coulomb forces are long range terms, and have to be handled differently. As is standard, the Ewald sum technique is used in DL_POLY_3. This splits the evaluation into two terms – one short ranged, one long ranged. The former can be handled in a very similar way to the VDW terms, and hence is evaluated in real space. The long range term, however, has to be handled differently. DL_POLY_3 uses the Smooth Particle Mesh Ewald (SPME) algorithm[6], the key feature of which is a Fast Fourier Transform (FFT). For this DaFT is used, a package written at Daresbury. This has some features in common with the various volumetric transforms that have been developed, but is novel in that it avoids performing "all-to-all" operations by parallelising the individual 1D FFTs[7].

The use of an FFT implies a considerable amount of communication, so one would expect that the scaling is ultimately controlled by this portion of the code. One of the main reasons for our attendance at the workshop was to examine this.

Results

Once ported, excepting one problem, described below, the code ran and scaled very well "out of the box". The scaling for MD of the test system described above is shown in Table 1. The figure shows speed-up values which are calculated relative to 2048 processors as below that insufficient memory is available to run the simulation.



Strong scaling parallel performance of DL_POLY_3 on the BG/L system. The speed-up of various computional intensive components of DL_POLY_3 is plotted as a function of processor count. The speed-up is taken with respect to 2048 CPU cores.

It can be seen that the scaling for the various elements of the force field is good. The VDW and short range Ewald terms both scale almost perfectly, and at least for VDW terms the expected deviation from ideal behaviour at 16384 processors is not very apparent.

As expected, the long range Ewald terms, i.e. those terms that require an FFT, scale less well. However, given the comparatively small size of the FFT grid, 512*512*512, the scaling is still good.

Table 1 reports the time per timestep for each of the components of the execution. It can be seen that, at these processor counts and for this system, the dominant term is the long range component of the Ewald summation. Whilst not totally unexpected, the margin by which this dominates is surprising, as for smaller systems it was found that the time taken for this term was roughly comparable with that for the VDW and short ranged Ewald terms. The behaviour is probably a reflection of the system size scaling of the long range Ewald term being O(NLog(N)), as compared to O(N) for the other sections.

The most important time in Table 1 is that for an MD timestep on 16384 processors. This is sufficiently small to allow full simulations to be performed in a realistic amount of time, or to put it another way, the code runs fast enough to allow science to be done. This must be the ultimate criterion of performance!

The one major problem that was experienced was I/O. Reading the input file took 10 minutes, and dumping the final results ½ hour.

While these are not too bad, as each has only to be done once, it was found impossible to perform periodic dumping of the atomic coordinates. To give an idea of how bad this problem is: it takes about 4 minutes to perform 500 time steps, and 10 minutes to dump the coordinates. As a full simulation would take over 70,000 timesteps and require coordinate dumping every 500 timesteps it is clear that the total time taken would be prohibitive. As such, in all the above figures the periodic dumping of coordinates has not been included, and the time reported is only for the MD steps, not the initialisation and finalisation.

The reason for this bottleneck is probably in the way I/O is implemented in DL_POLY_3 rather than the Blue Gene I/O subsystem, at least at present. In DL_POLY_3 all I/O is performed:
1. in serial, i.e. all through one processor
2. to/from formatted files.
The reasons that it is done in such a simple way are simplicity and portability of the files, and up until now the time taken for I/O has just not been an issue in the development of the code as the time taken has been small compared to the compute time. However it is now very clear that if such system sizes are to be regularly simulated the I/O performance must be investigated.

Summary

DL_POLY_3 has been shown to scale well out to 16384 processors on BG/L. Porting was very straightforward, and the code scaled very well "out of the box". It was shown that the code runs fast enough on 16384 processors to allow a detailed scientific study of the system were time permitting. The one major problem was I/O.

References

1. Todorov IT and Smith W, 2004, *Phil Trans R Soc Lond, A* **362**, 1835

2. Todorov IT, Purton, JA, Allan, NL, Dove M.T 2006 *J. Phys. Cond. Mat.* **18**, 2217

3. E.g., Trachenko K 2004 *J. Phys. Condens. Matter* **16** R1491 and references therein

4. Trachenko K, Pruneda J M, Artacho E and Dove M T 2005 *Phys. Rev. B* **71** 184104

5. M.R.S. Pinches, D. Tildesley, W. Smith, 1991, *Mol Simulation*, **6**, 51

6. U. Essmann, L. Perera, M.L. Berkowtz, T. Darden, H. Lee and L.G. Pedersen, 1995, *J Chem Phys,* **103**, 8577

7. Bush IJ, Todorov IT, Smith W, *Comp. Phys. Commun.*, **175**, 323

HPC-Europa

EC-funded collaborative research visits for computational scientists

See what we can offer you:

- Access to the best high performance computing systems in Europe
- Technical support and consultancy
- Scientific collaboration with a host researcher
- Travel and living expenses
- Logistical and administrative support
- Minimal paperwork
- Open to scientists in all disciplines, of all levels, who work in the EU and Associated States.

Next HPC-EUROPA deadline for applications:
31 August 2007 for visits taking place between October 2007 and December 200

See http://www.hpc-europa.org/ta.html for more information, eligibility conditions and online application form.

Current success rate for applications: around 65%

UK scientists can apply to one of the Transnational Access centres: CINECA (Italy), BSC (Spain), HLRS (Germany), IDRIS (France), SARA (The Netherlands).

PIs & PhD supervisors: APPLY NOW

High End Computing Training Centre

- Combine a PhD in Physics, Astronomy, Chemistry or Engineering from any UK university with a one-year Masters in High Performance Computing from the University of Edinburgh
- EPSRC and STFC offer ways to extend a PhD studentship to four years to incorporate the training in High End Computing. Funding for MSc fees is also available for students funded from DTAs.

www.epcc.ed.ac.uk/hec

Students: APPLY NOW

MSc in High Performance Computing

Limited places are still available for entry in September 2007.

The MSc is recognised by SAAS under their Postgraduate Student Allowance Scheme, and funded places are therefore available for Scottish students and EU students studying in Scotland. EPSRC studentships are also available.

www.epcc.ed.ac.uk/msc

Forthcoming Events

22nd International Supercomputing Conference ISC'07
26–29 June 2007, Dresden, Germany
http://www.isc07.org/

13th Annual Meeting of ScicomP and the Summer 2007 Meeting of SP-XXL
16–20 July 2007, Computer Center Garching (RZG), Garching, Germany
http://scicomp.rzg.mpg.de/

Euro-Par 2007 Conference
28–31 Aug 2007, Rennes, France
http://europar2007.irisa.fr/

Parallel Computing 2007
4–7 Sep 2007, Forschungszentrum Jüelich and RWTH Aachen University, Germany
http://www.fz-juelich.de/conference/parco2007/

CCP2007 Conference on Computational Physics
5–8 Sep 2007, Brussels, Belgium
http://ccp2007.ulb.ac.be/

UK e-Science 2007 All Hands Meeting
10–13 Sep 2007, Nottingham, UK
http://www.allhands.org.uk/

14th European PVM/MPI Users' Group conference, EuroPVMMPI 2007
30 Sep–3 Oct 2007, Paris, France
http://pvmmpi07.lri.fr/

Supercomputing 2007: the International Conference for High-Performance Computing, Networks, Storage and Analysis
10–16 Nov 2007, Reno, Nevada, USA
http://sc07.supercomputing.org/

5th HPCx Annual Seminar
26 Nov 2007

18th Machine Evaluation Workshop
27–29 Nov 2007, STFC Daresbury Laboratory, UK
http://www.hpcx.ac.uk/about/events/annual2007/
http://www.cse.stfc.ac.uk/disco/mew18/