

Patient-specific simulation as a basis for clinical decision-making

BY S. KASHIF SADIQ¹, MARCO D. MAZZEO¹, STEFAN J. ZASADA¹,
STEVEN MANOS¹, ILEANA STOICA¹, CATHERINE V. GALE¹, SIMON
J. WATSON², PAUL KELLAM², STEFAN BREW³ AND PETER V. COVENEY^{1,*}

¹*Centre for Computational Science, Department of Chemistry,
University College London, London WC1H 0AJ, UK*

²*Centre for Virology, Windeyer Institute of Medical Science,
University College London, London W1T 4JF, UK*

³*The National Hospital for Neurology and Neurosurgery,
Queen's Square, London WC1N 3BG, UK*

Patient-specific medical simulation holds the promise of determining tailored medical treatment based on the characteristics of an individual patient (for example, using a genotypic assay of a sequence of DNA). Decision-support systems based on patient-specific simulation can potentially revolutionize the way that clinicians plan courses of treatment for various conditions, ranging from viral infections to arterial abnormalities. Basing medical decisions on the results of simulations that use models derived from data specific to the patient in question means that the effectiveness of a range of potential treatments can be assessed before they are actually administered, preventing the patient from experiencing unnecessary or ineffective treatments. We illustrate the potential for patient-specific simulation by first discussing the scale of the evolving international grid infrastructure that is now available to underpin such applications. We then consider two case studies, one concerned with the treatment of patients with HIV/AIDS and the other addressing neuropathologies associated with the intracranial vasculature. Such patient-specific medical simulations require access to both appropriate patient data and the computational resources on which to perform potentially very large simulations. Computational infrastructure providers need to furnish access to a wide range of different types of resource, typically made available through heterogeneous computational grids, and to institute policies that facilitate the performance of patient-specific simulations on those resources. To support these kinds of simulations, where life and death decisions are being made, computational resource providers must give urgent priority to such jobs, for example by allowing them to pre-empt the queue on a machine and run straight away. We describe systems that enable such priority computing.

Keywords: patient specific; HIV; resistance; binding; blood flow; urgent computing

* Author for correspondence (p.v.coveney@ucl.ac.uk).

One contribution of 12 to a Theme Issue 'The virtual physiological human: building a framework for computational biomedicine I'.

1. Introduction

The exponential increase of computational power, facilitated by the advance of stand-alone ‘island’ high-performance computational (HPC) resources, has allowed the proliferation of new research in many fields. The development of heterogeneous supercomputing grid technology and infrastructure has provided a ‘step jump’ relative to single island resources for a wide range of problems. Not only is such computation advantageous from widespread research perspectives, it can now be used to address, in real time, the outcome of many diverse emergency situations, such as the development and impact of hurricanes and earthquakes (Manos *et al.* 2008).

In the biological domain, high-performance computation has been used in a research capacity to investigate the interactions at many spatio-temporal scales and extensively with regard to biomolecular function and its interference due to disease. These recent advances in computing mean that for the first time deductive computational modelling can be envisaged in more than a scientific research capacity so far as biomedicine is concerned. It is our claim in this paper that the application of large-scale computation to offer real-time support for clinical decision-making is becoming feasible. Furthermore, the ability to use the distinctive biomedical data available from a patient to optimize patient-specific treatment using simulation means that, in future, the effectiveness of a range of potential treatments could be assessed before they are actually administered, preventing the patient from experiencing unnecessary or ineffective treatments. This will provide a substantial benefit to the medical world and hence to the quality of life of human beings. Additionally, in the context of research and development, it will reduce both the bench to bedside time for drug development and the need for animal testing. We define patient-specific simulation as a deductive computational simulation that uses biomedical data, unique to a given patient, to determine any distinctive biophysical or biochemical property relating to that patient.

In this paper, we discuss the motivations for conducting patient-specific computational simulations that confer real-time clinical advantage, as well as the technological infrastructure that needs to be in place to realize such a goal. To illustrate the diversity of this new paradigm, we outline two very different case studies by means of which we discuss the use of the patient-specific approach and the limitations of current computational infrastructure. We do not pretend to provide an exhaustive overview. In these case studies, we have made inroads that furnish the opportunity to enhance clinical treatment, directly following diagnosis: first in the antiretroviral inhibitor (ARV) treatment of HIV/AIDS and second in the identification and subsequent treatment of abnormal blood flow in the brain. While grid technology allows such novel studies to be performed, several issues arise with regard to the optimal administration of computational resources. In the light of the prioritization and ‘urgent’ status that some computations require in order to meet the needs of the clinical environment, we discuss the ramifications of urgent computing for the evolution of grid technology and high-performance computing.

2. The basis for patient-specific simulation approaches

Traditional medical practice requires a physician to use knowledge, judgement and experience to decide on the course of treatment best suited for an individual patient's condition. While the training and experience of the physician hones an ability to decide the most effective treatment for a particular ailment from the range available, this decision-making process often does not take into account all the data potentially available. For example, in the treatment of HIV/AIDS, the complex variation inherent within the data generated by the analysis of viral genotype resulting in a prediction of phenotype (in terms of viral sensitivity to a number of treatments) makes the selection of treatment for a particular patient based on these predictions fairly subjective.

One step up from the physician as the sole decision-maker is the use of so-called expert systems to predict a suitable course of treatment for a patient. Often such systems are developed using knowledge engineering methodologies or data mining techniques predicated on statistical inference. These use induction to generate rules for a given situation based on the data present in one or more relational databases. Further support for clinical decision-making is likely to become available from the emerging field of patient-specific medical simulation, being pursued by the Virtual Physiological Human initiative (Viceconti *et al.* 2007) and the International Physiome Project (Bassingthwaight 2000) among others.

Patient-specific medical simulation holds the promise of evaluating tailored medical treatment based on the particular characteristics of an individual patient and/or an associated pathogen. Approaches using simulation are based on the development of theories and models from which deductions can be made, as is the standard approach in the physical sciences and engineering. In reality, biology and medicine are still too poorly understood for deductive approaches to replace inductive ones so, in the foreseeable future, both will continue to sit side by side (Coveney & Fowler 2005).

Inductive decision-support systems that incorporate deductive patient-specific simulation therefore have the potential to revolutionize the way that clinicians plan courses of treatment for various conditions. While the details vary widely between medical conditions, several basic elements are common to all fields of patient-specific medical simulation in support of clinical decision-making. Data are obtained from the patient concerned, for example from an angiography scan or genotypic assay, which are used to construct a computational model. This model is then used to perform a complex workflow of simulations of a proposed course of treatment, for example molecular dynamics simulations of drugs interacting with a range of viral proteins, and the results of the simulations are then interpreted to assess the efficacy of the treatment under consideration. The use of simulation to assess a range of possible treatments based on the data derived from the patient who is to be treated will give the physician the ability to choose a treatment based on prior knowledge of how the patient will respond to it.

Such patient-specific medical simulations require access to both appropriate patient data and the infrastructure on which to perform potentially very large numbers of complex and demanding simulations. Computational infrastructure providers need to furnish access to a wide range of different types of resources, typically made available through a computational grid, and to institute policies that facilitate the performance of patient-specific simulations on those resources.

In order to make patient-specific simulations useful to a physician, results need to be obtained within a clinically useful time frame that could range from instantaneously to a few weeks. In addition to the expediency of access to patient data, consideration must also be given to policy and procedures that will ensure the maintenance of patient confidentiality.

3. Infrastructure and technology required for patient-specific simulation

For applications such as the patient-specific technologies discussed in this paper, in which a number of computationally intensive simulations may need to be performed routinely for thousands of patients, exploiting the power of large-scale grid computing will be essential. We define grid computing (Coveney 2005) as distributed computing conducted transparently across the multiple administrative domains.

For such an enterprise to succeed, grid computing will need to focus not only on the provision of large island compute machines but also on the performance characteristics of the networks connecting them. The process of clinical decision-making, requiring access to relevant data, timely availability of computational results, visualization, data storage, etc., needs an infrastructure that can facilitate the transfer of many gigabytes of data within clinically relevant time frames.

Clinicians who require interactive access to machines, for example for steering and visualization, as well as cross-site applications, also need to be able to schedule time on specific resources—both compute and networking—as well as tools to easily co-reserve them, so that the required resources are available when needed. This in turn leads to a demand on resource providers to implement policies and tools that allow such reservations to be made as and when required, so that such methodologies can be incorporated into a user's normal research activities, rather than just providing such facilities on an ad hoc basis. Moreover, the resources provided by a single grid may not always be sufficiently powerful or appropriate to run large-scale distributed models, and resources provided by multiple grids have to be federated in order for a particular investigation to be conducted. This compounds the problems of interactive resource use; each grid has its own policies and systems for making advanced reservations, if it has any at all. Additionally, the high-performance network provision between grids may also be limited or non-existent. Nevertheless, such obstacles must be overcome to make efficient use of the available federated systems.

(a) *Resource reservation and urgent computing*

The on-demand availability of networks, high-performance computers, data storage facilities and visualization engines in this 'urgent computing' scenario is imperative if the tools described in this paper are to be used in the day-to-day clinical decision-making process. Grid applications such as the Highly Available Resource Co-allocator (HARC; MacLaren 2007) and Grid Universal Remote (Yoshimoto et al. 2005) allow the co-allocation of separate resources to achieve these grid-based simulations when needed, so that for example patient-specific medical simulations can be booked in advance much like the scheduling of pathology laboratory tests. On the other hand, infrastructures such as Special

PRiority and Urgent Computing Environment (SPRUCE; Beckman *et al.* 2007) support urgent computing requirements such that time-critical simulations can be run quickly.

HARC is an open-source system that allows users to reserve multiple distributed resources in a single step. These resources can be of different types, including multiprocessor machines and visualization engines, dedicated network connections, storage, the use of a scientific or clinical instrument, and so on. HARC can be used to allocate resources for use at the same time, for example, within a scenario in which a clinical instrument is transferring data over a high-speed network link to remote computational resources for a real-time processing. It can also be used to reserve resources at different times, for the scheduling of workflow applications. We envisage clinical scenarios where patient-specific simulations can be timetabled in advance, via the booking of an instrument, the reservation of network links and storage facilities, followed by high-end compute resources to process data, and finally the use of visualization facilities to interpret the data for critical clinical decisions to be made.

Currently, HARC can be used to book the computing resources and lightpaths across networks based on Generalized Multiprotocol Label Switching with simple topologies, such as a dedicated direct optical link between two geographically separated computing resources. HARC is also designed to be extensible, so new types of resources can be easily added; it is this that differentiates HARC from other co-allocation solutions. There are multiple deployments of HARC in use today, including the US TeraGrid, the EnLIGHTened testbed in the United States, the regional North West Grid in England and the National Grid Service (NGS) in the UK. We use HARC on a regular basis to make single and multiple machine reservations, within which we are able to run numerous applications including HemeLB (see §5*a*). The availability of resources for advance reservation is dependent on individual site policy and site usage. For example, some sites limit user batch jobs to a run time of 24 hours, which roughly corresponds to the time a user would have to wait for an advance reservation slot to become available. The recent version 2 release of HARC also has a feature that allows users to find the ‘next available’ reservation slot. In the case where there is a resource shortage, a different approach needs to be taken.

SPRUCE (Beckman *et al.* 2007) is an urgent computing platform that has been developed to address the growing number of problem domains where critical decisions must be made quickly with the aid of large-scale computation. SPRUCE uses simple authentication mechanisms by means of transferable ‘right-of-way’ tokens. These tokens allow selected users to begin an urgent computing session on a set of pre-defined resources, where they can request an elevated priority for jobs. The computations can be run at different levels of urgency; for example, as a ‘next-to-run’ priority, such that the computation is run once a current job on the machine completes, or ‘run immediately’, such that the existing jobs on the system are removed, making way for ‘emergency’ computation in a pre-emptive fashion, the most extreme form of urgent computing. In the context of the cerebral blood flow study, the HemeLB code (§5*a*) has been used with SPRUCE in a next-to-run fashion on the large-scale Lonestar cluster at the Texas Advanced Computing Center (TACC), and demonstrated live on the show floor at Supercomputing 2007, where real-time visualization and steering were used to control HemeLB within an urgent computing session.

(b) Grid middleware

Within the last 5 years, much work has been done to enable the transparent sharing of HPC resources via grids (Coveney 2005). The goal of grid computing is to give users transparent, uniform access to resources including such high-performance computing platforms owned and operated by disparate organizations that may have different security and access policies at an institutional level, while at the same time giving reasonable security assurances to the people and institutions making use of the grid.

Fundamental to allowing the inter-institutional sharing of resources in a grid is the middleware that enables resources to be accessed in a consistent and transparent way, essential to support the high volume of compute intensive jobs that need to be run as part of patient-specific medical simulation. Several different middleware technologies exist to facilitate the sharing of resources, the most widespread being Globus (Foster 2006), gLite (Middleware, <http://glite.web.cern.ch/glite>) and UNICORE (Almond & Snelling 1999). Many of these software stacks are focused at the server-side resource providers, and take little account of user needs.

The Application Hosting Environment (AHE; Coveney *et al.* 2007) that has been developed in response to end-user concerns at the usability of grid middleware features an interaction model based on scientific user applications rather than user–resource interaction. Specifically, AHE provides a mechanism to represent a scientific application as a stateful web service that a user interacts with in order to run that application on a set of grid resources. AHE abstracts much of the detail of interacting with remote grid resources from users, so that they do not have to learn access mechanisms for each and every resource they want to use. By providing a simple interface that sits on top of the lower level middleware such as Globus and UNICORE, with which a user interacts when running an application on the grid, AHE hides the complexities of the underlying middleware deployed on remote resources. Indeed, since AHE can be used to launch applications on a number of different middleware systems, it can be used to provide a single interface to both local computational resources and federated grid machines (Coveney *et al.* 2007). Importantly, it enables seamless interoperability between federations of grids using Globus and UNICORE, which otherwise would not interoperate.

(c) Infrastructure

Until now grid computing applications have mostly made use of best-effort, shared TCP/IP networks: the network has simply been the glue that holds the middleware-enabled computational resources together. By contrast, by using the switched optical networks, the networks themselves become schedulable, ‘first-class’ grid resources. These types of high Quality of Service (QoS) networks form the basis of the next generation of more network-centric applications. Lightpaths provide several features that are not achievable using regular, production, best-effort networks, but which are needed for many high-performance grid applications. These include higher bandwidth connections (Hirano *et al.* 2006), user-defined networks and implementation of novel protocols which all provide essentially contention-free, high QoS links. Such low-latency, high-bandwidth links are ideal for the rapid shifting of large

quantities of data from high-end resources to local data storage and data processing facilities. Beyond that, however, they are ideal to support distributed and/or interactive simulations. For example, the GENIUS project makes use of a purpose-built dedicated 1 Gb s^{-1} lightpath network, spanning the UK and USA, which is designed to culminate on a desktop workstation within the National Hospital for Neurology and Neurosurgery (NHNN) at Queen's Square, London. Here it will be used by clinicians to interactively run, steer and visualize HemeLB in real time on remote resources using the dedicated contention-free lightpath network.

4. Case study 1: simulation-based decision support in HIV treatment

From the perspective of clinical medicine, there are two fundamental problems that generate an imposing barrier in the quest to prescribe the most effective treatment for patients infected with HIV. These stem from the innate genetic instability of HIV, coupled with its rapid replication rate which, in the presence of selective pressure exerted by ARVs, lead to the rapid emergence of mutations associated with drug resistance. A range of viral proteins targeted by ARVs, including the protease and reverse transcriptase enzymes, among others, develop patterns of mutations to different drugs or classes of drugs (Johnson *et al.* 2006).

One prevalent clinical problem remains the determination of which drug regimen best treats a patient's viral genotype. While genotypic assaying of individuals infected with HIV is a standard procedure implemented to obtain portions of the viral genome sequence (Snoek *et al.* 2005), the interpretation of such information, given the complexity of emergent mutational patterns (Wu *et al.* 2003), both in treatment-naïve and treated individuals, often means that clinicians have to resort to 'decision-support' software for assistance (Kantor *et al.* 2001). Such decision-support applications use existing clinical databases as well as phenotypic information based on the inhibition studies to infer the susceptibility of a range of inhibitors to a particular viral sequence.

Moreover, even though an initially optimal drug regimen may be determined through such a process, the mutational response to inhibitor treatment, as well as predicting the future mutational pathway for a given genotype and a given inhibitor, remain largely unsolved problems. Invariably, the effect of treatment is the emergence of drug-resistant mutations that impair inhibitor efficacy, causing a subsequent evolution of the viral genotype within the host and escape from drug pressure, resulting in a rebound in viral load. Monitoring the viral load as a marker of treatment efficacy leads to a re-evaluation of the optimal drug regimen based on the evolved viral genotype, which can in turn lead to an alteration in treatment. Unfortunately, evolution of the viral genotype may result in mutational pathways that lead to multidrug-resistant viruses. The efficacy of any decision-support software is ultimately dependent on the extent of available clinical or phenotypic information relevant to a particular sequence. Given that the vast majority of drug-resistant mutations are non-pleiotropic (Wu *et al.* 2003; Johnson *et al.* 2006), a patient-specific viral genotype can be reduced to an assessment of the amino acid sequence of key enzymatic proteins, such as the protease and reverse transcriptase, for which a range of Food and Drug Administration (FDA)-approved inhibitors have been developed.

Drug-resistant mutations are especially pronounced in the case of the viral protease. HIV-1 protease is a homodimer with C2-symmetry, composed of 99 amino acids in each chain (Wlodawer & Erickson 1993). Up to half of the amino acid positions have been shown to tolerate mutation (Schinazi *et al.* 1997), while the protease maintains a functional three-dimensional structure (Zoete *et al.* 2002). Alongside primary mutations that cause resistance, there are patterns corresponding to the build up of several accessory mutations for which significant resistance may be exhibited only when several of these mutations are present in the individual.

At the molecular level, it is the change in the binding affinity of enzyme–inhibitor complexes across a range of mutants that provides a key measure of the drug resistance afforded by such mutants. Binding affinities are determinable via both experimental (Maschera *et al.* 1996; Velazquez-Campoy *et al.* 2001) and computational (Wang & Kollman 2001; Lepsik *et al.* 2004) means; they have been reported for proteases and reverse transcriptases across a range of inhibitors and deposited in a web-based database known as BindingDB (Chen *et al.* 2001).

Unfortunately, the determination of drug-binding affinities by either experimental or computational means is not trivial and has conventionally taken far too long to be of immediate use in clinical response. Instead, binding affinity studies are constrained to provide information only in retrospect, once drug-resistant mutations have evolved in viral populations and have been characterized clinically in the event of treatment failure. While such studies are invaluable for optimizing treatment in clinical response to characterized mutations, they are not informative about mutations that have not been characterized, but which may exist in infected individuals. A gap therefore exists between scientifically rigorous approaches that determine drug resistance at the molecular level and their applicability in a real-time clinical environment to confer support for the patient-specific treatment.

An attractive goal in the medical treatment of HIV is therefore the accurate and fast determination of the binding affinity of a range of inhibitors for the unique viral genotypic constitution of an infected individual. Such information, if available on a clinically relevant time scale, would assist in the optimization of treatment on a patient-specific basis. Furthermore, a tool that could use molecular simulation-based free energy methods would provide the natural scientific methodology for generating such data and assist in bridging the molecular–clinical gap.

There are several requirements for a tool that can bridge this gap. First, it should be predictive, while not purely based on inductive techniques that rely solely on the existing clinical records. Second, it should be accurate when ranking the susceptibility of inhibitors to a variety of viral genotypes. Furthermore, it should be automated so that the clinician need not be concerned with the specific methodology of a calculation and, finally, it should return its results on the appropriate time scales (less than two weeks) to be of use in assisting with a clinical response.

Such a ranking tool, provided it can meet these requirements, in principle would be a stand-alone entity. However, integration within the existing decision-support software to provide a combined inductive/deductive approach to optimizing drug treatment at the molecular–clinical interface would be desirable.

An example of where such a framework could be adopted is the ViroLab project (Sloot *et al.* 2005). In addition to providing a common virtual environment for the collective accessibility of a range of previously independent clinical cohorts, one aim of ViroLab is to incorporate computational data at the molecular level, complementing the existing clinical and phenotypic data. Any lack of significant data that may exist for the interpretation of an optimal treatment against a unique viral genotype may then be bridged by conducting a suitable computational study on that particular genotype. The §4*a* describes the steps we have made towards realizing such a tool that we call the binding affinity calculator (BAC).

If we envisage a scenario in which automated molecular simulation, as encapsulated by the BAC, is routinely used as a component of effective clinical support, then on a global level this involves the simultaneous processing of data from potentially thousands of patients, each of whom would require of the order of 100 000 CPU hours in order to determine the optimal treatment, all within a two-week time frame. The millions of CPU hours that would need to be available over such a short time scale using current state-of-the-art computers represent an urgent computing scenario along the lines of those discussed in §3*a*. Urgent computing is thus an essential prerequisite for patient-specific molecular simulation to deliver clinical impact.

(*a*) *Rapid and accurate ranking of binding free energy differences using molecular simulation*

Several methods exist for determining computational binding free energies (Wang *et al.* 2001) ranging from accurate methods that are often overwhelmingly computationally intensive to approximate methods that may not be sensitive enough to discriminate slightly varying binding affinities. In the present context, the molecular mechanics Poisson–Boltzmann solvent-accessible surface area (MMPBSA) method appears very attractive (Wang & Kollman 2001) as no *a priori* fitting to experimental data is required (Wan *et al.* 2005) while its computing requirements are low compared with the exact methods that are 10 times more computationally intensive.

Previous studies on HIV-1 protease using the MMPBSA method have highlighted the difficulties associated with attaining accurate absolute values of binding (Lepsik *et al.* 2004) and have not been sensitive enough to ‘explicitly’ rank drug-resistant mutations (Wang & Kollman 2001). Furthermore, other research has shown that changes in the configurational entropy of the system are non-negligible in attaining accurate absolute free energy differences of binding (Chang *et al.* 2007).

We have developed a robust and accurate methodology for determining both the absolute and relative free energy differences of binding of inhibitors to the HIV-1 protease using the MMPBSA method in molecular simulation (Stoica *et al.* 2008). Using molecular dynamics simulations in explicit water alongside the MMPBSA method, including entropic considerations from normal mode analysis, it is possible to determine the absolute free energies of binding of the protease inhibitor saquinavir to the wild-type, the G48V, L90M and G48V/L90M mutants of HIV-1 protease. Furthermore, we explored the applicability of the MMPBSA approach combined with the analysis of the entropic contribution

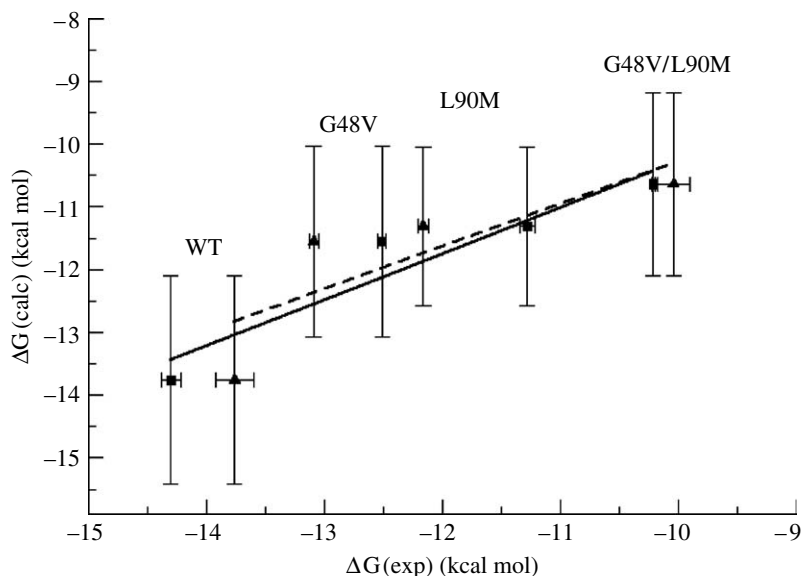


Figure 1. Correlation of relative free energies of binding to experiment for wild-type (WT) HIV-1 protease and the G48V, L90M and G48V/L90M mutants. Correlations were evaluated using MMPBSA including configurational entropy for the last 4 ns of 10 ns duration time-averaged datasets. Calculated values are compared against two experimental datasets ϵ_1 (squares) and ϵ_2 (triangles) from Maschera *et al.* (1996) and Ermolieff *et al.* (1997), respectively. Binding free energies were converted from inhibition constants at $T=298.15$ and 310.15 K for ϵ_1 and ϵ_2 , respectively. $Cc(\epsilon_1)=0.96$, $Cc(\epsilon_2)=0.81$.

in determining absolute free energies of binding as well as predictively and accurately ranking drug-resistant mutants of the HIV-1 protease through an alteration in the drug-binding affinity.

The full details of the method are reported in Stoica *et al.* (2008); however, in summary, our method uses a long and robust multistage equilibration protocol lasting 2 nanoseconds (ns) followed by a fully unrestrained production phase of 10 ns for single-trajectory MMPBSA calculations. This duration allows the convergence of the free energies of binding, calculated via the MMPBSA method including the contribution of changes in the configurational entropy upon binding. A single calculation of this type currently takes 54 hours at an optimum processor number of 32 CPUs; accounting for a further 20 hours for post-processing on a cluster of serial CPUs, the resulting total turnaround time of 74 hours (approx. 3 days) is well within the time scale required to afford benefit to clinical decision support.

We demonstrated that the inclusion of configurational entropy is required not only to obtain accurate absolute values of binding but also to accurately rank mutants of the HIV-1 protease with respect to the wild-type. In doing so, we obtain an excellent correlation with experiment (figure 1), with correlation coefficients of 0.96 and 0.81 with respect to experimental datasets ϵ_1 (Maschera *et al.* 1996) and ϵ_2 (Ermolieff *et al.* 1997), respectively. These are remarkably good considering that the correlation coefficient between the two experimental datasets themselves is 0.93. Furthermore, the binding of saquinavir to each mutant protease is correctly ordered relative to the wild-type.

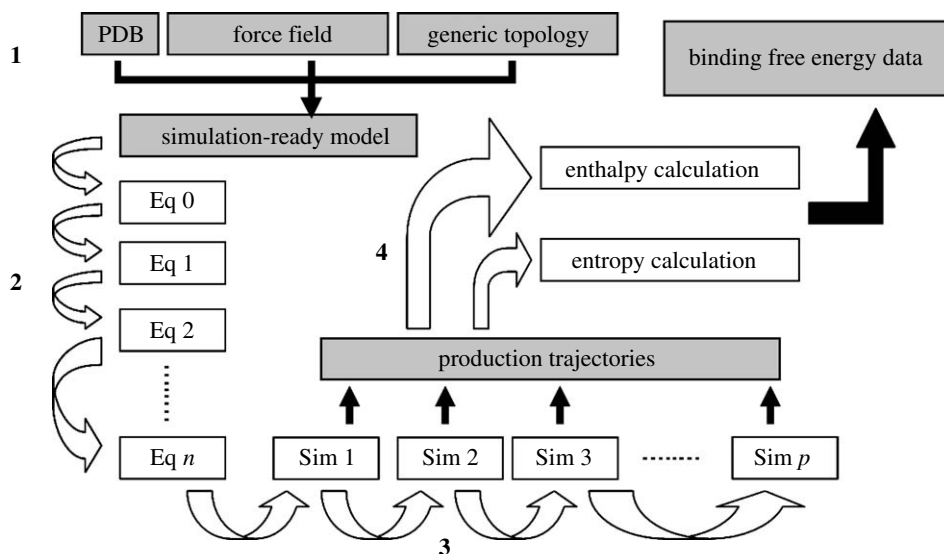


Figure 2. The workflow of an MMPBSA free energy calculation comprising four sequential stages. 1. Preparation of a simulation-ready model from the protein data bank (PDB), force field parameters and generic topology information. 2. A linear chain of equilibration simulations. 3. A linear chain of production simulations each generating trajectories for analysis. 4. Post-production execution of the enthalpy and entropy calculations leading to a determination of the binding free energy.

The excellent quantitative ranking of drug-resistant mutants exhibited in this study is encouraging for studies, currently underway, on a large array of drug-bound protease variants. Furthermore, the methodology adopted by this study, due to its trade-off between speed and accuracy, can readily form the template for a simulation-based tool that would be used in clinical decision support.

(b) Automation of simulation and calculation workflow

Figure 2 shows the various steps required for the execution of the workflow involved in a binding free energy calculation of a HIV-1 protease–ligand variant, which uses the MMPBSA method, mentioned above. We assume that a starting crystal structure of the complex exists and that force field and charge parameters for the protein and ligand are also provided.

Prior to performing any molecular dynamics simulation, a simulation-ready model has to be generated from the coordinate, generic topology and force field parameter information. This includes the extraction of suitable protease and ligand coordinates, incorporation of any mutations, the addition of neutralizing ions and solvation of the target structure. System-specific simulation-ready topology and coordinate files then have to be generated.

The next step involves an array of sequential equilibration runs incorporating the stages of minimization, annealing the system, gradual relaxation of constraints (which vary based on the mutations that have been incorporated) and, finally, unrestrained equilibration in a desired thermodynamic ensemble. Each stage uses a separate configuration file containing the exact instructions for that simulation. The production phase is very similar to equilibration and also

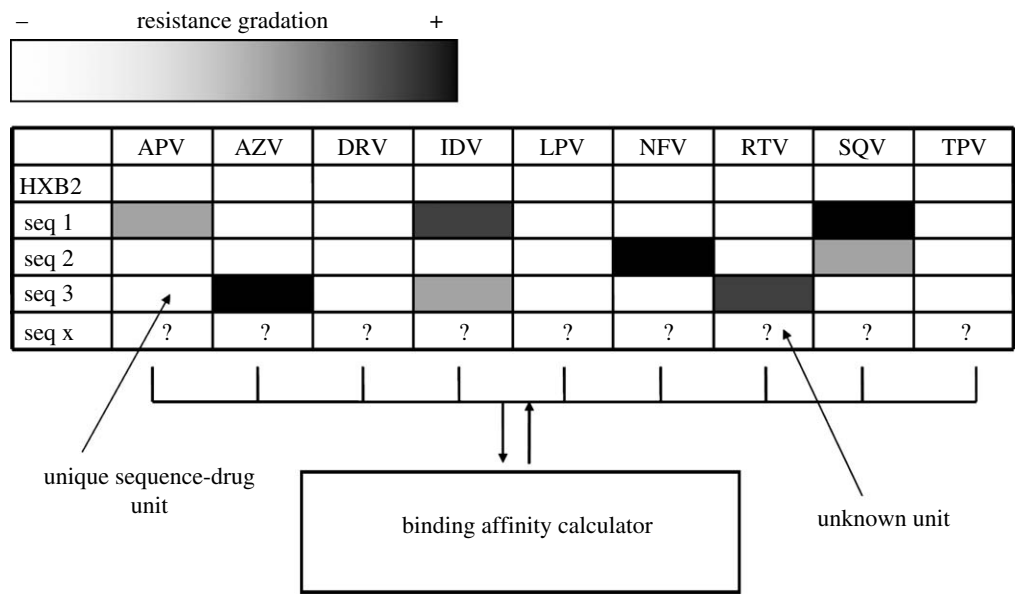


Figure 3. Illustration of a BAC determinable resistance profile. The resistance conferred by a unique protease–drug combination relative to a wild-type structure (in this case the clinical consensus HXB2 sequence) is determined from a single BAC unit. Calculation of a set of units leads to a resistance profile for each sequence across the array of available drugs. To facilitate an interpretation of drug efficacy, resistance gradation can then be colour coded, similar to schemes used by existing clinical decision-support systems (Kantor *et al.* 2001). Uniquely, however, the schema presented here derives from deductive, quantitative free energy calculations performed routinely in an automated fashion.

consists of a chain of sequentially executed simulations. Finally, the trajectory information that is output in the production phase is then post-processed in order to calculate the enthalpies and entropies of binding. Each part of the calculation uses separate configuration files that contain the specific instructions pertaining to the energy-calculation method.

Automation of such a workflow when studying an array of varying complexes thus saves considerable time and is an essential ingredient of any tool that repeatedly uses such a process to confer decision support. The main obstacles to such automation are the requirements that the entire pre-simulation model including all simulation configuration files be generated automatically as well as automated marshalling of simulations to and from relevant computational resources.

(c) *The binding affinity calculator*

We have developed a tool, called the ‘BAC’, for the automated, fast and accurate ranking of drug-resistant mutants of HIV-1 protease for an array of FDA-approved inhibitors. BAC automates the entire workflow outlined above, required to implement a molecular simulation-based free energy calculation of protease–ligand binding. The workflow is decomposed into three main components: (i) building a model, (ii) MD equilibration and simulation of the model, and (iii) post-production analysis through which the free energy is calculated.

Full automation is enabled by the integration of BAC with the AHE (Coveney *et al.* 2007), which manages the workflow around various computational resources. Perl scripts using the AHE command line interface are then used to construct workflows to manage the order in which a series of simulations are conducted, and are responsible for managing the calculation of a single, uniquely defined protease–ligand sequence, termed a ‘unit’. A detailed description of BAC is given in Sadiq *et al.* (2008).

An illustration of how the BAC could, in the future, be used to grade the resistance conferred by a unique drug–protease sequence is shown in figure 3. The determination of the resistance conferred by a unique protease–drug combination relative to a wild-type structure is naturally determined from a single BAC unit. Although represented similar to the grading schemes used by the existing decision-support systems (Kantor *et al.* 2001), the fundamental difference is that grading here is determined deductively and quantitatively from free energy calculations, in a routine and automated fashion.

5. Case study 2: patient-specific investigation of cerebral blood flow

Cardiovascular disease is the cause of a large number of deaths in the developed world (The World Health Report 2002). Cerebral blood flow behaviour plays a crucial role in understanding, diagnosis and treatment of the disease; problems are often due to anomalous blood flow behaviour in the neighbourhood of bifurcations and aneurysms within the brain although the details are not well understood. Experimental studies are frequently impractical owing to the difficulty of measuring behaviour in humans; however, X-ray and magnetic resonance angiography (MRA) enable non-invasive static and dynamical data acquisition (Goyen *et al.* 2001). Indeed, some studies reveal relationships between specific flow patterns around walls and cardiovascular diseases such as atherosclerosis (Thubrikar & Robicsek 1995).

Today such imaging methods represent a very important tool for the diagnosis of various cardiovascular diseases, together with the design of cardiovascular reconstructions and devices for the enhancement of blood flow. Notwithstanding these advances in measurement methods, modelling and simulation undoubtedly have a crucial role to play in haemodynamics. Simulation, for example, offers the clinician the possibility of performing non-invasive virtual experiments to plan and study the effects of certain courses of (surgical) treatment with no danger to the patient, offering support for diagnosis, therapy and planning of vascular treatment (Taylor *et al.* 1999). Modelling and simulation also offer the prospect of providing clinicians with virtual patient-specific analysis and treatments, and it is in this context that we describe the following study.

Realizing the goal of blood flow simulation of the entire neurovasculature is dependent on the availability of computational models that can efficiently simulate haemodynamics in complex structures such as vascular systems. In the project described here, the aforementioned imaging techniques are used to provide patient-specific input data for such simulations. Furthermore, the computational fluid ‘solver’ used must itself be numerically highly efficient and provide scientists and neurosurgeons with the ability to manipulate and visualize the associated large datasets. While conventional continuum fluid solvers based

on finite-difference, finite-volume and finite-element codes certainly do exist (Botnar *et al.* 2000), they are beset with problems in three spatial dimensions due to the computational costs of mesh generation, the need to solve the auxiliary Poisson equation for the pressure field and the various approximations associated with the calculation of the shear stress from the flow velocity field. For large systems, such as those we are concerned with in addressing the cerebral blood flow, it is essential to develop and use scalable, high-performance parallel codes that further complicate the use of continuum models. The intricate geometry of the fluid vessels and treatment of fluid boundary conditions at such walls are also difficult for continuum fluid dynamics models to handle. The lattice-Boltzmann (LB) method offers an attractive alternative, and our approach is based on this modelling and simulation method.

Our approach is intended to yield patient-specific information that helps to plan embolization of arteriovenous malformations and aneurysms *inter alia*. The prospect of patient-specific three-dimensional mathematical models with temporal resolution to address issues of pulsatile flow, phase differences and effects of treatment is very compelling both in terms of understanding neurovascular pathophysiology and in planning patient treatment. Existing grid-based interactive models for the visualization of medical images have barely begun to address these issues (Sloot *et al.* 2003). However, in the context of grid-based haemodynamics simulations, the LB method has already played a significant role. Indeed, it has provisionally been used with grid facilities and, alongside the finite-element method, incorporated in problem-solving environments (PSEs; Sloot *et al.* 2003; Zudilova & Sloot 2005). Here, the PSE combines the access to medical image data (e.g. obtained with angiography), their three-dimensional reconstruction by means of segmentation algorithms and manipulation through three-dimensional graphical editing tools, the simulation stage and the visualization and steering of the flow (pressure, velocity and shear stress fields).

While these important developments in three-dimensional modelling have been taking place, there have also been rapid developments in three-dimensional imaging capabilities, e.g. recent installation of new equipment within the National Hospital (NHNN), which comprises

- (i) two new angiography suites (capable of three-dimensional rotational angiography (RA), in which a volumetric image is created by obtaining many two-dimensional projections of the brain during rapid injection of intra-arterial contrast fluid into the carotid arteries),
- (ii) a new 64 slice CT scanner capable of rapid acquisition of volumetric data during intravenous injection of contrast into a peripheral vein, and
- (iii) two new MRI scanners capable of the generation of MRA images, also able to be processed volumetrically.

Large-scale patient-specific modelling of cerebral blood flow is however largely without precedent. To fully exploit its potential, it is crucial to engage with neuroradiologists and neurosurgeons.

Our work draws on a number of cutting-edge grid and high-performance computing technologies. In the area of patient-specific medical treatment, the time taken to turn around a simulation is of crucial importance. The goal of

the present project is to demonstrate the ability to perform whole-brain blood flow simulations distributed around the globe, steered and visualized in real time by a surgeon, in order to treat a patient within a time frame that is clinically useful.

(a) *HemeLB*

We have developed a new LB algorithm and associated code, named HemeLB (Mazzeo & Coveney 2008), which is specifically designed to address fluid flow in sparse geometries, that is situations in which the fluid channels occupy a relatively small (typically 15–20%) volume fraction of the simulation cell defined by the scanned brain image, the remainder of which is treated as ‘void’ from the point of view of the fluid flow solver. Performance benchmarks show that HemeLB scales essentially linearly with processor count up to at least 1024 processors on HPCx (the full production partition on this capability computing resource, which puts HemeLB in the highest performance class on this machine). The algorithm is based on a data layout that automatically handles lattices with large unstructured regions covered by void (i.e. non-fluid) nodes. In the parallel code, spatial domain decomposition is handled by a simple and fast cluster algorithm; partial overlapping of communication between different processor domains and on-processor computation is also exploited to enhance performance. Indeed, the parallel performance of this code outstrips comparable codes that we are aware of in the literature, for example Artoli *et al.* (2006). Two additional attractive features should be pointed out. HemeLB has been combined with MPIg, the successor to MPICH-G2 (Karonis *et al.* 2003), which hides latency much more effectively than earlier versions: we find that the communication costs (being overlapped with computation) become almost negligible, meaning that cross-site runs with HemeLB on very large models run very efficiently. The topology-aware data communications implemented in HemeLB take advantage of the asynchronous communication capability provided by MPIg.

The HemeLB mesh generator receives a dataset produced by the MRA technique in the form of slices of two-dimensional Portable Pixel Map images. If the dataset is not in this format, the program MRICRO (<http://www.sph.sc.edu/comd/rorden/mricro.html>) allows proper data conversion. The three-dimensional RA images are used to reconstruct the vascular tree on the basis of assigning bright pixels with blood. In our scientific work, we also make use of the publicly available anonymized MRI brain scans available at: <http://www.ixi.org.uk>. The LB formulation adopted in HemeLB is based on regular grids and thus our mesh generator takes advantage of the same topology associated with three-dimensional RA datasets.

We have developed a fast and robust parallel ray-tracing technique that is used initially to automate the construction of inlet/outlet boundary sites and conditions based on an efficient clustering method. The reconstructed system computed in this way is used to specify the input configuration file for HemeLB. HemeLB has been instrumented with the RealityGrid steering library (Pickles *et al.* 2005) in order to permit interactive control of HemeLB simulations and associated visualizations.

(b) Fluid flow visualization and rendering

An effective fluid flow simulation environment should employ both a high-performance parallel fluid solver and the computational resources to compute, render and visualize the progress of a simulation. In addition to the individual compute, rendering and visualization components, the communication between these three components has a major impact on performance. In order to minimize the time taken for data exchange between computation and rendering, we perform both on the same compute resource (figure 4).

Figure 5 illustrates some of the features of our PSE. The computational mesh generator reads pre-filters and optionally increases the resolution of the dataset with trilinear interpolations. Here, the original resolution is increased and the resulting bounding box comprises $2560 \times 2560 \times 861$ voxels but only some of them are stored through a two-level grid. Input reading, pre-filtering and two-level grid construction takes a few minutes, in particular due to the high cost of trilinear interpolations. By selecting a pixel associated with a blood voxel, the whole cerebral vascular tree is segmented by means of an efficient clustering algorithm. Only a few seconds are needed to produce this high-resolution system that can be rendered and visualized at an interactive rate. The boundary conditions are determined by rapid construction of triangles representing the fluid input/output boundaries using an efficient and robust ray-tracing-based algorithm. The boundary condition set-up of our cerebral vascular tree, depicted in figure 5b with different intensities of red to enhance visualization and including the input/output boundary triangles, requires a few minutes to set up. The manipulated system comprising approximately 8 million fluid lattice sites is simulated and visualized by HemeLB, within which we have incorporated a parallel ray tracer, at a speed of approximately 25 time steps per second with the use of 100 processors. A snapshot of the velocity flow field is shown in figure 5c using a volume-rendering technique.

6. Discussion

Patient-specific computer simulation is a compelling and inevitable prospect for the enhancement and optimization of medical treatment. The paradigm of patient-specific simulation is based on our ability to use digital biomedical data, unique to a patient, in one or more computations that result in the determination of relevant biophysical and/or biochemical properties of the patient. The tailoring of medical treatment based on the outcome of such simulations has clear advantages over treatment that does not distinguish between one patient and another, but demands that patient-specific simulations must be informative within clinically relevant time scales.

Here, we outline the computational infrastructure that is required to realize such a goal. Furthermore, we have illustrated, by way of two specific and diverse applications, the inroads we have made into realizing such a paradigm. In case study 1, we demonstrated that the determination and automation of well-defined and robust molecular simulation protocols and computation parameters, in combination with HPC resources and grid technology, can be employed to rapidly determine accurate free energies of binding of a set of inhibitors to an array of HIV-1 proteases and thus rank the resistance conferred by particular

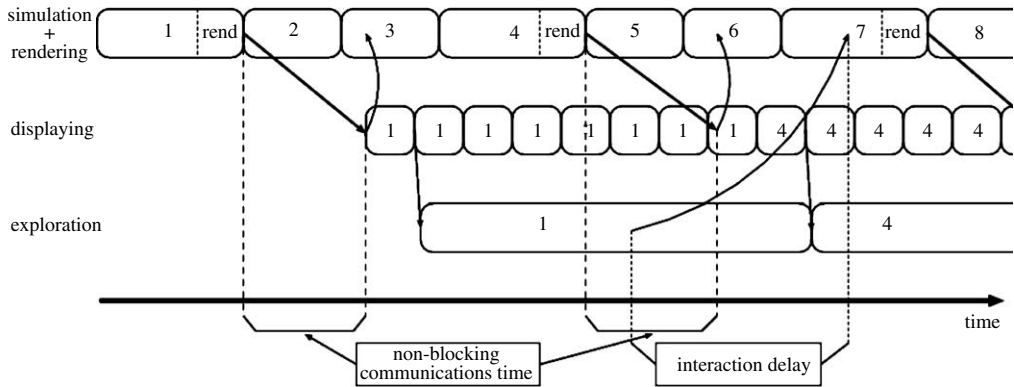


Figure 4. Representation of the time frames of the interactive simulation environment used to model cerebral blood flow using HemeLB (boxed numbers indicate frame numbers corresponding to the rendered simulation). The simulation and the ray-tracing-based rendering (both parallel) are performed using one or multiple grid-based computational resources ('simulation+rendering'). The image representing the simulation flow field is visualized on the client ('displaying'). Assuming that the visualization time is negligible, the frequency with which the ray tracing is performed depends only on the server-client communication bandwidth and the speed of the simulation plus visualization component. The effect of the depicted interaction frame is simulated at the eighth time step due to the interaction-communication delay.

mutants that may exist in a patient. Any such patient-specific mutants are already routinely determinable from the current genotypic assay methodologies (partnered by decision-support systems); we have developed a tool, called the BAC, that automates the determination of the free energy of binding of any given patient-specific protease sequence with an array of FDA-approved inhibitors, returning results well within the clinical time scales.

In case study 2, we show how patient-specific three-dimensional RA-based imaging of the brain can be used to generate simulation conditions from which arterial blood flow in the brain can be determined via LB fluid simulations. We have developed a grid-based high-performance interactive simulation environment to investigate cerebral blood flow of patient-specific systems on a clinically relevant time scale. We have adopted an efficient fluid solver to simulate blood behaviour in complex systems either on a single multiprocessor machine or on a loosely coupled, geographically distributed set of platforms. Our computational mesh generator reads and pre-filters, increasing the original resolution and allowing us to reconstruct a very large simulation system in a few seconds. Our interactive simulation environment relies on an efficient rendering approach and permits the extremely rapid simulation and visualization of a large-scale cerebral patient-specific system. We can effectively distribute jobs around the available computational resources, monitor as well as interact with them, all enabled by the development of a simple graphical interface.

Inevitably, the adoption of patient-specific simulation by the medical community will have several ramifications for the way computational resources are administered. First, the potential scale of such an enterprise will not only necessitate access to large-scale HPC resources and computational grids with fast network connections and seamless, interoperable middleware, but also will require changes to usage policies across federated grid resources. But there is also a moral

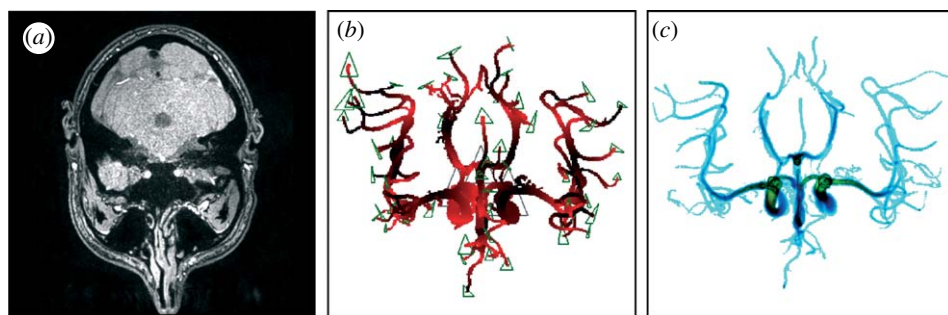


Figure 5. Illustration of several features of the PSE. (a) One of the 440 slices of 512^2 pixels each of a three-dimensional RA patient-specific dataset. (b) The boundary condition set-up of our cerebral vascular tree with different intensities of red to enhance visualization. (c) A snapshot of the velocity flow field using a volume-rendering technique with increasing velocities depicted from blue to green.

dimension to be considered: with limited access to resources, who will be able to benefit from such patient-specific facilities? Current grid accounting and scheduling models are not capable of supporting the routine medical simulation. Compute jobs that are within this class would certainly need to have some kind of urgent status on general purpose grids to ensure that the turnaround time is able to meet clinical requirements and thus be prioritized in queuing systems. Indeed, it is quite possible that, if these techniques become routine, the demand will not be met solely by general purpose computational grids, and may require additionally provisioned infrastructure, either through the relevant medical service providers and/or private companies that sell compute power for such use.

Another issue concerns matters of privacy and confidentiality. In the case of patient-derived data for case study 1, for example, the viral genotypic data in question (required for drug sensitivity prediction and thus patient-specific treatment guidance) are likely to be held locally within hospital databases protected by virtue of their own local policy and procedure. Addressing data-privacy issues in the eventuality of using these data for patient-specific simulations, however, is not going to be merely a hospital issue. Rather it is likely to be a national and international issue, since patient-specific data may be regularly crossing international borders in order to access the required computational resources. Clearly, an important consideration here is the protection and privacy of sensitive medical data, which will require considerable review and policy development if patient-specific simulation for medical intervention is to become commonplace. For case study 2, for example, one may envisage a scenario where the transfer of the three-dimensional RA data from a private hospital network to a network infrastructure, such as the UK's JANET Light-path network, raises tangible issues of data privacy and the requirement of data anonymization. In this case, the data associated with a three-dimensional RA can be broken up into two broad parts: metadata to be distinguished from the actual scan data and associated physical parameters. These can also be described as anonymized data and patient-specific confidential data, respectively. Metadata relate to patient-specific data, such as name, age, weight, etc. The anonymized data relate to the scan volume data, along with the physical parameters, such as voxel size, which are required by HemeLB for accurate fluid flow modelling. In

this example, a simple solution would see the metadata kept on the private hospital network, while the anonymized three-dimensional RA is sent to remote compute resources, with both sets of data related by a database key. For such a neurosurgical patient-specific simulation environment to work on a regular basis, such systems of anonymization and de-anonymization of data must become automated.

It is clear that, if agreed methods on data anonymization are not available, this may jeopardize the entire enterprise of patient-specific simulation.¹ Furthermore, ethical issues arise. Who would bear responsibility for mistakes made using such treatment methodologies? In this area, the Virtual Physiological Human initiative (Viceconti *et al.* 2007) will play a pivotal role in publicizing and communicating the legal and ethical issues that beset this developing field of research. The broad remit of this and other physiological modelling initiatives that rely heavily on access to large amounts of patient data (often collected as part of routine clinical practice) will require that significant efforts are made in the development of the tools, methods, data standards, communication strategies, data and model curation strategies and associated policy for ethical access to and use of medical data.

Finally, as and when patient-specific simulation becomes established, there will need to be a shift in the way ‘grid computing’ and ‘supercomputing’ are perceived. The concept of patient-specific simulation adumbrates an era when, instead of being the playground of a privileged few scientists, large-scale high-performance computing resources will be routinely used by many people vast numbers of times per day. It will result not only in overturning our perceptions of what supercomputing constitutes but also the economics of the entire enterprise.

The work described in this paper has used medical data from anonymized publicly available online repositories. For current and future work in patient-specific studies, ethics approval is currently being sought.

We thank various people who have contributed to this work at the NGS (UK), HPCx, HECToR and the US TeraGrid. The GENIUS project is funded under EPSRC grants EP/F00561X/1, EP/C536452/1 (RealityGrid) and by the National Science Foundation MRAC grants TG-DMR070013N and TG-DMR070014N. This research has also been partially supported by the EU-funded ViroLab project (EU IST STREP Project 027446). Our work in these areas has benefitted greatly through involvement of P.V.C. and C.V.G. in the EU-funded STEP coordination action.

References

- Almond, J. & Snelling, D. 1999 UNICORE: uniform access to supercomputing as an element of electronic commerce. *Future Generation Comput. Syst.* **15**, 539–548. (doi:10.1016/S0167-739x(99)00007-2)
- Artoli, A. M., Hoekstra, A. G. & Sloot, P. M. 2006 Mesoscopic simulations of systolic flow in the human abdominal aorta. *J. Biomech.* **39**, 873–874. (doi:10.1016/j.jbiomech.2005.01.033)
- Bassingthwaighe, J. B. 2000 Strategies for the physiome project. *Ann. Biomed. Eng.* **28**, 1043–1058. (doi:10.1114/1.1313771)
- Beckman, P., Beschastnikh, I., Nadella, S. & Trebon, N. 2007 *Building an infrastructure for urgent computing*. Amsterdam, The Netherlands: IOS.

¹It could be argued that in a life-or-death situation, the processes needed to make a life-saving decision would take precedence over the issues of privacy and confidentiality.

- Botnar, R., Rappitsch, G., Scheidegger, M. B., Liepsch, D., Perktold, K. & Boesiger, P. 2000 Hemodynamics in the carotid artery bifurcation: a comparison between numerical simulations and *in vitro* MRI measurements. *J. Biomech.* **33**, 137–144. (doi:10.1016/S0021-9290(99)00164-5)
- Chang, C. A., Chen, W. & Gilson, M. K. 2007 Ligand configurational entropy and protein binding. *Proc. Natl Acad. Sci. USA* **104**, 1534–1539. (doi:10.1073/pnas.0610494104)
- Chen, X., Lin, Y. & Gilson, M. K. 2001 BindingDB: a web-accessible molecular recognition database. *J. Comb. Chem. High Throughput Screen.* **4**, 719–725.
- Coveney, P. V. 2005 Scientific grid computing. *Phil. Trans. R. Soc. A* **363**, 1707–1713. (doi:10.1098/rsta.2005.1632)
- Coveney, P. V. & Fowler, P. W. 2005 Modelling biological complexity: a physical scientist's perspective. *J. R. Soc. Interface* **2**, 267–280. (doi:10.1098/rsif.2005.0045)
- Coveney, P. V., Saksena, R. S., Zasada, S. J., McKeown, M. & Pickles, S. 2007 The application hosting environment: lightweight middleware for grid-based computational science. *Comput. Phys. Commun.* **176**, 406–418. (doi:10.1016/j.cpc.2006.11.011)
- Ermolieff, J., Lin, X. & Tang, J. 1997 Kinetic properties of saquinavir-resistant mutants of human immunodeficiency virus type 1 protease and their implications in drug resistance *in vivo*. *Biochemistry* **36**, 12 364–12 370. (doi:10.1021/bi971072e)
- Foster, I. 2006 GLOBUS toolkit version 4: software for service-oriented systems. In *IFIP International Conference on Network and Parallel Computing*, pp. 2–13. Berlin, Germany: Springer.
- Goyen, M., Laubb, G., Ladd, M. E., Debatin, J. F., Barkhausen, J., Truemmler, K.-H., Bosk, S. & Ruehm, S. G. 2001 Dynamic 3D MR angiography of the pulmonary arteries in under four seconds. *J. Magn. Reson.* **13**, 372–377.
- Hirano, A. et al. 2006 The first functional demonstration of optical virtual concatenation as a technique for achieving terabit networking. *Future Generation Comput. Syst.* **22**, 876–883. (doi:10.1016/j.future.2006.03.027)
- Johnson, V. A., Brun-Vezinet, F., Clotet, B., Conway, B., Kuritzkes, D. R., Pillay, D., Schapiro, J. & Richman, D. 2006 Update of the drug resistance mutations in HIV-1: fall 2006. *Int. AIDS Soc.—USA* **14**, 125–130.
- Kantor, R., Machekano, R., Gonzales, M. J., Dupnik, K., Schapiro, J. M. & Shafer, R. W. 2001 Human immunodeficiency virus reverse transcriptase and protease sequence database: an expanded data model integrating natural language text and sequence analysis programs. *Nucleic Acids Res.* **29**, 296–299. (doi:10.1093/nar/29.1.296)
- Karonis, N. T., Toonen, B. & Foster, I. 2003 MPICH-G2: a grid-enabled implementation of the message passing interface. *J. Parallel Distrib. Comput.* **63**, 551–563. (doi:10.1016/S0743-7315(03)00002-9)
- Lepsik, M., Kriz, Z. & Havlas, Z. 2004 Efficiency of a second-generation HIV-1 protease inhibitor studied by molecular dynamics and absolute binding free energy calculations. *Proteins: Struct. Funct. Bioinform.* **57**, 279–293. (doi:10.1002/prot.20192)
- MacLaren, J. 2007 HARC: the highly-available resource co-allocator. In *Proc. GADA'07, LNCS 4804 (OTM Conferences 2007, Part II)*. Berlin, Germany: Springer.
- Manos, S., Zasada, S. & Coveney, P. V. 2008 Life or death decision-making: the medical case for large-scale, on-demand grid computing. *CTWatch Q. J.* **4**, 35–45.
- Maschera, B., Darby, G., Palu, G., Wright, L. L., Tisdale, M., Myers, R., Blair, E. D. & Furfine, E. S. 1996 Human immunodeficiency virus mutations in the viral protease that confer resistance to saquinavir increase the dissociation rate constant of the protease-saquinavir complex. *J. Biol. Chem.* **271**, 33 231–33 235. (doi:10.1074/jbc.271.52.33231)
- Mazzeo, M. D. & Coveney, P. V. 2008 HemeLB: a high performance parallel lattice-Boltzmann code for large scale fluid flow in complex geometries. *Comput. Phys. Commun.* **178**, 894–914. (doi:10.1016/j.cpc.2008.02.013)
- Pickles, S. M., Haines, R., Pinning, R. L. & Porter, A. R. 2005 A practical toolkit for computational steering. *Phil. Trans. R. Soc. A* **363**, 1843–1853. (doi:10.1098/rsta.2005.1611)
- Sadiq, S. K., Zasada, S. J., Wright, D., Stoica, I. & Coveney, P. V. 2008 An automated molecular simulation-based binding affinity calculator for ligand-bound HIV-1 proteases. Preprint.

- Schinazi, R. F., Larder, B. A. & Mellors, J. W. 1997 Mutations in retroviral genes associated with drug resistance. *Antivir. News* **5**, 129–142.
- Sloot, P. M. A., van Albada, G. D. & Zudilova, E. V. 2003 A grid-based interactive visualisation of medical images. In *Proc. First European HealthGrid Conference. Commission of the European Communities, Information Society Directorate-General*.
- Sloot, P. M. A., Boukhanovsky, A. V., Keulen, W., Tirado-Ramos, A. & Boucher, C. A. 2005 A grid-based HIV expert system. *J. Clin. Monit. Comput.* **19**, 263–278. (doi:10.1007/s10877-005-0673-2)
- Snoek, J., Riva, C., Steegen, K., Schrooten, Y., Maes, B., Vergne, L., Van Laethem, K., Peeters, M. & Vandamme, A.-M. 2005 Optimization of a genotypic assay applicable to all human immunodeficiency virus type 1 protease and reverse transcriptase subtypes. *J. Virol. Methods* **128**, 47–53. (doi:10.1016/j.jviromet.2005.04.001)
- Stoica, I., Sadiq, S. K. & Coveney, P. V. 2008 Rapid and accurate prediction of binding free energies for saquinavir-bound HIV-1 proteases. *J. Am. Chem. Soc.* **130**, 2639–2648. (doi:10.1021/ja0779250)
- Taylor, C. A., Draney, M. T., Ku, J. P., Parker, D., Steele, B. N., Wang, K. & Zarins, C. K. 1999 Predictive medicine: computational techniques in therapeutic decision-making. *Comput. Aided Surg.* **4**, 231–247.
- The World Health Report 2002 Reducing risks, promoting healthy life. See <http://www.who.int/whr/en>.
- Thubrikar, M. J. & Robicsek, F. 1995 Pressure-induced arterial wall stress and atherosclerosis. *Ann. Thoracic Surg.* **59**, 1594–1603. (doi:10.1016/0003-4975(94)01037-D)
- Velazquez-Campoy, A., Kiso, Y. & Freire, E. 2001 The binding energetics of first- and second-generation HIV-1 protease inhibitors: implications for drug design. *Arch. Biochem. Biophys.* **390**, 169–175. (doi:10.1006/abbi.2001.2333)
- Viceconti, M. *et al.* 2007 Seeding the EuroPhysiome: a roadmap to the virtual physiological human. See <http://www.europhysiome.org>.
- Wan, S., Coveney, P. V. & Flower, D. R. 2005 Peptide recognition by the T cell receptor: comparison of binding free energies from thermodynamic integration, poisson–Boltzmann and linear inberation energy approximations. *Phil. Trans. R. Soc. A* **363**, 2037–2053. (doi:10.1098/rsta.2005.1627)
- Wang, W. & Kollman, P. A. 2001 Computational study of protein specificity: the molecular basis of HIV-1 protease drug resistance. *Proc. Natl Acad. Sci. USA* **98**, 14 937–14 942. (doi:10.1073/pnas.251265598)
- Wang, W., Donini, O., Reyes, C. M. & Kollman, P. A. 2001 Biomolecular simulations: recent developments in force fields, simulations of enzyme catalysis, protein–ligand, protein–protein, and protein–nucleic acid noncovalent interactions. *Annu. Rev. Biophys. Biomol. Struct.* **30**, 211–243. (doi:10.1146/annurev.biophys.30.1.211)
- Wlodawer, A. & Erickson, J. W. 1993 Structure-based inhibitors of HIV-1 protease. *Annu. Rev. Biochem.* **62**, 543–585. (doi:10.1146/annurev.bi.62.070193.002551)
- Wu, T. D. *et al.* 2003 Mutation patterns and structural correlates in human immunodeficiency virus type 1 protease following different protease inhibitor treatments. *J. Virol.* **77**, 4836–4847. (doi:10.1128/JVI.77.8.4836-4847.2003)
- Yoshimoto, K., Kovatch, P. & Andrews, P. 2005 Co-scheduling with user-settable reservations. In *Job scheduling strategies for parallel processing* (eds D. G. Feitelson, E. Frachtenberg, L. Rudolph, & U. Schwiegelshohn). Lecture Notes in Computer Sciences, no. 3834, pp. 146–156. Berlin, Germany: Springer.
- Zoete, V., Michielin, O. & Karplus, M. 2002 Relation between sequence and structure of HIV-1 protease inhibitor complexes: a model system for the analysis of protein flexibility. *J. Mol. Biol.* **315**, 21–52. (doi:10.1006/jmbi.2001.5173)
- Zudilova, E. V. & Sloot, P. M. A. 2005 Bringing combined interaction to a problem solving environment for vascular reconstruction. *Future Generation Comput. Syst.* **21**, 1167–1176. (doi:10.1016/j.future.2004.04.004)