

MEETING REPORT

SupraBiology 2014: Promoting UK-China collaboration on Systems Biology and High Performance Computing

Ettore Murabito¹, Riccardo Colombo^{2,3}, Chengkun Wu⁴, Malkhey Verma⁴, Samrina Rehman⁴, Jacky Snoep⁴, Shao-Liang Peng⁵, Naiyang Guan⁵, Xiangke Liao^{5,*} and Hans V. Westerhoff^{4,*}

¹ Manchester Institute of Biotechnology, School of Computer Science, Faculty of Engineering and Physical Sciences, Manchester Centre for Integrative Systems Biology, The University of Manchester, Manchester, M139PL, United Kingdom.

² Department of Informatics, Systems and Communication, University of Milan-Bicocca, Milan, 20126, Italy.

³ SYSBIO – Centre of Systems Biology, Milan, 20126, Italy.

⁴ Manchester Institute of Biotechnology, School of Chemical Engineering and Analytical Sciences, Manchester Centre for Integrative Systems Biology, The University of Manchester, Manchester, M139PL, United Kingdom.

⁵ School of Computer Science, National University of Defence Technology, Changsha 410073, China

* Correspondence: xkliao@nudt.edu.cn, hans.westerhoff@manchester.ac.uk

Received January 27, 2015

In the second half of 2014 the Manchester Institute of Biotechnology, based in Manchester (UK), hosted the first SupraBiology congress, an event attended by representatives of different academic institutions and industry based in both the UK and China. The congress was aimed to serve as a platform to discuss and promote potential collaborations between the UK and China on the subject of Systems Biology and High Performance Computing. The event, sponsored by the “BBSRC China Partnering Awards” and ISBE, was organised as a sequence of talks addressing the different aspects of Systems Biology that can benefit from High Performance Computing. A general discussion session followed where the scientific, technical, and logistic aspects of the prospected UK-China collaborations were examined.

In what follows we summarize the contributions of the different attendees to the congress. In particular, the material hereby presented is organized around five main areas of interest:

- Systems Biology in medicine
- Computational methods of Systems Biology
- Computational Biology for Industry
- Computational Systems Biology in Education
- Computational Biology on the Tianhe-2 Supercomputer

The report ends with a discussion section relating the

ideas that have been pushed forward to establish collaborations between UK and China.

INTRODUCTION

Systems Biology is an interdisciplinary field that studies the complex interactions occurring within biological systems, and how such interactions give rise to the emergent properties of life. Systems Biology deals with overwhelming amounts of data, harvested through different high-throughput techniques (e.g., proteomics, metabolomics, genomics). Because of the inherent complexity of living organisms — resulting from the tight interconnectedness of their components, Systems Biology needs to invoke modelling and systems theory so as to understand the mechanism of emergence. The challenge that Systems Biology rises to consists of discovering the principles that govern the behaviour of biological systems based on such corpuses of data of such complexity. This has profound implications in every biology-related field: from industrial production of biomaterials to personalized medicine to bioenergy.

Some important research-lines are today hampered by the lack of adequate computing power within the relevant institutions. The ability to solve advanced computation problems in an expendable time-scale makes high-

performance computing (HPC) particularly appealing to systems biologists. Under this perspective, project collaborations between the two fields of Systems Biology and HPC can harvest the potential for a breakthrough in a number of very hot topics, both in academy and industry.

Systems Biology in Medicine

Hans V. Westerhoff (University of Manchester, UK)

The old paradigm of disease was molecule-oriented. The consensus was that any medical condition was ultimately ascribable to the malfunctioning of a single molecule. In recent years, Systems Biology has brought about a new paradigm, where the entire biochemical network has to be considered rather than the single components [1]. The concept of systems medicine is based upon this new perspective where the emergence of a condition has to be inscribed in a multi-factorial scenario. The target of a drug is now rightly understood as the entire biochemical network rather than a single molecule. This paradigm shift exposes us to a new level of complexity in medicine.

In the domain of systems medicine, personalized medicine aims to design drugs that are specific to individual patients (or at least classes of patients) [2]. To achieve such goal one line of research consists of creating a “person simulator” [3] able to produce a virtual twin of the patient, based on multiple biomarkers subject to individual variability. Such a virtual twin would then be used to predict and test *in silico* an array of drugs designed to elicit in the patient the desired effect.

The challenges of such a project are multiple:

- The “virtual twin” implies a multi-scale approach where computer models of cells, tissues, organ and eventually the entire organism are integrated into one sole framework;
- Although the topology of many biochemical systems (and indeed of essentially the entire metabolism of the human) is known, there is a substantial lack of knowledge about the dynamics of their behaviour and the parameter values that determine those dynamics;
- The spatial location of different biochemical networks may play an important role, but this has not yet been investigated in the context of systems medicine;
- Computing the effect of a drug on one individual and relating the result to the actual clinical outcome can test the model’s prediction and lead to an improved model’s parameterization. This then increases the predictive strength of the computation for a second individual. Because every new individual represents an experiment that can be used to test/train the models for all the already existing patients, the predictive strength of the computations will increase supralinearly. Hence, there is virtually

no limit to the computer power that can be used sensibly here; any additional computer power will greatly enhance each individual prediction.

Andrew Narracott (University of Sheffield, UK)

An important step toward the “virtual twin” is represented by the Virtual Physiological Human (VPH), which is being developed by the VPH Institute [4,5]. While the many VPH projects target scales from organism to tissue, some integrate down to the molecular scale, with application in informing clinical practice and assisting medical doctors in decision-problem scenarios.

For example, if a patient suffers from stenosed coronary artery and needs a stent, what is the best location to implant the device? A complete examination to assess where to implant the stent is long, costly and unpleasant for the patient. Through a 3D representation of the blood vessels and using clinical information on blood pressure, the VPH can help to shorten the procedure and predict the best location for the implant, thus substantially lessening the cost and discomfort associated with a traditional exam.

The VPH can perform complex analysis on the human physiology over a range of physical and temporal scales. However the delivery time of truly informative models must be compatible with clinical time-scale. This means that for a truly personalized approach, HPC is required to deliver the volume of analyses required to consider multiple treatment strategies and parameter sensitivity assessment over large patient cohorts.

Jacky Snoep (University of Manchester, UK)

The multi-scale factor and the interactions between different models are crucial aspects in personalised medicine. Attempts to address the issue of integrating models describing systems at a biochemical level have already been made. An example is provided by a study on malaria where mathematical models were constructed for the central carbon metabolism of the parasite (*Plasmodium falciparum*) and the infected red blood cell during the different life stages of the pathogen. These models are stored in JWS online, a repository of curated kinetic models which also provides tools for online simulations and analyses [6].

Currently these models are being extended to the whole body level for a rat animal model and for simulation of glucose metabolism in malaria patients. This further step implies the integration and interactions of different hierarchical levels of organization. The resulting complexity faced when enlarging the scope of such mathematical models makes the resort to high computing

power a very important if not decisive factor.

Peter Coveney (University College of London, UK)

The trend in biomedicine is to rely increasingly on computer modelling. For computer models to be profitably used in clinical practice, they need to deliver relevant predictions in a timely manner. The use of large-scale distributed computing is essential to ensure these requirements to be fulfilled.

HPC is already being used in some relevant research lines. A notable example is provided by drug discovery, where simulations of macromolecular interactions and quantifications of binding free energy are key [7,8]. Such simulations are computationally very demanding, yet they must be reproducible to increase the statistical quality of their predictions. HPC resources are being used to meet this need in the study of the HIV protease system and the protein kinases in cancer treatment.

Another important area where HPC is applied is the study of brain blood flow. The HemeLB simulation environment, developed by the University College of London [9], is capable of simulating large vascular networks with the aim of assisting clinicians in diagnosing the effects of aneurysms and choosing appropriate interventions. HemeLB runs on thousands of compute cores, this making a strong scientific case for the use of HPC in biomedicine.

Computational methods of Systems Biology

Pedro Mendes (University of Manchester, UK)

It is fundamental in Computational Systems Biology to have reliable tools for the analysis of *in silico* representations of biochemical systems. Depending on the spatial scale considered and/or the level of detail of the model, different analysis may be put in place to describe/predict the behaviour of the underlying system. There are two main approaches that are used to create informative models of biological systems.

The “bottom-up” approach relies on the characterization of singular molecular interactions. In this case the behaviour of the system emerges as the result of the combined single events occurring in the network. Different software package are available today to perform analysis on models generated through a bottom-up approach. Copasi is one of the most popular in the Systems Biology community as it implements several analytical tools and has extensive support for parameter estimation and optimization [10].

The “top-down” approach, in contrast, adopts a data-driven perspective and seeks to construct models from large scale ‘omics data’. In this case, machine-learning

methods and network theory are applied to infer the relevant properties that most affect or determine the behaviour of the system in a given set of experimental conditions.

Both these two approaches can benefit from high-performance and parallel computing. One of the main issues in the bottom-up approach is the estimation of the model’s unknown parameters. In general there is an infinite set of parameter values that can give rise to the same system’s behaviour for a given condition. Many algorithms have been designed to estimate the most likely set of parameter values according to some biological criterion (see [11] for a recent review). All these algorithms however require a computing time that increases exponentially with the number of unknown parameters. Parallelizing such computation and making it run in a supercomputer would speed the process of parameter evaluation up to some order of magnitude. In the top-down approach, the size of the data routinely produced is such that big-data tools and methods, such as HPC, become increasingly necessary.

Neil Swainston (University of Manchester, UK)

The sequencing of entire genomes, their functional annotations and the vast knowledge already available about enzyme-reaction relationships allow computational systems biologists to generate reconstructions of large-scale metabolic networks. Although a curated reconstruction implies an intensive, manual curation process, the generation of a first draft reconstruction is done automatically through web-services or client-based software packages [12,13]. The automatic generation of a metabolic map can start either from prior knowledge of the pathways involved in the network (as in Path2Models [14]), or just from the organism genome. In this latter case the computational requirements are much higher and potentially needing HPC.

These genome-scale reconstructions can be seen as a first step in the endeavour of creating fully-characterized large-scale kinetic models. Despite the lack of knowledge about the parameter values associated with each reaction in the network, such reconstructions can still prove informative. In fact they are widely used to understand the metabolic capabilities of the organism and identify which processes can or cannot be carried out based on the stoichiometry of network. This kind of analysis does not necessitate a high computing power, but it fails to capture the highly non-linear nature of the system’s dynamics. On the contrary, to reach the goal of building and using a fully characterised kinetic model the use of HPC may likely be required with respect to two main issues:

- Parameter estimation of the network;
- Dynamic simulation of one or more interacting

genome-scale kinetic models.

Andrzej M. Kierzek (University of Surrey, UK)

The main challenge in producing fully characterised *in silico* representations of large biochemical systems is the experimental quantification of all the parameters involved in the underlying set of molecular interactions. When only a small subset of the parameters is known, a possible way around consists of integrating different modelling techniques. On the one hand Flux Balance Analysis (FBA) is already widely used to produce semi-quantitative predictions of cellular metabolic phenotypes on a genome-wide scale without knowledge of kinetic constants and molecular concentrations. On the other hand, stochastic and ODE-based kinetic models already exists for small portions of the biochemical network, for which the relevant parameters have been estimated. An important direction in which Computational Systems Biology is currently moving is the integration of different kinds of models that rely on different analytical approaches. The Quasi Steady State Petri Net (QSSPN) is a hybrid simulation algorithm developed by Andrzej Kierzek's group [15]. QSSPN integrates qualitative rule-based models, stochastic kinetic models, deterministic kinetic models and FBA models in one single framework. QSSPN allows for the iterative improvement of genome-scale networks by plugging in newly measured kinetic parameters thus increasing the level of detail of the model.

QSSPN would greatly benefit from being implemented in a supercomputing environment with respect to three main applications:

- Exploration of alternative molecular events in qualitative simulations leading to the same outcome;
- Exploration of alternative parameter sets in hybrid models leading to the same predicted behaviour;
- Support for large research communities performing simulations in an interactive online environment.

Chengkun Wu (University of Manchester, UK)

Apart from the bottom-up and top-down approaches aforementioned there is a third way that is being exploited to attempt the automatic reconstruction of biochemical networks. This alternative approach is based on text-mining techniques and consists of automatically discovering and extracting relevant information from literature about the relationships between different factors pertaining to a given biological system [16]. These relationships may be very diverse in nature, from molecular interactions to causative events in the emergence of a medical condition. This approach turns out to be very useful, for example, when complementing other automatic techni-

ques of model generation, such as in genome-scale metabolic reconstructions. Another application of large-scale text mining aims to discover the molecular bases of different diseases and builds a comprehensive molecular interaction map of any disease or biological process of interest based on literature data.

Text mining requires significant computational resources but, by its nature, can be easily parallelized. Some tasks however may require months if applied on the whole corpus of the MEDLINE repository and all open-access full-texts. That is why text mining would greatly benefit from harnessing the power of HPC.

Computational Biology for Industry

Steve Marciniak (SimCyp, Sheffield, UK) and Amin Rostami (Manchester Pharmacy School, UK)

The SimCyp consortium has developed a platform designed to conduct physiologically based pharmacokinetic modelling (PBPK) as well as pharmacokinetic/pharmacodynamic (PK/PD) studies on virtual patient populations. This platform includes numerous databases containing human physiological, genetic and epidemiological information. By mapping single patients' clinical data against this corpus of information, the SimCyp Simulator allows for the prediction of PK/PD behaviour in 'real-world' relevant contexts [17]. Another benefit of using virtual populations instead of a single virtual human reference is that individuals at higher-than-average risk may be identified.

The platform is the result of an ensemble of models that have been developed since 2001. Each year the consortium identify an area where to focus R&D efforts. SymCyp exerts an important influence on regulatory agencies. The Food and Drug Administration, for example, now regards *in silico* simulations as a sound validation tool to inform decision-making policies [18]. The projections for 2020 show that many R&D activities in the pharma industry will occur *in silico*.

To enable the spread of the platform, while maintaining or even improving its performance, Grid and Cloud computing are key factors and represent a direction towards which SimCyp is interested to move.

Janette L. Jones (Unilever, UK)

Systems Biology is becoming an integral part of personal care product development in Unilever. The biochemical interactions between microbes and their human substrate needs to be understood in order to optimize product efficiency and safety. Today products are expected to satisfy higher and higher safety standard and this implies to also account for personal variability. As a consequence

the underlying biological problem becomes extremely complex and brings about unprecedented challenges. Industry predicts that the integration of biological big-data with Systems Biology analytical approaches will transform the way R&D is carried out in the next decade. Unilever has already adopted *in silico* modelling approaches. A multi-layered model of epidermis, for example, has been developed to allow *in silico* clinical trials for surfactant-induced skin damage.

To enlarge the scope of such models and also allow *in silico* analysis based on personal variability HPC is regarded as a promising area to direct and invest resources.

Douglas B. Kell (University of Manchester, UK)

In the development of new drugs, the system properties of the targeted biological network have to be considered with many respects. Although today there is rightly a big emphasis on putting the molecular drug target in the context of its interactome, the mechanism through which drugs enter the cells is usually overlooked. There is abundant evidence in literature that the uptake of drugs is carrier-mediated rather than 'passive' [19–21]. This implies that drugs enter the cells using carriers that are normally involved in the transport of natural metabolites. The identification of such carriers and the study of this competition mechanism is essential to understand how the drug intake affects the host. This is a Systems Biology problem of high relevance for the pharma industry. To address this problem it is necessary to create and analyse *in silico* representations of the underlying biochemical networks.

Not only network modelling, but molecular modelling of docking of drugs onto their transporters and then of course their integration, are complex enough to require supercomputing.

Computational Systems Biology in Education

Gerold Baier (SysMIC and University College of London, UK)

Given the interdisciplinarity of Systems Biology it is pivotal to train a new generation of scientists who have an understanding of the computational tools and approaches available today in life science. SysMIC offers a comprehensive online course covering a wide array of mathematical, computational and engineering notions relevant to Systems Biology.

This educational platform is currently used by more than 700 UK-based researchers, mainly doctoral students and principal investigators, and is increasingly gaining the attention of bio-industry professionals. SysMIC has a

modular structure comprehensive of basic skills, advanced skills and project-based work. The training approach consists of: (i) presenting a biological problem, (ii) providing the relevant computational knowledge to address the problem and (iii) making the student apply the learned techniques to carry out a proper solution strategy.

The institutions involved in SysMIC are the University College London, the Open University, the Birkbeck College and the University of Edinburgh.

Computational Biology on the Tianhe-2 Supercomputer

Shao-Liang Peng (National University of Defence Technology, Changsha, CHINA)

With high-throughput technology becoming more affordable each year, virtually every biology lab has the potential to become a big-data generator. Existing biology tools cannot cope with the rapid growth of data scale, as the figures are in petabytes. The need to handle and extract information from such an enormous amount of data leaves no doubt about the necessity to resort to high-performance computing.

China has emerged as an important player in HPC [22,23]. Built by China's National University of Defence Technology (NUDT) in collaboration with the Chinese IT firm Inspur, Tianhe-2 is today the fastest supercomputer in the world [24]. With a peak speed of 55 petaflops, 1 petabyte of memory and 12 petabyte of storage capacity, Tianhe-2 can deal with computation-, memory-, and communication-intensive workloads.

Tianhe-2 is already used in computational biology applications, mainly related with sequence analysis/assembly and genome wide annotation studies, and runs well-known bioinformatics software such as BWA, SOAP3-dp, SOAP-denovo 2, SOAP-fuse, SOAP-snp [25–28]. Tianhe-2 staff can also design software solutions to make a given computational problem take full advantage of the computing power of the facility.

In the future, Tianhe-2 is expected to be open for online use to researchers based all over the world.

Naiyang Guan (National University of Defence Technology, Changsha, CHINA)

The availability of HPC resources should not come at the expense of implementing well-designed algorithms to solve computationally demanding problems. In fact these two aspects (computing power and efficient algorithms) have to be used synergistically to obtain fast and reliable results. An example of such synergy is provided by the application of a non-negative matrix factorization (NMF) solver on supercomputers. NMF is a dimension reduction

method used to approximate a non-negative matrix by the product of two lower-rank non-negative matrices. This method is particularly suitable for microarray clustering, modular pattern discovery and noise-filtering of data. There exist many NMF algorithms, but they all suffer from at least one of the following drawbacks: (i) slow convergence, (ii) high computational complexity, (iii) numerical instability. The algorithm proposed by Dr. Naiyang Guan has proved to be much faster than previous approaches and has been used successfully in gene-expression clustering [29]. In particular NeNMF has proved to be superior to other representative NMF solvers in terms of efficiency as well as approximation accuracy. It also overcomes the numerical instability problem suffered by previously proposed algorithms.

DISCUSSION

How can Systems Biology benefit from supercomputing?

The following main areas related to computational biology were identified where HPC can be profitably applied:

- Model construction;
- Model parameterization and sensitivity analysis;
- Extensive analysis of large biochemical networks;
- Integration of models at different spatio-temporal scales;
- Software development;
- Visualization of complex biological data;
- Accessibility of high-performance computing resources;
- Computation of millions of individual virtual humans in concert.

Although there may be some overlap, the goals set in these macro-areas can greatly differ in the time-scale required to harvest expendable results. Some problems are virtually ready to be submitted to an HPC facility while others require more time to find a formal definition and identify a rigorous solution strategy. In terms of “HPC-readiness” the most tractable problems are probably those related with model (re)construction, parameter estimation, and with parallel computation of multiple individuals with integration of the results. Data structures and efficient algorithms already exist for these procedures and what is often needed is just more computing power. In the case of generation of biochemical networks from genomes, the required data are already stored, formatted and retrievable in a usable way, while tested numerical approaches are regularly used to build the interactome from the extracted information. The refinement of first-draft reconstructions through text-mining is also an established procedure, although computationally

demanding, and is easy by its very nature to be parallelized.

To deal with the lack of knowledge about the values of the system parameters there are two main numerical approaches. One consists of making an estimation of the parameters based on the known outcome of the system under specific conditions. There is an abundance of algorithms that address this problem but the computing power needed increases very rapidly with the number of independent parameters. There is great interest in using HPC for this type of parameter estimation as this represents an important step for bridging the gap between biochemical reconstructions and fully characterized kinetic models. The second approach consists of exploring the region of the parameter space that is compliant with the known outcome of the system. In this case Monte-Carlo sampling is commonly used to gain a probabilistic insight of some relevant system properties such as sensitivity coefficients [30–32]. Although random sampling can often get away with the combinatorial explosion of a systematic parameter screening without compromising prediction accuracy, its application to genome-wide networks or even multi-scale models is likely to greatly benefit from HPC resources. The very nonlinearity of the relevant networks that causes functional emergence, may not be captured by low-density parameter sampling. Hence HPC for higher density sampling would be beneficial.

Another way to deal with the unknown quantities of a system is to explore the range of its possible outcomes within the topological restrictions imposed by the underlying network. This is what constraint based modelling techniques, such as FBA or flux variability analysis (FVA), are commonly used for. A problem of interest in FBA is to assess the effect that multiple gene-knockouts have on the network’s capability to perform specific biological functions. For this kind of study a systematic gene-knockout screening is preferable to a random sampling approach, making the problem combinatorial. For multiple knockouts involving n genes the number of FBA runs (1 for each resulting model) is $(\#Genes)^n$. In the case of triple knockouts on *recon2*, the most comprehensive reconstruction of human cells metabolism [33] today available, the number of FBA runs is $(2194)^3 = 1 \times 10^{10}$, making the problem complex enough to require HPC resources.

The integration of models characterised by processes operating at different spatio-temporal or organizational scales is a problem of different nature. In general, integrating different models (regardless of the scale of the underlying system), requires a standardized approach to make their inputs and outputs “talk” with each other. Unfortunately different models are very often characterized by different degrees of detail, making this task not

trivial. Algorithms to identify conjunction points that allow for the best juxtaposition (with minimal, if at all, loss of information due to the less detailed model) have to be formalized and put in place. Other issues are related to the harmonization of the units and the consistency of some parameters such as thermodynamic constants or kinetic constants being restricted by those. A more ready-to-use approach consists of simulating a culture or a tissue at a cellular level. An ODE model would be created to describe the different cellular functions. This would then be replicated to obtain the number of cells that one wants to consider. In this case the models would be identical (apart from some variability in the value of parameters and initial conditions) and they can be easily integrated to study the system dynamics at a supra-cellular level. For a cell model of only 10 ODEs (hence only capturing the basic cellular functions) and a system of a million cells, the number of ODEs to be solved is already big enough to justify the use of HPC. More detailed cellular models are already available and the tendency is to go towards genome-scale kinetic models, e.g., for the application of individualized medicine. This leads to a degree of complexity that makes the use of supercomputers not only beneficial but essential.

Another aspect to be considered is that there exist already a number of well-established software packages designed to perform a wide range of analyses relevant to Computational Systems Biology. An example is provided by Copasi. A possible area of interest is to enable these programs to take full advantage of HPC by writing their procedures in ways that allow for parallelization. Copasi has been already expanded to some extent in that direction. It currently allows user to run simulations on the Condor High Throughput Computing environment [34]. More work may be undertaken to make it available on other HPC environments such as Tianhe-2.

Promoting UK-China collaborations

To initiate the collaboration between UK and China, the different partners agreed to write project proposals involving aspects of Computational Systems Biology that are readily transferable to a high-performance computing environment. Two of such projects, involving FBA and systematic screening of gene-knockouts, have been already discussed in the previous section. UK post-doctoral research associates will also be visiting the National University of Defence Technology in Changsha, China, to familiarize themselves with the facilities and the research staff. These visits would provide a concrete chance to use Tianhe-2 hands-on.

Further, educational relations will be established by sending students from UK to China and vice versa and promoting the co-authorship of publications on HPC and

Systems Biology.

Periodic workshops will also be organized to keep the different partners updated on the current activities and future plans of the SupraBiology community.

REFERENCES

1. Westerhoff, H. V. and Palsson, B. O. (2004) The evolution of molecular biology into systems biology. *Nat. Biotechnol.*, 22, 1249–1252
2. Liao, W.-L. and Tsai, F.-J. (2013) Personalized medicine: A paradigm shift in healthcare. *Biomedicine*, 3, 66–72
3. Kolodkin, A., Boogerd, F. C., Plant, N., Bruggeman, F. J., Goncharuk, V., Lunshof, J., Moreno-Sanchez, R., Yilmaz, N., Bakker, B. M., Snoep, J. L., et al. (2012) Emergence of the silicon human and network targeting drugs. *Eur. J. Pharm. Sci.*, 46, 190–197
4. Lawford, P. V., Narracott, A. V., McCormack, K., Bisbal, J., Martin, C., Brook, B., Zachariou, M., Kohl, P., Fletcher, K., Diaz-Zuccarini, V. (1921) Virtual physiological human: training challenges. *Philos. T. Roy. Soc.*, 2010, 368
5. Viceconti, M., Clapworthy, G. and Jan, S.V.S. (2008) The Virtual Physiological Human — A European Initiative for in silico Human Modelling —. *J. Physiol.*, 58, 441–446
6. Olivier, B. G. and Snoep, J. L. (2004) Web-based kinetic modelling using JWS Online. *Bioinformatics*, 20, 2143–2144
7. Stoica, I., Sadiq, S.K. and 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
8. Hou, T., Wang, J., Li, Y. and Wang, W. (2011) Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.*, 51, 69–82
9. Mazzeo, M. D. and 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
10. Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., Singhal, M., Xu, L., Mendes, P. and Kummer, U. (2006) COPASI—a COMplex PATHway SIMulator. *Bioinformatics*, 22, 3067–3074
11. Cazzaniga, P., Damiani, C., Besozzi, D., Colombo, R., Nobile, M.S., Gaglio, D., Pescini, D., Molinari, S., Mauri, G., Alberghina L., et al. (2014) Computational strategies for a system-level understanding of metabolism. *Metabolites*, 4, 1034–1087
12. Karp, P. D., Paley, S. and Romero, P. (2002) The pathway tools software. *Bioinformatics*, 18, Suppl. 1, S225–S232
13. Pinney, J. W., Shirley, M. W., McConkey, G. A. and Westhead, D. R. (2005) metaSHARK: software for automated metabolic network prediction from DNA sequence and its application to the genomes of *Plasmodium falciparum* and *Eimeria tenella*. *Nucleic Acids Res.*, 33, 1399–1409
14. Büchel, F., Rodriguez, N., Swainston, N., Wrzodek, C., Czauderna, T., Keller, R., Mittag, F., Schubert, M., Glont, M., Golebiewski, M., et al. (2013) Path2Models: large-scale generation of computational models from biochemical pathway maps. *BMC Syst. Biol.*, 7, 116
15. Fisher, C. P., Plant, N. J., Moore, J. B. and Kierzek, A. M. (2013) QSSPN: dynamic simulation of molecular interaction networks describing gene regulation, signalling and whole-cell metabolism in human cells. *Bioinformatics*, 29, 3181–3190
16. Wu, C., Schwartz, J.-M. and Nenadic, G. (2013) PathNER: a tool for

- systematic identification of biological pathway mentions in the literature. *BMC Syst. Biol.*, 7, Suppl. 3,S2
17. Chetty M., Rose, R.H., Abduljalil, K., Patel, N., Lu, G., Cain, T., Jamei, M. and Rostami-Hodjegan, A. (2014) Applications of linking PBPK and PD models to predict the impact of genotypic variability, formulation differences, differences in target binding capacity and target site drug concentrations on drug responses and variability. *Front. Pharmacol.*, 5, 258
 18. Valerio, L. G. Jr. (2011) *In silico* toxicology models and databases as FDA Critical Path Initiative toolkits. *Hum. Genomics*, 5, 200–207
 19. Dobson, P. D. and Kell, D. B. (2008) Carrier-mediated cellular uptake of pharmaceutical drugs: an exception or the rule? *Nat. Rev. Drug Discov.*, 7, 205–220
 20. Kell, D. B., Dobson, P. D., Bilsland, E. and Oliver, S. G. (2013) The promiscuous binding of pharmaceutical drugs and their transporter-mediated uptake into cells: what we (need to) know and how we can do so. *Drug Discov. Today*, 18, 218–239
 21. Lanthaler, K., Bilsland, E., Dobson, P. D., Moss, H. J., Pir, P., Kell, D.B. and Oliver, S.G.(2011) Genome-wide assessment of the carriers involved in the cellular uptake of drugs: a model system in yeast. *BMC Biol.*, 9, 70
 22. Yang, X., Liao, X., Xu, W., Song, J., Hu, Q., Su, J., Xiao, L., Lu, K., Dou, Q., Jiang, J., et al. (2010) TH-1: China's first petaflop supercomputer. *Front. Comput. Sci. China*, 4, 445–455
 23. Yang, X., Liao, X.-K., Lu, K., Hu, Q.-F., Song, J.-Q. and Su, J.-S. (2011) The TianHe-1. A supercomputer: Its hardware and software. *J. Comput. Sci. Technol.*, 26, 344–351
 24. Tianhe-2 (MilkyWay-2) Supercomputer. Available from: <http://www.nssc-gz.cn/>
 25. Luo, R. B., Liu, B.H., Xie, Y. L., Li, Z.Y., Huang, W.H., Yuan, J.Y., He, G. Z., Chen, Y. X., Pan, Q., Liu, Y. J. et al. (2012) SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience*, 1, 18.
 26. Luo, R., Wong, T., Zhu, J., Liu, C. M., Zhu, X., Wu, E., Lee, L. K., Lin, H., Zhu, W., Cheung, D. W., et al. (2013) SOAP3-dp: fast, accurate and sensitive GPU-based short read aligner. *PLoS One*, 8, e65632
 27. Jia, W., Qiu, K., He, M., Song, P., Zhou, Q., Zhou, F., Yu, Y., Zhu, D., Nickerson, M. L., Wan, S., et al. (2013) SOAPfuse: an algorithm for identifying fusion transcripts from paired-end RNA-Seq data. *Genome Biol.*, 14, R12
 28. Cui, Y., Liao, X.-K., Zhu, X. Q. and Wang, B. Q. (2014) mBWA: A massively parallel sequence reads aligner, in 8th international conference on practical applications of computational biology & bioinformatics (PACBB 2014). Springer International Publishing, 113–120
 29. Guan, N.Y., Tao, D. C., Luo, Z. G. and Yuan, B. (2012) NeNMF: An optimal gradient method for Nonnegative Matrix Factorization. *IEEE Trans. Signal Process.*, 60, 2882–2898
 30. Murabito, E., Smallbone, K., Swinton, J., Westerhoff, H. V. and Steuer, R. (2011) A probabilistic approach to identify putative drug targets in biochemical networks. *J. R. Soc. Interface*, 8, 880–895
 31. Murabito, E., (2013) Targeting breast cancer metabolism: A metabolic control analysis approach. *Curr. Synthetic Sys. Biol.*, 1, 104
 32. Murabito, E., Verma, M., Bekker, M., Bellomo, D., Westerhoff, H. V., Teusink, B. and Steuer, R. (2014) Monte-Carlo modeling of the central carbon metabolism of *Lactococcus lactis*: insights into metabolic regulation. *PLoS One*, 9, e106453
 33. Thiele, I., Swainston, N., Fleming, R. M., Hoppe, A., Sahoo, S., Aurich, M. K., Haraldsdottir, H., Mo, M. L., Rolfsson, O., Stobbe, M. D., et al. (2013) A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.*, 31, 419–425
 34. Kent, E., Hoops, S. and Mendes, P. (2012) Condor-COPASI: high-throughput computing for biochemical networks. *BMC Syst. Biol.*, 6, 91