from ~5,000 to 33,000 but the overlap is low. In addition, no single method has implicated >60% of known yeast proteins in these interactions. As shown by von Mering *et al*., greater coverage of protein–protein interaction space by any method corresponds to lower accuracy and vice versa. Thus, combining different methods leads to increased accuracy (when compared with carefully annotated reference sets of protein–protein interactions). Another interesting observation is that most of the interactions identified come from experiments. Only ~7,500 interactions have been predicted using computational methods. This can be rationalized by the fact that the algorithms currently used are comparative in nature or largely dependent on orthology, as mentioned above. Thus, interactions between proteins that do not share such relationships would be difficult to predict.

Clearly, the systematic study of protein–protein interactions and networks is expected to add another dimension to target identification, despite current limitations. To date, most studies have by necessity been carried out in prokaryotes or yeast and are not readily transferable to the human proteome. Also, the overall accuracy of high-throughput methods is still limited (as indicated, for example, by their low overlap). However, an attractive feature of considering protein–protein interactions and networks is that this might greatly help to identify 'beyond homology' targets outside established protein (super)families that would be difficult to find using other *in silico* approaches.

References

1 Duckworth, D.M. and Sanseau, P. (2002) *In silico* identification of novel therapeutic targets. *Drug Discov. Today* 7, S64–S69

2 Schächter, V. (2002) Protein interaction networks: from experiments to analysis. *Drug Discov. Today* 7, S48–S54

3 von Mering, C. *et al.* (2002) Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* 417, 399–403

Jürgen Bajorath
Computer-Aided Drug Discovery
Albany Molecular Research
Bothell Research Center
18804 North Creek Pkwy
Bothell, Washington 98011, USA

From screen-saver to virtual screener: harnessing latent PC power through distributed computing

A typical desktop PC spends greater than 90% of its time doing nothing, yet today's PC computer processing unit (CPU) technology is highly competitive with that of high performance Unix workstations. The potential exists, therefore, to release many more CPU cycles than have generally been available for compute-intensive computational chemistry calculations. As a consequence, distributed (grid-based) computing stands to provide the next great leap forward in computing power for those engaged in computer-aided molecular design (CAMD).

Davies and Richards have eloquently summarized initial efforts in this regard in their recent review on the subject [1]. The discourse concentrates primarily on virtual screening (VS) applications running on individually donated screen-saver time across the Internet. Contrasting two of the highlighted examples, Dockcrunch [2] and CAN-DDO [1], illustrates the potential of this technique. Using a 64 processor Silicon Graphics server (which only a couple of years ago would have cost millions of dollars), Dockcrunch screened 1.1 million compounds in six days. By contrast, CAN-DDO screened 3.5 billion compounds across eight targets inside nine months using only spare PC CPU cycles.

These numbers clearly illustrate the attractive nature of distributed computing. There are, however, specific technical issues that need to be addressed before the technology can be used to its full advantage. First and most importantly, a distributed computing application must run transparently on any computer to which it is assigned. Individuals will only donate CPU cycles willingly when they see no cost to themselves. This restriction leads into the problem of application selection and design. It is essential that applications have a memory footprint able to run cleanly on any assigned PC. The software must also be compiled to run on the operating systems of these PCs, and should have a low IO (input-output) requirement to permit coarse grain parallelization, because distributed computing is by its nature network bound. IO issues extend to the server, which must have a queuing system able to cope with the large and discontinuous amounts of data being pushed in and out of the attached PCs. It must also deal with failed jobs resulting from PC shutdown, and so on,with comprehensive restart and job monitoring facilities.

For pharmaceutical companies interested in leveraging distributed computing technology, many similar issues exist, although with subtle differences. The inevitable need for confidentiality is such that most within the industry will only countenance the use of intranet PCs. This places an intrinsic limit on available resources, although for larger companies these are still considerable (probably 1–2 orders of magnitude increase in available CPU cycles). There are certain advantages to such an approach, however. Tighter control of PC resources permits the use of more flexible queuing. This in turn allows the use of applications with larger memory footprints to access only those PCs able to cope with the additional RAM requirement. It is also possible to

upgrade certain elements of a PC inventory to meet these needs. In addition intranet IO is generally superior to that on the Internet, somewhat reducing application IO limitations.

VS technology is the first family of CAMD applications to see use in a distributed computing environment, because it fits in naturally with the technological demands. The examples described by Davies and Richards focus on the 'more is better' approach in its exploitation. This raises a significant issue, however, because in a typical virtual screen, at least as much time is spent post-processing the results as is in obtaining them [2]. Using today's VS technology with its inherent approximations [3], larger database screens are liable to bring with them larger hit sets with many false positives, making post-screen analysis an increasingly intimidating prospect. If we are to fully harness the potential of distributed computing over the long term, more headway will probably need to be made using the CPU power to run better algorithms, rather than simply running the existing ones faster.

CAMD distributed computing is still in its infancy, with major players in the distributed computing software arena (e.g. entropia: http://www.entropia.com; Platform Computing: http://www.platform.com; and United Devices: http://www.ud.com) continuing to develop code to meet the technological challenges. Further, client side software vendors are still wrestling with the selection of and licensing costs for ported applications in the new environment. Nevertheless, the potential that exists within this computing paradigm is such that it will likely usher in a new and exciting era in CAMD calculations.

References

1 Davies, E.K. and Richards, W.G. (2002) The potential of Internet computing for drug discovery. *Drug Discov. Today* 7, S99–S103

2 Waskowycz, B. *et al.* (2002) Receptor-based screening of very large chemical datasets. In *Rational Approaches to Drug Design*. (Holtje, H-D. and Sipple, W., eds), pp. 372–381, Prous Science

3 Good, A.C. (2001) Structure-based virtual screening protocols. *Curr. Opin. Drug. Discov. Devel.* 4, 301–307

Andrew Good
Bristol-Myers Squibb
5 Research Parkway
Wallingford
CT 06492, USA

Systems biology: the new darling of drug discovery?

Eric Werner, President, Cellnomica, Fort Myers, Florida, and Munich, Germany, e-mail: eric.werner@cellnomica.com, http://www.cellnomica.com

Systems biology is gaining momentum. Japan showed great vision in entering the field early, swiftly followed by other countries including the USA, Canada, the UK and, more recently, other European nations such as France and Germany. It is also the new darling of the investment community; reports on systems biology by Cambridge Healthtech Institute (CHI; http://www.healthtech.com) and Frost and Sullivan (http://www.frost.com) emphasize the financial and commercial promise of this new field. But why is systems biology gaining the attention of governments and the financial sector?

Astronomical amounts of genomic and proteomic data coming in from research laboratories lies dormant, not really understood. Systems biology promises to change all that. It attempts to understand the data by integrating it into computational formal theories and models that explain and predict data thereby making sense of it all. Systems biology merges computer science with mathematics, physics and biology. It is a field where interdisciplinary cooperation and research is a defining feature. The *In Silico Biology* conference (part of CHI's *Beyond Genome 2002* triconference; 2–3 June 2002, San Diego, CA, USA) conference displayed the vigor and the promise of this exciting field.

It was a conference not without controversy. There were claims and counter-claims, scepticism and vision, but such are the hallmarks of a burgeoning field. The *In Silico Biology* conference consisted of four sessions: Networks (modeling and cellular networks and pathways), Systems Modeling (modeling whole systems including organs and tissue with the aim of a global integration of networks with data), Cell Modeling (mathematical and computational simulation of

1359-6446/02/$ – see front matter PII: S1359-6446(02)02403-0