12 Kirton, M.J. (1976) Adaptors and Innovators: a description and a measure. *J. Appl. Psychol.* 61, 622–629
13 Kirton, M.J. (1987) Adaptors and Innovators: cognitive style and personality. In *Frontiers of Creativity Research: Beyond the Basics* (Isaksen, S.G. ed.), pp. 282–304, Bearly Limited, New York
14 Dale, M. (1994). Learning organizations. In *Managing Learning* (Mabey, C. and Iles, P., eds), The Open University, Thomson Business Press
15 Stork, D. (1998) Not all differences are created equal: Not all should be managed the same: the diversity challenge in pharmaceutical R&D. *Drug Dev. Res.* 43, 174–181

The *Discussion Forum* provides a medium for airing your views on any issues related to the pharmaceutical industry and obtaining feedback and discussion on these views from others in the field. You can discuss issues that get you hot under the collar, practical problems at the bench, recently published literature, or just something bizarre or humorous that you wish to share. Publication of letters in this section is subject to editorial discretion and company-promotional letters will be rejected immediately. Furthermore, the views provided are those of the authors and are not intended to represent the views of the companies they work for. Moreover, these views do not reflect those of Elsevier, *Drug Discovery Today* or its editorial team. Please submit all letters to Joanna Owens, Acting News & Features Editor, *Drug Discovery Today*, e-mail: Joanna.Owens@elsevier.com

Predicting novel proteins and their interactions

Different experimental and computational approaches are experiencing significant interest in the hunt for new proteins as potential therapeutic targets. Two contributions in the recent *Information Biotechnology* supplement to *Drug Discovery Today* focus on *in silico* identification of target proteins [1] and the study of protein interaction networks [2]. In addition, an excellent analysis of currently available yeast protein interaction data has recently been published in *Nature* [3]. Taken together, these reports highlight the potential of such approaches for target discovery.

The *in silico* approach to target identification is largely based on (increasingly sophisticated) detection of sequence homology, both at the DNA and protein level [1]. Typically, genome sequences are scanned for detectable similarity to protein domains of known target families or superfamilies (such G-protein-coupled receptors or protein kinases). By contrast, the basic idea of protein–protein interaction analysis is to identify novel proteins that interact with partners implicated in disease states or cellular networks involving disease-related proteins [2]. Computer methods applied in this context generally do not predict protein–protein interactions directly but attempt to establish functional relationships between proteins, which are thought to often involve actual binding events. Essentially, these methods are based on the analysis of gene fusion and proximity or orthology and phylogenetic profiles and have thus far mostly relied on the comparison of prokaryotic genomes.

Recently, von Mering *et al*. have compared protein–protein interactions in yeast identified by, or inferred from, different high-throughput technologies [3]. Some of their data are summarized in Figure 1. A total of ~80,000 yeast protein interactions have been detected, to date. The number of interactions identified by different methods ranges

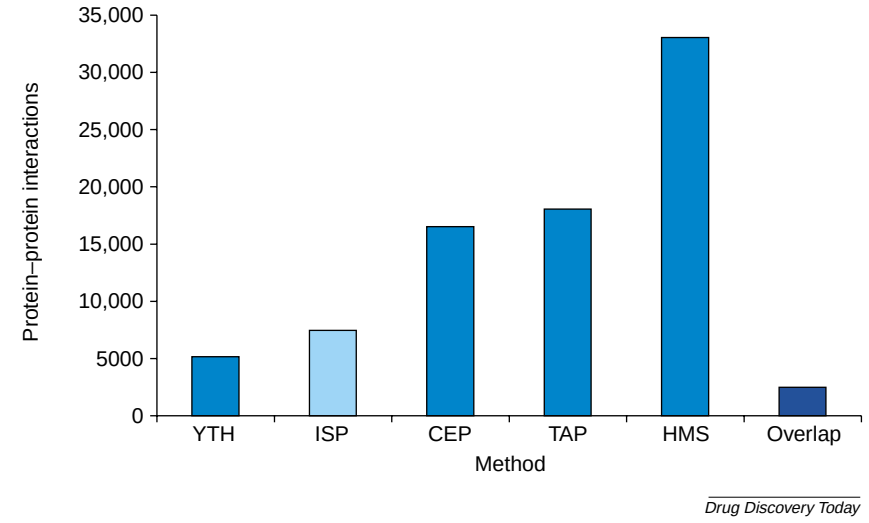

Figure 1. Protein–protein interactions in yeast determined by different high-throughput methods. Abbreviations: YTH, yeast-two-hybrid; ISP, *in silico* predictions; CEP, correlated expression profiling (mRNA arrays); TAP, tandem affinity purification; HMS, high-performance mass spectrometry. 'Overlap' is defined here as the number of interactions identified by more than one method. Data were taken from reference [3].

  PII: S1359-6446(02)02446-7

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