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 Rebecca Lawrence, News & Features Editor, *Drug Discovery Today*, e-mail: Rebecca.Lawrence@drugdiscoverytoday.com

Toxicoproteomics: learning to walk before it can run

The genomics revolution has spawned a new lexicon of 'omics' terms, each representing a technology either claiming to be the next 'big thing' or, more realistically, claiming to complement those already in place. Toxicoproteomics can be regarded as the daughter of toxicogenomics and proteomics. It enters an uncertain world, its parents not yet being fully established as high value contributors in the world of pharmaceutical R&D. However, toxicoproteomics has potential; as long as it doesn't claim it can run before it can walk.

The study of proteins, rather than genes or messenger RNA, has appeal because it is the proteins that do most of the work in (and out of) the cell, as well as comprise much of the structural components of cells, tissues and organs. However, unlike DNA and RNA, which are relatively simple chemicals to work with, proteins are highly heterogeneous both in their primary composition as well as their potential for secondary modification (e.g. glycosylation and phosphorylation). This heterogeneity has created many technical hurdles to the study of the 'proteome'. As proteomics (and toxicoproteomics) sets out to understand the biological significance of changes in the proteome it is important that the proteome can first be resolved and then the individual components be characterized. Currently, technologies have been developed that go some way to achieving these objectives. Resolution of the proteome can be achieved through approaches such as two-dimensional (2D) gel electrophoresis. However, the heterogeneity of size and physical properties (the isoelectric point in this case) requires a large number of gels to be run to attempt to resolve the proteome of any biological sample (and this is assuming suitably sensitive detection methods are available). Basically, 2D gels can identify changes in protein composition but cannot address the biological meaning of the changes. To even attempt to do this, the identity of the 'changed' proteins needs to be determined. This can be done by subjecting individual protein spots on the 2D gel to MS to determine protein or peptide sequence followed by the analysis of this sequence against protein databases.

Armed with the knowledge that an identified protein has changed in amount or composition in a sample, compared with other proteins, one might be able to form a hypothesis as to the cause and consequences of such a change. However, in many cases this is going to be a speculative hypothesis that will require the development of much corroborating information. An alternative would be to forego the mechanistic approach (at least temporarily) and focus on associations. Without understanding the biological mechanisms at play, one might be able to associate certain changes with a biological outcome. Although intellectually less satisfying, it is probable that the young field of toxicoproteomics will initially develop along these lines.

Pharmaceutical executives like to see their technology investments pay off (in the short term as well as in the long term). To mature into a hypothesis and mechanism-based science, proteomics and toxicoproteomics will probably have to demonstrate their use in the application of specific phenomenological problems. This will be the walk – the run will come as the technology, databases and analytical tools gain more sophistication.

Paul Spence
Vice President, Biotechnology
Pharmacia
700 Chesterfield Parkway North
Chesterfield, MO 63198, USA

The feedback cycle of conventional and virtual screening

Developments in virtual screening methods are helping to establish these computational techniques as an important complement to conventional screening in the drug discovery process [1]. There are, however, several practical aspects to their application.

For the foreseeable future, it seems inevitable that we shall require a biological assay to test compounds. The scoring functions used by the computational methods are not sufficiently reliable to provide accurate predictions of potency. It is also

1359-6446/02/$ – see front matter PII: S1359-6446(02)02315-2

important to appreciate that virtual methods, by definition, cannot model unpredictable behaviour. An unmodelled (and perhaps previously unobserved) conformational change in the protein target, for instance, could provide important new modes of binding that could be detected through experimental methods, but would not be captured through theoretical methods. The real strength of virtual methods lies in their ability to search compounds that do not yet exist.

It is now routine to use similarity-searching methods to pull out analogues of screening hits to determine rapid SARs and also to capture any compounds that might have been missed through HTS. Successful generalization of the model for binding (e.g. through a pharmacophore) can enable the identification of alternative lead series, providing lead optimization programmes with a choice and helping circumvent issues further downstream (e.g. ADME liabilities). It can also be highly advantageous to use virtual screening methods to provide a focused set for screening ahead of HTS, potentially providing early discovery of lead series for chemical optimization while awaiting the results of a more exhaustive search of the whole screening collection.

The successful large-scale application of virtual screening methods requires intelligent prioritization of filtering and selection techniques dependent on computational efficiency. Simple property filters, such as those used to ensure adherence to Lipinski's Rules or substructural searches used to eliminate reactive groups, should be applied early. Substructural searches can also be used as a precursor to a pharmacophore query, enforcing the presence of a particular combination of features before using the more computationally expensive pharmacophore search to ensure that the features are correctly disposed relative to one another in three dimensions. If required, more complex prediction algorithms (e.g. blood–brain barrier models, cytochrome P450 models) can be applied either before or after ligand-docking procedures, depending on the time required for the calculation.

For many, the use of docking procedures is limited by the lack of crystal structures. This is probably most acute in GPCR-targeted work, where the only currently available crystal structure is that of bovine rhodopsin, although here, as with other targets, the use of high-quality homology models can bridge the gap.

Conventionally, compound selection and design were somewhat distinct approaches. Known available compounds could be docked into a target protein and assessed for probable potency. Alternatively, compounds could be assembled *de novo* to fit the constraints of the binding site. Now, however, we have more of a spectrum of approaches available. Vast virtual libraries are emerging for validated chemistries [2] that enable the use of computational methods to match chemistry to targets of interest (in essence, the reverse of *de novo* drug design because we are starting with known chemistry and working to the target). Moreover, experimental techniques such as NMR [3] or X-ray crystallography [4] can be employed to detect molecular fragments that bind well to targets of interest, and these could be used as seeds in the virtual screening process.

Some authors have advocated the use of fragment frequency data derived from previous experience in the selection of arrays for synthesis (e.g. RECAP [5]). More generally, a knowledgebase is starting to evolve regarding 'privileged' fragments for particular target classes [6]. The use of these approaches seems likely to increase the success rate in providing potent molecules, but will naturally lead to a lack of diversity in the hits obtained. However, most importantly these approaches could prove problematic if, for instance, ADME liabilities have been an issue in the past – these could be reinvented for future programmes.

Consequently, it is important to track successful approaches and use the results to improve future selection work. The performance of conventional screening should be analyzed, seeking to identify frequently occurring false positives and determine if promiscuity is assay-format dependent. There should be an attempt to rationalize 'dirty' behaviour and to remove compounds from the screening collection if appropriate. This might be inappropriate because the compound is only a problem in certain screens, or it could be impractical to remove the compound because of plating issues. Under those circumstances, it is particularly important to provide computational tools as a means to quickly filter screening hits electronically and to help chemists identify the best hits for subsequent optimization.

References

1 Schneider, G. and Böhm, H-J. (2002) Virtual screening and fast automated docking methods. *Drug Discov. Today* 7, 64–69

2 Leach, A.R. and Hann, M.M. (2000) The *in silico* world of virtual libraries. *Drug Discov. Today* 5, 326–336

3 Hajduk, P.J. *et al.* (2000) Privileged molecules for protein binding identified from NMR-based screening. *J. Med. Chem.* 43, 3443–3447

4 Nienaber, V.L. *et al.* (2000) Discovering novel ligands for macromolecules using X-ray crystallographic screening. *Nature* 18, 1105–1108

5 Lewell, X.Q. *et al.* (1998) RECAP – retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* 38, 511–522

6 Patchett, A.A. and Nargund, R.P. (2000) Privileged structures – an update. In *Annual Reports in Medicinal Chemistry.* (Vol. 35), (Trainor, G.L. ed.), pp. 289–298, Academic Press

Stephen L. Garland
Senior Investigator,
7TM Systems Modelling Group,
Computational, Analytical and Structural Sciences Department,
GlaxoSmithKline Pharmaceuticals
New Frontiers Science Park (North),
Harlow, Essex, UK CM19 5AW

1359-6446/02/$ – see front matter PII: S1359-6446(02)02316-4