in silico structure-based drug design —

SHIFTING THE BOTTLENECK?

The use of *in silico* methods in drug design has grown significantly in popularity over the past couple of years. Indeed many pharma companies have already adopted some type of virtual screening capability to complement HTS methods. Will we eventually see this technology replacing HTS and removing the burden of discovery chemistry?

The current post genomic era has been characterised by a large increase in the number of potential therapeutic targets amenable to investigation. This growth is, in turn, increasing pressure on the ability of the pharmaceutical industry to prioritise programmes and conduct lead discovery in a highly efficient manner. Such pressure is evident in a recent statistic: over the past 10 years only 25% of quality targets have yielded a quality lead series[1]. Logistical reasons could be put forward to explain this, such as inefficient programme management, but the overriding significance of this figure is that the technology used to generate leads from a given target is not functioning efficiently. The financial implications of this high attrition rate, even to the lead identification stage, are huge. There is an urgent need therefore to review the technologies currently employed in lead identification and critically assess which methodologies are likely to increase productivity at the early discovery stages.

Hit finding the old way

High Throughput Screening (HTS) has traditionally been the most widely used methodology at the hit finding stage of the drug discovery process. To this end, many of the current market drugs have had their precursors identified from compound collections using HTS. However, HTS has some serious weaknesses. In particular, the success of this technique relies upon several important assumptions being satisfied. The first of these is that the compound collection used for screening contains sufficient diversity that novel high affinity active compounds can be found. This is arguably the Achilles' Heal of HTS, in that even a library of one million compounds is a factor somewhere in the region of 10^{59} short of true diversity in terms chemically accessible space. This is interesting given that a large amount of the increase in the lead discovery R&D spend in this phase has been attributed to increased robotisation and parellelisation of the HTS systems to afford the processing of larger numbers of compounds, but that the absolute increase in compound diversity is almost negligible. A second key assumption of HTS is that active compounds can be discriminated from non-actives. This is not necessarily straightforward and many false positives are expected in a modern HTS result set. Indeed, certain chemical structures have recently been termed 'frequent hitters'[2] as they are found to turn up as positives in most assays irrespective of the target.

What's the alternative?

In contrast to HTS, lead discovery using a target structure as a starting point for computational techniques to screen, design and prioritise compounds is promising to be a much more efficient process. This *in silico* structure-based design is

By Dr Jonathan Heal

Cheminformatics

TOOL	VENDOR
DOCK	University of California at San Francisco
Gold	Cambridge Crystallographic Data Centre
AutoDock	Accelrys, Inc
FlexX	Tripos, Inc
Glide	Schrodinger, Inc
ICM	Molsoft, Inc
LigandFit	Accelrys, Inc

rapidly becoming the lead identification corner-stone of many drug discovery processes.

Structure-based *in silico* methods fall into two main categories – virtual screening (VS) and *de novo* design. Virtual screening can be thought of as an *in silico* version of HTS in which a virtual compound library is screened for predicted activity. The predicted actives, if identified, can then be prioritised and tested in a suitable assay for biological confirmation. The most well-known VS tool is DOCK[3], although there are many other commercially available software programs, each encapsulating slightly different theories of the most accurate way of representing a ligand binding to its receptor (**Table 1**). 3D pharmacophore searching is another flavour of VS and can be applied when structural information on the protein target is not available.

De novo design can be distinguished from VS in that the process computationally builds com-

Figure 1
Crystal structure of thrombin and inhibitor

pounds (usually novel) directly into the active site of a target. Again, predicted compounds can then be synthesised and tested in a suitable assay. *De novo* design programmes are also available commercially, such as LUDI[4], but much of the more advanced software remains proprietary to specialised drug discovery companies such as Locus Discovery Inc and De Novo Pharmaceuticals Ltd.

As with any competing technologies, *de novo* design and VS both have their respective strengths and weaknesses. VS has quickly gained popularity because it can be used to screen a company's current compound collection and hits resulting from this process can be tested in a time efficient manner (without synthesis usually). This is the simplest use case of VS and reflects its main application today. The downsides of VS are that it is still a 'screen' and so the problem of compound diversity and suitable library construction remains. It is worth noting that several companies are moving towards a chemical fragment VS screen, in which common molecular building blocks are evaluated for binding. The fragment combination space represents a large chemical space when compared to a ready-built compound library.
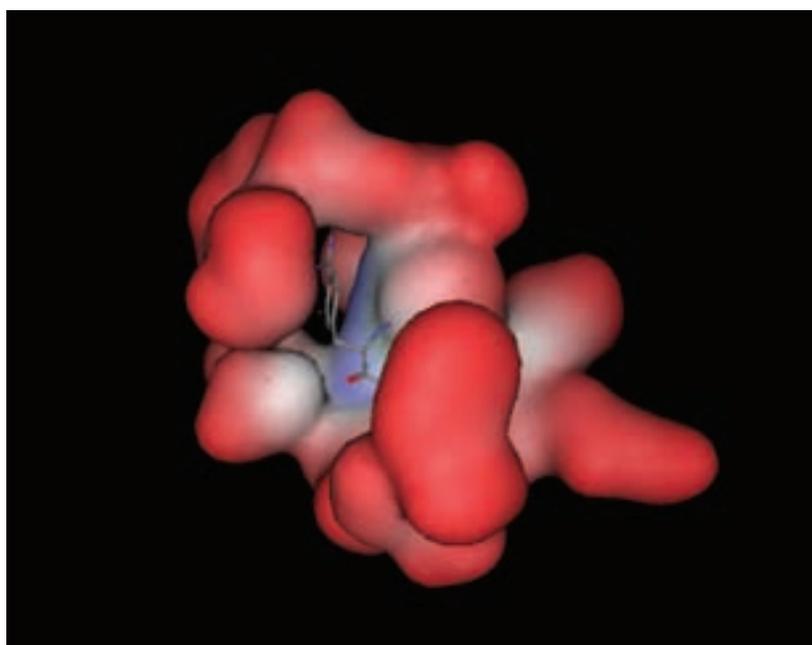
Another criticism of VS is the accuracy of the scoring. This is an ongoing area of research with many companies now adopting a consensus scoring approach in which the predictions of several scoring functions are averaged to provide a more robust prediction. Regardless of research, there is always likely to be a trade-off between accuracy of scoring and the genuine high throughput nature of VS.
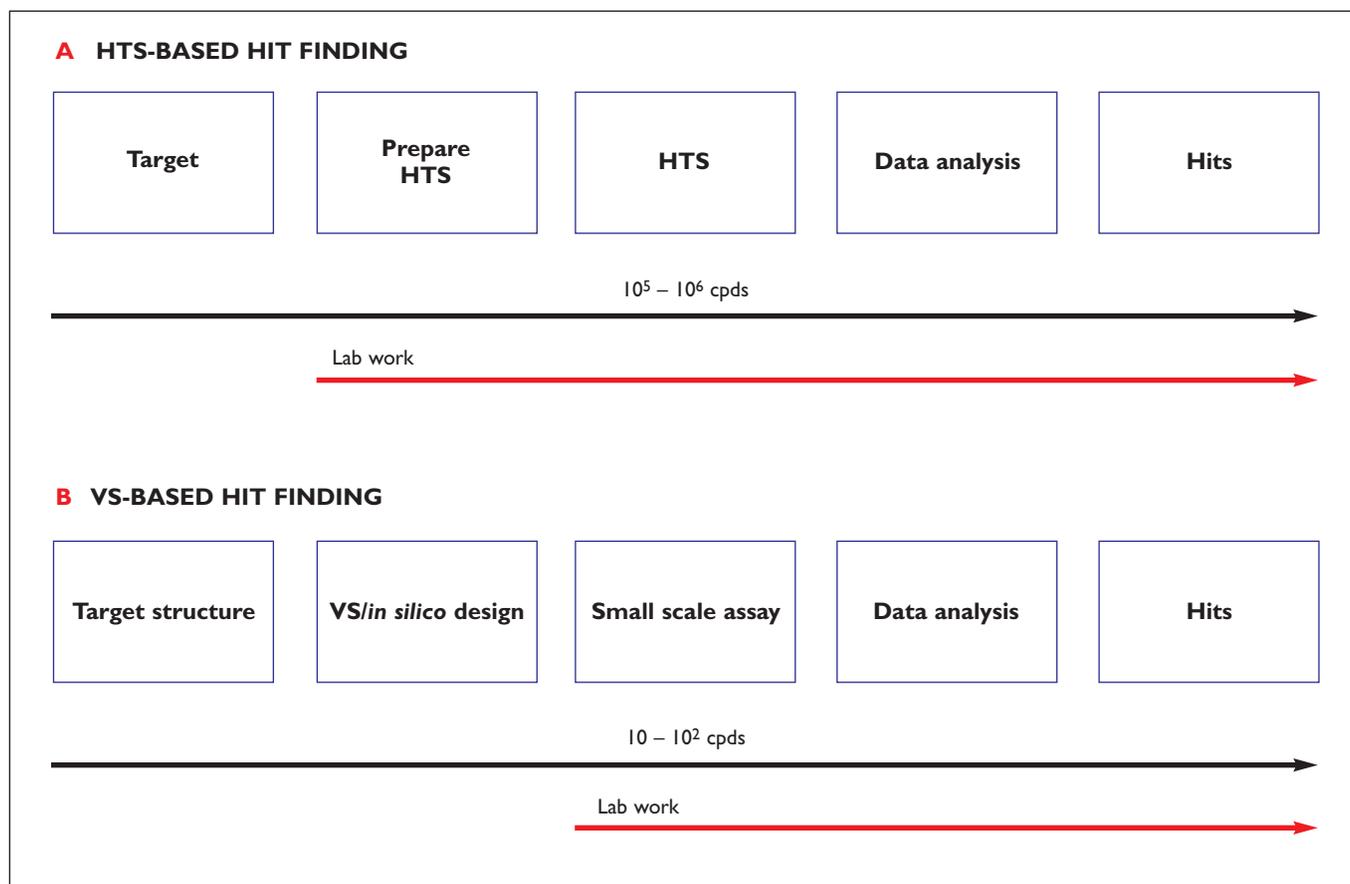
In contrast to VS, *de novo* design methods emphasise the search of novel structures. There are several embodiments of these techniques but the most popular employ a growing strategy in which small chemical fragments are built up within the protein active site. In principle, this approach accesses a vast virtual chemistry space, far in excess of what could ever be biologically screened. As such, this process is likely to locate many new scaffolds, critical in 'me too' drug discovery programmes where the IP coverage on a given target is likely to be heavy. The downside of the *de novo* approach is ensuring the chemical feasibility of the predicted structures which represents an on-going technical challenge.

Realising the productivity shift

The main advantage conferred by structure-based *in silico* techniques is speed. In particular, the time consuming set-up requirements of an HTS for a new target can be circumvented if an *in silico* technique can reliably predict less than 50 compounds

A   HTS-BASED HIT FINDING

Target	Prepare HTS	HTS	Data analysis	Hits

$10^5 - 10^6$ cpds

Lab work

B   VS-BASED HIT FINDING

Target structure	VS/*in silico* design	Small scale assay	Data analysis	Hits

$10 - 10^2$ cpds

Lab work

for testing. This is not an unrealistic expectation of these technologies in that there are a growing number of case studies in which semi-potent hits have been found from tens of predictions. Among these success stories are PTP-1B[5], Thrombin[6] (**Figure 1**) and Factor Xa[6].

The relative cost of these techniques represent another significant advantage over a traditional HTS driven discovery programme. In particular, it is widely accepted that using computer-based predictions prior to committing to lab work is a cost-effective measure. In addition, a computer experiment is much faster to set up than a biological one, so a major cost saving can result from increasing virtual throughput. Of course, nothing in life is free and some initial investment must be made in hardware and software to run a VS or *de novo* design environment, but the total cost of ownership of such solutions is falling rapidly and is certainly not prohibitive for a SME.

Thirdly, and in some respects the most interesting aspect of *de novo in silico* design, is that the virtual chemistry space can chart areas of chemistry unlikely to be featured in non-virtual library based experiments.

A leaner pipeline

The concept of using structural information in the lead optimisation process is not a new one. Traditional lead discovery involves the identification of a hit or series of hits by HTS and using a co-crystallised structure of these hits with their respective receptor to drive a lead optimisation programme. This has proven to be a successful model in certain programmes, but as is shown by the 75% attrition rate in lead discovery – arguably not a very efficient one. In contrast, *in silico*-driven lead discovery alters the pipeline workflow by replacing HTS with an *in silico* screening method and outputs directly into test assays and then into medicinal chemistry (*Figure 2*). Since 40% of research costs are attributed to the lead identification phase of discovery, the cost savings that follow from this leaner methodology should be significant.

In practice what is occurring now is a hybrid between the two extremes where VS is being used to complement HTS and indeed to focus library design for HTS.

An important aspect of this new process is to factor upfront as many desirable ADMET properties

Figure 2
HTS versus VS-based hit discovery workflow. VS-based method replaces HTS with a fast computer-based screen and lowers burden on laboratory resources by requiring a small scale assay

Cheminformatics

References

1 Robeson, BL (2002). Pharmatech, Business Briefing.
2 McGovern SL, Caselli E, Grigorieff N, Shoichet BK (2002). Journal of Medicinal Chemistry 45, 1712-1722.
3 Ewing, TJA et al (2001). J. Comput.-Aided Mol. Design 15,411-428.
4 Bohm, HJ (1992). J. Comput.Aided Mol. Design 8,243-256.
5 Doman, TN et al (2002). J. Med. Chem. 45,2213-2221.
6 Baxter, CA et al (2000). J. Chem. Inf. Comput. Sci. 40, 254-262.
7 Research Collaboratory for Structural Bioinformatics (2002). PDB Annual Report 2002.

into the virtual libraries of prospective leads. Although strictly speaking this is more likely to affect efficacy in clinical trials, rather than purely in lead discovery, the *in silico* process at the front end of lead discovery enables this pre-processing of compounds. Taking these methods together should enable the testing of fewer compounds that individually have a higher potential of making marketable drugs.

One significant caveat for the above revolution in lead finding pathways is that of the accessibility of suitable protein structure data to provide the starting point to a programme. Any company with access to their own X-ray crystallography facility (most medium pharma companies and all large pharma companies qualify here) is in a good position to solve a reasonable structure and begin the process. Many companies have also used NMR for full structural determination, although the more common use of this is to detect specific interactions and build up a picture of the active site. For the rest of us, there is the PDB, which contains nearly 20,000 structures and whose deposition rate is exponential. Given the power of homology modelling software to derive a good structure for a target protein which has a close homologue of known structure, a true structure-based design method should get easier over time. Indeed, it has been predicted that good quality structures for more than 90% of the human proteome will be available by 2005[7].

What does the future hold?

Computational methods such as VS have so far positively impacted resource allocation and expectations of lead identification within some companies' drug discovery efforts. A logical question is to ask to what extent one would expect these methods to completely replace large HTS environments in the future. Not surprisingly, companies with significant investments in HTS are cautious while several SME companies have so far been quite prolific combining only *in silico* approaches and simple assays to identify quality lead series.

In reality, *in silico* approaches are gaining credibility but there is still much work to be done in the area of scoring, correct modelling of protein flexibility and in simplifying the large scale informatics workflow required to routinely perform VS.

A sign that most of these problems are soluble is that the pharmaceutical industry is currently investing resources into all of the above areas.

One significant obstacle is to provide tools to the medicinal chemist such that they can own an *in silico*-driven process. Currently there are no off-the-shelf informatics solutions that provide

this and many companies have to slowly build up their own expertise in house. A change in the working ethic between informatics and chemistry teams is required to make an *in silico*-driven process function successfully. This is surely a rate determining step in the transition in workflow shown in **Figure 2**.

Summary

Because of the significant advantages outlined above, the use of *in silico* methods has grown significantly in popularity over the past couple of years. Specifically, most pharma companies have adopted some type of virtual screening capability to complement HTS methods and it is accepted that the predictions made by these techniques represent a fast method to enrich a biological screen. Some companies have already successfully adopted a far more radical discovery method that uses *in silico* methods to replace HTS.

Drug discovery is often thought of (quite negatively) as a series of never ending bottlenecks. Currently, the largest attrition rates are associated with lead discovery and first trial in man stages. If the attrition rate could be dramatically reduced at the lead discovery stage that would represent a significant step forward and a more cost-effective discovery process would result. The signs so far are that *in silico* structure-based drug design can help deliver this vision. **DDW**

Dr Jonathan Heal is a founder and Director of Informatics at Proteom Ltd, a drug discovery company focusing on novel computational approaches to drug design. He has a degree and PhD in Chemistry from Imperial College London and as a Microsoft Certified Solution Developer has worked in both financial and pharmaceutical sectors.