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

Hello Drug Discovery, I am from *Insilico*, take me to your President

In a recent review in the *Information Biotechnology 1* supplement to *Drug Discovery Today* [1], Darko Butina and colleagues catalogue the state-of-the-art in ADME data interpretation and prediction via *in silico* methods. The catalogue is broad and serves as a useful benchmark on the state of ADME computational methods in the world today. But which world?

Computational methods for the drug discovery and development process tend to exist in a parallel universe in a different time zone. The same is particularly true for those concentrating on ADME properties. All the methods outlined by Butina and colleagues show promise; some have demonstrated real, albeit limited, prediction capability. However, as the authors state: 'No single approach can be used to predict the full range of ADME properties that are desired'. This is fair enough, but a little later in the same paragraph a substantial caveat slips by almost unnoticed: 'In fact, a combination of two or more models for the same property, based on different principles, can give higher confidence in the results obtained for which they agree or identify areas of uncertainty where they differ'. Therefore, reviews of the performance of ADME methods and models often end with the statement that it is 'relatively simple to develop models that fit an entire data set, but that typically, such models do not predict new data sets well' [2].

What is needed, the reviewers and enthusiasts argue, is more and better data, better descriptors, better algorithms, more time. Being in a different universe and time zone is acceptable for the optional luxuries that so many computational methods appear to be: tomorrow is always another day, and the occasional promising signs and signals from the methodology mean that, one day, all the promises will be fulfilled.

Planet *Insilico* is a great place to show VIPs how clever the pharmaceutical industry is and to give these VIPs the unforgettable thrill of knowing what a substrate sees in 3D as an enzyme swallows it. *In silico* prediction is difficult, which explains why the different universe and time zone and lack of clear and visible success has been tolerated, and even encouraged; in addition, the let out clause is that the pharmaceutical industry has put the muscle into screening power. Automation and robotics drive hugely efficient 'assay to data capture' systems that mean real data is available in a timely manner. Interestingly, the robotics area is always the other port of call on the VIPs visit: brains and brawn in one day. The computational gurus (*insilicoids*), living in their darkened cave on *Insilico*, really have needed to be in contact for a small period of the day: minimal contact with the 'real world', minimum influence on the 'real world' and minimum influence of the 'real world' on 'them'. But suddenly 'them' are vitally important to the future. Companies have finally thought through how to use HTS and high-speed chemical synthesis (parallel synthesis). Now 'druggable targets' will be screened against a 'drug-like file' for lead material and then the sheer firepower of parallel combinatorial synthesis or multiple chemical analoging ensures rapid progress towards development candidates.

However, the sheer firepower of this branch of medicinal chemistry overwhelms conventional ADME screening. ADME screens can be made to be high(er) throughput, but their dependence on cells, human tissues, MS endpoints and so on, limits their use in the new world, both in terms of timeliness of data and cost. Calculations suggest that screening every compound synthesized in the 'new world' ways, in a battery of ADME permeability and stability screens,

1359-6446/02/$ – see front matter PII: S1359-6446(02)02485-6

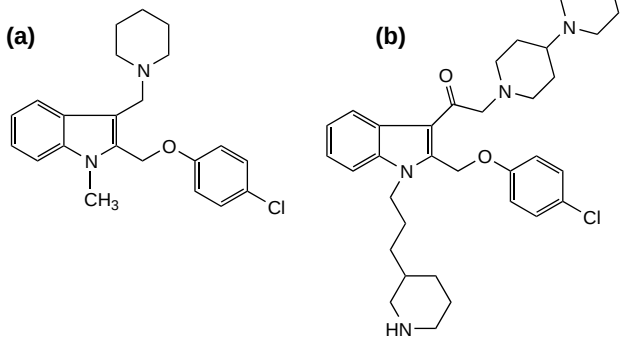

Figure 1. An example of potency led drug design. The original lead structure **(a)** appears to have excellent drug-like properties. As NPY Y-1 antagonist potency has been impressively increased **(b)** so oral drug properties have been attenuated.

would drive up the cost of delivering a development candidate out of discovery by a factor of 10–20. This is clearly not a sound business proposition. Moreover, ADME data is vital. Much of development candidate attrition is ADME-based and the new world demands that what begins as drug-like ends as a drug. The need for ADME property assessment is to ensure that the (oral) drug properties are maintained or produced as an early drug-like lead is advanced in potency by successive rounds of chemistry.

Figure 1 illustrates an example where this has not been done. The NPY Y-1 indole lead [3] would probably have good oral drug properties based on inspection of the structure, and its molecular weight (369), one measure of drug-like, complies with this. The resultant most potent compound of the series is 2000-fold more potent (around 1 nM) and this compound could be judged a great success on this single criteria. Unfortunately, it is now not a drug because it is unlikely to have good oral drug properties. As a measure of non drug-like, it has a molecular weight of 591 and also several positions of extreme metabolic vulnerability. When tested *in vivo*, the serum levels of the compound following oral administration were inadequate to evaluate the compound by this route [2]. So the cycles of parallel synthesis must be led by information on potency and selectivity against the pharmacological target and ADME properties. This is true SAR-led drug design or super-rational drug design; clearly differentiated from blind combinatorial chemistry or the single compound progression of traditional SAR-led rational drug design.

However, we have already established that conventional screening cannot stretch to providing the ADME information and guidance in a cost effective and timely manner, leaving the stage free for the *insilicoids* to rescue the 'real world'. The 'real world' needs 'them' like never before. The *insilicoids* have developed the inventory that Butina and colleagues have supplied. They now need to make the last tweaks to the software; to decide if it is 2D atom types, 3D atom–pair distances or polar surface area that are important; lock down on statistical regression, neural networks or genetic algorithms, and then come in force from *Insilico* and deliver in the new real world.

Whether it is pure *de novo in silico* or a partial use of screening information on selected subsets of each entire synthetic run, what is needed is a map of ADME space and target space to guide each cycle. The companies that succeed here will prosper; the companies that do not are at a disadvantage. Welcome *insilicoids* to the 'real world, real time zone'; get this right and do it now, and we'll make you the President.

References

1 Butina, D. *et al.* (2002) Predicting ADME properties *in silico*: methods and models. *Drug Discov. Today* 7, S83–S88
2 Bures, M. *et al.* (2000) Trials and tribulations of developing predictive models for ADME. Abstracts of *220th American Chemical Society* 20–24 August 2000, Washington, DC, USA (Abstract COMP-011)
3 Hipskind, P.A. (1997) Potent and selective 1,2,3-trisubstituted indole NPY Y-1 antagonists. *J. Med Chem.* 40, 3712–3714

Dennis A. Smith
Pharmacokinetics, Dynamics and Metabolism
Pfizer Global Research and Development
Sandwich, Kent
UK CT13 9NJ

Patent opportunities in bioinformatics

Biotechnology research is a wide and diverse field of scientific endeavor that spans the study and development of transgenic organisms, the dissection of molecular and cellular pathways, pharmacogenomics and genetically modified crops. However, a common vein is seen to run through these varied areas of research, namely bioinformatics.

A reason that bioinformatics is needed in these fields is that many organisms have been the subject of intense genome sequencing. Scientists are faced with the task of converting this mountain of raw data into informative and useful information by discovering what portions of the data are useful and what portions are extraneous.

According to the National Institute of Health's (http://www.nih.gov) Biomedical Information Science and Technology Initiative Consortium, bioinformatics is generally defined as: 'Research, development or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze or visualize such data.'

Although the classic term bioinformatics most often describes computerized technologies used to store,

  PII: S1359-6446(02)02486-8