400
300
200
100

Why is it Still Drug Discovery?

Matthew Segall at the ADMET Division, BioFocus DPI, explores the balance between luck and judgement in drug discovery

Matthew Segall is Senior Director of the ADMET group of BioFocus DPI. He has a MSc in Computation from the University of Oxford and a PhD in Theoretical Physics from the University of Cambridge, with three years postdoctoral experience in the Cavendish Laboratory pioneering the use of first principles quantum mechanical modelling in the study of biological systems. In 2001 he joined Camitro (UK) working on the development of predictive models of drug metabolism. Following a merger with ArQule, he led the informatics team in developing state-of-the-art intuitive ADME decision and visualisation tools, while continuing to work on quantum-based ADME models. After the Cambridge-based ADME group joined Inpharmatica in 2003, he led the *In Silico* Optimisation group in developing StarDrop™. Since January 2006, he has managed the ADME/PK business for BioFocus DPI (formerly Inpharmatica), which combines experimental and *in silico* approaches to ADME optimisation in drug discovery.

Estimates for average costs and durations vary, but a reasonable approximation puts the typical duration for bringing a drug to market at 12-15 years (1) and the average cost at over $800 million (2). The reason for this is that the process is dominated by failure, with a high attrition rate of compounds and projects, even in the later and most expensive clinical phases. Therefore, the majority of the cost is associated with compounds and projects that never result in a marketed drug.

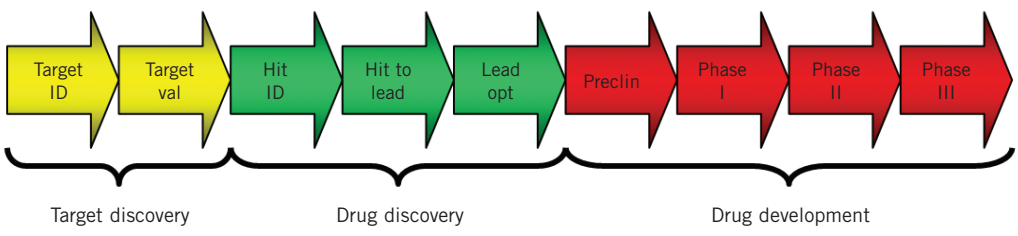
Pharmaceutical R&D is conventionally divided into three phases, as illustrated in Figure 1. Although the exact definitions of these phases may vary and the boundaries are blurred, they may broadly be described as; target discovery, drug discovery and drug development. In this article, we focus on the drug discovery phase for small molecule drugs – those with a molecular weight less than around 800Da.

DRUG DISCOVERY VERSUS DRUG DESIGN

It is interesting to note that the middle phase is termed drug discovery, which suggests a search for a solution among many possibilities and with a certain element of speculation. This is in contrast to a design process, which implies a rational process by which a small number of optimal solutions are constructed *ab initio*. The high attrition rates of the current state-of-the-art suggest that drug discovery has not yet become drug design. Why is this, given recent industry trends that have emphasised techniques such as 'rational drug design' (3), 'structure-based drug design' (4), and '*de novo* drug design' (5)?

Beresford *et al* (6) and others have drawn an interesting contrast between drug discovery and the design process used by Boeing when creating the 777 aircraft. The Boeing 777 was designed and 'tested' entirely on a computer before a full-scale prototype was built, which successfully flew on the first attempt. The drug discovery process is more like the historical approach to building an aircraft – building and crashing many prototypes before finding a design that flies. This is evidenced by the enormous number of compounds synthesised, tested and rejected for every successful drug reaching the market (7). The vision that Beresford *et al* espouse for drug discovery is to become more like the 777 design process, whereby most of the work is done on a computer (*in silico*), which is relatively inexpensive, before embarking on the resource intensive synthesis and testing of only a small number of compounds.

Figure 1: Diagram illustrating the process of pharmaceutical R&D



SO, WHY NOT DESIGN?

The vision of an *in silico* design process for drug molecules is certainly attractive, so why has this goal yet to be realised, despite an enormous effort over the past 10 years?

The first reason that one might suggest is the sheer complexity of the design task. A successful drug must have not only good potency against the target protein, but also appropriate properties to enable it to get to the location of the target in the body and then be safely dealt with by the body, usually termed absorption, distribution, metabolism and elimination (ADME) properties, with minimal side effects (a lack of toxicity). However, if we return to the analogy of aircraft design, we can see that we must also optimise a variety of properties – power-to-weight ratio, mechanical strength, capacity, range and fuel economy for example. Therefore, it is clear that design can be undertaken even for complex multi-parameter problems.

The underlying difficulty in implementing a rigorous design process for drugs is that the rules that link the design (the structure of the potential drug molecule) to many of the required properties are not well understood. In contrast, for aircraft, the underlying rules, expressed as mathematical equations, governing fluid dynamics and the behaviour of materials under the effects of temperature and stress have been known for many years. Although the behaviours that emerge may be complex, the equations may be solved with high precision with sufficient computational power.

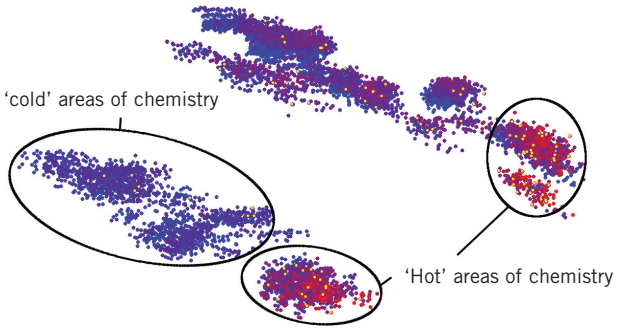
In the case of drug molecules, most effort has been applied to the prediction of potency, or binding affinity, of a molecule to its intended target protein. Even in the ideal case, where a three-dimensional structure of the binding site on the protein is available from X-ray crystallography, computational techniques, such as pharmacophore modelling (8) or docking (9), cannot quantitatively predict the effect of small modifications to the proposed molecule. Quantitative structure activity relationship (QSAR) models (10) can predict changes in potency that are correlated with structural features of chemically similar molecules. However, the uncertainties in these predictions are usually high, which limits their effectiveness in proposing specific structural modifications.

The rules linking chemical structure with ADME properties have also been widely studied (11). The simplest of these are rules-of-thumb, such as Lipinski's 'Rule of Five' which provides guidelines for common structural characteristics, molecular weight, logP and hydrogen-bond acceptors and donors of molecules, found to be well absorbed through the human intestine (12). QSAR models have been developed for properties such as aqueous solubility, affinities for enzymes responsible for drug metabolism and penetration of compounds into the brain (11). Quantum-mechanical modelling approaches enable simulation of the chemical reactions leading to drug metabolism and hence the vulnerability of molecules to some drug metabolising enzymes (13). However, all of these methods exhibit a significant degree of uncertainty in their prediction, which again limits their ability to quantitatively guide the design of individual molecules.

Figure 2: A 'chemical space map' generated by BioFocus DPI's StarDrop™

This visualisation represents each compound as a point coloured according to the 'quality' of the molecule from red (best) to blue (worst). The proximity of two dots represents the similarity of the corresponding compounds in terms of their molecular structure, so that closely grouped compounds are similar, while those distant from one-another have little in common.

The yellow crosses illustrate an example compound selection that focuses on the 'hot' regions of chemistry, while exploring intermediate regions and avoiding 'cold' regions, biasing the chances of success in the projects favour, while spreading risk across diverse chemistry.



Probably the least understood mechanisms are those leading to toxic side effects in potential drugs. In many cases, the molecular pathways leading to particular toxic effects are not even known, making it almost impossible to predict what molecular structures will trigger a toxicity. In those cases where interaction with a particular protein can be associated with a toxic effect, for example inhibition of the hERG ion channel, potentially leading to the heart arrhythmia *torsade de pointes*, QSAR or pharmacophore models can be developed (14). However, these have the same limitations as those models discussed above.

IS IT ALL JUST LUCK?

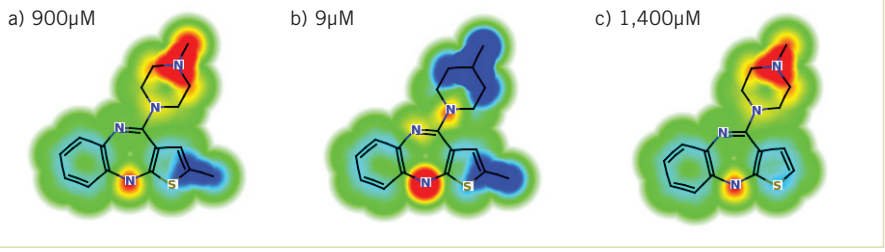
In light of the situation described above, it is tempting to believe that we are playing in a drug discovery casino! To some extent that is true, but it is well known that some games of chance can be beaten. As entertainingly described in the book *Bringing Down the House* (15) a casino can be beaten by a team of dedicated and skilful players who know how to bias the odds in their favour. By developing a card counting system and applying it systematically, the game of blackjack can be beaten by placing the largest bets on 'hot' decks biased in the player's favour, smaller bets on indifferent decks and walking away from decks that are 'cold'. There is no guarantee of winning any one hand, but on average, over a number of games, the player can accumulate large winnings.

This analogy may be applied to drug discovery if a skilful, multi-disciplinary drug discovery team can consistently bias the

Figure 3: The 'glowing molecule™' in StarDrop™ aids visualisation of the rules captured by QSAR models to guide the design of improved molecules

In this simple example, the colours behind the molecular structures indicate those regions of the molecules exerting a positive influence on the predicted aqueous solubility (red), those exerting a negative influence (blue) and those which are neutral (green).

Molecule (a) is Olanzapine, molecule (b) shows the result of changing the piperazine into piperidine, and (c) illustrates the result of removing the thiophene methyl group from Olanzapine.



odds in favour of finding a successful drug candidate. Certainly, successful project teams do not use a random approach to finding a drug candidate, but are guided by their combined experience and careful interpretation of the available data to identify relationships between structure and biological activity. However, a systematic approach can improve the odds further, identifying areas of chemistry that are 'hot' (or have a high likelihood of success). The majority of resources should be applied to synthesis and testing of molecules in these areas, less to those areas with a lower chance of success and little, or none, to 'cold' areas. There is no guarantee that any individual molecule will be successful, but on average a project is more likely to quickly find a successful candidate. For some targets, there may be no chemistries that have a decent chance of success. In these cases, after considering a wide range of possibilities, the project team can 'walk away' before losses become too great.

The question remains, how to identify 'hot' and 'cold' areas of chemistry? An experienced medicinal chemist may have an instinctive feeling for the rules that govern some of the properties discussed above. However, while the computational models described above may not provide exact results, they are certainly capable of biasing the odds towards successful chemistries, and encoding the rules for more properties than even an experienced medicinal chemist can juggle in their head. Furthermore, computers can be used to explore a much larger number of molecules than a synthesise-and-test approach or even by visual inspection, increasing the chance of finding 'hot-spots'. When predictions indicate that no chemistry will provide a sufficient chance of success, it may be appropriate to test this finding by synthesising and testing a small number of compounds, but a long and fruitless search can be avoided.

Software platforms can be used to predict a wide range of properties and 'map' potential chemistry to effectively identify those areas with the highest chance of success. Figure 2 (see page 73) shows an example of such a 'chemical space map' along with an illustration of how compounds may be selected from this chemistry to bias the odds in favour of success. These maps are generated by estimating the likelihood that each compound will have an optimal balance across a wide range of properties, using a unique 'probabilistic scoring' approach that takes into account all of the uncertainties in the available data.

The maximum benefits can be obtained from this approach by applying it as early as possible, when the range of potential chemistries is largest. For example, in the hit-to-lead stage, chemistries around each hit may be efficiently explored and prioritised to focus experimental resources. However, even later, in lead optimisation, it is often necessary to consider a wider chemical space when attempting to resolve an issue or to 'scaffold hop' into a new chemical series (16).

The card-counting analogy begins to break down when we note that in blackjack it is not possible to carry learning from one deck to the next. Once a deck is shuffled, the cards are randomised and the count must begin again from scratch. Of course, in drug discovery we can learn from the results of previous projects or experiments to increase our chances of creating a 'hot' chemistry. This is the advantage that an experienced medicinal chemist brings to the table. However, *in silico* approaches can also provide an advantage. 'Machine learning' techniques may be applied to improve predictive models as experimental data are gathered on new compounds or to develop models 'tuned' to the effects of subtle structural differences between compounds within specific chemistries. These methods can identify new rules linking molecular structure with a property, typically encoded within QSAR models that may be applied to improve the confidence in predictions on untested molecules.

A common criticism of QSAR models is that, while they capture rules linking molecular structure to a property, this connection is often hidden within the complex mathematics of the models (so-called 'black box' models). However, novel visualisation techniques, such as the 'glowing molecule™' illustrated in Figure 3, can now provide this information in an intuitive form, transferring learning and helping to move closer to the efficiency of a true design process.

CONCLUSION

The vision of a comprehensive drug design paradigm remains elusive. However, significant progress has been made in understanding and capturing the rules that govern the properties of potential drugs. These can be used to focus resources on chemistry most likely to have optimal properties for the therapeutic objective of the project. The result is a reduction in the time and effort wasted and a much improved chance of finding a successful drug candidate. In the end, even if models are not perfect, this is enough to dramatically improve the efficiency and productivity of drug discovery. ◆

The author can be contacted at matt.segall@glpg.com

Acknowledgements
The author would like to thank the staff of BioFocus DPI's ADMET group who contributed to development of the concepts described in this article, in particular Alan Beresford for his helpful suggestions.

References

1. **Bolten BM and Degregorio T, From the analyst's couch, Trends in development cycles, *Nat Rev Drug Discov* 1: pp335-336, 2002**
2. **Dimasi JA, Hansen RW and Grabowski HG, The price of innovation: new estimates of drug development costs, *J Health Econ* 22: pp151-185, 2003**
3. **Truhlar DG, Howe WJ, Hopfinger AJ, Blaney J and Dammkoehler RE, *Rational Drug Design*, 1999**
4. **Veerpandian P, *Structure-Based Drug Design*, 1997**
5. **Dean PM, Lloyd DG and Todorov NP, De novo drug design: integration of structure-based and ligand-based methods, *Curr Opin Drug Discov Devel* 7: pp347-353, 2004**
6. **Selick HE, Beresford AP and Tarbit MH, The emerging importance of predictive ADME simulation in drug discovery, *Drug Discov Today* 7: pp109-116, 2002**
7. **Gardner J, Early Stage Drug Discovery 2004, *Combi-Chem & High Throughput Screening Revisited*, 2004**
8. **Guner OF, *Pharmacophore Perception, Development, and Use in Drug Design*, 2000**
9. **Jones G, Willett P, Glen RC, Leach AR and Taylor R, Development and validation of a genetic algorithm for flexible docking, *J Mol Biol* 267: pp727-748, 1997**
10. **Hansch C and Leo A, *Exploring QSAR Volume* 1: *Fundamentals and Applications in Chemistry and Biology*, 1995**
11. **Van de Waterbeemd H and Gifford E, ADMET in silico modelling: towards prediction paradise? *Nat Rev Drug Discov* 2: pp192-204, 2003**
12. **Lipinski CA, Lombardo F, Dominy BW and Feeney PJ, Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings, *Adv Drug Deliv Rev* 23: pp3-25, 1997**
13. **Jones JP, Mysinger M and Korzewka KR, Computational models for cytochrome P450: a predictive electronic model for aromatic oxidation and hydrogen atom abstraction, *Drug Metab Dispos* 30: pp7-12, 2002**
14. **Cianchetta G, Li Y, Kang J, Rampe D, Fravolini A, Cruciani G and Vaz RJ: Predictive models for hERG potassium channel blockers, *Bioorg Med Chem Lett* 15: pp3,637-3,642, 2005**
15. **Mezrich B, *Bringing Down the House: How Six Students Took Vegas for Millions*, 2004**
16. **Segall MD, Beresford AP, Gola JMR, Hawksley D and Tarbit MH, Focus on Success: Using a Probabilistic Approach to Achieve an Optimal Balance of Compound Properties in Drug Discovery, *Expert Opin Drug Metab Toxicology*, 2: pp325-337, 2006**