The emerging importance of predictive ADME simulation in drug discovery

Alan P. Beresford, Harold E. Selick and Michael H. Tarbit

Absorption, distribution, metabolism and excretion (ADME) studies, are widely used in drug discovery to optimize the balance of properties necessary to convert leads into good medicines. However, throughput using traditional methods is now too low to support recent developments in combinatorial and library chemistry, which have generated many more molecules of interest. To the more enlightened practitioners of ADME science, this situation is generating both the problem and the solution: an opportunity is now forming, with the use of higher throughput ADME screens and computational models, to access this wide chemical diversity and to dissect out the rules that dictate a pharmacokinetic or metabolic profile. In the future we could see ADME properties designed-in from the first principles in drug design.

Harold E. Selick
Camitro Corporation
4040 Campbell Avenue
Menlo Park, CA 94025, USA
Alan P. Beresford and Michael H. Tarbit*
Camitro UK
127 Science Park
Milton Road
Cambridge
UK CB4 0GD
*tel: +44 1223 424825
fax: +44 1223 425416
e-mail: mtarbit@camitro.com

▼ Imagine life in the design department of a modern aeronautics company. A design team has been asked to devise the specification for a new twin-engine long-haul aircraft that can hold 300 passengers and operate with a range of 10,000 miles. The team will have access to design and performance data on a whole range of engineering and electronic materials, as well as data on entire families of aircraft that have operated under similar conditions.

Using this data and a library of computational tools, with their inherent sets of rules, the team will prepare initial designs of both components and plane body: they will use computational programs to simulate the fit and function of the components and overall behaviour of the plane and evaluate its likely performance; they will design the entire manufacturing process so as to ensure that the plane can actually be constructed on time and within budget; and they will then carry out final adjustments before making a full-sized, fully functional prototype for flight testing in anticipation of production.

Associated with this process is a high degree of confidence that the final product will perform according to both specification and simulation.

Now compare this with pharmaceutical R&D, in which a research team must discover a drug that will interact with a novel target for therapeutic intervention in an important disease. The team will have access to data on related targets and existing drugs that interact with them. They might even have a crystal structure of the target protein. By using a library of computational tools, with their inherent sets of chemical rules, the team can make an informed assessment regarding the shape of molecules that might interact with the target. In modern companies, they will then be able to enumerate a focused library of possible actives using these methods. However, this is where their 'simulation' ends. They will have no computational ability to determine whether or not any of their active molecules will ultimately work as a suitable drug on the market. To discover this, they will have to commit major laboratory and production facilities, and associated high costs, within the company to conduct empirical, practical tests to find out whether any of the compounds, or their analogues, will make the grade and work in man.

The problem here might be more akin to our aeronautics company being told that the aircraft specification now demands a need to fly at 50,000 feet, a height that lies outside their collective experience. Suddenly, the equation is full of unknowns. But, given the fact that this is a passenger aircraft, the implications of product failure for the design

1359-6446/02/$ – see front matter PII: S1359-6446(01)02100-6

team are unthinkable. Consequently they will make every effort to simulate all conceivable situations – every combination of atmospheric, pressure and temperature change that their aircraft is likely to encounter on its journey. The data from each simulation – each success or failure – will be used to further improve the design. All of this will be achieved without having to manufacture any 'candidate' aircraft. In pharmaceutical R&D, there is currently no equivalent process. Every molecule needs to be synthesized and then subjected to a battery of *in vitro* and *in vivo* tests using animal or human-derived systems. Furthermore, little of the data on the success or failure of molecules in the various tests will be captured, analyzed and used in any sort of rigorous manner to facilitate the design of subsequent molecules. Indeed, most design constraints are imposed primarily by the availabilities of chemical reagents and the efficiency, or tractability, of the available chemistry reactions. This is rather akin to our aeronautics team making a component out of those parts they have closest to hand and hoping it will do the job! Consequently, the attrition of molecules in drug development is both high and extremely costly. Not surprisingly, the further a compound survives through the testing process, the more costly and damaging, in wasted time and competitive position, is the failure. Companies attempt to mitigate this potentially hazardous outcome by adopting a back-up, or follow-on strategy, whereby a project team will place another compound, usually with slightly different physicochemical properties, into the pipeline behind the lead. This is aimed at reducing the probability of a project's overall failure. However, it is still a perpetuation of the same process and the odds of success for the back-up are not altered materially over those for the lead.

So why do drugs fail?

Drugs fail at multiple points along their discovery and development track. Some might even make it to the market, only to be withdrawn following an initial successful launch, with the cost in lost revenue to the companies who have invested the time and money in their development being potentially crippling. The recent withdrawal of the anticholesterol drug Baycol™ (Bayer AG, Leverkusen, Germany) is the latest example of a compound with excellent potential in its therapeutic class, but which will now not realize any value for the discovering company. Often the withdrawal is for safety reasons, such as the unforeseen occurrence of an adverse reaction or cross-reaction that was only likely to become apparent with large-scale use, or abuse, of the compound in man [1].

Clearly, then, one of the keys to developing a successful medicine is in minimizing the unpredictable features of the molecule. This includes trying to predict the way in which the molecule affects the body, and vice versa, including interaction with any other drugs and assorted xenobiotics that might be present from time to time during its therapeutic use. Consequently, the development of a medicine is a long and complex business and, as we have seen, the failure rate before market is high [2]. Although less dramatic than a post-market withdrawal, this failure rate in which more than 90% of all development candidates fail to make it to market, still costs billions of US dollars and is a major contributing factor to the significant R&D budgets of pharmaceutical companies.

In the current climate of downward pressure on healthcare costs, combined with the need for expensive technology development and greater information on drugs in development, the more enlightened companies are starting to investigate and address the main reasons behind drug failures after a molecule is launched into the development process. Historically, the main qualification for a molecule to enter the process has been potency against the presumed therapeutic target, with any selectivity for that target an added bonus at that stage, combined with a lack of any overt toxicity in the animal disease model. Hence, for example, a CNS project team would begin with a potent molecule that they knew would affect the target responsible for triggering some required response in the brain, but they would have little knowledge of whether the compound would dissolve in the gut, be absorbed into the systemic circulation across the gut mucosal layer, survive passage through the gut and the liver, which guards the system against such xenobiotic challenge with powerful enzyme and efflux systems, and continue in the bloodstream at high enough free concentrations to be effective – all before attempting to cross the blood–brain barrier (BBB) to reach the target site. In addition, the compound would have to withstand the efforts of the body's clearance mechanisms in the liver and kidney, attempting to break down or excrete it, for long enough for it to exert its effect on the target and therefore be an effective medicine. It is perhaps small wonder that these processes of ADME, along with toxicity (Tox), account for a high proportion of drug development failures. Little is more frustrating for a project team than to see their carefully chosen, highly potent and potentially disease-altering compound, which will exhibit all the desired properties of selectivity for its pharmacological target, fail to have any effect whatsoever when administered to animal models, or humans, for these reasons.

Reducing the risk of failure

On occasion, the pharmaceutical industry has been so focused on potency that it has missed the warning signs of

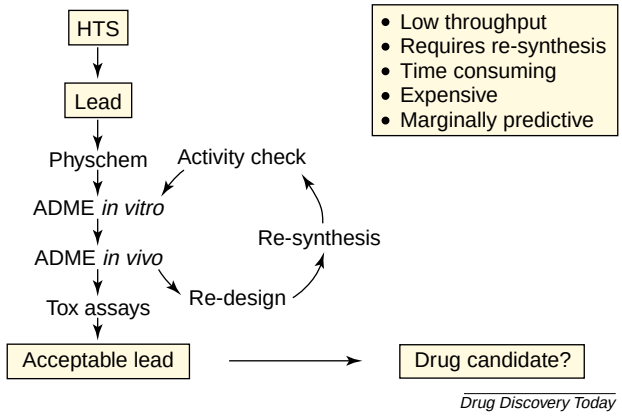

Figure 1. Serial-cyclical model of lead optimization. Compounds are tested for potency first, followed by serial examinations of their absorption, distribution, metabolism and excretion (ADME) properties and toxicity (Tox) profiles. Compounds failing ADME/Tox profiling require new analogues to be synthesized and resubmitted through the cycle. Each cycle takes weeks to months.

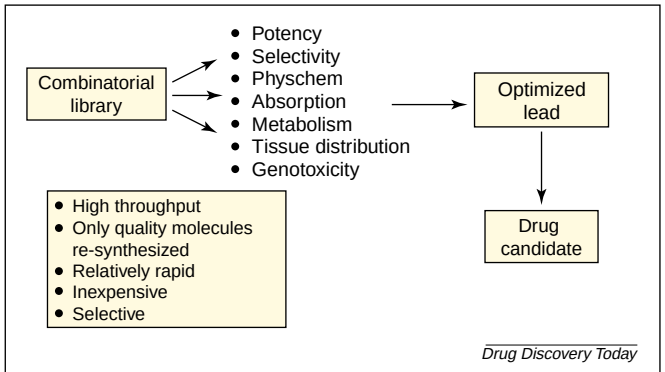

Figure 2. 'Parallel' model of lead optimization. Libraries of 'pre-synthesized' analogues are submitted to multiple tests in tandem, building a database of molecular profiles. Best-option molecules are selected for further final optimization and candidate status. Cycles take weeks, but afford much higher productivity.

potential problems ahead and persisted in pushing a molecule through to the market. In some cases, this has led to spectacular failures and serious consequences for patients [3,4]. Culturally, the industry is embedded in a 'serial' process whereby potency testing is carried out first, followed by the other tests that are essential to revealing a drug candidate's total profile. Even the more modern-thinking pharmaceutical companies, who seek to determine potential for failure by applying ADME testing at the earliest stages of the discovery exercise, still conduct this as a serial process (Fig. 1). There are recent signs of change, with some companies now attempting to adopt 'parallel' processing, so that potency tests and surrogate ADME screens (usually some moderate throughput *in vitro* systems) (Fig. 2) are carried out in close proximity, thereby allowing project teams the ability to select on multiple parameters [5,6]. The problem with this approach is that it is still largely empirical – the chemist still has to make a large series of molecules in a manner more dictated by reagent availability than by keen medicinal chemistry insight, and then carry out tests to determine whether they have the right properties or not. This will be followed by re-synthesis and re-test in an iterative cycle, which on average takes in excess of 12 months. The use of combinatorial synthesis is having some impact on this, in that, once a lead has been identified, larger numbers of compounds that are optimized for one or more properties can be synthesized per cycle.

Nevertheless, even with parallel processing of libraries, the cost and time in optimization cycling and testing can be significant. Our aeronautics design team would probably be staggered at this approach and, given the amount of data which must be available from decades of pharmaceutical research, might ask why molecules cannot be designed with inherently good properties from first principles? Why can we not simulate their *in vivo* ADME behaviour and use that to guide their design? Of course, it is argued that the physiology of the body is extremely complex and, therefore, not a facile exercise in modelling or simulation. However, the same could be argued about the construction of the modern jet airliner mentioned previously. Yet Boeing was able to design its 777 aircraft from first principles to production without the need to manufacture and test any intermediate prototypes at all! (See http://www.boeing.com/commercial/777family/cdfacts.html).

The real answer as to why we cannot do the same with a drug is because we just do not know the rules that determine ADME behaviour – we have little understanding of the relationships between a molecule, its physicochemical properties, and its likely fate in the body. But, with ADME behaviour, like any complex problem, the task of resolving it becomes less daunting if it can be broken into more manageable pieces. If we take the first stage, absorption, it is reasonable to assume that the majority of successful orally administered drugs in therapeutic use cross the gut epithelium effectively. If we assume a simple passive diffusion of these molecules from the gut through the mucosal cells of the intestine into the bloodstream, then it should be feasible to establish a relationship between their physicochemical properties and their ability to transit the gut epithelium. In other words, we can define the rules governing this process. This is precisely the approach encouraged by Lipinski *et al.* [7]. They did what our aircraft design team

would have done; they took the data already available from previous successes and failures and derived a set of guidelines to apply to new compounds. The novel feature of their work was not in the parameters used, as these had been applied empirically in the pharmaceutical industry for years, but in the simple fact that they were used to define a rule. Perhaps even more important for the industry is the enthusiasm with which the Lipinski 'Rule of 5' has been received, showing that there is a desire for change – a willingness to try and understand the rules of ADME and apply models and simulations to predict these properties and design better molecules. In some cases, companies are now applying Lipinski's rules as a selection filter for improving the potential 'drug-like' properties of compounds entered into their screening libraries in the hope that they will 'pre-select' compounds that will be absorbed. One drawback to this approach is the rigid limitations placed on a compound's lipophilicity and molecular weight, both of which tend to increase as the chemists try to optimize their initial 'hits' from a screening campaign. However, the principle is sound and can be applied to other ADME processes.

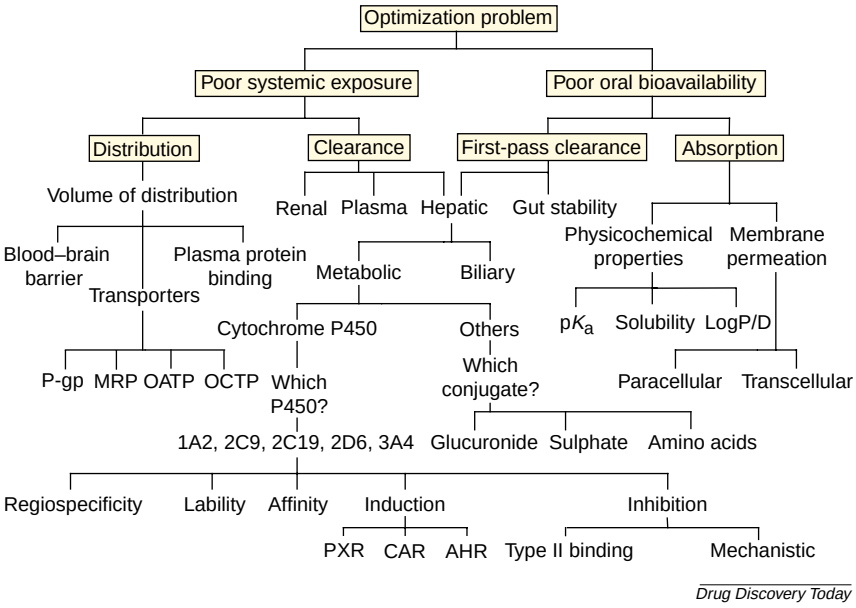

Figure 3. Reducing complex *in vivo* disposition processes to discrete mechanisms that can be modelled. This decision tree defines the analysis of an absorption, distribution, metabolism and excretion (ADME) optimization problem reduced to a specific mechanism. Predictive models need to be built at mechanistic levels to reduce the impact of conflicting processes. Abbreviations: AHR, aryl hydrocarbon receptor; CAR, constitutive androsterone receptor; MRP, multidrug resistance protein; OATP, organic anion transport protein; OCTP, organic cation transport protein; P-gp, P-glycoprotein; PXR, pregnane X receptor.

How reliable is ADME/Tox modelling?

The drivers for data acquisition

Even when split into its component 'mechanistic' parts, the task of predicting the ADME/Tox behaviour of a novel compound, and solving its optimization, remains fairly intimidating (Fig. 3). Assuming that they are absorbed, the body has evolved a myriad of systems for dealing with the systemic challenge that drugs, or xenobiotics in general, represent. Nevertheless, considerable progress is now being made in developing first-generation models for the major processes. What has aided these developments has been both a pressure in the system caused by the explosion of data emerging from combinatorial chemistry and HTS, combined with a deepening understanding of some of the fundamental mechanisms underlying processes as seemingly complex as drug metabolism.

Purveyors of ADME data for use in optimizing the design of a project team's compounds have long represented a bottleneck in the system, as their traditional experimental methods have never matched the ability of chemists to synthesize novel compounds. The quantum leap in the chemist's ability to synthesize libraries of thousands of molecules, where tens previously existed, has rendered the traditional methods of ADME support inadequate. Consequently, the ADME scientists have needed to adopt some of the HTS culture of their biology colleagues in designing screens and methods that are of much higher throughput. Success has been variable, but the processes are still evolving [5,8,9].

How has this aided model development? It is clear that the development of reliable models requires substantial amounts of good quality data. Any primordial models in the past were invariably poor in their predictability because they were based on small data sets of tens of compounds. If you compare that scenario with the current ADME laboratories in some major pharmaceutical companies, where upwards of 500 compounds can be processed through metabolism screens per week, it becomes apparent that there is now a major opportunity to study the behaviour of many compounds with a high degree of chemical diversity and to search this to understand the rules. In that sense the 'problem' has helped to provide the solution. In fact, as evidenced by numerous presentations in the current plethora of conferences that have focused

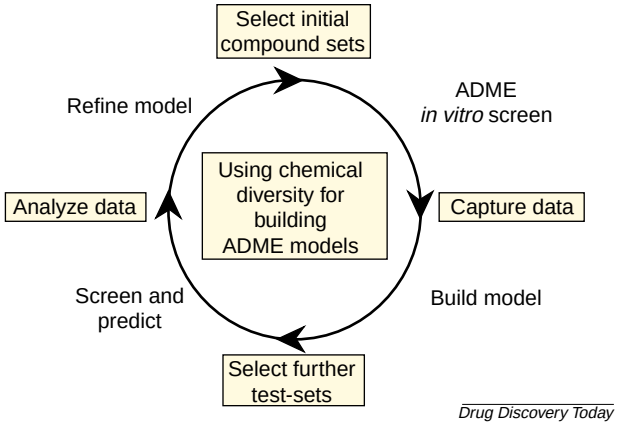

Figure 4. Screen and model paradigm for building absorption, distribution, metabolism and excretion (ADME) models *in silico*. Modern chemical libraries comprise millions of molecules covering extensive chemical diversity. Clustering into subsets and screening through *in vitro* ADME tests provides high quality data for building, testing and refining predictive models.

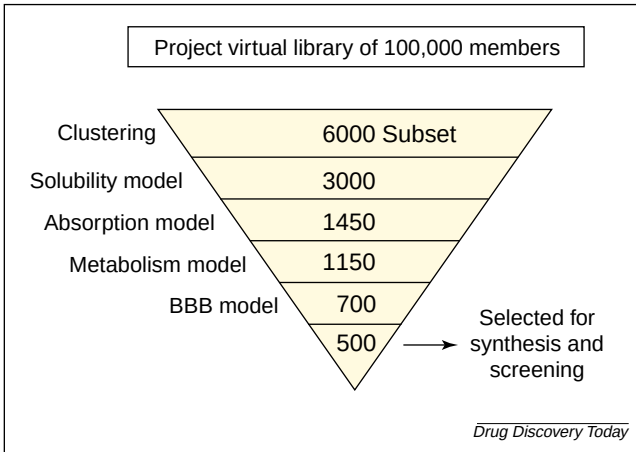

Figure 5. Using models as filters in library screening. Predictive models can be used as filters for pre-selecting 'optimal' focus libraries from large virtual arrays before synthesis. In some cases, in which potency tests are low throughput, the models can filter actual libraries to select 'best balance' molecules for potency determinations.

on ADME 'bottlenecks', enlightened chemists in some pharmaceutical companies are now working with their drug metabolism colleagues to select sets of molecules, from their chemical diversity, which are then tested specifically to aid understanding of the ADME mechanisms and facilitate model building, rather than just using the screens for selecting the best project leads [10]. This principle of clustering-screen-build-test-modify (Fig. 4) is proving successful in building descriptor-based models in several pharmaceutical company laboratories and model-building companies. These models can then be used as profiling filters for sieving out the best molecules in project libraries (Fig. 5). It is chastening to realize that some of the software tools needed to do this pattern recognition and data mining, such as Clementine™ (SPSS, Chicago, IL, USA) and decision-tree packages, have been used to analyze trends in the supermarket and banking businesses for years, and yet are still not widely used in pharmaceutical R&D.

The impact of rapidly increasing computer power

Another significant aid to the modern interest in model building has been the exponential increase in computer power available on the bench top or via web-servers. The seminal pioneering structure–activity work of Hansch and his peers in the 1970s relied on small datasets and on computers that lacked the computing power of a single modern PC [11]. Now PCs can be clustered together into hundreds of 1 GHz processors using a few wires, and the computationally intense calculations needed to understand the quantum mechanical, electronic and steric behaviour of molecules in biological systems suddenly become feasible [12]. These are interesting developments and offer the promise of deriving a first-principle understanding of how the complex proteins that transport and metabolize drugs function.

A further happy coincidence flows from the development of high-throughput ADME screens. Invariably, these screens are based on *in vitro* systems, using human tissue preparations or cell lines. This tends to make it easier to isolate specific mechanisms and study them using high throughput methods and access to chemical diversity mentioned previously. Most models that attempt to simulate or predict *in vivo* behaviour of drugs will fail if they are based solely on *in vivo* data. This is because drugs are affected by multiple processes *in vivo*, and the only process observable is the rate-limiting one. In many cases, the observer will not have knowledge or control over what specifically impacts on his data. Thus an important principle of model building is to reduce to mechanism as far as possible. Return to our example of a compound targeted to the brain mentioned above. The first hurdle for an orally administered drug will be to gain systemic bioavailability (i.e. to appear in the bloodstream). As previously discussed, to achieve this the compound has to dissolve at the pH of the intestine, permeate across the mucosal cell layer, being exposed to eliminating drug transporters and metabolizing enzymes while it does so, gain access to the portal blood supply, then survive a much higher hurdle of drug transporters and metabolizing enzymes in the liver, before reaching the heart and subsequent systemic distribution. Notwithstanding the general principles gleaned by Lipinski *et al.* on orally administered drugs, it is extremely

difficult, if not impossible, to gain accurate data on these processes *in vivo*, let alone attempt to model such a composite, complex situation [7]. However, we can isolate the cells of the gut, or the enzymes that take part in metabolism or transport, and study their treatment of large collections of diverse compounds with known characteristics. This, combined with a fundamental understanding of events, sometimes at the atomic level, then enables the construction of fairly accurate mechanistic models. Of course the corollary, and something of a penalty, is that we need to bolt any predictions from these models back together again if we are to gain a good 'predictive' insight into how a compound will behave when it is administered. There is also the major consideration of how applicable the *in vitro* ADME screens themselves are to the reality *in vivo*. There are several caveats on most of the screens used to test ADME properties *in vitro*, such as the validity of using transformed human intestinal carcinoma cells for permeation studies, the fact that *in vitro* metabolism screens usually do not use whole cells, or use cells that are isolated rather than morphologically arranged, or operate at oxygen concentrations that never exist in the liver, and so on. Thus the whole subject of *in vitro–in vivo* correlations needs to remain in focus as models are developed. Nevertheless, the evolution of *in vitro* data-based models will undoubtedly lead to much more informed insights into predicted *in vivo* behaviour of drugs.

One might ask the question of exactly how much data is needed in constructing these models before the resultant models are good enough to remove the need for any 'wet' screening at all. After all, the ultimate throughput of compounds in lead optimization support would be achieved on a computer, where millions of molecules can be examined in minutes, rather than in labour-intensive experimental work. Our own recent experiences in developing and using these models indicate that this can be achieved quite quickly in some cases. Thus, for example, there is increasing evidence that good computational models of passive permeation, which have been built on high-quality data relating structural features and properties of >300 known drugs to their absorption in humans, are now proving as predictive of human intestinal absorption as the often questioned value of permeability measurement through monolayers of Caco-2 or Madin–Darby Canine Kidney (MDCK) cells [13]. The availability of the models also presents another major benefit to the discovery process: for the first time in the evolution of pharmaceutical R&D, project teams can test their molecular hypotheses – virtual structures – for ADME properties before they even make them. In fact, given the throughput of some of the modern ADME computational models, they can test whole virtual libraries before synthesis. In that sense, these models have already impacted on that cost wastage mentioned in the introduction to this review. The chemist is now in possession of information that could tell him he is wasting his time in synthesizing specific members or chemotypes in a particular library of compounds – that they will never make good medicines. It is unlikely that *in silico* approaches to drug discovery will completely replace high quality experimental determinations in the short term. However, these computational tools will facilitate the design and interrogation of vast numbers of potential drug candidates, resulting in higher quality leads on which valuable experimental resources can then be focused, with an expected increase in the frequency of successful development of valuable new medicines.

We now begin to recognize a scenario in which project teams in discovery can generate 'parallel' information virtually ahead of synthesis – and make informed choices and decisions – rather than having to commit to synthesis and await the outcome of serial, iterative evaluations. We can begin to see the start of the more cost-efficient design and simulations process employed by our hypothetical aeronautics company.

How widely used is ADME/Tox modelling?

In light of the preceding paragraphs, and given that reasonable models now exist for several processes mentioned, such as intestinal absorption, BBB penetration, solubility and log P calculation, the naive reader might ask, why are these models not being used now throughout the pharmaceutical industry? A good question, and one that raises again the analogy of our aeronautics company. The models listed above are based on some relatively straightforward parameters and molecular properties, supported by a weight of validation data. Although they might not represent earth-shattering science, if the equivalent calculation in engineering terms had repeatedly been shown to improve the chances of an aeroplane staying in the air, we would not ignore them because of their simplicity; and undoubtedly the application of such ADME models is growing.

However, consider a single drug-metabolizing enzyme in the human liver, cytochrome P450 3A4. This enzyme can metabolize everything in shape and size, from a simple small molecule like toluene, all the way through to a large and complex molecule like cyclosporin. Moreover, it can do so by several different mechanisms. Indeed, together, the major human hepatic cytochrome P450 enzymes are capable of metabolizing virtually every xenobiotic with which they come into contact; a remarkable feat considering the specific way in which enzymes are traditionally thought to interact with their substrates. This processing

of xenobiotics is a role for which nature has adapted these versatile, complex enzyme systems, which also conserve homeostasis by regulating internal processes such as lipid- and hormone-signalling mechanisms. Attempting to understand this degree of complexity and interaction is akin to uprating the specification of our aircraft to allow it to leave the atmosphere and operate in space! Conducting the experiments and understanding the rules that these enzymes teach us is, therefore, taking significant time, even with the deliberate systematic approaches now occurring in the hands of key practitioners of the art [14].

What remains to be discovered?

In the next few years, we will undoubtedly achieve an even more comprehensive set of ADME models whose performance will continue to improve along with our understanding of these complex phenomena. Clearly, these will be combined with major throughput enhancements as a consequence of ever-evolving computer hardware and software architecture improvements. But significant challenges remain.

There are >100 enzymes that can metabolize drugs. In some cases, these involve major classes of molecules and represent major routes of elimination, such as glucuronosyltransferases or sulfotransferases. In addition, recent findings have uncovered the presence of a whole family of active drug-transport proteins, such as MDR-1 and OATP, which appear to have a major impact on absorption and drug disposition. These enzymes and transporters represent 'Jupiter and beyond' for our aeronautics company, because our knowledge is scant on these systems. Indeed, we have little or no data on how many there actually are, what their substrate specificity is, or even where they are expressed in the body. The building of successful models in these areas must await the development and conduct of suitable mechanistic screening tests and the datasets they will generate.

Another wholly understated challenge is that of building predictive models for toxicity. Some feel that this represents an impossible task, given the range and diversity of toxic phenomena. Again, however, a systematic approach of breaking the problem down into small pieces and developing data-driven models will undoubtedly pay dividends. Thus, for example, precedent-based models of genotoxicity have been under development for some time and, although not wholly accurate, are finding widening use in the industry [15]. In addition, the development of good models of metabolism will undoubtedly lead to the spin-off of models that predict whether a compound will form a reactive intermediate during metabolism. This is one serious toxicity problem that might, therefore, be identified and eliminated within a library or project series at an early stage. Similarly, although compounds that are overtly cytotoxic are being identified by fairly empirical cell-based screens at present, there are early signs that some structural determinants that cause this can be modelled [16]. There is much work under way in the field of toxicogenomics, which over the next three to five years will initially lead to a battery of gene markers that reflect developing toxicity, but which will eventually lead to structure–activity-based understanding of what determines a toxic event.

The ultimate scenario?

So, imagine life in a pharmaceutical company in ten years' time. A medicinal chemist is sitting down at a computer to design a new medicine for the chronic treatment of a CNS disease such as dementia. They are armed with the knowledge of what a suitable product profile would be for the treatment of dementia, including the desired pharmacokinetic (PK) properties. They know what other drugs are likely to be co-administered with their medicine, and they know the characteristics of their patient population. They know what their target looks like, because they have a crystal structure for it, with all of its molecular co-ordinates and binding sites. They have access to a large number, probably millions, of virtual molecules and fragments online, which enables them to sift through several chemical options to identify a series of molecules that would fit the target template – in fact, they have software that will do that for them, judging the degree of conformation and fit 'on-the-fly' to match their nominated specification. Following identification of a series of potentially active molecules, they link to their ADME/Tox modelling software, which automatically derives the predicted properties of their molecules and matches 'best fit' against the desired PK profile. The ADME/Tox software then checks for best balance by running an automated analysis that evaluates a series of potential substitution patterns that will improve the PK profile, resubmitting these to the target-fitting software in an iterative process until the optimum molecule emerges. They then conduct some final checks by submitting this molecule to software that tells them of any potential drug–drug interaction problems, or any major variability in the kinetics in their global patient populations. Assuming all is well, they submit the optimum structure to a suite of software that checks for patent status, orders the starting materials and programs the synthesis robot to make this resultant molecule.

Fanciful? No! All of this software exists now, or is under active development. The test for the intervening ten years will be to see just how comprehensive it will need to be for pharmaceutical companies, with a history of entrenched

practices and resistance to innovation in drug development, to embrace it. Doing so will require a fundamental change in the way that pharmaceutical drug discovery is conducted. However, given the productivity as well as quality challenges currently facing the industry, it will be those companies most willing and capable of embracing the type of change described herein that will flourish into the next decade. To achieve the ultimate scenario, as exemplified by our hypothetical aeronautics company, will take vision, drive and a determination within a company to overcome the inherent conservatism of the industry, and to break down the silos that still exist between all of the disciplines that currently comprise pharmaceutical drug discovery.

References

1 Fueng, M. *et al.* (2001) Evaluation of the characteristics of safety withdrawal of prescription drugs from worldwide pharmaceutical markets – 1960 to 1999. *Drug Inf. J.* 35, 293–317
2 DiMasi, J.A. (2001) Risks in new drug development: approval success rate for investigational drugs. *Clin. Pharmacol. Ther.* 69, 297–307
3 Honig, P.K. *et al.* (1993) Terfenadine–ketoconazole interaction. Pharmacokinetic and electrocardiographic consequences. *J. Am. Med. Assoc.* 269, 1513–1518
4 Griffin, J.P. (1998) The withdrawal of mibefradil (Posicor). *Adverse Drug React. Toxicol. Rev.* 17, 59–60
5 Tarbit, M.H. and Berman, J. (1998) High-throughput approaches for evaluating absorption, distribution, metabolism and excretion properties of lead compounds. *Curr. Opin. Chem. Biol.* 2, 411–416
6 Eddershaw, P.J. *et al.* (2000) ADME/PK as part of a rational approach to drug discovery. *Drug Discov. Today* 5, 409–414
7 Lipinski, C.A. *et al.* (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug. Deliv. Rev.* 23, 3–26
8 Rodrigues, A.D. (1997) Preclinical drug metabolism in the age of high-throughput screening: an industrial perspective. *Pharm. Res.* 14, 1504–1510
9 Eddershaw, P.J. and Dickins, M. (1999) Advances in drug metabolism screening. *Pharm. Sci. Technol. Today* 2, 13–19
10 Butina, D. (1999) Unsupervised database clustering based on Daylight's fingerprint and Tanimoto similarity: a fast and automated way to cluster small and large data sets. *J. Chem. Inf. Comput. Sci.* 39, 747–750
11 Hansch, C. *et al.* (1973) Strategy in drug design. Cluster analysis as an aid in the selection of substituents. *J. Med. Chem.* 11, 1217–1222
12 Segall, M.D. Applications of *ab initio* atomistic simulations to biology. *J. Phys. Condens. Matter*, (in press)
13 Ren, S. and Lien, E.J. (2000) Caco-2 cell permeability vs human gastrointestinal absorption: QSPR analysis. *Prog. Drug Res.* 54, 1–23
14 Higgins, L. *et al.* (2001) An assessment of the reaction energetics for cytochrome P450-mediated reactions. *Arch. Biochem. Biophys.* 385, 220–230
15 Ridings, J.E. (1996) Computer prediction of possible toxic action from chemical structure: an update on the DEREK system. *Toxicology* 106, 267–279
16 Pollack, N. *et al.* (1999) Chemical diversity approach for evaluating mechanistic relatedness among toxicological phenomena. *SAR QSAR Environ. Res.* 10, 533–543