Genomics

Exploiting the fruits of the genome: a strategic persp

It is almost axiomatic that the pharmaceutical R&D industry has not achieved in the last decade an increase in productivity (measured as high value NCEs) commensurate with the explosion of new biological information, chemistry techniques and information storage/retrieval/examination systems, let alone the real increase in industry R&D expenditure.

By Dr David C. U'Prichard

In particular, heads of discovery and R&D are under increasing pressure to 'exploit the fruits of the human genome' – the huge spending on genomics technology that the industry has made since 1990. What needs to be done? I believe the strategic answers relate to a much fuller understanding of (a) the economics of the discovery process, (b) the real sources of competitive advantage throughout the discovery process, and (c) the best way to manage a portfolio of discovery assets. I define discovery to include all activities from initial identification of targets of interest through to basic determination of clinical efficacy in Phase IIa trials.

In the pharmaceutical and biotechnology industries we are shortly entering a unique era of opportunity where the entire human genome including the roughly 100,000 genes therein will have been sequenced and available for drug discovery purposes, but the function, and pathological and therapeutic significance of even the subset of 5000-10,000 genes that are reckoned to be useful drug targets will not be known for years to come. This era of 'partial knowledge' represents the ripest conditions for fierce competition within the industry, where the big winners will be those companies that select good molecules at the clinically most attractive targets earliest.

This article will focus less on particular discovery technologies, which were well covered in the first issue of *DDW*, and more on the strategic utilisation of these technologies in the context of (a) the impact that the sequencing of the human genome is having on pharmaceutical R&D, (b) the finite, constrained nature of the R&D budget in the pharmaceutical industry, and (c) the great increase in intensity of competition within the industry. I also take as my perspective the discovery and development of orally active small molecules, as opposed to protein products, believing that for the foreseeable future the great majority of future commercial value in the pharmaceutical industry will come from oral drugs – classic 'pills'. As **Figure 1** shows, the mouth of the 'R&D funnel' is radically widening as a result of the genomics revolution, whereas the spout of the funnel is unchanged, whether it is taken to represent the number of NCEs in Phase III clinical development, or the overall industry spend in Phase III. This post-genomics shape of R&D together with increased competition means that discovery heads, faced with the huge increase in new target opportunities, need to make choices much earlier in discovery regarding which assets or projects to advance and which to shelve or kill. Obviously to maximise competitive position, the choices need to be made in as intelligent and informed a manner as possible. Although many categories of assembled information will be brought to bear, especially the degree to which the target has been 'validated' (the link that the target has to a disease and to a commercially attractive efficacy profile), first and foremost will be the quality of the chemical leads which have been generated for the target. More than one research head of a major pharmaceutical company has complained to me that

the human erspective

they have a superabundance of good targets in their organisation, but way too few good leads. The key prerequisites for success are therefore industrial-scale production of good leads and a robust process for managing and prioritising the portfolio of lead compounds.

THE NEW ECONOMICS OF DRUG DISCOVERY

Target identification

Most major pharmaceutical companies in the past few years have availed themselves of significant access to either proprietary or increasingly useful public cDNA databases to identify and select new targets. Many have attempted to secure large target patent estates. Each company is in a race to obtain as large a slice of the finite target universe as possible in the shortest time. Most of the target patent filing activity is still 'subterranean' and therefore we have not yet seen major litigation over target patents. In fact, it is commonly believed that at the end of the day there will be significant 'trading' of target IP between companies. Some competitive advantage is obtained when companies have the best bioinformatics tools for discovering new targets with putative generic biochemical function through a search for DNA sequences homologous to known therapeutically significant genes, and clearly some companies have used these 'in silico' techniques to try to establish dominant positions in particular families of target gene products, eg GPCRs, tyrosine kinases, cytokines and their receptors etc. However, I doubt that a target patent estate *per se* will be the major source of competitive advantage for a company.

High throughput screening

There is clearly a huge increase in the value of a new genomics target when good drug-like chemical leads have been discovered that are reasonably potent and selective for the target. Massively parallel high throughput screening
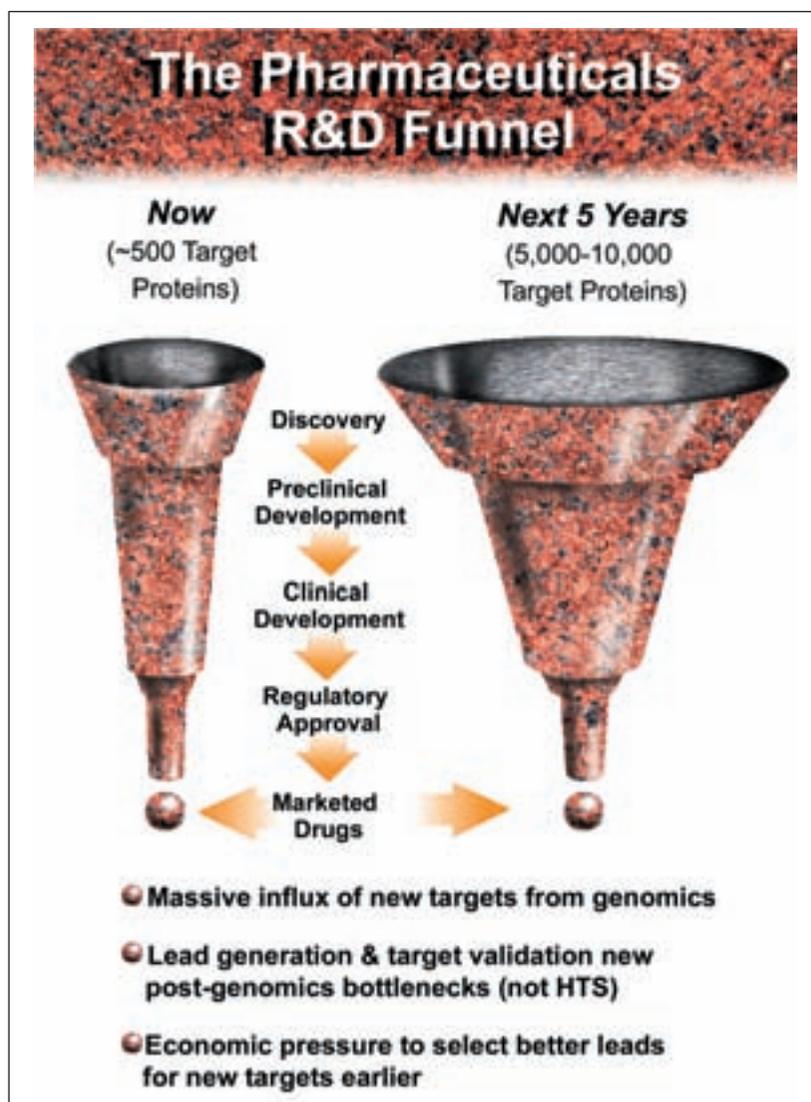
(HTS) has been the strategy of choice for almost a decade to discover initial hits at many targets in a short time. The advent of high volume, high throughput parallel synthetic chemistry ('combinatorial chemistry') that has finally reached the requisite level of chemical diversity means that in principle billions of compounds could be screened at thousands of targets on an industry-wide basis. This is neither an economic nor a scientifically sensible proposition. In fact, it is the rare company that currently is geared up to screen its compound library at more than 100 targets per year. Many companies are also taking a reasonable shortcut of screening across all their targets not their entire library of compounds, but a much smaller subset of compounds that represents the chemical diversity contained within the entire library.

Figure 1

HTS even at the scale at which it is currently practised is an expensive proposition. There are major resource bottlenecks both in the production of recombinant human target proteins, and in the establishment of biochemical HT screens which are necessarily individualised for each target protein. A typical HTS for a new genomics target can take up to six months to set up, and thus represents a major rate limiting step. Several attempts have been made to improve the economics of the HTS process in recent years.

One obvious way to cut expense and the time taken to produce target is ultraminiaturisation, and HTS specialist service companies such as EvoTec[1], Aurora[2] and Caliper[3] are leading the way in nanolitre technology. These same companies have led the effort to increase HTS throughput by miniaturising assay set-ups such that the industry generally is progressing from 96-well plates through the current standard of 384-well plates, to 1536-well plates. Many people believe that 1536-well assays will become the industry standard since in the overall economics of the discovery process it is not easy to see how greater throughput than this will confer advantage.

Another way to improve the economics and success rate of HTS-driven lead generation is to use cell-based rather than biochemical assays, on the premise that cell-based assays have more inherent information content, since there are now powerful techniques for engineering target-driven response systems into cells in such a way that selectivity as well as primary activity can be measured with little incremental resource, or alternatively that entire intracellular signalling cascades can be screened at one time. Multi-gene HTS is also beginning to come on stream. Industry scientists have extensively debated the pros and cons of cell-based versus biochemical assays from the standpoints of information content and discovery economics. I believe an overriding consideration is the desire of industry medicinal chemists to establish structure-activity relationships and generate leads from hits using the harder-edged quantitative affinity data that comes from biochemical screens. Thus typical hits in an initial cell-based assay are rescreened in a more target-specific biochemical assay to obtain more secure data to establish SAR.

A major breakthrough in the effort to reduce discovery cycle times and improve the economics of discovery in the post-genomics era is the advent of the 'any-target' screen. These assays in principle measure the ability of small molecule ligands to bind to a target protein by detecting a change in
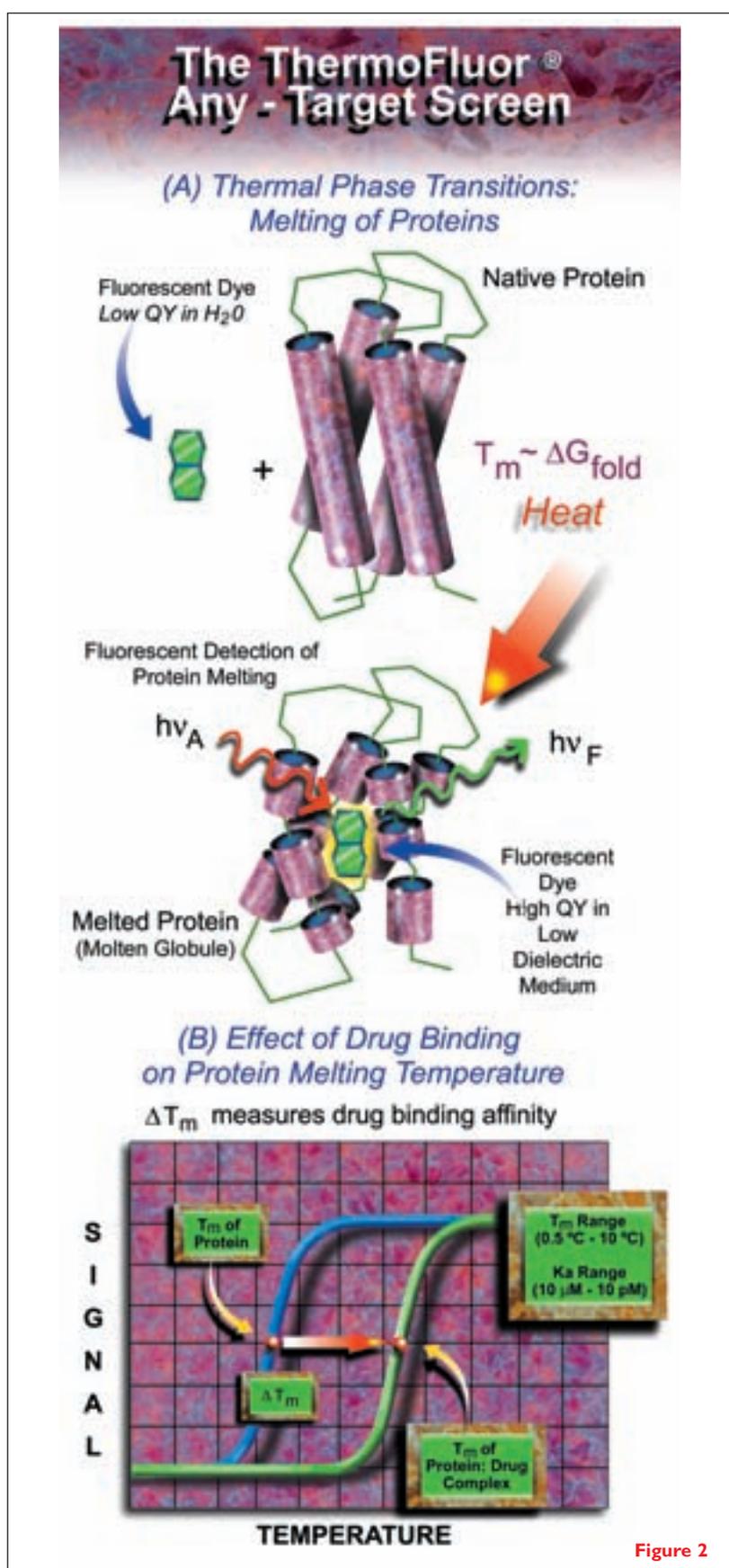

Figure 2

Genomics

Massively Parallel Post-Genomics Lead Generation for Many Genome Targets (with J.I.T. combinatorial chemistry)

the physicochemical properties of the protein, and therefore do not rely on foreknowledge of the biochemical function of the genomics target. The use of such a screen in HTS mode permits, if the screen is sufficiently robust and accurate and has a wide enough 'dynamic range', the generation of high quality lead compounds through iterative screening and SAR analysis in a relatively short time. Such leads can in turn drive the validation of the target through their use as biochemical and pharmacological probes. This 'chemistry-driven target validation' can be much more efficient and less time consuming than other target validation approaches, and plays into the natural predilection on the part of industry scientists to investigate the utility of a new target with a drug-like molecule. Any-target HTS can become a important strategy to progress many new genomics targets in parallel to determine which ones are inherently drugable, and which ones look most promising from the

standpoint of production of high quality leads for oral drugs.

In our laboratories at 3-Dimensional Pharmaceuticals, Inc (3DP)[4], we have developed a robust, quantitative any-target screen called ThermoFluor®, currently used in 384-well HTS mode (**Figure 2**), which relies on the ability of relatively tightly binding small molecule ligands to stabilise proteins to heat-induced melting. The assay detection system is the fluorescence emitted by a proprietary dye when the dye molecules quantitatively intercalate into increasingly exposed hydrophobic portions of the target protein as the latter melts and unfolds. The affinity constant of the ligand can be directly calculated from the extent of shift in the mid-point of the protein melting curve, greatly increasing the information content of what is otherwise a straightforward HTS. Other biotechnology companies employing any-target screening approaches include Anadys[5], Cetek[6] and NeoGenesis[7]. Novalon[8] uses a somewhat different any-target screening approach.

Lead generation

The last decade has seen real advances in the science of combinatorial chemistry, especially in regard to massively parallelising synthesis of discrete compounds from a much more diverse array of building blocks, conferring a significant increase in the chemical diversity of current combinatorial libraries. While many pharmaceutical companies have built or purchased large (10 exp[5] – 10 exp[6]) combinatorial libraries to augment their core compound collection, more attention has been paid to the production of much smaller 'focused libraries' constructed around particular SAR features, to aid lead optimisation. Increasingly, the concept is taking hold that the initial generation of strong leads is most efficiently achieved from a starting point of a moderately-sized diverse collection of compounds (several hundred thousand rather than millions) that occupies useful chemistry space quite comprehensively, and at sufficient density, to provide a reasonably high probability of obtaining several hits with different structural features at any target. Generating leads from these initial hits is most efficiently done by 'exploding' the hits, ie constructing focused libraries of analogues of the hits.

At 3DP, we carry this concept of 'just in time' synthesis to the next phase of efficiency by selecting focused libraries of one or several thousand compounds from a vast pre-designed virtual

Genomics
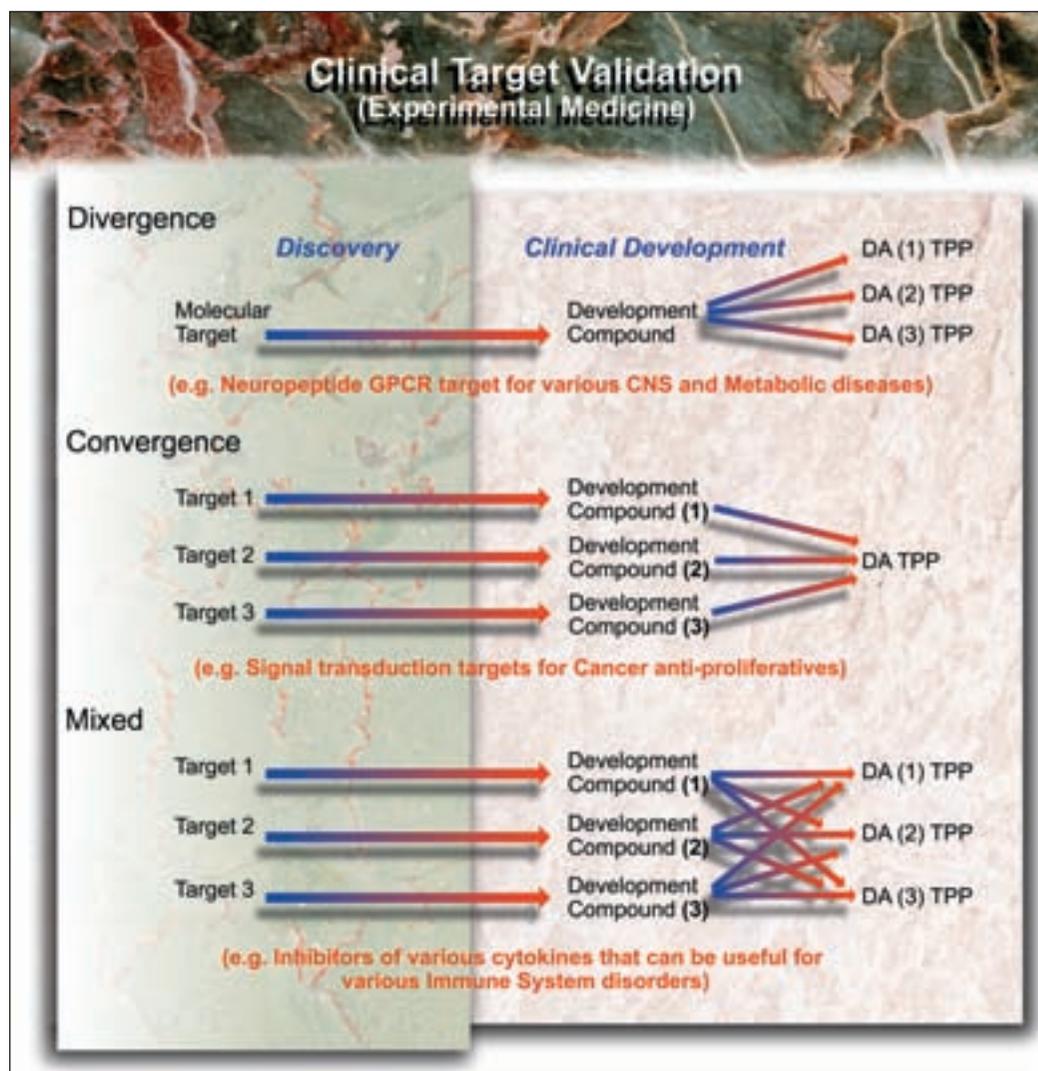
library of 2.5 billion rapidly synthesisable structures which represents the comprehensive analoging by combinatorial means of all the compounds in our initial screening library. The focused libraries are selected on the basis of the preceding SAR data and the desire to co-optimise many desirable drug-like properties in parallel. This lead generation cycle is illustrated schematically in **Figure 3**, and its essential virtue is that it allows the medicinal chemist to conduct very many optimisation experiments in parallel in a short (two to three weeks) cycle time. In consequence, many fewer iterations of this cycle convert raw initial screening hits at new, unvalidated genomics targets to potent, selective lead compounds. The efficiencies built into this process allow leads to be generated at many new targets in parallel, providing a major solution to the lead generation bottleneck.

Target validation

'Chemistry-driven target validation' is an integral component of what has been more generally termed 'ligand-driven discovery'. The global economics of the post-genomics discovery process mandate that, faced with very many new target opportunities to be exploited in a limited timeframe driven by competitive pressures, more emphasis must be placed initially on ascertaining which new targets are inherently drugable, and which of these can in a short time yield fruitful chemistry avenues, than on the comprehensive biological validation of these targets prior to instituting HTS and lead generation. Over the past 10 years many elegant molecular biological methods for target validation have come into play. They run the gamut from massively parallel differential expression (and soon proteomics) screens using array

technology, that yield relatively limited validation information on many targets in an efficient manner, through new pathway analysis approaches driven by protein/protein interaction readouts, to altered gene expression in transgenic animals, which may yield more robust, but still tentative, validation information in a temporally inefficient manner. Other target validation modes employ model organisms (fly, worm, zebra fish) to examine the function of gene products analogous to the target. These biological methods of target validation have their place in the post-genomics discovery process, but any uniform recipe for their use runs counter to the objective of maximising the efficiency of the discovery process. A 'horses for courses' approach to the use of biological target validation tools is best, ie more or fewer target validation experiments should be performed depending on the status of the particular target in the overall portfolio, its competitive position, its *ab initio* degree of association with the human disease, the ultimate commercial value of a drug acting at the target, etc. At SmithKline Beecham a few years ago, we established such a paradigm for employing target validation techniques to different extents at different times within the discovery timeline, depending on our weighting of each target.

The definition of 'target validation' is by no means yet uniform. I tend towards a more rigorous definition that the target must not only be shown to be therapeutically relevant, but that a drug acting selectively at the target has some likelihood of having a clinical profile that will confer competitive advantage and therefore commercial success. According to this definition, the global economics of the discovery process are such that an increasing number of drug candidates will enter the industry's clinical development pipelines that act at new genomics targets that have not themselves been completely validated. We are entering an R&D era where an early development pipeline of drug candidates will require inexpensive, timely, concurrent clinical validation of targets by means of small scale clinical experiments in volunteers and patients that yield high quality data affording reasonably accurate assessments of prospective competitive clinical profiles, to allow key portfolio decisions to be made about which compounds should proceed to Phase IIb and III clinical development. Several pharmaceutical companies are now investing in such new approaches to parallelising PhaseI/IIa or 'clinical proof of concept' work, under the rubric of 'experimental medicine',

Genomics
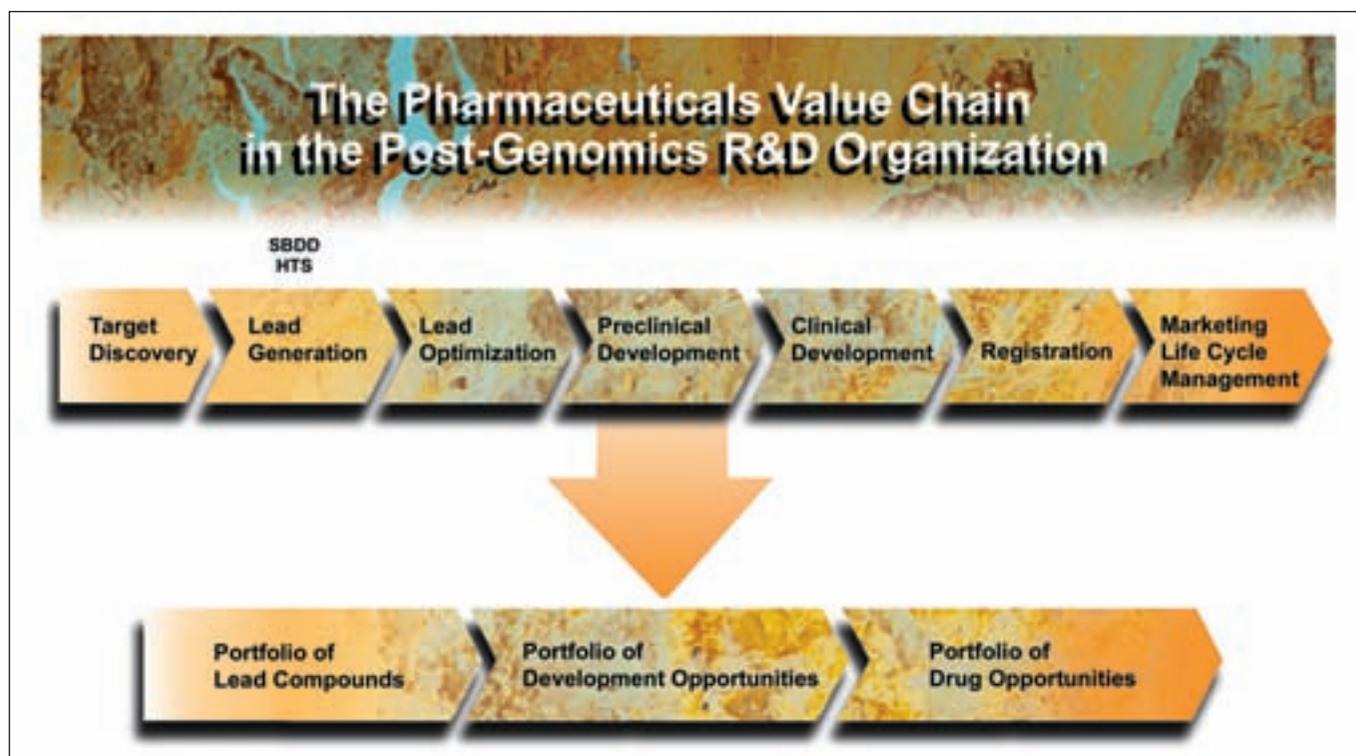
and some specialist service companies in this area have been formed (Predict, Inc)[9]. The experimental medicine approach relies heavily on new clinical informatics paradigms to allow real time recruitment of highly characterised and stratified patient groups for such proof-of-concept trials, and the continuous development of state-of-the-art biomarkers that increasingly predict clinical efficacy with statistical significance in small patient populations. Ultimately, pharmacogenomic stratification of the patient groups will be required. A quite current illustration of situations where the experimental medicine approach would be very beneficial in parallel proof-of-concept trials is shown in **Figure 4**, which illustrates 'divergence', where a particular molecular target is recognised to have multiple more or less probable therapeutic applications (different Target Product Profiles, TPPs); 'convergence', where several targets contemporaneously progressed through discovery and early development are associated with a particular TPP, and 'mixed convergence/divergence', exemplified best by immune system targets.

The new age of structure

More than 20 years ago, the advent of computational modelling software and emergence of structural models of target proteins swung the pendulum in the direction of rational drug design as the most logical and productive avenue for drug discovery, on the basis of 3-dimensional knowledge of the target active site and of small molecule pharmacophores. However, the impact of the approach was initially limited, since it was difficult to rapidly determine the 3D structures of many targets of interest. This limited the co-ordination of structural and chemical synthesis programs, and little attention was paid to designing compounds that would not only have the required target specificity, but also possessed all of the additional properties (eg solubility, low cost of goods, etc) required of a successful development candidate. In the late 1980s, HTS technology and combinatorial chemistry swung the pendulum of opinion back in the direction of relying on chance to discover drug molecules of interest if the mass of screening data was large enough to support a statistical likelihood of useful hits. The recent advances in the 3D structure determination enabled by advances in molecular biology for improved protein production, synchrotron radiation sources for x-ray data collection, and hi-field NMR instruments for solution determination of protein structures, have tremendously expanded the role of structure determination technology in drug discovery. The pendulum perhaps is now resting at an ideal midpoint as we see the data from target structure analysis becoming completely integrated into computational selection of focused libraries for

synthesis (3DP), and powerful docking software becoming available to examine the interaction of very large virtual libraries of compounds with the target active site. This trend will be strongly reinforced by new technologies that allow high throughput target protein crystallisation and structure determination (Structural GenomiX[10], Syrrx[11]). The integration of target structure analysis with HTS/combinatorial chemistry should be a powerful driver for the more routine generation of high quality lead compounds at genomics targets.

R&D STRATEGY – SOURCES OF COMPETITIVE ADVANTAGE

Portfolio management

In the post-genomics era with a likely abundance of good-looking compounds acting at a much larger number of targets than historically accessible to the industry, what are the sources of competitive advantage? Increasingly, companies are reorganising their R&D groups in recognition that there are three portfolios of assets that need to be very actively managed (**Figure 5**) – Lead Compounds, 'Development Opportunities' (compounds that have satisfied clinical proof of concept in Phase IIa studies) and 'Drug Opportunities' (Phase IIb/III compounds). To compete successfully, a company must aim to sustain portfolios of Lead Compounds and Development Opportunities that have the maximum risk-adjusted expected value (with an ideal balance of risk), and must aim to keep these portfolios as large as economically possible, so that there is always an active choice from many good options when moving projects through these three phases of the R&D process. Historically, few companies have had the luxury of this kind of active choice, but rather have had to take what comes along out of discovery, or Phase II. The difference is schematically illustrated in **Figure 6**. It is clear that the most efficient and economic path to successful, high value, competitive drug R&D is to produce good leads at many targets in parallel, and many good clinically validated compounds in parallel through early development. The technology advances driven by or accompanying the genomics revolution, described in the first issue of *DDW* and above, coupled to a clear-sighted strategy that focuses on the largest future return on present investment in the shortest time, should allow pharmaceutical companies to thoroughly exploit the human genome.

Scale and critical mass

Finally, it is worth touching on the strategic issues of scale and critical mass in the competitive industry race. In the current era of industry M&A activity, this topic has generated more heat than light. For the technology platform that underpins the R&D pipeline, critical mass is more important than scale – it is vital to be able to put up the 'table stakes' to keep your technology platform constantly state-of-the art for your R&D strategy.

On the other hand, for the R&D pipeline itself, scale is fundamental – the number and scope of R&D projects must be commensurate with the commercial requirements of the company. However the competitive status of an R&D pipeline is more cogently examined at the therapeutic area or disease area level rather than globally, taking into consideration all the technical and commercial therapeutic area talent in an organisation, as well as the R&D project portfolio in the area. A few years ago at Zeneca, we devised a simple system for evaluating and comparing R&D efforts in different disease areas (hierarchically a level lower than therapeutic area, eg asthma, psychiatric disease, rheumatoid arthritis, etc). This is shown in **Figure 7**, where, in a hierarchy of many molecular targets feeding into a smaller set of preclinical biological effects that in turn translate to the several clinical target product profiles that constitute the company's activities in a disease area, the expected value of
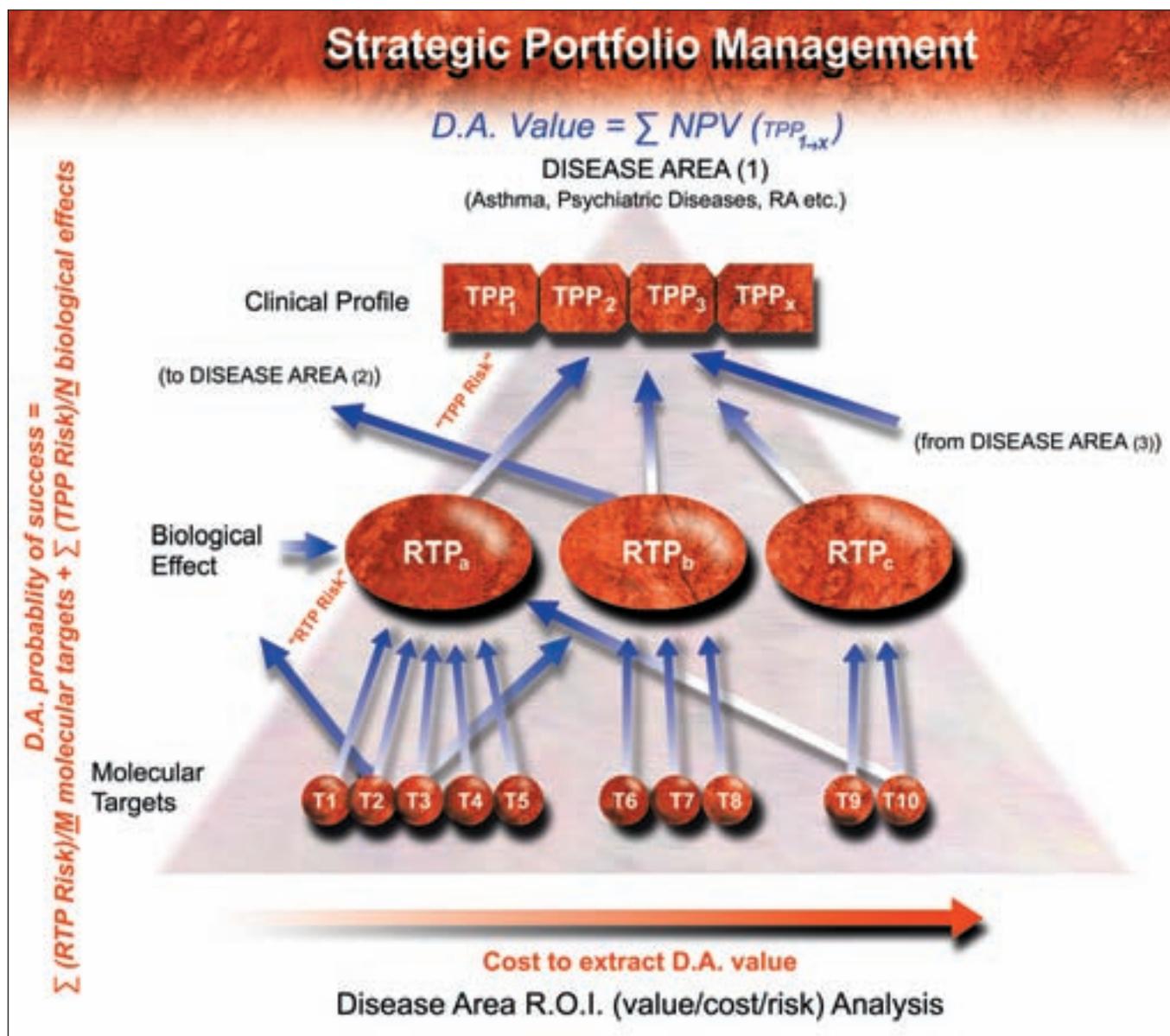
References
1 Details on EvoTec can be found through their website at URL www.evotec.com
2 www.aurorabio.com
3 www.calipertech.com
4 www.3dp.com
5 www.anadyspharma.com
6 www.cetek.com
7 www.neogenesis.com
8 www.novalon.com
9 www.predict-inc.com
10 www.stromix.com
11 www.syrrx.com

Figure 6

Genomics

a disease area is roughly measured as the sum of the value (PYS, NPV) of the TPPs, and the investment is the overall cost (technical and commercial) to extract the expected value. Two major types of risk are factored in, the purely technical risk that action at a particular target will not provide the desired biological effect ('research target profile'), and the techno-commercial risk that the biological effect itself does not fully translate to the competitive clinical target product profile. This way of looking at the different ROIs for the disease areas in the R&D portfolio allows for strategic prioritisation, focus and competitive excellence within and across disease areas, and is an essential strategic overlay to the above operational portfolio management concepts when managing one's business to fully exploit the fruits of the human genome in pharmaceutical research and development. **DDW**

Formerly President, Research and Development, SmithKline Beecham Pharmaceuticals and International Research Director of Zeneca, Dr David U'Prichard has overseen the entry of 10 compounds into clinical development and progressed the clinical development to NDA filing of Avandia, the blockbuster diabetes drug. Dr U'Prichard was instrumental in the launch of Nova Pharmaceuticals in 1983 and is the current Chief Executive Officer of 3D Pharmaceuticals, Inc.