A strategic vision for 21ST CENTURY DRUG DISCOVERY

Delivering genuine therapeutic innovation is more challenging and complex than ever. Despite major scientific breakthroughs and technological advances leading to a greater understanding of the aetiology of diseases, it is generally accepted that the current approach to creating new small molecule therapeutics remains fundamentally inefficient. This article explores whether adopting a more holistic drug-hunting approach may increase effectiveness and ultimately prove more sustainable, and provides a strategic vision of what needs to happen next.

Bringing a novel therapeutic to the market is a lengthy process requiring endurance and commitment. Scientific and technological advances and greater understanding of the underlying mechanisms of disease are needed to fuel the discovery of new treatments and cures, but the overall cost of research and development are being driven higher by greater regulatory stringency and longer times to market impact discovery funding.

Ideally, the drug-like properties promised by a discovered molecule will be safely delivered as the expected therapeutic response in a human patient during the development phase. How much time and cost should be expended on ensuring that the promise is actually fulfilled requires a careful balance of expediency and effectiveness, as there is nothing more wasteful than late-stage failure in the clinic due to impaired efficacy or safety (**Figure 1**)[1]. Being effective is far more important than being expedient[2]. Effectiveness depends not only on generating and demonstrating a potentially safe, effective and deliverable ligand to the target in model experiments, but also on the validity of the target as an essential mediator of the target disease. As a viable drug is the only certain validator of a target and *vice versa*, and target validity is unalterable, for as long as this paradox remains unresolved the best achievable project security is that of striving to identify an superlative rather than an adequate ligand.

In this article, we present an analysis of past and current drug discovery approaches, exposing their relative strengths and weaknesses. From this, a new strategic vision is suggested that can contribute to a resilient, sustainable biopharmaceutical industry to satisfy the demands of 21st century healthcare. After nearly 25 years of being a passive partner in the passive process of screening large collections of known compounds in the hope that happenstance will provide a desperately-needed cure for a significant unmet medical need, the vision is one in which the active application of accumulated and newly-emerging knowledge will force the direction of drug discovery towards these goals. The concept of exploratory and directed drug discovery is not new, quite the contrary, but during the fallow years of its use in large pharmaceutical companies, many new and relevant technologies have become available. Their incorporation into the exploratory drug discovery paradigm has the potential to be a disruptive tool for the discovery of novel and exclusive drugs.

By Dr Wayne Bowen, Dr Giles Sanders, Dr Andreas Werdich, Dr Brian Warrington and Dr Elizabeth Farrant
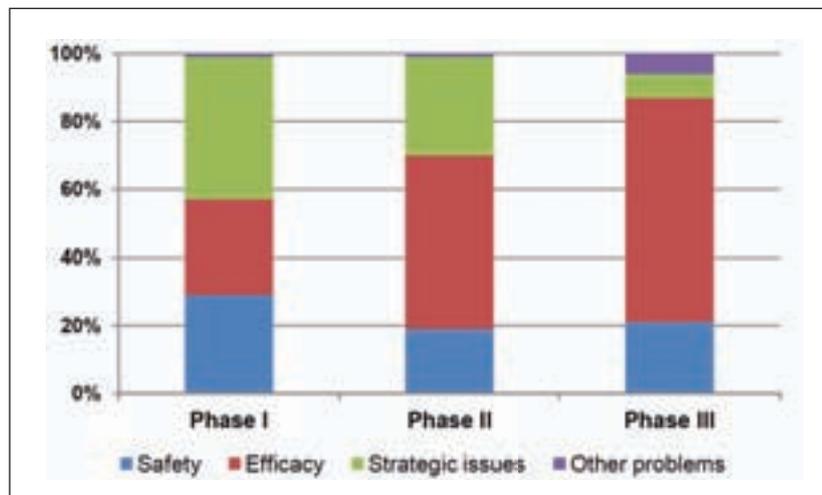
Figure 1: Overview of principle causes of failure at each clinical phase between 2000 and 2009 (based on 410 drug candidates). When combined, insufficient efficacy is the main obstacle to success, and dominates Phase III issues

A brief history of drug discovery

The time before high throughput

Man has always engaged in drug discovery. In early times the shaman sought out exogenous materials with recreational, curative, restorative, disinhibiting or poisonous properties. The knowledge extended to observations that certain natural products were best gathered during a particular season, that some could be purified by sublimation, and that active agents could be co-administered to synergise or mutually annihilate their pharmacological effects. Clearly, the shaman understood the role iterative closed-loop hypothesis-led experimentation ('think and try') as a means of extending knowledge to make real and novel progress in science and technology.

Only the nomenclature has changed, the shaman's basic mantra 'get', 'test', 'think', 'try again' being rephrased in terms of reflecting the technology, in this case drug discovery (**Figure 2**). Here we see it as a succession of hypotheses that can be tested experimentally by the production of an appropriate molecule, each new hypothesis being based on an intelligent assessment of the several biological test results obtained in the previous cycle[3]. In every cycle, the method will either find a solution better than that already known or fail to do so. Either result has the value in extending knowledge of the structural and electronic factors governing activity for all of the properties assayed. The process ends when a developable molecule is obtained or when consistent failure can no longer be tolerated. The cyclic process can be seeded in a variety of ways: screening hits or leads, natural molecules, active molecules marred by unwanted properties, active molecules with no known mechanism of action, or even molecules designed *de novo* with a theoretical chance of activity. The process is most effective when several assays that can measure key 'drug-like' properties (eg potency, selectivity, bioavailability and toxicity) are used in parallel so that the different structure-activity requirements for each of these properties can be considered jointly in the analysis and design steps to target molecules whose holistic property profile will be most likely to succeed in development.

### The 1980s and the shift in approach

A threat perceived by large Pharma during the 1980s was that genes identified by shotgun sequencing and reassembly of expressed sequence tags (ESTs) could be patented by third party sequencers, and this could prohibit them from being therapeutic targets except by paying a large licence fee. The success of the 'Moore's Law'-style increases in productivity in genomics, arising from automated high throughput sequencing and analysis of large sample and data pools, and that this could result in the complete disclosure of an entire human genome, also excited a view that the application of similar high throughput automation could energise other components of the drug discovery process, particularly chemistry.

From this developed the notion that if a sufficiently large collection of drug-like compounds could be assembled and screened, the drug discovery process could be turned into a single linear automated workflow notionally suitable for finding leads for all possible therapeutic targets (**Figure 3**). The foundation of this process was to be pre-validation of gene products of the human genome believed to be 'therapeutic' mediators. Each was to be subjected in turn to high throughput screening (HTS) to find drug leads in an existing compound collection. To improve the chance of success, the collection would be augmented in size and diversity using automated high throughput chemistry (HTC). The 'hits' found by HTS would then be verified and qualified by filtering through a succession of assays, in a so-called 'Hit to Lead' (H2L) process to find to any compounds meeting the more advanced biological and chemical criteria for a 'lead'. The lead could then seed a more intelligent process (reminiscent of that shown in **Figure 2**) to explore congeners within the lead chemotype, to find a molecule with the more optimal properties needed to meet the preclinical criteria of a development candidate.

The effects of the new approach

To those designing the new approach, compared to the automated élan and high throughput paradigms of genomics and screening, the iterative cycling of medicinal chemistry to identify a single highly customised drug candidate for only one particular therapeutic target was seen as inconvenient. It stood in the way of a more expedient lead discovery process that could serve all targets proteins generated from the human genome. On balance, it was decided that the opportunistic gains to be garnered by a more expeditious path of identifying novel therapeutic proteins and automatically matching them with a pre-prepared lead contained in a huge library of candidate leads outweighed the compromises made in terms of lost rigour, scope, novelty and would probably output a different candidate for development. The key limitations of the linear process were:

1. The axiom that a therapeutic target is only validated by a viable drug is replaced by a concept of pre-validation of a target based on relevance to the disease rather than 'essentiality'.

2. Chemical scope is limited. For practical reasons, such as maintaining target throughput and library management and curation, the size of the compound library that can be physically displayed for screening (of the order of 10^6 molecules) amounts to only an infinitesimal representation of the drug-like universe as defined by Lipinski's Rule of Five (RO5)[4] (up to 10^{60} molecules)[5] so there is no possibility of achieving comprehensive display drug-like chemical entities. 'Rumsfeld's Razor'[6] further restricts the choice of molecules to those known or designed to be active at historical targets and so cannot contain the unknown compound needed for an unknown target. These problems can only be solved by either an exhaustive lead library for HTS (a physical impossibility) or an active iterative closed-loop search supported by appropriate target-based or phenomenological (phenotypic) assays and intelligent choices (eg the iterative search defined in **Figure 2**).

3. The discovery strategy is less ambitious in its concept of 'lead'. In the iterative model, a 'lead' was a temporary term for the most favoured compound of the day in an unrestricted and open-ended search for a superlative molecule as the development candidate. In the linear model (**Figure 3**), 'The Lead' is the molecule from the screening library best able to satisfy minimal pre-set 'Lead Criteria'. The subsequent exploration of the lead chemotype for a congener meeting the requirements of a development candidate unlikely to pro-
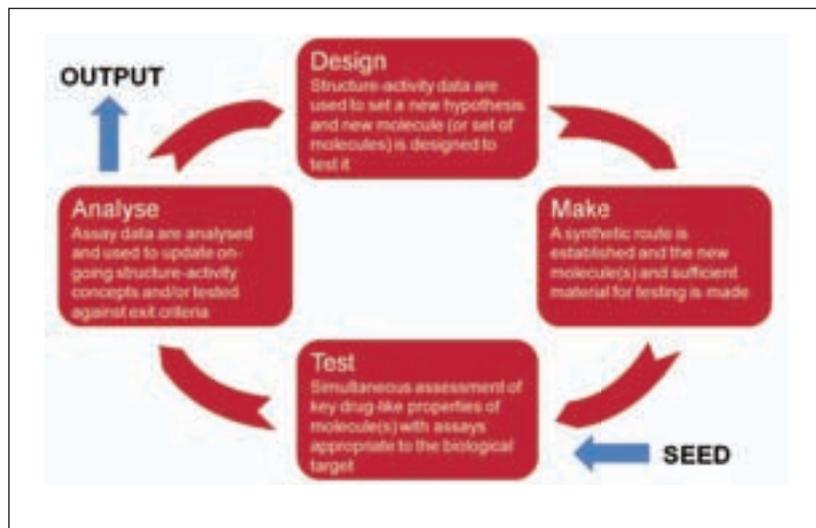

Figure 2: An iterative closed-loop discovery paradigm. Conceptually, a drug is evolved stepwise in each iteration by setting a new structure-activity hypothesis and testing it by synthesis of a new compound (or set of compounds). The design of this test molecule(s) incorporates all of the previously-accumulated structure-activity concepts for all assays used to assess 'drug-like' properties up to the analysis step of the previous cycle. The process can be seeded from target or phenotypic information such as a hit, lead or observation of activity associated with a chemical class or *de novo* based on knowledge of the target protein. As Analysis and Design can reflect the choices of individual users or algorithms, the process is stochastic and structure and quality of the final output will be dependent on the quality of these inputs. In a similar way, the process can be tuned towards diversification or optimisation

vide a better molecule than an exploratory search of more chemotypes.

4. The rigour and scope of the assay cascade is another casualty of the Rumsfeldian principle[6]. The results of each assay in the cascade alone determine the composition of the much smaller set of compounds the next assay will receive. Thus, a molecule showing modest, but adequate, potency which might achieve an excellent profile in subsequent assays can be rejected in favour of passing through a more potent molecule with poorer prospects. In addition, the fast, facile assays suitable for the automated liquid handling needed to achieve timely processing of large numbers of molecules restricts the type of targets that can be assayed. Generally, assay development for novel targets is more challenging than for existing assay families.

5. A side-effect was division of research into silos – at worst, biology and chemistry efforts comprised large independent teams without common goals and discourse.

With 25 years of hindsight, there is enough time separation for a fair comparison of the virtues and pitfalls of the two paradigms to be made. Most notably the cost defrayable against each launched
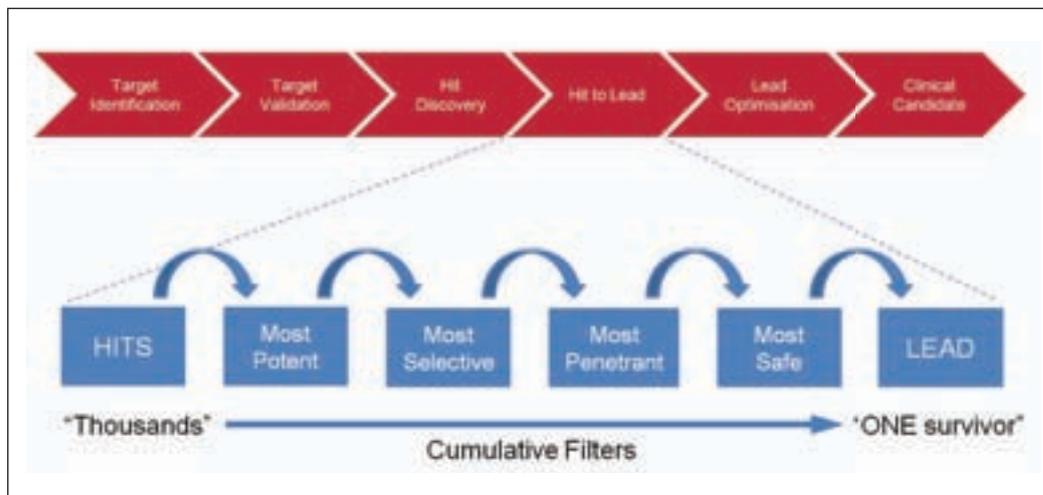
Figure 3: RED arrows: A linear 'industrial' drug discovery process embodying the notion that drug targets can be pre-validated without need of a proven drug and that a speculative sufficiently large collection of real, compounds, deemed 'drug-like' on the basis of historical findings, contains the leads for all new targets. A found lead is then optimised by exploring congeners of the lead within the scope of the lead's chemotype. BLUE arrows: finding the lead for a particular target is an analytical process using a succession of filter screens. The survival rate at each stage is set to cull an intake of hits that may amount to many thousands to yield one lead. The approach is opportunistic in that, for expediency, a molecule with adequate potency and an excellent profile in later assays can be excluded in favour of an extremely potent molecule performing poorly in subsequent assays

drug has risen from considerably less than $1 billion to more than $3 billion, but the rate of output of new drugs has not increased, the proportion of late stage failures has increased and drug gestation time remains about 12 years. In the opinion of regulators, the novelty of small molecule output has decreased and there has been little impact on important and urgent social targets such as dementia and novel antibiotics.

The industry itself has not fared well during this era having moved from a period of great growth with individual companies making breakthrough drugs that enjoyed many years of market exclusivity to those same companies turning out poorly differentiated products trying to compete in shared markets and undergoing consolidation and downsizing. Many, even within the industry, have seen Pharma's dulled performance as a reflection of the change in its discovery strategy rather than a coincidence, as evidenced by Pharma's new practices of seeking novelty and innovation through open innovation and precompetitive collaborations.

Recent approaches to improve return
While the hardware and software of high throughput automation and the scope and facility of secondary assays have grown steadily, the focus of effort in chemistry has been to try to fix shortfalls in the chemical library serving the linear paradigm rather than instituting any alternative strategy, not

least because of the large investment made in the overall linear strategy. Thus, the incompatible objectives of creating bigger and better libraries remain in place and are only ameliorated by efforts to improve chemical library diversity through cheminformatics, or increasing the number of molecules displayed for screening by the implementation of improved automation for high-throughput chemistry and more efficient workflows.

Chemistry: the numbers game
Success in opportunistic HTS relies on the timely screening of each of a succession of biological targets. Even with fast assays and good library curation, the economically optimal limit seems to be in the order of a million notionally 'drug-like' molecules. Approaches to enrich the diversity of these screening libraries to improve hit rates with novel targets have been based on multiple regression analysis to 'score' the relative contribution to biological activity of substituents in fixed scaffolds[7], or the physiochemical character of their substituents, or global descriptors of hydrophobic, electronic, steric and other effects[8] through to the energetics-based plethora of descriptors of ligand efficiency used and argued about today[9,10]. The concept behind all of these descriptors, however, is that existing molecules can inform the design of future molecules. While it is true that these tools can find better molecules within known drug space,

again it is the Rumsfeldian principle that prevents the finding and filling of 'holes' in unfound-drug space where the unknown drugs for unknown targets might lie. These can only be found by active exploration of new chemical space (**Figure 4**), not passive screening or the synthesis of molecules claimed to be diverse based only on an index that references similarity to historical molecules[11].

Fragments: A Trojan horse?

As an example of recent discovery approaches, fragment-based lead discovery (FBLD) recognises the paradox that exists in HTS screening libraries in that molecules included on the basis of their drug-like selectivity and lack of toxicity at historical drug targets are actually less likely to bind to unrelated novel targets. To overcome this, FBLD seeks to find from a library of 'fragments' a low molecular weight 'fragment', or combination of 'fragments', that bind the novel target with only marginal affinity. These can then be intelligently elaborated into larger 'RO5' molecule leads with high affinity and selectivity for the target using an iterative method emulating that shown in **Figure 2**. A key difference is the *modus operandi* in determining the affinity of the seed 'fragment' and early derivatives as their binding is too weak, and the concentrations required too high to permit the use of routine biological assays.

The perceived advantages of FBLD are that the necessary hydrophilic character (for solubility during screening) also ensures high ligand efficiency due to the high enthalpic binding in the elaborated product which can be further enhanced by the entropic contributions of hydrophobic character without rendering the final compound too hydrophobic for use. In addition, combination of multiple fragments magnifies the screening power of the library by the square or cube relative to the same number of molecules in an RO5 set. The practical downside of the method is the complex biophysical technology needed to detect the marginal affinity of fragments. In this respect, the 2D-protein detected NMR used in the original 'SAR by NMR' work at Abbott to measure the changes in specific amino acids of the target disturbed by the fragment[12] or fragments[13], has now been joined by a host of no less simple screening techniques[14]. The need to obtain structural models of the protein fragment complex by x-ray crystallography is a further encumberment[15].

With just a few reported successes and little knowledge of the failure rate, it is probably too soon to say how versatile and reliable FBLD can be as a tool for lead discovery. However, although there is some likelihood that the fragment libraries and targets of competitors overlap, it should not be a concern as the fragment loses its identity in the very first step of a stochastic iterative elaboration to a full lead or even a product. In effect, it may be said that FBLD is just another way of seeding an iterative process. As iterative processes are stochastic, ie where they end up is highly dependent on what is measured and how it is used in the molecule design, it is probable that in different hands the same fragment/protein combination will deliver distinctive solutions reflecting the quality of the assessment and design steps. In fact, the same fragment could seed solutions for other targets in different therapeutic areas. However, in all cases, the versatility and the quality of the found asset derives at least as much from iterative process as much as from the seed, including a propensity for obliquity[16] when a discovery team is allowed to exercise their unfettered skills in pursuit of a superlative solution from even a hint of activity.

Where are we now?

Large Pharma is in the odd position of being about the only industry that has invested in high throughput automation without gaining the expected increase in output and lower costs that became expressed as 'Moore's Law'[17]. That is not to say that performance output did not increase for individual tasks. In fact, the productivity measured as the number of discovery items completed, compounds screened, data points plotted, compounds synthesised, hits identified and leads or candidates delivered, increased rapidly on a year-on-year basis. By the measure of the earlier iterative process (**Figure 2**), which could identify breakthrough drugs by preparing and testing less than 1,000 compounds, under the new regime a better than 1,000-fold drug output might have been expected. In fact, despite a huge hike in the cost of R&D, the rate of output of drugs did not substantially change, the time to market has remained at 12 years and there was good evidence that their market novelty and exclusivity was often impaired. Some might say that an active stochastic intelligent iterative search was better than unguided compound synthesis and fatalistic screening. However, the response to this deficiency has been mainly to seek to further increase HTS and HTC throughputs, seeking a library size where random chemical profligacy will actually enhance corporate productivity, particularly in respect of novelty. As discussed above, it is unlikely that this position can be reached.
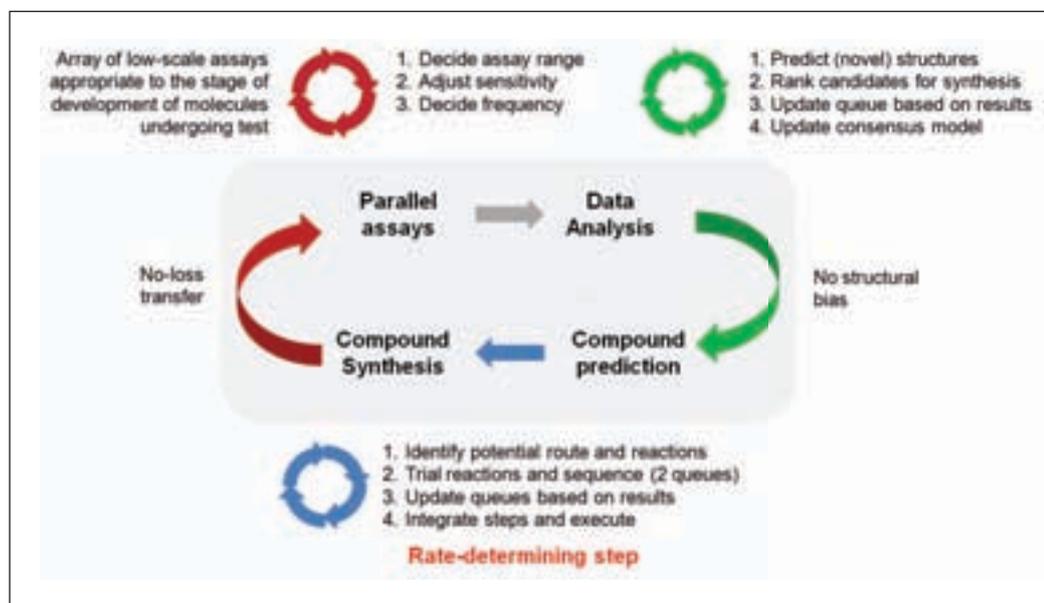
Figure 4: The iterative discovery paradigm is necessarily rate-limited by the 'Make' step because the predicted target-specific molecule (or small set) is extremely likely to be novel. Therefore, its synthesis will need to be investigated. This investigation is another example of an iterative discovery process (Blue arrows) as is the compound prediction process (Green arrows) . However, the long duration of the 'Make' step allows time for a detailed, multiparameter biological analysis of a test molecule (or small set of molecules) and the optimisation and modelling of putatively improved molecules

Attempts to improve hardware and software for the iterative paradigm have been held back by the absence of resources and the difficulties encountered in trying to conduct exploratory chemistry using array synthesisers designed to carry out cheminformatic-based exploitation of known reaction types. It is somewhat ironic that after more than two decades of development of the linear strategy that the state-of-the-art in lead and candidate identification is the iterative learning process supporting FBLD. It is now surely time that the shaman's art received the benefit of modern tools fitted to that purpose so that in-house knowledge, experience and practices of corporate discovery teams can be built into products to generate durable exclusivity in the market based around those important medical needs inaccessible by screening. In fact, with corporate medicinal chemistry know-how as the competitive arena, the development of appropriate and better tools for iterative discovery should be, like the test tube, a mutually beneficial, precompetitive activity.

Back to the future: rational design and iterative discovery in a modern world

The four tasks of an iterative paradigm of 'Design', 'Make', 'Test', 'Analyse' (**Figure 2**) are necessarily rate-limited by the 'Make' step because the target-specific molecule (or a small set of molecules) predicted by unfettered application of structure-activity relationships, as updated in the previous cycle, is extremely likely to be novel. Therefore, a route for its synthesis will need to be investigated. This is itself an iterative 'Design', 'Test', 'Analyse' sub-loop (**Figure 4**). It is the longer duration of this process, however, that allows time for detailed, multiparameter biological analysis of molecules and the modelling of putatively improved molecules. In addition, several different stochastic chemistry threads directed to the same target could be pursued simultaneously using the same assay complement for 'Test' (**Figure 4**).

Due to the novelty of each molecule in the 'Make' phase, a wide range of synthetic methods and technologies need to be available to the chemist. Previous attempts to automate chemistry have focused on trying to force-fit chemical endeavours into a limited range of chemistries carried out using inappropriate high throughput synthesis systems. This has limited the available chemistry space to those reactions that can be performed in this manner[18]. To fully explore chemistry space, a chemist must use the full range of reactions combined with a full range of chemical technologies including batch, flow, microwave,

photochemical and electrochemical reactors along with the relevant analytical and downstream processing and isolation technologies. In the past, this has been a challenge due to the lack of integration and software flexibility. Advances in the Internet of Things and anthropomorphic automation, however, may provide new opportunities in this area[19].

Another challenge when accessing entirely novel chemistry space is reactant supply. Often reactants for novel molecules are in themselves novel and have to be custom synthesised or are available only in small quantities. In the discovery phase, material sparing technologies are key to fast progress. However, multi-well HTC synthesisers based on liquid handling robots can be useful in relaying the production of a congeneric range intermediates prepared via a single reaction type.

There have been attempts to reduce the material requirement and time needed for exploratory chemistry. The shortened reaction and purification times, analyte frugality and low loss transfers possible in nano and micro flow systems were an attractive basis for exploration of fast cycling lead and synthesis discovery systems, and efforts in this area have been expertly reviewed by Rodrigues et al[20]. In these it has been possible to adopt the many analyte-sparing protein-based or cell-based assays formatted for HTS. The fusion of solvent-

based colliding or printed droplets may also be a new technology worth developing for route scouting (**Figure 5**). Clearly, larger working scales and equipment can become relevant if the supply chain for intermediates can support them without limiting rate. Whatever hardware path is taken, key accessories requiring development are equipment with robust operating and recovery procedures, a self-educating reactions prediction and methods database and the operational platform, sensors and actuators that are essential to a goal-seeking automated system.

From a chemistry optimisation standpoint, embracing knowledge from every reaction is critical (eg the use of knowledge of what also is produced by side reactions to enable improved route scouting for any future synthesis). There is a need to integrate synthesis prediction and machine learning and this should be a focus for future work, as should integration of the many new information sources that have grown with the internet that could inform the piloting of an automated closed loop methodology or be presented for the consideration of a human operator. Machine learning can also have an impact on post-reaction processing and analysis of reaction products.

The model: nuclei or electrons?

If the full potential of the iterative model is to be achieved, predictive modelling techniques are required that can at extremes either propose a scaffold hop or display the full electronic, steric and conformational effects arising from the adjustment of just one H-atom.

Molecular recognition at the site of biological interaction is a complex process, involving both thermodynamic and kinetic effects. Modelling these complex effects requires a series of trade-off decisions to be made in order to enable calculations to run in reasonable time. This is especially true when dealing with large numbers of molecules (or potential molecules), highly flexible molecules (eg short peptides) or when modelling complexes of drug molecule and biological target. Molecular recognition and associated biological activity is related to both the 3D shape of the molecule and the electronic structure of the molecule. In a commonly-used simplification, the properties of the ligand are dealt with in isolation from any inductive or co-operative effects due to binding. Suitable computational systems then need to be able to model and compare the 3D shape and electronic properties of whole molecules; and in the context of an iterative closed loop system, also need to be able to run calculations in reasonable time. This
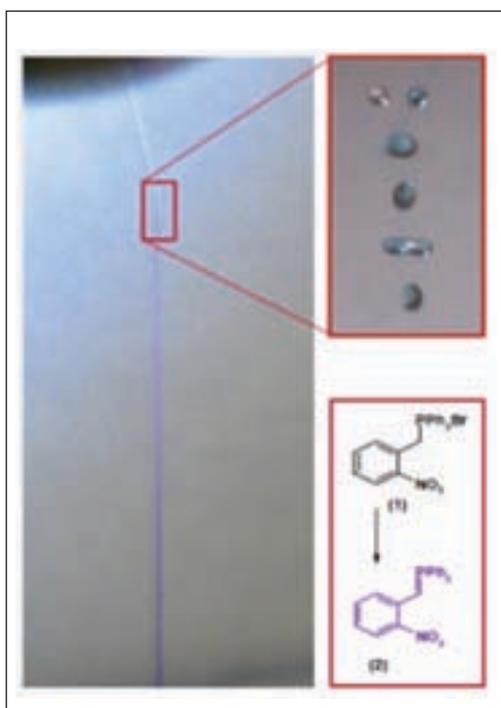
Figure 5: Chemical synthesis using nanofluidic droplet mixing

Drug Discovery

References

1 Tufts Center for the Study of Drug Development, Data Analysis. Reasons for clinical failures by phase. Appl. Clin. Trials. 2014 (December 2013/January);12.

2 Elebring, T. What is the most important approach in current drug discovery: doing the right things or doing the things right? Drug Discovery Today. 2012;17:1166-1169.

3 Plowright, AT, Johnstone, C, Kihlberg, J, Pettersson, J, Robb, G, Thompson, RA. Hypothesis driven drug design: improving quality and effectiveness of the design-make-test-analyse cycle. Drug Discovery Today. 2012;17(1-2):56-62.

4 Lipinski, C, Hopkins, A. Navigating chemical space for biology and medicine. Nature 2004;432:855-861.

5 Virshup, AM, Contreras-García, J, Wipf, P, Yang, W, Beratan, DN. Stochastic voyages into uncharted chemical space produce a representative library of all possible drug-like compounds. J. Am. Chem. Soc. 2013;135:7296-7303.

6 "There are known knowns. There are things we know that we know. There are known unknowns. That is to say, there are things that we now know we don't know." United States Secretary of Defense Donald Rumsfeld gave to a question at a US Department of Defense news briefing on February 12, 2002.

7 Free, SMJ, Wilson, JW. A mathematical contribution to structure activity studies. J. Med. Chem. 1964;7:395-399.

8 Hansch, C. Quantitative approach to biochemical structure-activity relationships. Acc Chem Res.1969;2: 232-239.

9 Shultz, MD. Setting expectations in molecular optimizations: Strengths and limitations of commonly used composite parameters. Bioorg Med Chem Lett. 2013; 23(21): 5980-5991.
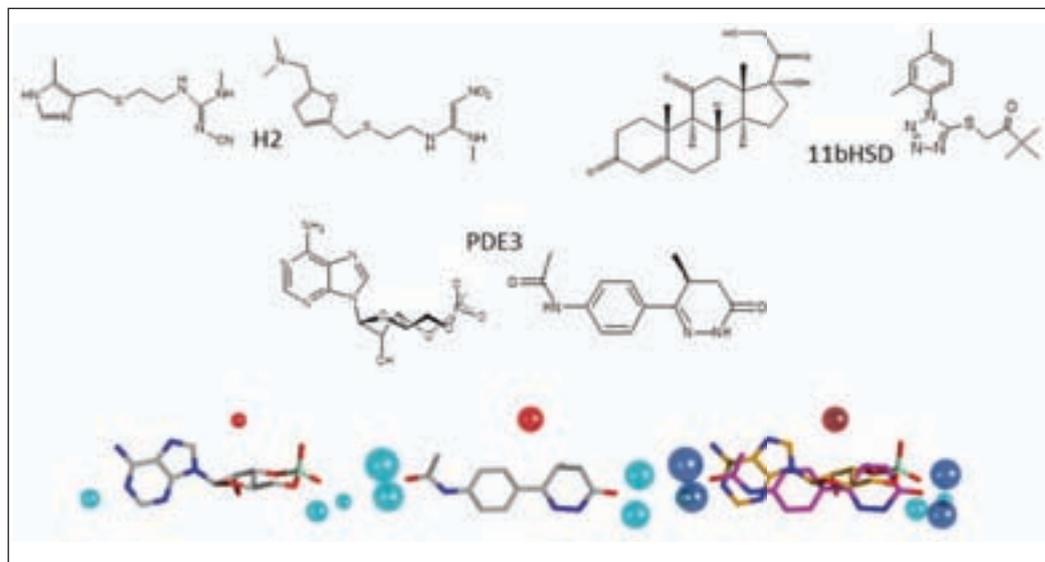
Figure 6: Three examples of diverse chemotype pairs acting at the same biological site. The 3D pharmacophores of the PDE3 interactors overlay in structure and electrostatic field space (Red centres are +ve minima, blue centres are -ve minima)[31]

limits choice to modelling systems that can view a molecule holistically as a 3D electronic pharmacophore and at the same time retain essential structural size and shape, yet deal with the computational challenge with sufficient approximation that the calculation time is tractable. A number of computational chemistry approaches are available to meet this challenge, each with its own set of compromises. As an example, the software from Cresset[21] has been designed explicitly to deal with the challenges of creating a population of reasonable molecular conformations and then efficiently assessing and comparing these using both shape and electronic properties. A consequence of this is that the Cresset tools are '2D structure agnostic' and are therefore extremely well-suited to the task of moving between different molecular scaffolds while retaining features essential for molecular recognition and biological activity.

It is well known that very different structures can bind at the same biological site (**Figure 6**). The 3D holistic pharmacophore is very similar across the diverse structures and can successfully hop from one chemotype to another while retaining biological effect. Several of the modelling software suppliers provide tools that can be used to achieve this result, eg Schrodinger, CCG and OpenEye, but none of these is integrated nor accurate enough because each relies on crude atom-centred charges which cannot describe experimentally observed electronic distribution. The Cresset technology is founded on the principle of 3D molecu-

lar holisticism and has experience in this area. It would now also be possible to draw on ontological and collected data pools to influence the direction of iteration such as PubChem, ChemSpider, DrugBank and UniProt.

Testing – effective use of time and technologies

Molecular assays underpin hit discovery in the linear paradigm due to their simplicity and cost-effectiveness, providing primarily insight on target engagement. In the iterative closed loop model, however, there is an opportunity to leverage information-rich, disease-relevant cellular approaches due to the low number of molecules requiring profiling (typically <1,000) and their phased synthesis. For example, it enables testing on primary cells derived from human tissue biopsy at the primary stage despite limiting numbers. Furthermore, these biomaterials can be transformed using advanced gene editing tools into isogenic cell lines to enable the study of both disease and 'control' responses to a new molecule against an otherwise identical genetic background. Emerging 3D cell technologies such as organoids and organ-on-chips can also be exploited to mimic human physiology on a microscale[22].

Finally, the use of animal models for phenotype-based screening in the initial stages could circumvent many of the failures associated with an exclusively target-centred drug discovery strategy. The zebrafish is a vertebrate whose organs are remarkably similar to their human counterparts at the

Drug Discovery

10 Shultz, MD. The thermodynamic basis for the use of lipophilic efficiency (LipE) in enthalpic optimizations. Bioorg Med Chem Lett. 2013; 23(21): 5992-6000.

11 Bajusz, D, Rácz, A, Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? Journal of Cheminformatics. 2015;7(20):1-13.

12 Shuker, SB. Hajduk, PJ, Meadows, RP, Fesik, SW. Discovering high-affinity ligands for proteins: SAR by NMR. Science 1996;274:1531-1534.

13 Hajduk, PJ et al. Discovery of potent nonpeptide inhibitors of stromelysin using SAR by NMR. J. Am. Chem. Soc. 1997;119:5818-5827.

14 Erlanson, DA, Fesik, SW, Hubbard, RE, Jahnke, W, Jhoti, H. Twenty years on: the impact of fragments on drug discovery. Nat Rev Drug Disc. 2016;15:605-619.

15 de Kloe, GE, Bailey, D, Leurs, R, de Esch, IJ. Transforming fragments into candidates: small becomes big in medicinal chemistry. Drug Discovery Today. 2009;14:630-646.

16 "In business as in science, it seems that you are often most successful in achieving something when you are trying to do something else. I think of it as the principle of obliquity." Sir James Black, FRS OM Nobel Laureate; 14 June 1924-21 March 2010. John Kay in Obliquity (Profile Books Ltd, London 2010).

17 Scannell, JW, Blanckley, A, Boldon, H, Warrington, B. Nature Reviews Drug Discovery. 2012;11:191-200.

18 Nadin, A, Hattotuwagama, C, Churcher, I. Lead-oriented synthesis: a new opportunity for synthetic chemistry. Angew Chem Int Ed Engl. 2012;51:1114-1122.

anatomical, physiological and molecular levels[23]. Recently, gene-editing techniques have been employed in combination with fluorescence-based reporter systems to create a new generation of genetic and chemical screens[24]. These techniques have demonstrated great potential in uncovering novel modulators of human disease pathologies.

The future

In 2006, a review of FDA-approved drugs showed that of 324 targeted proteins, 266 (82%) were human genome-derived proteins and of these 48% fell into three classes: GPCRs, nuclear receptors and ion channels and that 56% of drugs under investigation were for targets in these fields[25]. The remaining 52% of targets and 44% of drugs were split at levels of less than 5% over 120 domain families or singletons. A recent (2017) review showed that the number of recorded human-genome-derived targets had increased to 667 and 34% fall into the same three classes: GPCRs, nuclear receptors and ion channels and 67% of the 700 molecules in investigation are for these fields[26]. However, the remaining 66% of genome-derived targets now contains significant (>10%) targeting of cancer-related proteins (eg kinases, proteases, etc) but the number of compounds in study is low, presumably because they are promiscuous. In both cases, the targets not accounted as human genome-derived were for pathogenic targets.

Overall, these data present a picture of a pharmaceutical industry focused almost exclusively on human-genome-derived proteins as targets, particularly those dubbed on historical grounds to be 'privileged protein families' and modulated by structures described, based on long experience, as being somehow 'permissive'. As these targets are likely to be downstream of genetic causes this pushes discovery towards the treatment of symptoms rather than cures. In addition, the approach excludes non-protein human targets and mitochondrial DNA and its gene products. As the sole energy source of human cells, mitochondrial dysfunction could be, and indeed seems to be, the cause of many of today's sporadically occurring 'incurables' such as various dementias and metabolic diseases. To use high throughput screening of huge chemical libraries as the primary filter to find leads and drugs, confines drug discovery to those target classes privileged to have simple microplate-based fast assay procedures. It is no surprise permissive structures that modulate this privileged class achieve the classification of 'drug-like' and thus define library selection criteria.

It is unfortunate that the practices and influences of large Pharma for economic reasons defines the availability of hardware and software tools of medicinal chemistry for everybody, as this impedes the progress of academic and small biotech companies pursuing novel cures and treatments that cannot be launched by random screening and require a more rational iterative approach. Our own review brings us to the opinion that its preference for screening-based discovery may be self-harming. The performance of an earlier generation of researchers, typified by such as Sir James Black, who actively managed the course of discovery to its goal by use of in-loop feedback from a contemporaneous panel of assays, has proved to be no worse than the linear paradigm in terms of Pharma's drug output. In fact, based on its ability to support compound-lean exploration of large areas of chemical space to find novel drug chemotypes, some would say it was a better tool for lead discovery. The recent return of the iterative paradigm as the enabler for FBLD probably seals that view.

A strategic vision

'Knowledge is power' is a popular misquote of Francis Bacon, recognising that information-led approaches can achieve great results. In the modern age, life science is populated by a plethora of 'omics' with each spawning results into the so-called 'Big Data' pool. Given there is no apparent end to knowledge, its generation can become self-fulfilling with a consequential loss of ability to comprehend its meaning; indeed the great challenge in a modern world of Big Data is its conversion to useful and actionable data. In the context of drug discovery, it is therefore imperative to take advantage of modern instrumentation and data processing in a manner that provides a route to undertaking discovery whereby the information is pertinent, valuable and grows useful actionable knowledge.

A strategic vision is proposed based on a return to iterative discovery enabled by:

● Use of modern technology (such as microfluidics and image processing) enabling faster synthesis, route finding, analysis of complex bio-assays and decision making.

● Harnessing information at all points from multiple cycles of 'Make' and 'Test' to enable better decisions and more effective future iterations. In addition, create potential starting points for new discovery areas by complete and valuable 'Big Data' to generate useful leads and drugs (**Figure 7**).
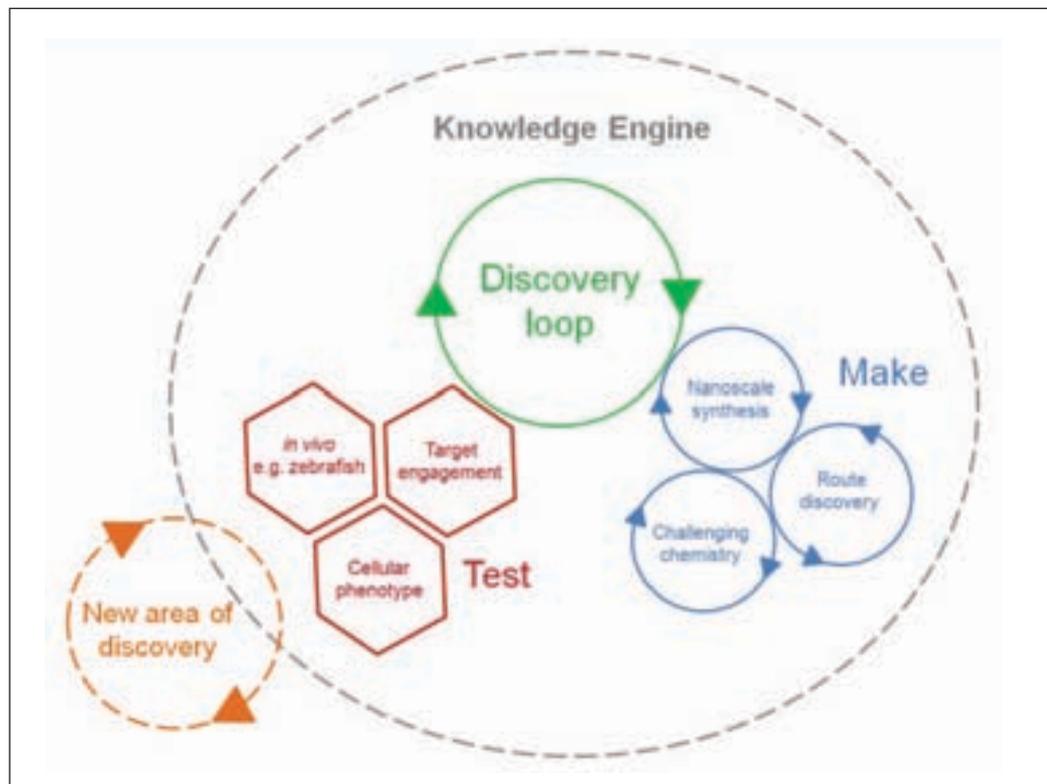
Figure 7: A strategic vision for the discovery process enabled by a complete and valuable Knowledge Engine to generate useful leads and drugs

Central to its successful implementation are five guiding principles:

1. **Iterative cycles of discovery.** Make and consider cycles of discovery.

2. **Information at every point.** Given the interdependence of each process there must be full integration of thinking between biology and chemistry (be this virtual or real). From a chemistry standpoint, know the failures and embrace knowledge of every reaction (eg, use knowledge of what also is produced by side reactions to enable improved route scouting for any future syntheses). In turn, apply complex assays and use all the information – positive, negative and unexpected. Chemical and biological space should not be sacrificed for automation.

3. **Pragmatism.** During implementation do not attempt to link every piece of equipment unless there is a real advantage. Engineer new solutions where the benefit of doing so is clear. Consider carefully the real requirement at each step (eg the need to purify a new molecule). If the synthesis is reproducible (for example, as typically enabled by flow regimes) then remake and purify only when required. Understand that chemistry is typically the rate determining step – use this as an opportunity to explore more complex biology. Moreover, moving from millions of compounds to typically less than 1,000 enables ready adoption of use of patient-derived cells and simple animal models at a statistically relevant number throughout the discovery process.

4. **Blur boundaries of thinking.** Continually challenge the 'known'. Integrate target, system and phenotype profiling of new molecules. Think beyond chemical structure to the pharmacophores; be willing to jump structures entirely (electrons not nuclei run chemistry).

5. **Embrace suitable tools.** For make anything goes – flow chemistry, library expansion, fragment screening. Chemistry is complex so do not over simplify to enable simpler instrumentation at the cost of chemical space. Selected HTS tools can help, for example, route finding in 1,536 microplates[27]. Screen new molecules early in complex organisms such as zebrafish as relevant, scalable *in vivo* systems to complement mammalian studies[28].

The authors are the first to admit the vision is not necessarily new, in fact, it is close to the drug discovery approaches of the past. However, it engages new data analysis, better screening tools and rapid

19 Ley, SV, Fitzpatrick, DE, Ingham, RJ, Myers, RM. Organic synthesis: march of the machines. Angew Chem Int Ed Engl. 2015;54:3449-3464.

20 Rodrigues T, Schneider P, Schneider G. Accessing New Chemical Entities through Microfluidic Systems, Angew. Chem. Int. Ed. 2014; 53:5750-5758.

21 Cheeseright, T, Mackey, M, Rose, S, Vinter, A. Molecular field extrema as descriptors of biological activity: Definition and validation. J. Chem. Inf. Model. 2006;46:665-676.

22 Kingwell, K. 3D cell technologies head to the R&D assembly line. Nature Reviews Drug Discovery. 2017;16: 6-7.

23 Lieschke, GJ, Currie, PD. Animal models of human disease: zebrafish swim into view. Nature reviews. Genetics. 2007;8:353-367.

24 Li, M, Zhao, L, Page-McCaw, PS, Chen, W. Zebrafish Genome Engineering Using the CRISPR-Cas9 System. Trends Genet. 2016;32:815-827.

25 Overington, JP, Al-Lazikani, B, Hopkins, AL. How many Drug Targets Are there? Nature Rev. Drug Discov. 2006;5:993-996.

26 Santos, R, Ursu, O, Gaulton, A, Bento, AP, Donadi, RS, Bologa, CG, Karlsson, A, Al-Lazikani, B, Hersey, A, Oprea, TI, Overington, JP. A comprehensive map of molecular drug targets. Nature Rev. Drug Discov. 2017; 16:19-34.

27 Santanilla, AB et al. Organic chemistry. Nanomole-scale high-throughput chemistry for the synthesis of complex molecules. Science. 2015;347:49-53.

28 MacRae, CA, Peterson, RT. Zebrafish as tools for drug discovery. Nature Reviews Drug Discovery. 2015; 14:721-731.

Drug Discovery

29 Agarwal, P, Sanseau, P, Cardon, LR. Novelty in the target landscape of the pharmaceutical industry. Nature Rev. Drug Discov. 2013;12:575-576.

30 DiMasi, JA, Faden, LB. Competitiveness in follow-on drug R&D: a race or imitation? Nature Rev.Drug Discov. 2011;10:23-27.

31 Davis, A, Warrington, B, Vinter, JG (1987). Strategies in Drug Design II – Modelling studies on phosphodiesterase substrates and inhibitors. J Comp-Aid Mol Design. 1987;1:97-120.

synthetic approaches (where pragmatic) to enable faster cycles of iteration and a holistic growth of pertinent knowledge. Our vision is that a Pharma-supported campaign to provide a 20-years' overdue upgrade of the hardware and software would provide a far superior engine for lead discovery efforts, not only by increasing drug output but also as a route to distinctive products with durable exclusivity reflecting corporate talent[29,30].   **DDW**

Dr Wayne Bowen is a Life Science Consultant at The Technology Partnership (TTP). His PhD on receptor biochemistry was received from the University of Glasgow. He subsequently worked in the Neuroscience Department at SmithKline Beecham developing novel therapeutics for the treatment of depression, Parkinson's Disease and stroke. In 1996, Dr Bowen co-founded Pharmagene plc and co-ordinated biochemical research on human tissue. Prior to joining TTP, he spent 13 years as Chief Scientific Officer at TTP Labtech.

Dr Giles Sanders is responsible for Business Development and Project Management in the Diagnostics and Life Sciences market sectors at The Technology Partnership. He has a degree in Chemistry and a DPhil in reactions at the solid-liquid interface under hydrodynamic conditions from Oxford University. Prior to joining TTP, Dr Sanders worked for a biotechnology start-up company where he was responsible for its microfluidics and microfabrication developments.

Dr Andreas Werdich is a Life Science Consultant at The Technology Partnership. He received his PhD in biological physics from Vanderbilt University in 2006 and completed postgraduate training in genetics, molecular biology and cardiac electro-physiology at Harvard Medical School and Case Western Reserve University. More recently, Dr Werdich was Assistant Professor in cardiovascular research at Brigham and Women's Hospital/ Harvard Medical School.

Dr Brian Warrington joined Smith Kline & French in 1965 as a medicinal chemist and has contributed to a large number of therapeutic targets while developing an increasing interest in the methodology of early drug discovery. He retired from his position as Vice President, Technology Development for GlaxoSmithKline Discovery Research in 2005 and now pursues the same interests as an independent consultant. Dr Warrington holds a PhD from

the University of London, is a Fellow of the Royal Society of Chemistry, a Chartered Chemist and was awarded the Royal Society of Chemistry Millennium Medal for leadership in the area of microfluidic synthesis and screening.

Dr Elizabeth Farrant is Chief Executive Officer at New Path Molecular Research. She has a degree in chemistry from Salford University and a PhD in natural product synthesis from Reading University. Dr Farrant has pioneered the application of microfluidics to chemistry, protein crystallisation, separations science and cell based assays during employment at GlaxoSmithKline, Pfizer and Cyclofluidic.