08:36:02.
Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032

CHAPTER 2

Parallel Synthesis and Library Design

ANDY MERRITT

MRCT Centre for Therapeutics Discovery, 1-3 Burtonhole Lane, Mill Hill, London NW7 1AD, UK
E-mail: Andy.Merritt@tech.mrc.ac.uk

2.1 INTRODUCTION

Why do companies build and maintain large chemical libraries? Often described as the most important asset of a pharmaceutical company's research arm (the 'crown jewels'[1]) they are typically the product of a large number of man years for internal synthesis plus a significant (multimillion dollar) spend on external compound acquisition from an ever widening range of commercial sources. Significant overhead is spent annually on selecting, acquiring, synthesising, maintaining and analysing compounds and the investment in facilities to curate, protect and distribute collections,[2] with direct equipment costs estimated to fall in the $1–2 million range.[3]

The answer to the initial question in the previous paragraph is clear—to increase the chance of finding something novel. For any company progressing lead discovery and optimisation programmes, if enough is known about a particular target and the type of molecules capable of interacting with it (in a pharmacologically relevant manner) then as long as there is novelty inherent in that knowledge, a curated and diverse collection of compounds is not required. All that is demanded of a compound management process in that situation is the shepherding of new compounds through any required assays to support project progression. However for other targets, often early stage and novel, but also fast follower targets where an organisation is trying to catch and overtake the known state of discovery, a compound collection becomes an invaluable source to potentially find something novel (usually a small molecule start point, but also target validation tools) that can be used to initiate a medicinal chemistry discovery programme. Where little or nothing is known about the requirements of the target active site in terms of preferred interacting molecules then that search may most likely be based on complete sampling of as much variety of chemical space as possible. If there is predetermined knowledge of the target (specific protein structure or knowledge of closely related proteins) then it may be possible to sample the compound collection to produce a set of compounds with a predetermined bias towards that target. However, whether the target structure is known or not, the main

The Handbook of Medicinal Chemistry: Principles and Practice
Edited by Andrew Davis and Simon E Ward
© The Royal Society of Chemistry 2015
Published by the Royal Society of Chemistry, www.rsc.org

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

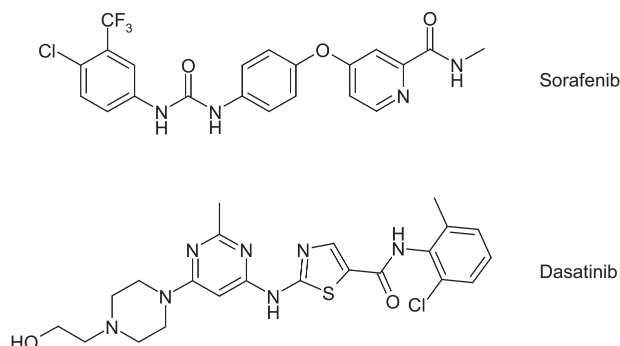


Figure 2.1 Examples of marketed drugs where high throughput screening and optimisation were part of the discovery process.

element of the screening exercise is the same—to find something new by accessing and sampling the best possible selection of hit, lead or drug like compounds available. Of course such an approach should not be considered in isolation for any new target, and should be carefully considered alongside other more rational design and established empirical medicinal chemistry approaches that are highlighted in many sections of this book. But it is true to say that there are now many examples where well designed arrays associated with high throughput screening methods have led to compounds that are now into the clinic.[4] For example, Sorafenib (Bayer/Onyx)[5] and Dasatinib (BMS)[6] are both tyrosine kinase inhibitors discovered in part from initial high throughput screening (in 1994 and 1997, respectively) followed by targeted array synthesis (Figure 2.1).

So what could be defined as a well-designed library? Like the old story of economists, put five chemists into a room and ask that question and the likelihood is that you'll get at least six separate answers. Indeed, studies have shown in some cases just asking the same chemist twice on separate occasions can provide differing results. In a 2004 study by Pharmacia[7] when chemists were asked to select/reject compounds from a set of 2000, the average pairwise agreement within the 13 chemists included was only 28%. Moreover nine chemists were subsequently given the same set of compounds to repeat the rejection process, with a result of only 51% consistency. But putting personal subjectivity aside there are some underlying principles that can be applied and are independent of any one favourite algorithm for selecting A over B or grouping X with Y instead of Z. It was during the initial development of combinatorial chemistry approaches to collection design in the 1990s, and the associated explosion in chemical technologies (design, synthesis, purification and analysis) that many key concepts of successful application of library design to drug discovery were learnt (or in many cases relearnt after having been forgotten!). The next sections of this chapter will briefly review the historical development of key approaches to library synthesis and construction to illustrate how we have got to our current stages of compound library and screening collection design. The remainder of the chapter will then focus on design strategies for general compound libraries and larger targeted arrays aimed towards specific protein classes. The development of the concepts and specific technologies of combinatorial chemistry and the application of combinatorial approaches to specific target prosecution (sometimes through equally large libraries) is beyond the scope of this chapter and has been covered in detail elsewhere.[8]

2.2 THE START OF COMBICHEM IN DRUG DISCOVERY

The development of miniaturised screening leading to high throughput approaches was a significant advancement of drug discovery.[9] The standardisation of assay format into microtitre

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

plates, initially 96 well format, alongside the development of automated processing, radically changed the opportunity for screening to deliver new leads into drug discovery programmes. Automation of plate movement, liquid handling and plate reading processes meant that where a few 10s of compounds may have been tested in a day by manual techniques, suddenly 1000s were possible in enzyme, (membrane bound) receptor and even whole cell assay format. Further enhanced by the miniaturisation of wells on the plates, from 96 to 384 (and subsequently 1536), high throughput screening of compound collections of 100 000s or more became clearly feasible, and when run alongside mechanism and knowledge/structural based targeted screening approaches provided much greater opportunity to identify novel lead series and structural classes.

As high throughput screening developed rapidly in the late 1980s and early 1990s attention was turned to the feedstock for such efforts—company compound collections. These had typically built up by a combination of 'file' compounds from previous and ongoing lead optimisation programmes and natural products, sourced either from in house fermentation or through external acquisition of samples, be they soil, microbe or plant derived. A compound collection of one to two hundred thousand such compounds was not atypical, but the potential for further growth through these traditional routes would always be limited. A 'traditional' medicinal chemist was likely to add no more than 40–50 compounds in any year, and perhaps even more significantly any file collection built on past programmes would clearly only represent those chemical areas that had been of interest. Many collections were significantly populated by specific structural classes, for example β-lactams or steroids. Meanwhile natural products were often complex structures, difficult to work with in lead optimisation, and becoming harder to source with exclusivity. International treaties correctly limited the ability to source natural products from countries without due regard to intellectual property ownership[10] and even when novel active natural products were identified, it was possible for more than one company to independently and concurrently identify the same structural series.[11,12]

So if high throughput screening presented the opportunity to screen 100 000s of compounds in a matter of days whilst collection sizes were still limited, alternative mechanisms to grow the collections were targeted. Collection sharing deals were struck between companies[13] and this concept was effectively continued in the mergers of the 1990s[14] where the formation of combinations such as GlaxoWellcome, Smithkline Beecham, AstraZeneca, Novartis, and Aventis, for example, provided immediate increases in corporate collection size. A recent analysis[15] of the combination of compound libraries from Bayer Healthcare AG and Schering AG following the takeover of the latter by Bayer showed a very low direct overlap of chemical structures between the two organisations (0.04% for in house synthesised and 1.5% in total) and reached the conclusion that collaborative screening efforts between companies (either through consortia or the result of more commercial takeovers) would be an effective means of increasing diversity coverage of screening libraries.

In addition, acquisition of compounds from external sources was increased, both from commercial and academic sources. Commercial suppliers provided compounds that could be added to screening collections, though these were available to all companies, thus raising concern over intellectual property control, and at that time were limited to only a few suppliers of fine chemicals. Access to more varied chemistry was available through academic collaborations, and many academic groups found they could fund several aspects of their research with money from compound selling, however a combination of structural integrity, purity, and sustainability of resupply were all potential issues for the pharmaceutical companies using this approach.

The optimum solution for companies appeared to be a combination of the above, but enhanced with an even greater component derived from a significant increase of productivity from their own chemists. Such internally derived compounds would be proprietary, exclusive and

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

could be targeted if necessary to areas of most interest to the company concerned. Knowledge would be retained for further synthesis, follow up and analogue work thus providing confidence downstream of any initial positive results. The rapid development of high throughput screening had demonstrated that technology and rethinking of strategies could in combination provide major increases in productivity, and drug companies began to consider whether this could be also true for chemistry.

Fortunately such ideas and approaches had already been developed, though not in the field of synthetic organic chemistry but in peptide chemistry. The technology and methodology of solid phase chemistry had been developed by Merrifield[16] in the 1960s and subsequent automation of the approach, maximising the advantages of forcing conditions (through excess reagent) and purification (through filtering), was well developed by this time.[17] Indeed, some solid phase work with non-peptide structures had been developed by the 1970s[18] though had not achieved widespread use in mainstream synthetic chemistry.

The ability to carry out peptide chemistry on support in parallel was demonstrated by Geysen[19] with the development of polystyrene coated pins. Using this methodology synthesis could be carried out in spatially addressed arrays so that common steps (deprotection and activation steps for example) could be performed using bulk reagents and reaction vessels. At around the same time Furka[20] was developing the approach of split and mix using resin beads to allow synthesis of large numbers of peptides (albeit as mixtures) in very few reactions (Figure 2.2). Houghten[21] introduced the compartmentalisation of resin beads as "teabags," thus allowing a more efficient and scaled up handling of the process, and introducing the idea that packaged resin could then be traced through the synthetic sequence thus allowing identification of the resulting compound (or compound mixture depending on the approach adopted).

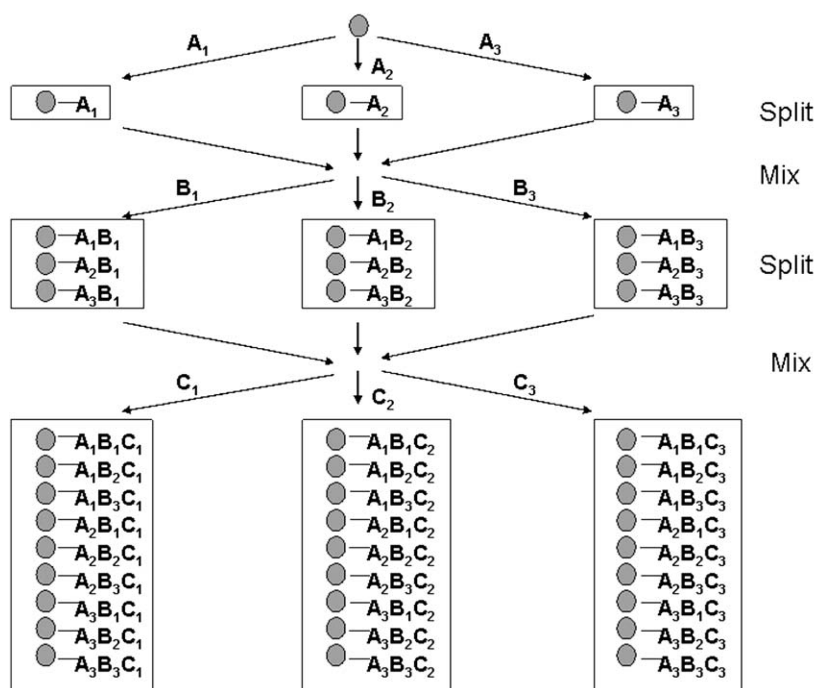
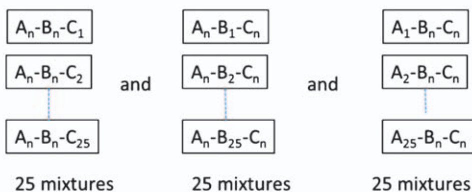
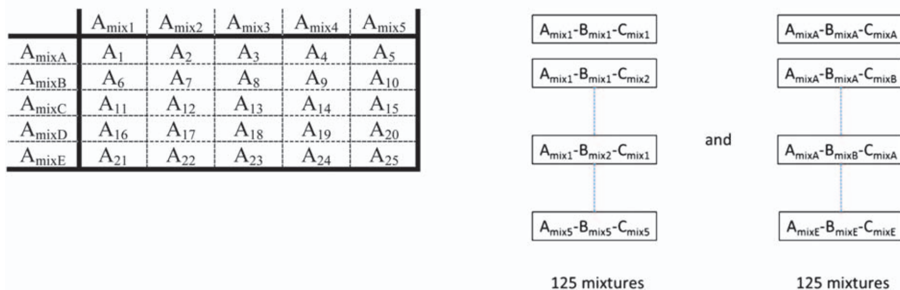


Figure 2.2 Polymer supported strategy of split-mix synthesis in the production of screening compound libraries.

View Online

These initial developments focused on manufacturing large numbers of small peptide fragments, used for example to evaluate protein–protein interactions (epitope mapping)[22] or enzyme[23,24] and antibody[25] specificities. The mixtures produced using the split mix approach needed to be deconvoluted to single active compounds, and a number of methods were developed, including iterative deconvolution[26] (fixed positions in mixtures and subsequent sub-library synthesis), positional scanning[27] (replicated synthesis of same library but with a different fixed position in each mixture) and orthogonal pooling strategies[28] (replicated synthesis with orthogonal chemistries allowing different pooling strategies) (Figure 2.3).





	A_{mix1}	A_{mix2}	A_{mix3}	A_{mix4}	A_{mix5}
A_{mixA}	A_1	A_2	A_3	A_4	A_5
A_{mixB}	A_6	A_7	A_8	A_9	A_{10}
A_{mixC}	A_{11}	A_{12}	A_{13}	A_{14}	A_{15}
A_{mixD}	A_{16}	A_{17}	A_{18}	A_{19}	A_{20}
A_{mixE}	A_{21}	A_{22}	A_{23}	A_{24}	A_{25}

Figure 2.3 Deconvolution strategies to identify single compound hits from pooled samples out of split-mix libraries.

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

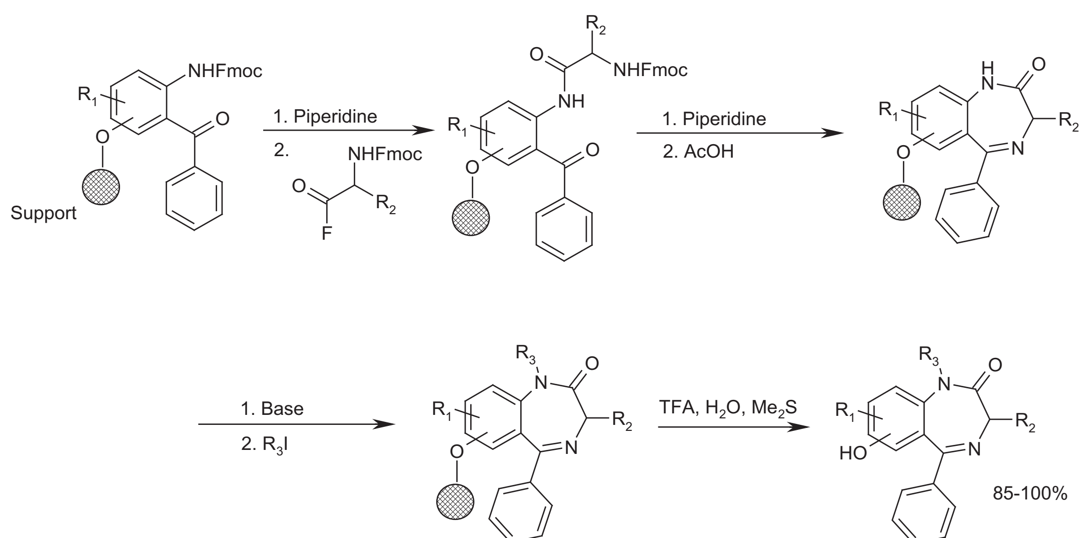
For readers interested in the statistical effectiveness of such strategies and the subsequent development of these approaches for HTS, especially with implementation into post synthesis pooling, the review by Kainkaryam and Woolf is worth reading.[29] The authors explore the concepts of adaptive *versus* non adaptive pooling strategies and provide several further examples of the effectiveness of the approaches.

2.3 FROM PEPTIDES TO SMALL MOLECULES

The ability of combinatorial chemistry to make large numbers of peptides, combined with various new screening approaches did not escape the attention of those involved in early hit identification programmes. Although peptides were not suitable compounds for lead identification, analysis of drug discovery literature confirmed what many practitioners were aware of, that the large majority of drug discovery programmes involved amide bond formation or related reactions (including heterocycle formation through subsequent dehydration). As such, many of the drug discovery compounds should be accessible using similar chemistries to those of peptide synthesis.

The first 'small molecule' combinatorial library was published by Ellman,[30] who demonstrated that a library of 40 benzodiazepines could be produced using solid phase approaches, with three points of diversity, or variation, on the core structure (Scheme 2.1). Ellman expanded this work, using the pin method of Geysen to give 192 compounds,[31] and further expanded this to several thousand compounds in later publications.[32] De Witt described the preparation of array compounds on solid phase using the 'Diversomer' approach,[33] coupled with simple automation that was the first of many automated synthetic approaches to be introduced. That De Witt was based in industry was significant—the approach of combinatorial chemistry was clearly applicable to issues of drug discovery where obtaining data to make the next structural series decisions was the driving component of the research rather than the development of the core discipline.

Over the following few years the two main strategies of split and mix (to generate large libraries using solid phase approaches) and parallel synthesis (focused on smaller libraries) were



Scheme 2.1 The first published example of a small molecule array synthesised on solid phase—Ellman's benzodiazepine synthesis.[30]

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

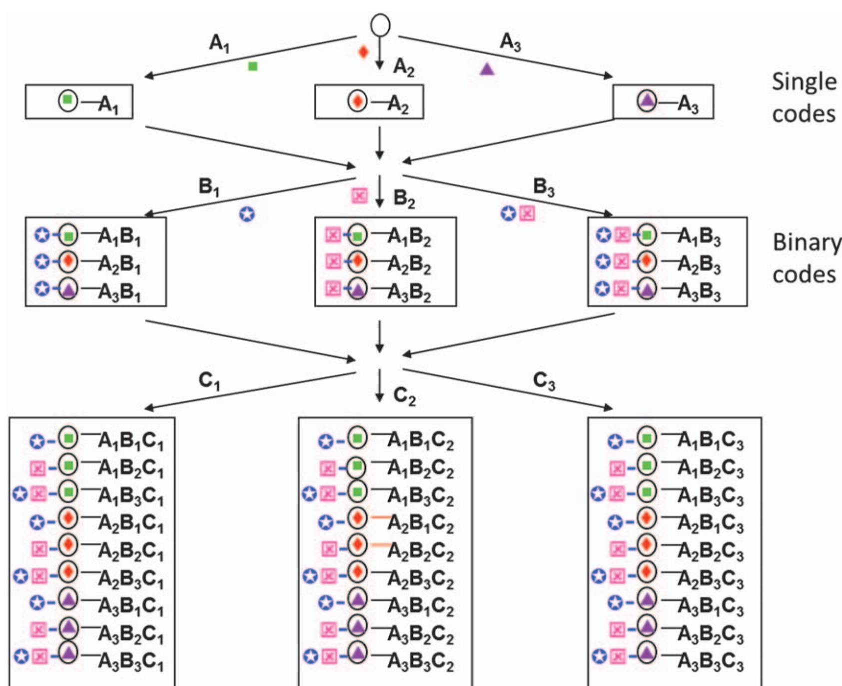


Figure 2.4 Introduction of encoding tags and strategies to split-mix synthesis.

refined and developed. The main focus for lead discovery split and mix approaches was on means of identifying compounds without the need for resynthesis or deconvolution stages, which typically took too long for fast moving lead discovery projects to allow simple mixture libraries to have an impact.[34] Tagging approaches were developed, where the solid phase was orthogonally reacted with molecules that could be 'read', typically using mass spectrometric approaches (Figure 2.4).[35] At the same time the ''teabag'' concept of Houghten was further developed, both with advancements of the container system, but more importantly with the inclusion of inert radiofrequency tags.[36] These then allowed the synthetic history of any container to be either tracked or directed, thus combining the potential of split and mix with both the potential scale and single product outcome of parallel methods.

At the same time there were also rapid developments in both the range of chemistry applicable to solid phase and in alternative approaches looking to maximise the advantages of solid phase techniques whilst keeping those of solution phase. The range of chemistries on solid phase became almost as broad as traditional solution chemistry,[37–39] though in the context of this review it is worth noting (perhaps discouragingly) that a recent review[40] of the current 'medicinal chemistry toolbox' showed a similar prevalence of amide chemistry in drug programmes. Considering just 'constructive' reactions (excluding protection/deprotection and oxidation/reduction processes) then 24.4% of reactions were simple acylations, whilst an additional 11.3% were N-heterocycle formations primarily through dehydration of intermediate acylated amines. N/O alkylation (including reductive amination methods) accounted for an additional 26% of reactions.

Attempts to get the solid phase 'in solution' included soluble polymers (*e.g.* polyethylene glycol monomethyl ethers,[41] non-cross-linked polystyrenes[42]) that could be precipitated for purification purposes, and the combination of fluorocarbon fluids and perfluorinated substrates[43] to allow separation from both aqueous and organic solution when required. The most applicable development to address the combination of solid and solution phase approaches was

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

in supported reagents, either as scavengers to remove excess reagents or unreacted substrates[44] or as removable reagents to catalyse specific reaction steps.[45] These approaches have achieved widespread use in mainstream synthetic chemistry as well as in the combinatorial research area, and have been extensively reviewed elsewhere.[46,47] Further examples specifically associated with array library design will be discussed later in the chapter in the section on realising a collection.

2.4 MY LIBRARY'S BIGGER THAN YOUR LIBRARY—THE 'UNIVERSAL' LIBRARY

Before considering current best practices and use of high throughput and parallel chemistry in drug discovery and lead optimisation it is important to understand how the initial promise of combinatorial chemistry failed to deliver, and the subsequent backlash against large combinatorial approaches that heralded the start of the 21st century. As has been described above, high throughput screening had rapidly developed as a key component of drug discovery, to be utilised where possible alongside other lead seeking strategies to maximise the chances of new serendipitous results. The need for 'feedstock' for the screening regime was compelling a push to maximise the scale of compound collections. New elements of diversity driven design were exploring a whole range of new ideas on compound structures.[48–53] In this light the power of combinatorial chemistry to generate potentially millions of compounds could not be overlooked. Pharmaceutical companies rapidly followed each other in building in-house combinatorial groups, whilst external new companies were developed to focus on the technology of delivering large numbers of compounds. Many of these were subsequently acquired by pharmaceutical companies, often accompanied with the expressed intent to allow these new technology companies to continue to operate independently of the mainstream world of drug discovery.

Thus by the mid to late 1990s there were many groups using combinatorial chemistry to generate large numbers of compounds, either within pharmaceutical companies or standalone companies operating fee for service provision of libraries. The range of chemistry and structural motifs expanded, and groups were able to make libraries of hundreds of thousands of compounds with a wide variety of structures, extremely rich in functionality.

The pinnacle of such approaches were the 'Universal Libraries', a concept that developed under a range of titles in many groups.[54,55] The hypothesis was a simple and powerful one. By using a set of core templates with several differentially protected functionalities and decorating these in a comprehensive combinatorial fashion with sets of compounds rich in potentially pharmacologically relevant functional groups displayed in directionally controlled manners, it should be possible to devise a single library that would cover all of 'pharmacological space' as relevant to target proteins in drug discovery. Some groups suggested this could be achieved with only a small number of cores series, whilst others argued that greater central variety would be needed. However all had one thing in common—the technology of synthesis, the concepts of spatial design of the molecules and the power of combinatorial numbers had driven the development rather than any real consideration of the nature of the resultant structures, which had to be viable structures for drug discovery optimisation programmes. Indeed at that time the belief was expressed by some that the need for optimisation itself would be mostly eliminated—after all, from such a large and comprehensive library surely the drug itself would be present in the first screening.

2.5 FROM COMBICHEM TO HIGH THROUGHPUT CHEMISTRY—REMEMBERING IT'S ALL ABOUT DRUGS

"The pharmaceutical industry has benefited...from rapid access to a large number of novel compounds and related biological data though combinatorial chemistry and high throughput screening. However this plethora of data has yet to translate into clinical success."

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

The above extract from Oprea's review[56] of the impact of combinatorial chemistry is just one of many that could be used at this point. Clearly the generation of millions of compounds, not to mention the investment of significant resources into developing technologies, strategies and expertise had not reaped the hyped dividends so readily promised in the early days of combinatorial chemistry. So where did it go wrong?

One of the most fundamental issues was a misconception around the scale of synthetic compound numbers as they related to all of potential chemical (or biological chemistry) space. Traditional medicinal chemistry and drug discovery had been a discipline where, once biological data had pointed the direction, the next compound for test used to take a week to prepare, and a medicinal chemist was seen as prolific if they added 100 test compounds over the lifetime of a particular project. The promise of 100 000 or more compounds from a small team and a few weeks' effort was therefore clearly a step change. Multiply that by concerted planning and the promise of hits every time from a library of maybe 1–2 million compounds appeared to be a reasonable supposition. In short, the naive view was that this step up in compound productivity was bound to yield success in screening campaigns and optimisation work. However, as computational chemists had been pointing out all along, the reality of druggable chemical space was in a completely different dimension. Final numbers vary between advocates of different techniques, but certainly the number of potential compounds to fill that space can be measured in numbers vastly greater than could ever be made (indeed greater than the number of atoms in the universe).[57,58] In a conceptual world of perhaps 10^{70} potential drug molecules then 10^6 is never going to deliver every time!

Even if the design of a library meant the potential blockbuster drug compound was intended to be in the library, the possibility of it actually being present was limited by the quality of the chemistry of the early libraries, and moreover the means of assessing whether it was in there did not exist. Although analytical (and purification) tools and capabilities have become much more powerful (*vide infra*) in the early days it was only possible to assess quality through extensive validation of the chemistry on sample sets and then build confidence by sampling a subset of final compounds, though even this step was not viable if split mix approaches yielding mixtures of compounds were being pursued. Solid phase methods especially were prone to producing varied yields in parallel steps, and the final cleavage of compounds often could generate unexpected and indefinable products due to the often forcing nature of cleavage conditions.[59]

The combinatorial chemists of the 1990s set themselves up as the new force in drug discovery. Although other areas of chemistry saw and utilised the potential of combinatorial approaches[60] it was in drug discovery that the practitioners viewed their way as revolutionary, leading as it would to a complete change in approaches to lead identification. As such, those who got involved in the field were often excellent scientists who were driven by the development of technology and the strategies of maximising the value of those technologies. Attempts to spread combinatorial approaches into mainstream drug discovery were at best of limited impact.[61] The belief that they were developing a whole new, and more effective, science for drug discovery is well illustrated by the publication challenges and how they were overcome. As the early practitioners of combinatorial chemistry looked to publish work they found the mainstream journals reluctant to accept manuscripts, demanding as they did levels of quality assurance and data than were not only not being gathered but due to the nature of the techniques of the day were not even feasible. Rather than work within the established literature constraints to refine how combinatorial chemistry could be adapted the result was the establishment of new journals dedicated to the science of Combichem.[62]

The separation of combinatorial technology approaches from mainstream drug discovery had a most significant impact on the design of libraries. Driven as it was by the desire to produce large numbers and to make maximum use of the associated technologies, it was almost inevitable that the libraries produced would have large, highly functionalised structures.[63]

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

In addition, the production of large numbers of compounds around similar core structures created an illusion of diversity but in reality exacerbated the issue identified so much earlier within compound collections being dominated by common core motifs.

The rehabilitation of combinatorial chemistry (as high throughput chemistry) was enabled by a number of analyses of problems associated with the earlier approaches (leading to several strategies such as described below), alongside the more widespread development of understanding of factors critical in limiting attrition in potential drugs across all aspects of drug discovery. Three particular strategies are worth noting here as they have had major impact on the design of combinatorial approaches; the physicochemical properties of drug structures and their ability to cross biological membranes; the size of lead molecules and subsequent optimisation impact; and the incorporation of experience and knowledge into targeted library approaches.

The first of these is the seminal publication of Lipinski,[64] outlining the 'rule of 5' as criteria to determine to likelihood that a particular compound will pass through biological membranes, and therefore have potential to act as a drug substance. Early library structures typically had a profile of properties with mean molecular weight well above the Lipinski limits of 500, and high functionality counts (especially amide bonds) that inevitably leads to too high a level of both H-bond acceptors and donors.[65] Therefore screening such libraries in any lead discovery phase, or using such design templates in lead pursuit and optimisation is fraught with developability issues and, not surprisingly, initial results from such libraries did not become successful development candidates. As all the Lipinski parameters can be calculated from compound structures it was simple to incorporate such factors into any design approach, for example using weighted penalties in a design strategy or just setting hard limits on molecular weight and other properties.

Extending the physicochemical property limitation further, Teague and colleagues from AstraZeneca published an analysis that showed that for lead compounds these parameters needed to be even stricter,[66] as lead optimisation consistently added both molecular weight and lipophilicity to any series as it progressed towards development candidate status. On a similar note, Hann[67] demonstrated that success rate of lead discovery was inversely related to the complexity of the screening structures, and that for more complex designs the likelihood of finding a successful hit against a target were very low.

Finally, the application of knowledge of past success has been brought into the design of libraries, most effectively for large targeted libraries for protein family screening. One example of this is the work of Lewell and Judd[68] where the knowledge of known active compounds against classes of related 7-TM structures was used to design library building block sets incorporating 'privileged' substructures. Computational algorithms looked for common feature motifs across a range of active structures, using chemically intelligent fragmentation approaches to identify real substructures that could be introduced into new designs.

2.6 REALISING A COLLECTION—TECHNOLOGY DEVELOPMENT AND COMMERCIAL OFFERINGS

Alongside the development of strategies of design and selection the development of combinatorial chemistry and subsequent movement to high throughput chemistry approaches has driven a number of technological advances. Many of these have been 'of the moment'; for example a number of high level automation approaches were extremely effective in producing large numbers of compounds but now exist only in archives of scientific equipment. Others however have become commonplace approaches, as have many of the developments in parallel analysis and purification, initially driven by the challenge of large number synthetic approaches.

View Online

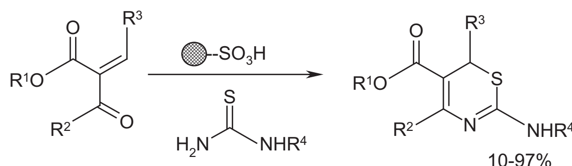
Synthetic automation is perhaps the most notable example of such short lifetime technologies. As in other sections of this review, fully comprehensive reviews of the wide range of synthetic automation equipment are available elsewhere,[69] and only illustrative examples are used here. For example, three synthetic automated technologies were in use within GlaxoWellcome in the late 1990s, all of which are now 'retired' (and indeed examples of all have been donated to the Science Museum in London). Initial solid phase work was driven by 'Advanced Chemtech' ACT machines.[70] Based around liquid handling robotics, and using proprietary designed reaction blocks there were a number of designs supporting solid phase chemistry. At the same time split mix approaches incorporated through the acquisition of Affymax by GlaxoWellcome were carried out on Encoded Synthetic Library (ESL) synthesisers,[71] automation based around adaption of peptide synthesisers with the ability to mix and redistribute resin to reaction vessels. Finally an arm of solution phase based work was supported by the development of synthetic robotics on a Tecan liquid handling bed with adaption for solvent removal through gas enhanced evaporation.[72] Between these three technologies millions of compounds were synthesised during the late 1990s, however all were to be subsequently overtaken by the development of RF-encoded encapsulated resin in the IRORI system.[73] Using automated directed sorting with capacity for up to 10 000 vessels this became the workhorse of large number synthesis, but was itself superseded by IRORI development of the X-Kan,[74] with 2D bar-coding replacing the RF tag approach. In the period of only 10 years, within just one company therefore we have seen the introduction and subsequent displacement of over four separate automated synthesisers, and in reality several more systems (*e.g.* Myriad,[75] Zinnser Sophas,[76] Argonaut Trident and Quest systems[77]) were also in use during the same period, again most of which are now no longer in use.

The type of automated synthetic equipment outlined above has typically remained as tools of the dedicated diversity chemist, with the development of expertise around synthetic automation technology, and several groups continue to develop extensions to these approaches.[78] Of much greater impact and lasting effect was the development of simpler parallel reaction equipment, much of which was developed in pharmaceutical laboratories and subsequently commercialised through equipment manufacturer partnerships.[72] Many examples are available and in use today, but examples include parallel tube based reaction blocks introduced by companies such as STEM,[79] allowing controlled stirring and heating of arrays of solution based reactions at significant scale, whilst Radleys introduced equipment based on commercialising the common practices of having several reactions on a single stirrer hotplate.[80] The carousel took advantage of the magnetic field created by a stirrer, whilst the greenhouse allowed reactions to be carried out readily under inert conditions. For solid phase chemistry a number of block based clamped filter based systems were introduced, including Bohdan Miniblocks,[81] which took advantage of a layout format identical to microtitre plates, thus facilitating subsequent transfer to assay plates.

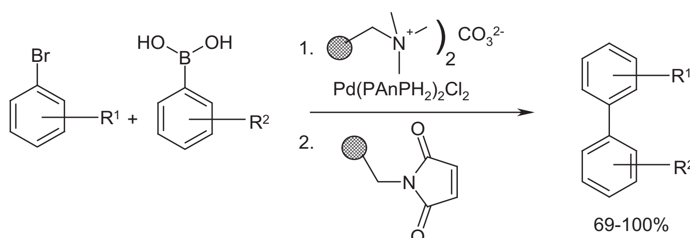
As discussed earlier, the development of polymer supported reagents and sequestration agents has made solution phase approaches to parallel chemistry viable, allowing filtration and work up approaches to be used in parallel using filtration reagent blocks. This area has recently been reviewed[82] and includes resin capture and release approaches, tagged reagents and substrates. The following examples illustrate how these approaches have been applied in library syntheses. Strohmeier and Kappe[83] used resin capture and release steps in the preparation of 1,3 thiazine libraries (Scheme 2.2). Parlow[84] reports the use of 2 different tagged reagents to support purification by removal of reagent by-products in Suzuki coupling reactions (Scheme 2.3). Wang[85] describes the use of polymer supported phosphines in the wide ranging syntheses of triazolopyridines (Scheme 2.4). Perhaps the ultimate demonstration of the power and flexibility of polymer supported reagents and reactions is in the synthetic work of the Ley group, who has produced several publications of total syntheses of natural products (Scheme 2.5)[86] as well as a number of approaches to library and array syntheses.[87]

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

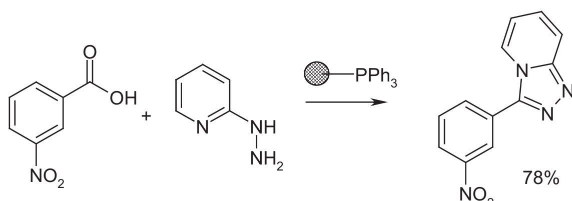
View Online



Scheme 2.2 Using polymer bound reagents to capture an intermediate with subsequent release into solution on further reaction.



Scheme 2.3 Using polymer bound reagents to scavenge excess reagents and reagent biproducts.



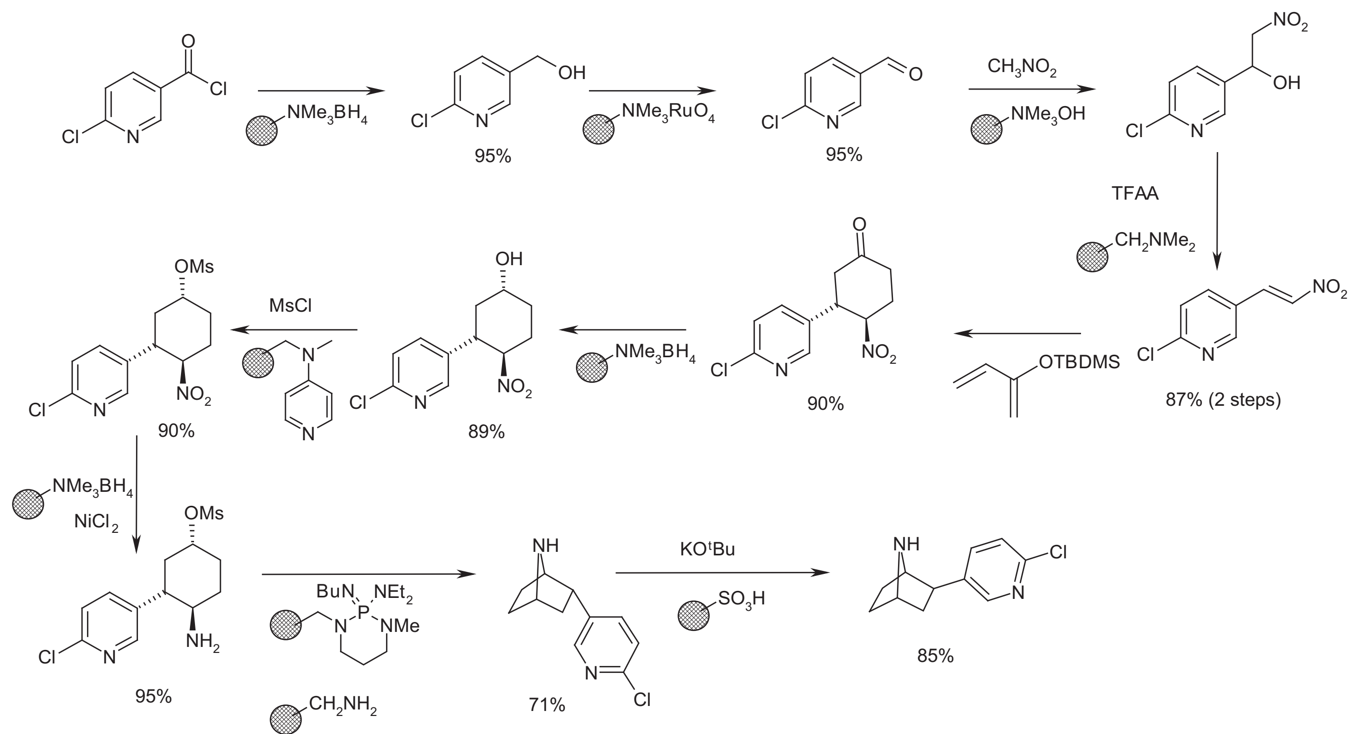
Scheme 2.4 Example of supported reagents in solution phase synthesis.

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

One now commonplace technique that developed alongside the high throughput chemistry techniques has been the use of microwaves to heat and accelerate reactions.[88] Although it was initially thought that microwaves could have a specific effect on reaction trajectories and rates, it is now generally agreed that the primary impact is the same as thermal acceleration, albeit a much faster and energy efficient one.[89] There are specific exceptions where homogeneous reactions may be affected by localised heating of solid catalysts[90] and recent designs of reaction vessels incorporate microwave absorbing materials to maximise the effectiveness of microwave heating.[91] However generally microwave technology has the main advantage of rapid heating, combined with being linked to automatic processing equipment that allows array chemistry to use this approach as a very specific tool for rapid compound synthesis. For example, a recent synthesis of dihydropyrimidone libraries using stepwise multi component Biginelli chemistry and Pd/Cu mediated cross coupling reactions, both accelerated and in high yield, illustrates some of the range and impact of microwave assisted synthesis (Scheme 2.6).[92]

Alongside parallel synthesis developments, the ability to analyse and purify large number of compounds has also developed extensively. The use of scavenger reagents and supported sequestration approaches, alongside catch and release methodologies certainly improved the purity and quality of combinatorial chemistry reactions. However it has been the development of fast, automated LC-MS analysis systems[93] and the more recent development of fast, parallel, mass directed preparative LC[94] that has allowed the approach of purifying all synthesised compounds to take over from previous triage processes,[95] whereby moderate to good purity compounds were typically progressed into screening without additional purification, and only

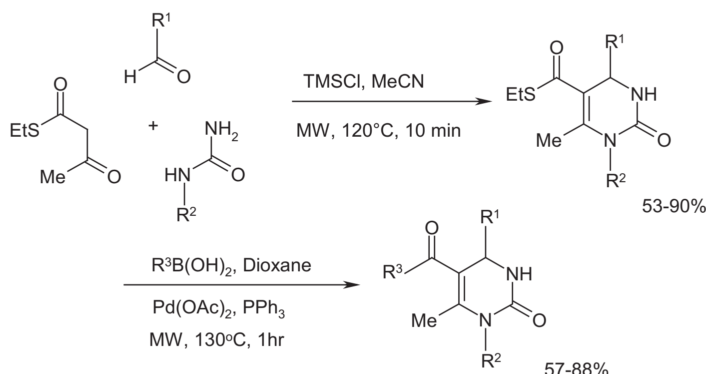
Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621
08:36:02.



Scheme 2.5 Multi step total synthesis of (±)epibatidine using supported reagents and/or scavengers at every step, and with no additional purification steps.

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.



Scheme 2.6 Microwave assisted synthesis of a library of dihydropyrimidones.

the less successful reactions were purified. The ability to estimate concentration using LC methods[96,97] has added a further level of quality into library compound in screening, as assay level concentrations can now also be determined with greater confidence rather than assuming only a single concentration across an entire collection.

The development of equipment and technologies to deliver novel chemical libraries within pharmaceutical companies as illustrated above was driven by the intent to create novelty within a company's collections. However, the lack of a wide variety of chemistry applicable to such approaches, combined with a limited internal resource meant that this could only be partially successful. As indicated above, some companies addressed this further through the acquisition of small synthesis companies, whilst others commissioned the external creation of companies to focus more closely on pure synthesis of collection compounds. However, the opportunity for alternative external support of collection development was not missed by others, and companies offering compounds for sale, often acquired through collaborative approaches with academic institutions, became more prominent. Initial issues of compound quality and of access to material for follow up studies plagued this process, but as more companies became involved and began to offer higher quality assured products the option to build collections though purchase of compounds became more prominent. In particular the opening up of the former Soviet states in Eastern Europe and Russia allowed the rapid development of a number of companies offering compounds, initially brokering academic sourced material but rapidly moving towards commissioned and designed libraries. Today there exists a highly competitive supply market built on a long tradition of good organic chemistry in these regions that offer screening sets targeted to particular proteins, general screening sets with the ability to cherry pick bespoke selections, custom synthesis of novel structures around array formats and full contract research services.[98] This market has developed in response to the needs and quality demands of the customers, typically the large pharma organisations, though many academic grant applications have been built around a component of library purchase for novel target investigation in universities. Many of the companies are now correctly regarded as design leaders in their own right, developing algorithms and approaches to defining novel chemistry and chemical space. As an illustration of the most well-known and valued companies, Table 2.1 shows a breakdown in percentage coverage terms of the top eight suppliers that MRCT used in an exercise in 2011 to build a representative diversity set suitable for screening in academic laboratories with modest HTS capacity.

In addition, the suppliers have continued to take note of the developing understanding of undesirable structural types and properties. Knowledge of many of these has only developed through screening and (failed) follow up across a number of targets, and publications such as

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

Table 2.1 Top eight suppliers of compounds used in a recent 10K library construction at MRCT.

Supplier	Compounds	% of library
Enamine Ltd.	3117	31
ChemDiv Inc.	1530	15
ChemBridge Corp.	1459	15
Vitas-M Laboratory Ltd.	1026	10
Maybridge (Thermo Fisher Scientific)	960	10
Life Technologies Corp	619	6
Asinex	606	6
InterBioScreen Ltd.	382	4

the PAINS paper from Baell and Holloway[99] are now used to help define desirable chemical space by suppliers. However, the onus still rests with the purchaser (*caveat emptor*) to ensure that any process of compound acquisition has robust mechanisms of analysis and filtering that are maintained to current knowledge to maximise the likelihood of quality hits from any screen, and such approaches will be outlined in the following sections.

At the time of writing the total commercial offering of screening compounds is in the excess of 20 million compounds and the efficiency of compound delivery, cherry pick selection and variable quantity supply at modest cost per unit item means that purchase of compounds is the preferred route for rapid collection development for the majority of parties interested in developing libraries for screening. Sub-selection from the commercial offering for initial library preparation also supports initial follow up of screening results through purchase of similar structures—'analogue by catalogue' (*vide infra*). There is still some question however over the absolute breadth of coverage of commercial offerings, in the same way that combinatorial 'universal libraries' were once believed to represent all of chemical space. In a follow up paper to the PAINS publication Baell has postulated that the commercial offering of millions of compounds is a 'shallow pool' that can be represented with fewer than 350 000 compounds,[100] and undoubtedly there are many more regions of chemical space that could be opened up and explored by alternative chemistries (and indeed technologies, such as the DNA encoding approaches that can increase the compound count to billions—more on that later)—the question beyond this chapter (though picked up in other sections of this book) is how much of that is truly 'drug space'.

2.7 DESIGN STRATEGIES

As highlighted in the introduction to this chapter, a compound collection can be broadly categorised as supporting two types of investigation—either looking for start points for discovery programmes where little or nothing is known about the target or for maximising the chances of finding novel results for targets where we believe we do have some understanding of the underlying requirements for that target (usually but not exclusively structurally derived). The first approach requires a focus on diversity and as wide a selection of compounds exploring chemical (or drug like) space as possible, though there are limitations that need to be considered around structures carrying unwanted liabilities that may potentially limit the developability of a particular series. The second approach needs to rely on structure guided knowledge to allow the library to focus down to advantageous regions of chemical space (often around 'privileged' motifs and structures). A good chemical collection design strategy can and should encompass both components (at least for collections intended for broad usage—an organisation whose whole focus is on one particular protein target class would do best to focus

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032 08:36:02.

towards maximising coverage around knowledge based design for that target class). However for the sake of clarity the two types of use (and relevant design) will be considered sequentially in this review.

2.8 DIVERSITY COLLECTIONS

Approaches to diversity, with extensive focus on the algorithms behind computational, cheminformatic and mathematical modelling, and the comparative analysis of the effectiveness of such design strategies, have been extensively discussed elsewhere[101–104] and are covered in other chapters of this book; a relatively high level appreciation is all that is necessary to illustrate the issues relevant to this chapter. Indeed, the effectiveness of any particular diversity model and design strategy is difficult to quantify objectively, as by the very nature of the use of such diversity libraries sparse data sets are created as primary outcomes, with the majority of data based on single point biological measurements that even when positive are not exhaustively followed through to confirmation. Potential chemical structures of interest are typically rapidly reduced down to a small number of compounds through cascades of counter screening (again often single point), calculated physical properties, and structural evaluation based on chemist intuition and it is this smaller subset that then may have more detailed data measured and evaluated. The objective of any novel target drug discovery screening campaign is to reach a decision point on 'hit identification' as fast as possible preferably generating a few good structural series for further medicinal chemistry development rather than on statistical validation of the effectiveness of any particular design strategy. A screening campaign is expensive in both time and consumable costs (often under close scrutiny to remain tight to budget and timelines) even when focused as rapidly as possible towards the hit finding goal and it is not surprising that experiments to understand why a design strategy may be effective or otherwise are not considered worth pursuing.

Although objective result based evaluation of a collection design is therefore rare, the importance of applying good design principles, especially around those learnt through the mistakes of the earlier days of chemical libraries, is clearly understandable. A good design must first and foremost try to limit the presence of undesirable compounds (why have them if they will never be followed up) whilst working to the physical constraints of the collection itself and the means of generating such collections (for example, it is easier to achieve a very favourable property profile through individual purchase of compounds but may be more cost effective and efficient to generate compounds through a combinatorial synthesis approach that will generate a wider profile of properties, even though this will produce a small number of undesirable structures). The overall size of a collection should reflect the constraints set by physical capacity (how many compounds can be held and processed by a particular group or organisation), level of automation in both compound handling and screening (what capacity of screening is actually achievable) and the cost per well of particular screening targets (is the focus on high throughput, low cost biochemical assays or higher cost cell based or complex reagent based).

Once the realistic limits of the size of a collection have been decided, then the next stage is to determine the highest quality selection of compounds to use as the potential library selection set. This could be defined from commercially available compounds and/or based on virtual libraries from available and reliable synthetic transformations available to the library designers (again the synthesis of these libraries may be from commercial suppliers as well as in house resources). The use of virtual libraries based on parallel and combinatorial synthesis approaches brings additional constraints of practicality, as sparse synthesis of representatives of a matrix of compounds is often much more labour intensive than blanket synthesis of all, but again this is a balance that needs to be considered for each case in turn and cost/value analysis carried out for each synthesis design and sub-selection strategy. For the sake of this process

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

however this review will focus on the commercial purchase strategy, but all the decision processes can and should be equally applied if using a synthesis driven approach.

All commercial compounds available need to be collated together and then analysed to remove those compounds that would not be wanted in a general diversity screening campaign—that is anything that can be predicted to have toxicity or promiscuity effects, plus all compounds where the molecular properties and physicochemical parameters mean that the compounds would be very unlikely to be successfully progressed to a small molecule drug candidate. Once again the actual values of for example molecular weight cut-off or polar surface area value are subjective decisions based on the strategy and previous experience of the designer and the organisation involved, and many different views will exist, though in a study on using the wisdom of crowds to develop compound libraries Agrafiotis[105] found some strong common held understanding of important properties and parameters. This study involved medicinal chemists from six different Johnson and Johnson Pharmaceuticals R&D sites across the US and Europe, and examples of the findings are shown in Table 2.2. One overall finding was that it appears easier for chemists to agree on molecules that they do not like rather than on those that they may all favour.

Many exclusion filters applied at this stage are well understood and include historic considerations such as those based on known reactive and toxic functionailities;[106,107] these will include structures such as alkylating agents (epoxides, aziridines, activated alkyl halides *etc.*), acylating agents (acyl halides, anhydrides, sulfonyl halides *etc.*) reactive carbon–heteroatom multiple bonds (aldehydes, ketones, imines, 1,2-dicarbonyls *etc.*) and heteroatom–heteroatom single bonds (disulfides, perethers *etc.*). Others are those that have been learnt through several years of high throughput screening[99] (and having been the repeated subject of failed hit to lead prosecutions). The actual mechanisms by which these compounds may interfere with screens are varied, and will include off target effects through promiscuous activity, false screening results due to protein aggregation, non-stoichiometric non-specific binding or interference with assay read out technologies. It should also be remembered that knowledge around problem structural types and functionalities is continually developing and regular reanalysis of a screening collection based on updated filters is to be recommended. A single structural type to illustrate these issues is shown in Figure 2.5.

Rhodanines and related thiohydantoin structures have been identified in a multiple of screens across a wide range of target classes. A comprehensive review[108] of this class of compound concluded that many of the results could be put down to aggregation effects, reactivity to proteins (through a conjugate addition mechanism) and the generation of reactive compounds through photochemistry during assays. However in another study,[109] a large library of rhodanines and other related structures were synthesised and analysed for their propensity to hit multiple targets. Structures containing an exocyclic sulfur double bond and with a benzylidene substitution (thus creating an extended aromatic system) were identified as 'frequent hitters'. However further studies precluded both aggregation and protein reactions as mechanisms

Table 2.2 Preferred compound properties identified by 'crowd sourcing' methodology within Johnson and Johnson.

Property	Preferred	Disliked
Molecular weight	300–400	<250 or >425
Rotatable bonds	4	<3 or >6
H-Bond donors	1	>2
H-Bond acceptor	3	<2 or >4
ALogP	Dependant on target location	>4

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

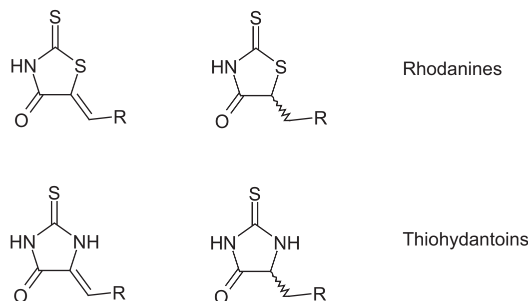


Figure 2.5 Rhodanine and related heterocycles—frequent hitting PAINS or viable hit to lead start points?

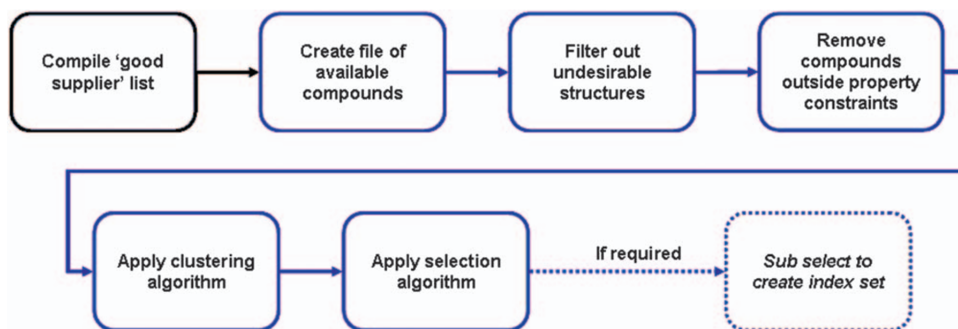


Figure 2.6 Typical process flow in the design and selection of a screening library from commercial sources.

involved—the authors concluded that the electronic and hydrogen bonding potential of these structures led to the formation of a wide range of weak/moderate molecular interactions. As such the compounds were acting in a specific manner, but were able to do so with a wide range of proteins.

A typical process to define a diversity based screening collection can therefore be summarised in Figure 2.6.

In brief, an initial collection of commercially available compounds is drawn together from reliable suppliers (another aspect that in reality is developed through experience of ordering compounds—it is important to use suppliers that have the track record of delivering compounds to order, as any design will be compromised if suppliers cannot supply their catalogue compounds). This set is then subjected to filters to remove undesirable and reactive structures and functionalities. Additional property limits will also be applied, such as molecular weight and lipophilicity ranges to ensure the set falls within a desired range of properties (again dependant on the type and use of the set—for example, focus on CNS targets may have a different lipophilicity profile), yielding a final set of available and acceptable structures. At this point the application of a selection method is required to reduce the numbers down to a final level that can be accommodated in the physical limitations of the collection housing. This will typically be a two-step process, with firstly some form of structural based algorithm applied which places the available compounds into clusters of 'like' compounds—the definition of like being based on the algorithm applied (most commonly based on Tanimoto similarity indices).[110,111] A second stage sampling method is then applied to the clusters to select representation of the clusters. Statistical methods[112] have been developed to consider the optimum number of compounds that need to be selected to maximise the potential of finding a hit from a cluster

View Online

(assuming that there is a hit in that cluster), but in simplistic terms the intention is to obtain a sufficient density of coverage of each cluster to maximise the chances of finding a positive from that cluster whilst at the same time allowing reasonable sampling across all the clusters. Within the MRCT collection, for example, the range is variable but an aim during the initial construction and subsequent refinement has always been to achieve between 10 and 20 representative cluster members for each substructure.

The advantage of using the cluster-selection approach is that it allows the design of an optimum representation of a full set of compounds within the constraints applied to the screening set (most typically size and cost driven). Though used typically to generate screening libraries that range from 10's of thousands up to millions of compounds, the approach can be used to support other diversity based strategies. For example, if the ability to carry out a large (100 000 plus compound) HTS campaign is limited for a particular target (possibly by reagent cost or technical capability) then sampling of the clustered diversity screening library can be used to define smaller 'index' sets of compounds, based on a sparser selection of compounds from the original clusters in the initial screening collection build. Clearly such an approach may limit the statistical likelihood of finding hits for any given cluster (as sampling is much more sparse) but it may be a pragmatic necessity that allows for a target to still be explored through screening approaches, and should any hits be found then there is immediate available follow up from the larger screening set which may then allow rapid assessment of the potential of a particular structural series.

Sampling of the wider cluster should also be the preferred first steps for following up any hits resulting from a full diversity HTS. Any hits from a screen should be analysed to identify whether there are multiple hits from a particular structural cluster. Follow up confirmation screening should be applied not only to the compounds that were found in the initial screen but also to other nearest neighbours in the same cluster. Assuming this initial rescreening confirms interest in the structural series then the next stage should be to return to the larger commercially available cluster that the screening set was drawn from and near neighbours should then be purchased for screening and the establishing of any early SAR trends (often referred to as 'analogue by catalogue'). Although there will never be perfect coverage of all the substituted analogues a good SAR design would demand, this approach is a highly efficient and time effective means of rapidly assessing a series' potential for further SAR development and allows for comparison between series rapidly. Testament to the (cost) effectiveness of such an approach (initial screen, confirm close neighbours, explore SAR by accessing wider cluster compounds) is the adoption of similar approaches by many large pharmaceutical companies to sample their in house diversity compound decks. Often several million compounds in size, the economics of screening such large collections has become challenging, and smaller sets representing the large set are now often screened, with the equivalent of 'analogue by catalogue' then being carried out on their in house large set (and also externally with commercial compounds). In one example an exercise based primarily on elimination of molecular redundancy within a screening file has allowed Pfizer to reduce their primary screening deck by almost 1.5 million compounds.[113]

Building a diverse screening set based on commercially available 'lead like' small molecules is the most common approach adopted for building and developing a screening capacity.[114] However alternative approaches towards diverse sets are used and these can be particularly valuable for some classes of targets, especially if the outcome of a screen is focused towards the generation of tool compounds for target validation and biological understanding of a particular target or pathway rather than the specific identification of a potential small molecule drug discovery start point. A number of smallish sets of pharmacologically active compound sets are available, some from commercial sources (*e.g.* Sigma Aldrich Lopac[115] or the Prestwick screening collection[116]) whilst others have been built by charitable groups to support specific

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

areas of disease biology (*e.g.* NINDS set[117]). Such sets can be used to interrogate biological pathways, for example in whole cell or *in vivo* phenotypic screens, as many of the compounds in the set will have well described pharmacology and underlying target information (albeit often developed in alternative indication studies). Such sets can also be used for studies looking at repositioning known compounds for new indications, identifying previous unseen pharmacology in novel systems that have the potential to become therapeutic intervention points.

Sets of natural products can serve a similar purpose to the pharmacologically active sets, and many are available as well defined and characterised sets of isolated compounds (as opposed to earlier natural product screening approaches based on fractionated uncharacterised extracts). It has been argued[118] that natural products should have a greater success rate in screening for biologically active compounds given that they are generated in a 'biological selection environment' and certainly hits found from natural products can often serve as tool compounds for valuable target validation exercises (assuming they have a good level of selectivity) even if they are not obvious start points for drug discovery programmes due to complexity or limited availability. In an extension to just accessing and screening natural products, the biosynthetic pathways themselves can in some cases be used to generate diversity compounds. For example, the polyketide biosynthesis pathways have been studied and modified to generate compound sets for biological evaluation.[119] Taking the idea one step further and into the synthetic chemistry domain, the concept of Diversity Oriented Synthesis (DOS)[120,121] has been developed to allow synthetic chemists to focus on delivering structures similar to natural products in their complexity. This approach allows synthetic chemistry groups to exploit their established chemistry methodology to generate screening compounds, though it is worth noting that to date the majority of these efforts have been limited to screening within the local generating groups.

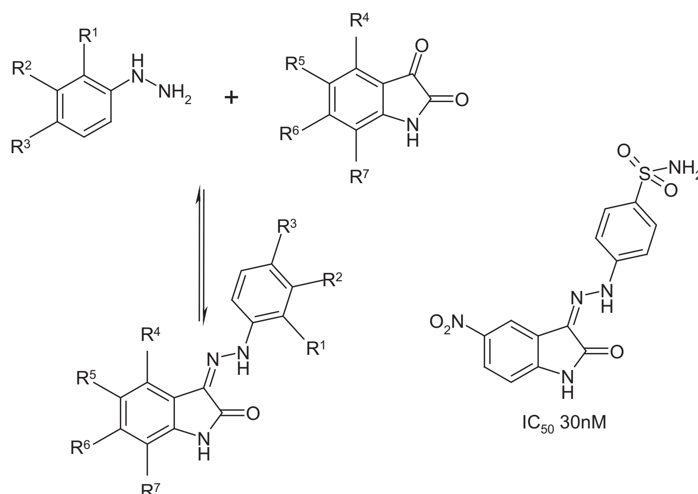
Before moving on to more focused library design it is worth considering the concept of fragment screening and dynamic combinatorial libraries as parts of the whole spectrum of diversity screening approaches.

Fragment screening represents an alternative approach to lead discovery and subsequent optimisation by accessing the widest possible diversity of chemical space using smaller molecules. Rather than sampling chemical space through a large number of discrete compounds, fragment screening covers wide areas with few compounds in a very nebulous manner, with any interactions likely to be weak but still identifiable, and that can subsequently be optimised through structure guided synthesis and design. Fragment screening has become a significant approach to diversity screening[122] and is covered in much greater detail in the dedicated chapter of this volume.

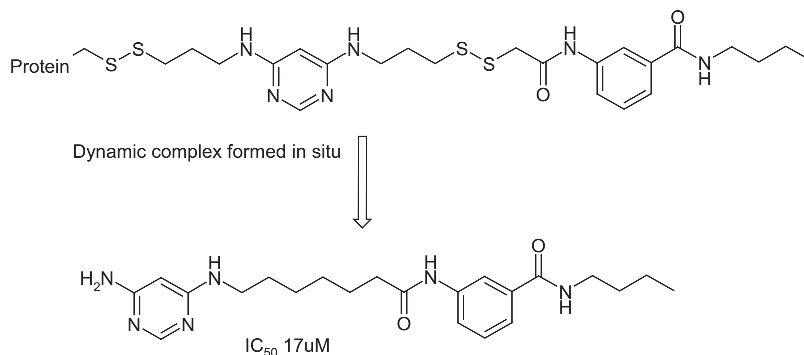
The related approach of dynamic combinatorial libraries typically uses a fragment like approach, but instead of following lead discovery programmes using iterative structural based design, the concept relies on the protein target itself building directed larger molecules of interest from a cocktail of fragments present in an assay mixture, using reversible bond forming reactions.[123] Astex have described the extension of fragment screening to generate larger lead like molecules bound to CDK2.[124] Mixtures of aryl hydrazines and isatins were soaked into individual crystals of CDK2, and under equilibrating conditions reacted in a condensation reaction to form hydrazones. These were then examined *in situ* using X-ray crystallography, before the most promising compounds were resynthesised and fully profiled in typical assays. The best compound had an IC_{50} of 30 nM (Scheme 2.7).

Another dynamic combinatorial approach using the target protein to template the chemistry was described by Sunesis pharmaceuticals.[125] In this example, the target Aurora Kinase was initially modified by site directed mutagenesis to present a cysteine SH close to the putative binding site. This handle was then exploited in a dynamic combinatorial chemistry strategy using mixtures of disulphide building blocks, which under the equilibrating conditions underwent S–S cleavage and reformation of disulphide bonds. Any building block favoured to fit

View Online



Scheme 2.7 Using reversible dynamic combinatorial chemistry in screening an extended fragment library *in situ* during crystal soaking experiments against the protein kinase CDK2.



Scheme 2.8 Disulfide equilibration in dynamic combinatorial chemistry and subsequent optimisation of compounds active at Aurora Kinases.

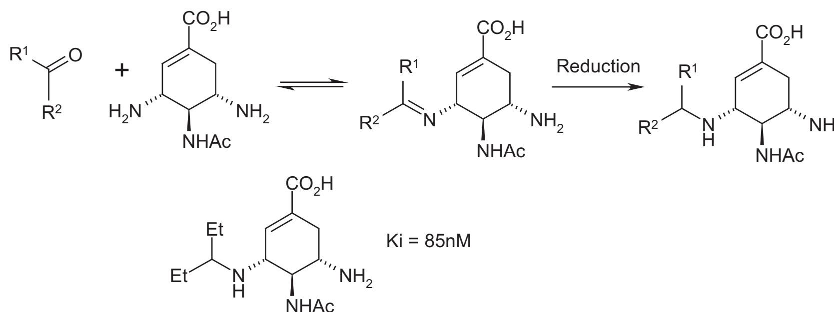
Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

in the binding site of the kinase was therefore held close to the cysteine handle in a favourable position to form a disulphide bridge. The initial series of monomers incorporated a second set of disulphide links, thus allowing a second round of equilibration with another set of monomers, and finally yielding compounds with activity in the single micromolar range (Scheme 2.8).

An alternative approach to performing reactions directly in the presence of a protein is to allow the dynamic mixture to equilibrate before introducing the target protein. Therascope have described such an approach to target novel neuraminidase inhibitors, using reductive amination chemistry on a scaffold related to known inhibitors.[126] In this example the initial imine formation was performed in the absence of the neuraminidase, and the resulting mixture reduced to yield a set of amines that could be profiled by LC-MS. The same reaction sequence was then repeated, this time with the introduction of neuraminidase during the initial imine equilibration and following the reduction step the amine profile was again analysed. A specific number of ketone examples were dramatically amplified by the addition of the enzyme, with all subsequently shown by resynthesis to be potent inhibitors of the enzyme, the most potent having a Ki of 85 nM (Scheme 2.9).

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.



Scheme 2.9 Target protein driven amplification of dynamic combinatorial chemistry to identify favoured constructs for inhibiting neuraminidase following reduction to non-reversible products.

2.9 TARGETED LIBRARIES

The design of targeted sets of compounds against specific protein classes can be approached from two complementary approaches akin to other design methods used in computational approaches to single target drug discovery, either protein structure based or ligand/substrate based. However, with the intent in designing such sets being to target future as yet unknown family members of the same protein family, additional care needs to be taken in selecting the most experimentally validated targets and ligands to build from, and to ensure any alignment of structures (protein or ligand based) is undertaken with an eye to class related common features wherever possible. However, even the best designed arrays can provide surprising results as demonstrated in the design of novel arrays targeting the kinase CDK2[127] and related family members. Based on crystallographic structural data using active oxindole compounds novel structures were designed around an aza indole structure, with the intent to maximise the interaction with the hinge region of the active site by picking up additional binding. However, analysis of the active compounds from the subsequent arrays using crystallography showed that rather than just picking up additional binding, the novel structures had adopted two new binding poses, with one even having reversed the direction of the core scaffold in its binding to the hinge region (Figure 2.7).

Rational design of targeted libraries based on protein structures is likely to be most effective where a significant amount of structural information around the protein family is known. The kinase family[128] and several protease family sub-classes[129] are good candidates for such approaches, sharing as they do significant active site structural homology between members of the same family. Perhaps surprisingly, another class of drug targets that has received similar structural attention is protein–protein interaction targets. In this case it is structural motifs that are the product of the protein secondary structure that are targeted, for example mimetics of the spatial arrangement of alpha helix side chains.[130]

Where limited structural knowledge of the protein class is available then ligand based screening sets have often been designed around trying to define privileged structures and motifs from active ligands and inhibitors of related proteins which can then be incorporated into targeted library designs, approaches that have been outlined by both AstraZeneca[131] and GSK.[68] As a simple illustration of the concept, the selection of a tricyclic biphenyl motif to include as a monomer building block would come from the analysis of known serotonin/adrenergic reuptake inhibitors (Figure 2.8).

There are however some issues with using this approach to define targeted sets and libraries. By designing a library around known structural motifs then there is a built in risk that any results from that library will sit in close structural space to known compounds, which in turn

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

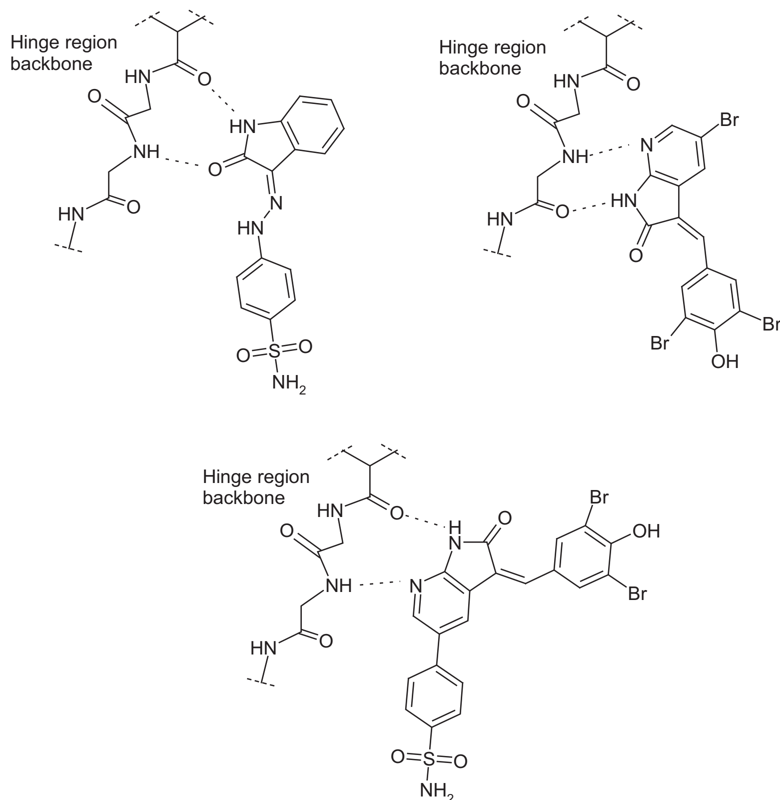


Figure 2.7 Three distinct modes identified for (aza)oxindoles binding to the hinge region of protein kinases.

brings issues around patentability and freedom to operate. Moreover the grouping of some common protein structures is often at a superficial and broad level, leading to very generic designs (for example, libraries purportedly focused for ion channels or GPCRs rather than subclasses of these targets). A limited number of studies have shown that for many of these 'targeted' library approaches, outside of kinase and some family A GPCR focused libraries, the statistical results of screening campaigns are no different from using general diversity based high throughput screening,[101] which if done at full scale also brings the potential for moving into truly novel chemical space for a target class. There are however published examples where targeted libraries have been used to successfully identify novel start points for drug discovery programmes, including kinases,[129] voltage gated ion channels[132,133] and serine/cysteine proteases.[129]

2.10 COMBINATORIAL POWER IN DESIGN

The original drive behind the development of combinatorial chemistry to support high throughput screening was the potential to access very large numbers of compounds using efficient synthetic paradigms. The limited success of early screening approaches using mixtures, either synthesised as such or through pooling strategies, combined with the increasing capacity of screening campaigns meant the focus shifted significantly to one assay one compound strategies, and this has remained the standard approach to screening in most companies to

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

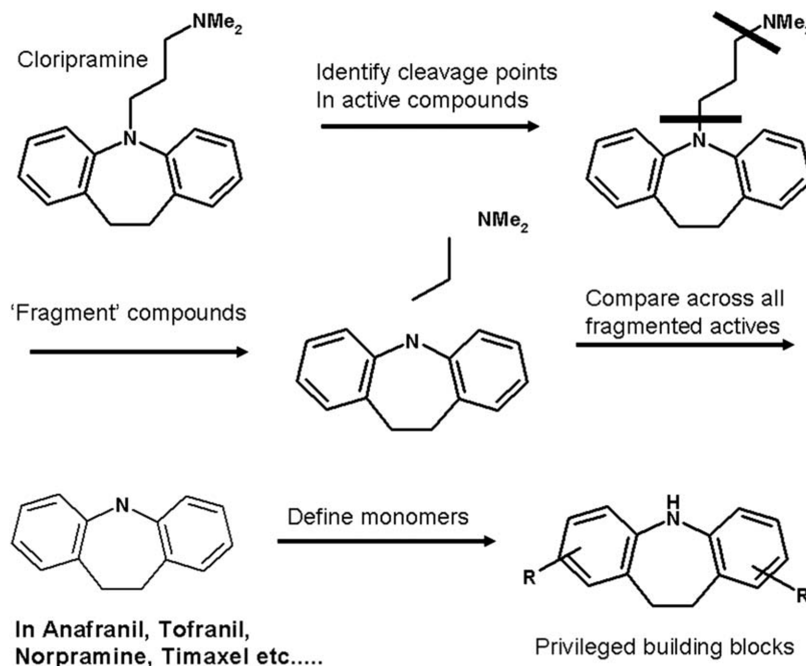


Figure 2.8 Illustration of the process that ReCap uses to identify privileged monomers from known active compounds.

date. However as illustrated in the historical review earlier in this article, a significant proportion of screening has shifted to more targeted approaches, either to specific protein families or to more refined means of interrogating diverse chemical space. This has been the result of a number of factors, with the cost of screening large numbers of compounds being a key component.

The alternative approach to controlling costs whilst still maximising the scale of combinatorial chemistry would be to return to the screening of compounds as mixtures, and approaches to support this have continued to develop in parallel to the mainstream developments in single compound screening. Houghten has continued to use and develop combinatorial libraries screening methodologies to maximise the effectiveness of compound mixture screening,[134] with extensive use of computational[135] and mathematical[136] modelling to support lead identification.

Finally, in an approach that takes the art of combinatorial synthesis back to its initial beginnings and focus on very large numbers, the use of DNA encoding to allow the rapid screening of extremely large mixtures of compounds has been described by several groups,[137,138] illustrated here by the production of 7 million triazine compounds by the Praecis group (Figure 2.9).[139] Making use of the sensitivity of PCR approaches to rapidly amplify a particular code, libraries of multiple millions of compounds have been prepared, screened and deconvoluted, though the range of chemistry associated with the small molecule still carries the same potential restrictions and liabilities associated with previous polymer bound and tagged approaches. Now part of GSK, this group recently published[140] the discovery of highly potent and selective ADAMTS-5 inhibitors using this technology, with the identification of compounds that did not carry the usual type of zinc binding motif (*e.g.* hydroxamic acids) that typically create selectivity issues across other zinc containing metalloproteinases (Figure 2.10).

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

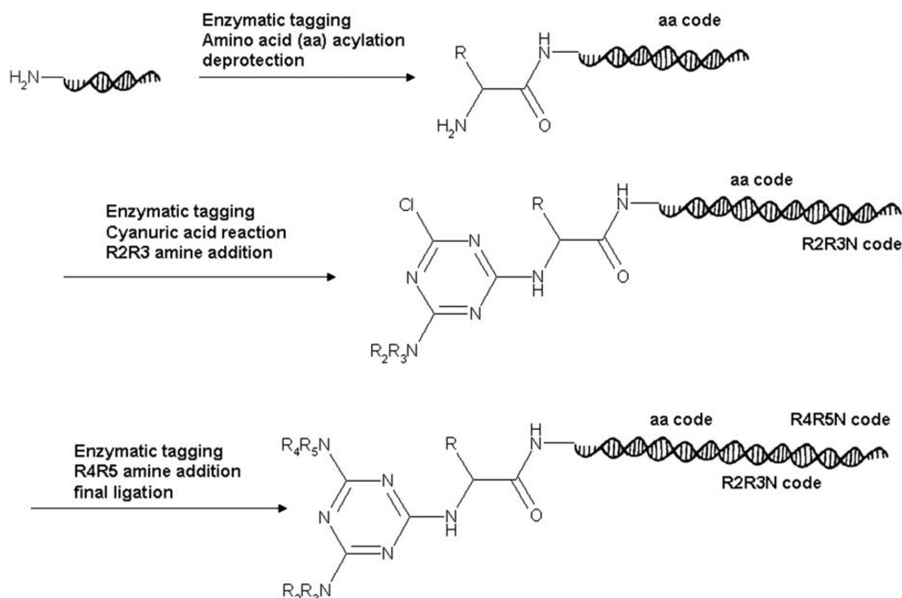
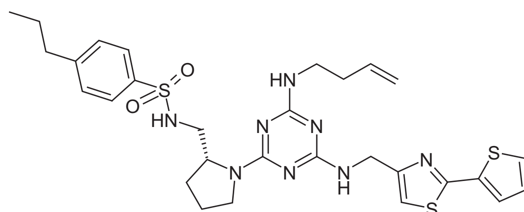


Figure 2.9 Preparation of a three point diversity triazine library with DNA encoding.



ADAMTS-5 IC_{50} = 0.030uM

ADAMTS-4 IC_{50} = 1.5uM

MMP-13 IC_{50} > 25uM

Figure 2.10 Selective and potent ADAMTS-5 inhibitor identified through the screening of a 4 billion member DNA encoded triazine library.

2.11 CONCLUSION

Combinatorial practices, be they large library purchase, syntheses or focused efforts of parallel chemistry around SAR generation, have become widespread throughout the drug discovery process. The initial promise of Combichem, leading as it did in the 1990s to the development of specialist teams and companies, has gone through a process of expansion, realisation, disappointment and reassessment, to reach a point where it is a valuable tool, part of the overall armoury of drug discovery to be used alongside other approaches. Compound collection numbers, very much the initial driver of the combinatorial explosion, are still significant factors in defining how drug discovery can be prosecuted. However, rather than the in house (or commissioned) combinatorial approach it is as much through purchase of compounds that these numbers are built.[114] Whereas 15 years ago purchasing compounds was very often a lottery of quality, availability and pharmaceutical relevance, it is now possible to build very large high quality diverse screening sets from commercial sources.

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

So what should be the take home messages from the last 15–20 years of combinatorial approaches to drug discovery? Well established understanding and due reflection on the multi parameter complexity of drug discovery should never be displaced by the technological challenges of a new strategy—after all it is still about finding an active compound that will elicit the correct response in a physiological system, not about the technology. Universal approaches will almost certainly never exist, and application of knowledge about the target proteins can maximise the effectiveness of one design over another. The application of derived drug/lead like properties at the start of any design strategy will save a lot of time at the screening stages of any program.

Combinatorial chemistry began as a tool for understanding biological processes. The application to drug discovery and the generation of small molecule drug compounds became a dream that for many developed into a nightmare of over investment and limited return. But 20 : 20 hindsight is always right, and we should not be so quick to condemn the work of the earlier combinatorial pioneers. Without those pushing the boundaries of the science we wouldn't now have an approach that when applied correctly can enormously shorten the discovery cycle and maximise the opportunity to optimise in parallel across a wider range of parameters than could ever have been imagined. Pick up any copy of the *Journal of Medicinal Chemistry* or *Biological and Medicinal Chemistry* and randomly open to an article—the odds are now very strong that one of the descriptors 'parallel', 'array', 'high throughput' or even 'combinatorial' will be prominent. The hype came and went but the processes embedded and stayed.

HINTS AND TIPS 1

Typical Equipment for the Parallel Synthesis Chemist

There are many varieties of parallel equipment available on the market, supporting synthesis, analysis, purification and final characterisation, and it would be impossible to provide a comprehensive listing here. The following suggestions are those which the author and his group have found to be useful, reliable and where relevant user friendly. This should not be viewed as a specific endorsement and many equivalent products are available. Moreover there is much to be recommended in a pragmatic approach of adaptation of normally available equipment—the development of some of these pieces of hardware that are now commercial products began life as elastic band and sealing tape prototypes in the author's laboratories back in the early 1990s.

Synthesis equipment: Adaptation of a normal magnetic stirrer to allow parallel reactions with varying levels of heating cooling and inert atmosphere control can be achieved using equipment such as the suite of reaction stations supplied by Radleys (www.radleys.com); including the simple Starfish multiple reaction station, the range of carousel stations and the greenhouse parallel synthesisers. The latter two series also have the advantage of companion workup stations that allow for simple work up procedures in a matching format to the reaction numbers. For more dedicated parallel reaction stations then either the Metz heater shaker system (also available from Radleys) or the STEM RS series reaction stations (www.electrothermal.com) provide dedicated parallel tube systems. More specialist equipment, typically relying on sequential flow reactions rather than parallel reactions, is also available, either for general reaction conditions (see www.syrris.com) or specialist gas reactions (H-Cube; www.thalesnano.com/h-cube). Finally the use of microwave heating has radically changed parallel chemistry capabilities over the past decade, and either CEM (www.cemmicrowave.co.uk) or Biotage (www.biotage.com) equipment is readily usable in the research laboratory.

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

Work up and purification: Parallel filtration and aqueous extraction processes are available to match the reaction stations as described above, and parallel evaporation systems such as the range of Genevac centrifuge systems (www.genevac.com) or gas evaporation systems such as the FlexiVap work stations (www.glascol.com) allow rapid concentration of multiple reactions. Filtration cartridges can be used either as simple clean up filters or by careful selection of resin content can allow selective removal or sequestration of various functionalities—for the range of solid phase extraction (SPE) and filtration cartridges available see suppliers such as Agilent (www.chem.agilent.com). For more complex purification and separation then parallel column chromatography such as the Isolera system from Biotage (www.biotage.com) can process up to four samples in parallel through automated column chromatography.

HINTS AND TIPS 2

Compounds and Functionalities to Avoid—the PAINS of Hit Discovery

As discussed in the main chapter, there have been a number of publications describing the issues of 'frequent hitters' and trying to understand why some structural motifs are best avoided in any hit to lead follow up program. There are a number of grey areas, and it is not for this author to state categorically that any particular class of molecule, even if pursued carefully, will never yield a viable lead compound for further optimisation. However, there are several classes of compounds that, through repeated evaluation, have clearly been identified as 'problem' structures that should be progressed only with open eyes and an awareness of the history of similar strategies. That list is ever changing and, as more is understood about underlying mechanisms of biological non-specific activity, changes in both increasing and reducing directions. However, the types of core structures highlighted in the PAINS paper[99] are always worth treating with care (Figure 2.11):

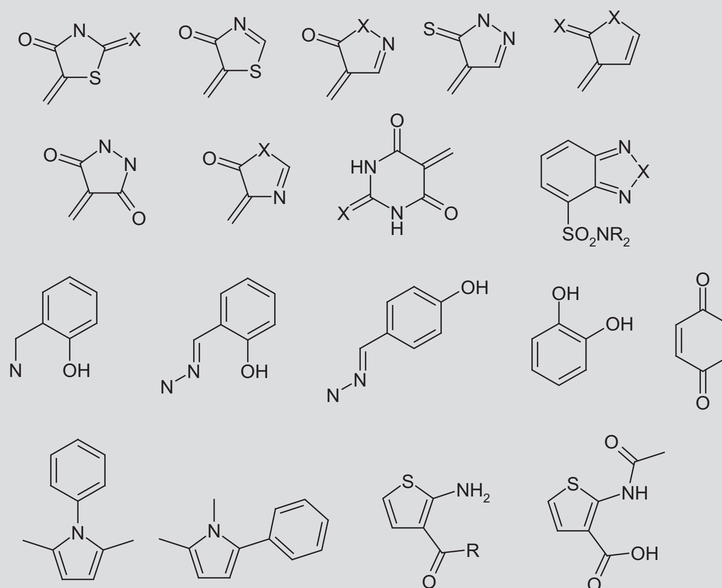


Figure 2.11 Illustrative core cyclic and heterocyclic structures identified as having 'frequent hitter' potential by Baell and Holloway in the PAINS paper.[99]

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

HINTS AND TIPS 3

Compounds and Functionalities to Avoid—the AZ approach

Perhaps easier to identify and avoid are more specific functional groups and reactive structures. AstraZeneca has identified a wide range of such functionalities in a recent paper,[141] including several of the following motifs:

Class 1: Bland structures

- Compounds containing atoms other than hydrogen, carbon, nitrogen, oxygen, sulfur, fluorine, chlorine, bromine and iodine
- Fewer than four carbon atoms
- Fewer than 12 heavy atoms
- No polar atoms (nitrogen, oxygen, sulfur)
- Straight or unbranched structures
- Positively charged atoms (for example, quaternary nitrogen)
- Compounds with three or more acidic groups
- Alkyl or aryl amine (with no other heteroatom)
- Hydroxyl or thiol (with no other heteroatom)
- Only hetero atom is one acid or derivatives

Class 2: reactive structures

- Michael acceptors: C=C–C=O, C=C–CN, C=C–SO_2, C=C–NO_2
- Reactive ester or thioester
- Anhydride
- Alpha halo ketone
- Halo methylene ether
- Acid halide and thio acid halide
- Aliphatic and aromatic aldehyde
- Peroxide
- Epoxide, aziridine, thiirane or oxazirane
- Thiocyanate
- Isocyanate, isothiocyanate
- Isocyanide, isonitrile

Class 3: frequent hitters

- More than two nitro groups
- Dihydroxybenzene
- Nitrophenols

Class 4: dye-like structures

- Two nitro groups on same aromatic ring, including naphthalene
- Diphenyl ethylene cyclohexadiene

Class 5: unlikely drug candidates or unsuitable fragments

- Large ring $\geq$C9
- C9 chain not in any rings
- Crown ethers
- Multi-alkene chain: C=CC=CC=C or N=CC=CC=C
- Diyne: –C≡C–C≡C–
- Annelated rings such as phenanthrene, anthracene and phenalene
- Two sulfur atoms (not sulfones) in 5-membered rings or 6-membered rings
- Triphenylmethyl

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

Class 6: difficult series or natural compounds
- Steroids
- Penicillin or cephalosporin
- Prostaglandins

Class 7: general 'ugly' halogenated structures
- Di- or trivalent halogens
- N–, S–, P– and O–halogens
- Sulfonyl halides
- Triflates: SO_3CX_3

Class 8: general 'ugly' oxygen
- Five or more hydroxyl groups
- p-,p′-dihydroxybiphenyl
- p-,p′-dihydroxystilbene
- Formic acid esters

Class 9: general 'ugly' nitrogen
- Hydrazine (not in ring)
- Three or more guanidines
- Two or more N-oxides
- Azo (N=N) or diazonium (N≡N)
- Carbodiimide
- N-nitroso groups
- Aromatic nitroso groups
- Cyanohydrin or (thio)acylcyanide
- Nitrite
- Nitramine
- Oxime

Class 10: general 'ugly' sulfur
- Five or more sulfur atoms
- Disulfide
- Sulfate
- Sulfonic acid
- Thioketone
- Sulfonic ester (except for aryl or alkyl–SO_3–aryl groups)
- Sulfanylamino groups
- 1,2-thiazol-3-one
- Dithiocarbamate
- Thiourea, isothiourea, thiocarbamic acid or thiocarbonate
- Isocyanate or isothiocyanate
- Thiocyanate
- Thiol
- Dithioic or thioic acid

HINTS AND TIPS 4

How to Build a Diverse Collection–The 5 Minute Practical Guide:

The main body of text contains a significant discussion on the steps and processes in designing a collection–but the basic principles and steps can be summarised in a few short

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

bullets (and referring back to Figure 2.6 in the main chapter):

- Identify as many viable commercial compound collections as possible
- Collate into a single data set in a format best suited to downstream analysis
- Apply initial exclusion filters based on strategic considerations to reduce to only compounds that would be acceptable in the collection
- Identify any cost and storage constraints to which the overall collection must comply to determine any size and/or format limitations
- Analyse the compound set using a relevant similarity and clustering protocol to generate a clustered available dataset
- Select representation from the cluster analysis to meet the number constraints of the compound collection whilst maximising coverage of the clustered dataset
- Allow chemist analysis of the selected compounds to ensure best options from cluster selections have been chosen by algorithm
- Spend the money.........

KEY REFERENCES

M. D. Hack, D. N. Rassokhin, C. Buyck, M. Seierstad, A. Skalkin, P. Ten Holte, T. K. Jones, T. Mirzadegan and D. K. Agrafiotis, *J. Chem. Inf. Model.*, 2011, **51**, 3275.
M. A. Snowden and D. V. S. Green, *Curr. Op. Drug Disc. Dev.*, 2008, **11**, 553.
J. B. Baell and G. A. Holloway, *J. Med. Chem.*, 2010, **53**, 2719.

REFERENCES

1. I. Yates, *Drug Discovery World*, 2003, 35.
2. W. S. Fillers, *Drug Discovery World*, 2004, 86.
3. D. Booth, http://www.dddmag.com/articles/2013/04/tackling-challenges-compound-management.
4. R. Macarron, M. N. Banks, D. Bojanic, D. J. Burns, D. A. Cirovic, T. Garyantes, D. V. S. Green, R. P. Hertzberg, W. P. Janzen, J. W. Paslay, U. Schopfer and G. S. Sittampalam, *Nat. Rev. Drug Discovery*, 2011, **10**, 188.
5. S. Willhelm, C. Carter, M. Lynch, T. Lowinger, J. Dumas, R. A. Smith, B. Schwartz, R. Simantov and S. Kelley, *Nat. Rev. Drug Discovery*, 2006, **5**, 835.
6. J. Das, P. Chen, D. Norris, R. Padmanabha, J. Lin, R. V. Moquin, Z. Shen, L. S. Cook, A. M. Doweyko, S. Pitt, S. Pang, D. R. Shen, Q. Fang, H. F. De Fex, K. W. McIntyre, D. J. Shuster, K. M. Gillooly, K. Behnia, G. L. Schieven, J. Wityak and J. C. Barrish, *J. Med. Chem.*, 2006, **49**, 6819.
7. M. S. Lajiness, G. M. Maggiora and V. Shanmugasundaram, *J. Med. Chem.*, 2004, **47**, 4891.
8. A. Merritt, *New Synthetic Technologies in Medicinal Chemistry*, ed. E. Farrant, RSC Publishing, Cambridge, 2012, p. 6.
9. G. R. Nakayama, *Curr. Op. Drug Disc. Dev.*, 1998, **1**, 85.
10. T. D. Mays and K. D. Mazan, *J. Ethnopharm.*, 1996, **51**, 93.
11. J. D. Bergstrom, M. M. Kurtz, D. J. Rew, A. M. Amend, J. D. Karkas, R. G. Bostedor, V. S. Bansal, C. Dufresne, F. L. Van Middlesworth, O. D. Hensens, J. M. Liesch, D. L. Zink, K. E. Wilson, J. Onishi, J. A. Milligan, G. Bills, L. Kaplan, M. N. Omstead, R. G. Jenkins, L.

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

Huang, M. S. Meinz, L. Quinn, R. W. Burg, Y. L. Kong, S. Mochales, M. Mojena, I. Martin, F. Pelaez, M. T. Diez and A. W. Alberts, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 80.

12. A. Baxter, B. J. Fitzgerald, J. L. Hutson, A. D. McCarthy, J. M. Motteram, B. C. Ross, M. Sapra, M. A. Snowden, N. S. Watson, R. J. Williams and C. Wright, *J. Biol. Chem.*, 1992, **267**, 11705.
13. R. Thiericke, *Modern Methods of Drug Discovery*, ed. A. Hillisch and R. Hilgenfeld, Birkhauser, Switzerland, 2003, p. 71.
14. O. Amedee-Manesmee, *Therapie*, 1999, **54**, 419.
15. J. Schamberger, M. Grimm, A. Steinmeyer and A. Hillisch, *Drug Discovery Today*, 2011, **16**, 636.
16. R. B. Merrifield, *J. Am. Chem. Soc.*, 1963, **85**, 2149.
17. B. Merrifield, *Methods Enzymol.*, 1995, **289**, 3.
18. C. C. Leznoff, *Acc. Chem. Res.*, 1978, **11**, 327.
19. H. M. Geysen, R. H. Meloen and S. J. Barteling, *Proc. Natl. Acad. Sci. U. S. A.*, 1984, **81**, 3998.
20. A. Furka, F. Sebestyen, M. Asgedom and G. Dibo, *Int. J. Pept. Prot. Res.*, 1991, **37**, 487.
21. R. A. Houghten, *Proc. Natl. Acad. Sci. U. S. A.*, 1985, **82**, 5131.
22. S. Rodda, G. Tribbick and M. Geysen, *Combinatorial peptide and nonpeptide libraries*, ed. G. Jung, VCH, Weinheim, 1996, p. 303.
23. P. M. St Hilaire, M. Willert, M. A. Juliano, L. Juliano and M. Meldal, *J. Comb. Chem.*, 1999, **1**, 509.
24. E. Apletalina, J. Appel, N. S. Lamango, R. A. Houghten and I. Lindberg, *J. Biol. Chem.*, 1998, **273**, 26589.
25. J. R. Appel, J. Buencamino, R. A. Houghten and C. Pinilla, *Mol. Div.*, 1996, **2**, 29.
26. R. A. Houghten and C. T. Dooley, *Bioorg. Med. Chem. Lett.*, 1993, **3**, 405.
27. C. T. Dooley and R. A. Houghten, *Life Sci.*, 1993, **52**, 1509.
28. B Deprez, X. Williard, L. Bourel, H. Coste, F. Hyafil and A. Tartar, *J. Am. Chem. Soc.*, 1995, **117**, 5405.
29. R. M. Kainkaryam and P. J. Woolf, *Curr. Op. Drug Disc. Dev.*, 2009, **12**, 339.
30. B. A. Bunin and J. A. Ellman, *J. Am. Chem. Soc.*, 1992, **114**, 10997.
31. B. A. Bunin, M. J. Plunkett and J. A. Ellman, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 4708.
32. C. G. Boojamra, K. M. Burow, L. A. Thompson and J. A. Ellman, *J. Org. Chem.*, 1997, **62**, 1240.
33. S. H. DeWitt, J. S. Kiely, C. J. Stankovic, M. C. Schroeder, D. M. R. Cody and M. R. Pavia, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 6909.
34. C. Barnes and S. Balasubramanian, *Curr. Op. Chem. Biol.*, 2000, **4**, 346.
35. M. H. J. Ohlmeyer, R. N. Swanson, L. W. Dillard, J. C. Reader, G. Asouline, R. Kobayashi, M. Wigler and W. C. Still, *Proc. Natl. Acad. Sci. U. S. A.*, 1993, **90**, 10922.
36. K. C. Nicolaou, X.-Y. Xiao, Z. Parandoosh, A. Senyei and M. P. Nova, *Angew. Chem. Int. Ed. Engl.*, 1995, **34**, 2289.
37. P. H. H. Hermkens, H. C. J. Ottenheijm and D. Rees, *Tetrahedron*, 1996, **52**, 4527.
38. P. H. H. Hermkens, H. C. J. Ottenheijm and D. C. Rees, *Tetrahedron*, 1997, **53**, 5643.
39. S. Booth, P. H. H. Hermkens, H. C. J. Ottenheijm and D. C. Rees, *Tetrahedron*, 1998, **54**, 15385.
40. S. D. Roughley and A. M. Jordan, *J. Med. Chem.*, 2011, **54**, 3451.
41. H. Han, M. M. Wolfe, S. Brenner and K. D. Janda, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 6419.
42. S. Chen and K. D. Janda, *J. Am. Chem. Soc.*, 1997, **119**, 8724.
43. A. Studer, S. Hadida, R. Ferritto, S.-Y. Kim, P. Jeger, P. Wipf and D. P. Curran, *Science*, 1997, **275**, 823.
44. S. W. Kaldor, M. G. Siegel, J. E. Fritz, B. A. Dressman and P. J. Hahn, *Tetrahedron Lett.*, 1996, **37**, 7193.

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

45. S. V. Ley, O. Schucht, A. W. Thomas and P. J. Murray, *J. Chem. Soc., Perkin Trans. 1*, 1999, 1251.
46. A. Solinas and M. Taddei, *Synthesis*, 2007, 2409.
47. S. V. Ley, I. R. Baxendale and R. M. Myers, *Comprehensive Medicinal Chemistry 2*, ed. J. B. Taylor and D. J. Triggle, Elsevier, Amsterdam, 2006, vol. 3, p. 791.
48. E. J. Martin, J. M. Blaney, M. A. Siani, D. C. Spellmeyer, A. K. Wong and W. H. Moos, *J. Med. Chem.*, 1995, **38**, 1431.
49. J. H. Van Drie and M. S. Lajiness, *Drug Discovery Today*, 1998, **3**, 274.
50. J. S. Mason and M. A. Hermsmeier, *Curr. Op. Chem. Biol.*, 1999, **3**, 342.
51. D. H. Drewry and S. S. Young, *Chemom. Intell. Lab. Syst.*, 1999, **48**, 1.
52. D. K. Agrafiotis, J. C. Myslik and F. R. Salemme, *Mol. Div.*, 1999, **4**, 1.
53. A. R. Leach and M. M. Hann, *Drug Discovery Today*, 2000, **5**, 326.
54. M. R. Pavia, S. P. Hollinshead, H. V. Meyers and S. P. Hall, *Chimia*, 1997, **51**, 826.
55. M. J. Sofia, R. Hunter, T. Y. Chan, A. Vaughan, R. Dulina, H. Wang and D. Gange, *J. Org. Chem.*, 1998, **63**, 2802.
56. T. I. Oprea, *Curr. Op. Chem. Biol.*, 2002, **6**, 384.
57. A. W. Czarnik, *ChemtTracts Org. Chem.*, 1995, **8**, 13.
58. Y. C. Martin, *Perspect. Drug Disc. Des.*, 1997, **7**, 159.
59. F. Z. Dorwald, *Organic Synthesis on Solid Phase*, Wiley and Sons, Hoboken, 2002.
60. A. Merritt, *Rodd's Chemistry of Carbon Compounds, Topical Volumes, Asymmetric Catalysis*, ed. M Sainsbury, Elsevier, Amsterdam, 2nd edn, 2001, vol. 5, p. 259.
61. A. T. Merritt, *Drug Discovery Today*, 1998, **3**, 505.
62. *J. Comb. Chem.; Comb. Chem. High Throughput Screening; Mol. Diversity*.
63. J. Alper, *Science*, 1994, **264**, 1399.
64. C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv. Drug Del. Rev.*, 1997, **23**, 3.
65. R. A. Fecik, K. E. Frank, E. J. Gentry, S. R. Menon, L. A. Mitscher and H. Telikepalli, *Med. Res. Rev.*, 1998, **18**, 149.
66. S. J. Teague, A. M. Davis, P. D. Leeson and T. Oprea, *Angew. Chem. Int. Ed. Engl.*, 1999, **38**, 3743.
67. M. M. Hann, A. R. Leach and G. Harper, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 856.
68. X. Q. Lewell, D. B. Judd, S. P. Watson and M. M. Hann, *J. Chem. Inf. Comput. Sci.*, 1998, **38**, 511.
69. http://www.combinatorial.com/.
70. http://www.peptide.com/.
71. B. Evans, A. Pipe, L. Clark and M. Banks, *Bioorg. Med. Chem. Lett.*, 2001, **11**, 1297.
72. N. Bailey, A. W. J. Cooper, M. J. Deal, A. W. Dean, A. L. Gore, M. C. Hawes, D. B. Judd, A. T. Merritt, R. Storer, S. Travers and S. P. Watson, *Chimia*, 1997, **51**, 832.
73. X.-Y. Xiao, R. Li, H. Zhuang, B. Ewing, K. Karunaratne, J. Lillig, R. Brown and K. C. Nicolaou, *Biotech. Bioeng.*, 2000, **71**, 44.
74. http://www.nexusbio.com/.
75. N. Hird and B. MacLachlan, *Laboratory Automation in the Chemical Industries*, ed. D. G. Cork and T. Sugawara, Marcel Dekker, New York, 2002, p. 1.
76. http://www.zinsser-analytic.com/.
77. http://www.combichemlab.com/website/files/Combichem/Workstations/argonaut.htm.
78. N. Kuroda, N. Hird and D. G. Cork, *J. Comb. Chem.*, 2006, **8**, 505.
79. http://www.stemcorp.co.uk/s_index.cfm.
80. http://www.radleys.com/.
81. http://uk.mt.com/gb/en/home/products/L1_AutochemProducts/L3_Post-Synthesis-Work-up.html.

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

82. J. J. Parlow, *Curr. Op. Drug Disc. Dev.*, 2005, **8**, 757.
83. G. A. Strohmeier and C. O. Kappe, *Angew. Chem. Int. Ed. Engl.*, 2004, **43**, 621.
84. P. Lan, D. Berta, J. A. Porco, M. S. South and J. J. Parlow, *J. Org. Chem.*, 2003, **68**, 9678.
85. Y. Wang, K. Sarris, D. R. Sauer and S. W. Djuric, *Tetrahedron Lett.*, 2007, **48**, 2237.
86. J. Habermann, S. V. Ley and J. S. Scott, *J. Chem. Soc., Perkin Trans., 1*, 1999, 1253.
87. S. V. Ley, M. Ladlow and E. Vickerstaffe, *Exploiting Chemical Diversity for Drug Discovery*, ed. P. A. Bartlett and M. Entzeroth, RSC Publishing, London, 2006, p. 3.
88. C. O. Kappe and D. Dallinger, *Nat. Rev. Drug Discovery*, 2006, **5**, 51.
89. A. De La Hoz, A. Diaz-Ortiz and A. Moreno, *Chem. Soc. Rev.*, 2005, **34**, 164.
90. B. Desai and C. O. Kappe, *Topics in Curr. Chem.*, 2004, **242**, 177.
91. J. M. Kremsner and C. O. Kappe, *J. Org. Chem.*, 2006, **71**, 4651.
92. L. Pisani, H. Prokopcova, J. M. Kremsner and C. O. Kappe, *J. Comb. Chem.*, 2007, **9**, 415.
93. J. N. Kyranos, H. Lee, W. K. Goetzinger and L. Y. T. Li, *J. Comb. Chem.*, 2004, **6**, 796.
94. W. Leister, K Strauss, D. Wisnoski, Z. Zhao and C. Lindsley, *J. Comb. Chem.*, 2003, **5**, 322.
95. S. J. Lane, D. S. Eggleston, K. A. Brindred, J. C. Hollerton, N. L. Taylor and S. A. Readshaw, *Drug Discovery Today*, 2006, **11**, 267.
96. A. W. Squibb, M. R. Taylor, B. L. Parnas, G. Williams, R. Girdler, P. Waghorn, A. G. Wright and F. S. Pullen, *J. Chromatogr. A*, 2008, **1189**, 101.
97. S. Lane, B. Boughtflower, I. Mutton, C. Paterson, D. Farrant, N. Taylor, Z. Blaxill, C. Carmody and P. Borman, *LC-GC Europe*, 2006, **19**, 161.
98. R. Richardson, *Exploiting Chemical Diversity for Drug Discovery*, ed. P. A. Bartlett and M. Entzeroth, RSC Publishing, Cambridge, 206, p. 112.
99. J. B. Baell and G. A. Holloway, *J. Med. Chem.*, 2010, **53**, 2719.
100. J. B. Baell, *J. Chem. Inf. Model.*, 2013, **53**, 39.
101. D. V. S. Green, *Drug Design Strategies*, ed. D. J. Livingstone and A. M. Davis, RSC Publishing, Cambridge, 2012, p. 367.
102. D. J. Huggins, A. R. Venkitaraman and D. R. Spring, *ACS Chem. Biol.*, 2011, **6**, 208.
103. R. E. Dolle, *Methods Mol. Biol.*, 2011, **685**, 3.
104. D. M. Schnur, B. R. Beno, A. J. Tebben and C. Cavallaro, *Methods Mol. Biol.*, 2011, **672**, 387.
105. M. D. Hack, D. N. Rassokhin, C. Buyck, M. Seierstad, A. Skalkin, P. Ten Holte, T. K. Jones, T. Mirzadegan and D. K. Agrafiotis, *J. Chem. Inf. Model.*, 2011, **51**, 3275.
106. J. Ashby and R. W. Tennant, *Mutation Res.*, 1991, **257**, 229.
107. G. M. Rishton, *Drug Discovery Today*, 1997, **2**, 382.
108. T. Tomasic and L. P. Masic, *Expert Opin. Drug Discov.*, 2012, **7**, 549.
109. T. Mendgen, C. Steuer and C. D. Klein, *J. Med. Chem.*, 2012, **55**, 743.
110. Y. C. Martin, J. L. Kofron and L. M. Traphagen, *J. Med. Chem.*, 2002, **45**, 4350.
111. P. Willett, J. M. Barnard and G. M. Downs, *J. Chem. Inf. Comput. Sci.*, 1998, **38**, 983.
112. G. Harper, S. D. Pickett and D. V. S. Green, *Comb. Chem. High Throughput Screening*, 2004, **7**, 63.
113. G. A. Bakken, A. S. Bell, M. Boehm, J. R. Everett, R. Gonzales, D. Hepworth, J. L. Klug-McLeod, J. Lanfear, J. Loesel, J. Mathias and T. P. Wood, *J. Chem. Inf. Model.*, 2012, **52**, 2937.
114. M. A. Snowden and D. V. S. Green, *Curr. Op. Drug Disc. Dev.*, 2008, **11**, 553.
115. http://www.sigmaaldrich.com/catalog/product/sigma/lo1280?lang=en®ion=GB.
116. http://www.prestwickchemical.com/index.php?pa=26.
117. http://www.ninds.nih.gov/.
118. J. Rosen, J. Gottfries, S. Muresan, A. Backlund and T. I. Oprea, *J. Med. Chem.*, 2009, **52**, 1953.
119. P. Caffrey, J. F. Aparicio, F. Malpartida and S. B. Zotchev, *Curr. Top. Med. Chem.*, 2008, **8**, 639.

View Online

Published on 09 December 2014 on http://pubs.rsc.org | doi:10.1039/9781782621836-00032
08:36:02.

120. G. L. Thomas, E. E. Wyatt and D. R. Spring, *Curr. Op. Drug Disc. Dev.*, 2006, **9**, 700.
121. C. J. O'Connor, H. S. G. Beckmann and D. R. Spring, *Chem. Soc. Rev.*, 2012, **41**, 4444.
122. M. Baker, *Nat. Rev. Drug Discovery*, 2013, **12**, 5.
123. P. T. Corbett, J. Leclaire, L. Vial, K. R. West, J.-L. Wietor, J. K. M. Sanders and S. Otto, *Chem. Rev.*, 2006, **106**, 3652.
124. M. S. Congreve, D. J. Davis, L. Devine, C. Granata, M. O'Reilly, P. G. Wyatt and H. Jhoti, *Angew. Chem. Int. Ed.Engl.*, 2003, **42**, 4479.
125. M. T. Cancilla, M. M. He, N. Viswanathan, R. L. Simmons, M. Taylor, A. D. Fung, K. Cao and D. A. Erlanson, *Bioorg. Med. Chem. Lett.*, 2008, **18**, 3978.
126. M. Hochgurtel, R. Biesinger, H. Kroth, D. Piecha, M. W. Hofmann, S. Krause, O. Schaaf, C. Nicolau and A. V. Eliseev, *J. Med. Chem.*, 2003, **46**, 356.
127. E. R. Wood, L. Kuyper, K. G. Petrov, R. N. Hunter III, P. A. Harris and K. Lackey, *Biorg. Med. Chem. Lett.*, 2004, **14**, 953.
128. S. L. Posy, M. A. Hermsmeier, W. Vaccaro, K.-H. Ott, G. Todderud, J. S. Lippy, G. L. Trainor, D. A. Loughney and S. R. Johnson, *J. Med. Chem.*, 2011, **54**, 54.
129. C. J. Harris, R. D. Hill, D. W. Sheppard, M. J. Slater and P. F. W. Stouten, *Comb. Chem. High Throughput Screening*, 2011, **14**, 521.
130. L. R. Whitby and D. L. Boger, *Acc. Chem. Res.*, 2012, **45**, 1698.
131. J. G. Kettle, R. A. Ward and E. Griffin, *Med. Chem. Comm.*, 2010, **1**, 331.
132. S. C. Mayer, J. A. Butera, D. J. Diller, J. Dunlop, J. Ellingboe, K. Y. Fan, E. Kaftan, B. Mekonnen, D. Mobilio, J. Paslay, G. Tawa, D. Vasilyev and M. R. Bowlby, *Assay Drug Dev. Tech.*, 2010, **8**, 504.
133. N. Y. Mok and R. Brenk, *J. Chem. Inf. Model.*, 2011, **51**, 2449.
134. R. A. Houghten, C. Pinilla, M. A. Giulianotti, J. R. Appel, C. T. Dooley, A. Nefzi, J. M. Ostresh, Y. Yu, G. M. Maggiora, J. L. Medina-Franco, D. Brunner and J. Schneider, *J. Comb. Chem.*, 2008, **10**, 3.
135. A. B. Yongye, C. Pinilla, J. L. Medina-Franco, M. A. Giulanotti, C. T. Dooley, J. R. Appel, A. Nefzi, T. Scior, R. A. Houghten and K. Martinez-Mayorga, *J. Mol. Model.*, 2011, **17**, 1473.
136. R. G. Santos, M. A. Giulianotti, C. T. Dooley, C. Pinilla, J. R. Appel and R. A. Houghten, *ACS Comb. Sci.*, 2011, **13**, 337.
137. L. Mannocci, M. Leimbacher, M. Wichert, J. Scheuermann and D. Neri, *Chem. Commun.*, 2011, **47**, 12747.
138. R. E. Kleiner, C. E. Dumelinz and D. R. Liu, *Chem. Soc. Rev.*, 2011, **40**, 5707.
139. M. A. Clark, R. A. Acharya, C. C. Arico-Muendel, S. L. Belyanskaya, D. R. Benjamin, N. R. Carlson, P. A. Centrella, C. H. Chiu, S. P. Creaser, J. W. Cuozzo, C. P. Davie, Y. Ding, G. J. Franklin, K. D. Franzen, M. L. Gefter, S. P. Hale, N. J. V. Hansen, D. I. Israel, J. Jiang, M. J. Kavarana, M. S. Kelley, C. S. Kollmann, F. Li, K. Lind, S. Mataruse, P. F. Medeiros, J. A. Messer, P. Myers, H. O'Keefe, M. C. Oliff, C. E. Rise, A. L. Satz, S. R. Skinner, J. L. Svendsen, L. Tang, K. van Vloten, R. W. Wagner, G. Yao, B. Zhao and B. A. Morgan, *Nat. Chem. Biol.*, 2009, **5**, 647.
140. H. Deng, H. O'Keefe, C. P. Davie, K. E. Lind, R. A. Acharya, G. J. Franklin, J. Larkin, R. Matico, M. Neeb, M. M. Thompson, T. Lohr, J. W. Gross, P. A. Centrella, G. K. O'Donovan, K. L. Bedard, K. Van Vloten, S. Mataruse, S. R. Skinner, S. L. Belyanskaya, T. Y. Carpenter, T. W. Shearer, M. A. Clark, J. W. Cuozzo, C. C. Arico-Muendal and B. A. Morgan, *J. Med. Chem.*, 2012, **55**, 7061.
141. J. G. Cumming, A. M. Davis, S. Muresan, M. Haeberlein and H. Chen, *Nat. Rev. Drug Discovery*, 2013, **12**, 948.