Diversity-oriented synthesis; a spectrum of approaches and results

Richard J. Spandl,[a] Andreas Bender[b] and David R. Spring*[a]

Received 20th December 2007
First published as an Advance Article on the web 19th February 2008
DOI: 10.1039/b719372f

Since our emerging area article, *diversity-oriented synthesis* (DOS), which aims to prepare efficiently collections of skeletally diverse small molecules, has developed in the synthetic approaches it employs. This article describes three general strategies, highlighting some successful examples. The utility of DOS, in the interrogation of chemical space and in the identification of novel biologically active lead compounds, is also discussed.

Introduction

Nature 'sees' molecules as three-dimensional surfaces of charges, polarities and other specific bonding interactions. In natural products these interactions are displayed on a multitude of complex and diverse molecular architectures. *Diversity-oriented synthesis* (DOS) aims to prepare collections of skeletally diverse small molecules to mimic this variety. As a result of its non-focused nature, a DOS library displays a wide range of physical and biological properties and, as such, can be useful in assays to identify novel lead compounds.

As it is still in its infancy, the potential of DOS, as with many new technologies or concepts, can be easily overstated. Consequently, the short term results may not meet initial expectations. Eventually, however, a 'rational' analysis phase occurs when the realities, including where DOS can best be applied, are appreciated better. Although not a comprehensive review, the purpose of this perspective article is to highlight the synthetic concepts employed in DOS, particularly examples since our emerging area article.[1] We also comment on its application to explore chemical space and, in so doing, to identify biologically active compounds.

[a]*Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, UK CB1 1EW. E-mail: drspring@ch.cam.ac.uk*
[b]*Leiden/Amsterdam Center for Drug Research, Leiden University, Gorlaeus Laboratories, 2300 Leiden, The Netherlands*

Small molecules and chemical genetics

As an alternative to traditional genetic approaches, *chemical genetics* has provided scientists with a set of complementary chemical tools to investigate biological systems.[2–5] Chemical genetics uses small molecules to perturb the function of gene products (*e.g.* proteins), thus facilitating the dissection of biological processes by chemical intervention.[3] In *forward chemical genetics* (Fig. 1) it is common for the lead compound (or compounds) identified in the screening processes to be novel in structure and hence not predictable prior to experimentation. The ultimate goal of chemical genetics is to identify small molecules that perturb the function of every gene product specifically; this is known as '*chemical genomics*'.[6] *Reverse chemical genetics* (Fig. 1) on a genome-wide scale would allow the systematic use of small molecules to explore biological systems.[4]

The enormous challenge faced by chemical genomics highlights one of the major drawbacks of the chemical intervention method; their lack of generality. For all of the approved therapeutic drugs, only 324 validated biological targets have been identified.[8] It is estimated that only 10% of the human genome (estimated 25 000 genes) encodes proteins that will bind drug-like compounds (the 'druggable genome');[9] however, only approximately a thousand of these have known chemical modulator partners.[10] Therefore, an enormous number of small molecules that perturb protein function specifically are still required. Chemical modulators can

Richard J. Spandl

Richard Spandl was born in 1982 and brought up near the village of Eye in Suffolk, England. He obtained his MChem in Biological Chemistry from the University of Leicester in 2004 before joining the Spring Group as a PhD student. Richard's PhD, which was funded by the BBSRC and Eli Lilly, has focussed on the development of strategies for diversity-oriented synthesis.

Andreas Bender

Andreas Bender is an Assistant Professor for Cheminformatics with the Leiden/Amsterdam Center for Drug Research. Before joining the LACDR, he was a Presidential Postdoctoral Fellow with the Novartis Institutes for Biomedical Research in Cambridge/MA. He earned his PhD from the University of Cambridge working with Prof. Robert C Glen at the Unilever Center for Molecular Science Informatics.

This journal is © The Royal Society of Chemistry 2008

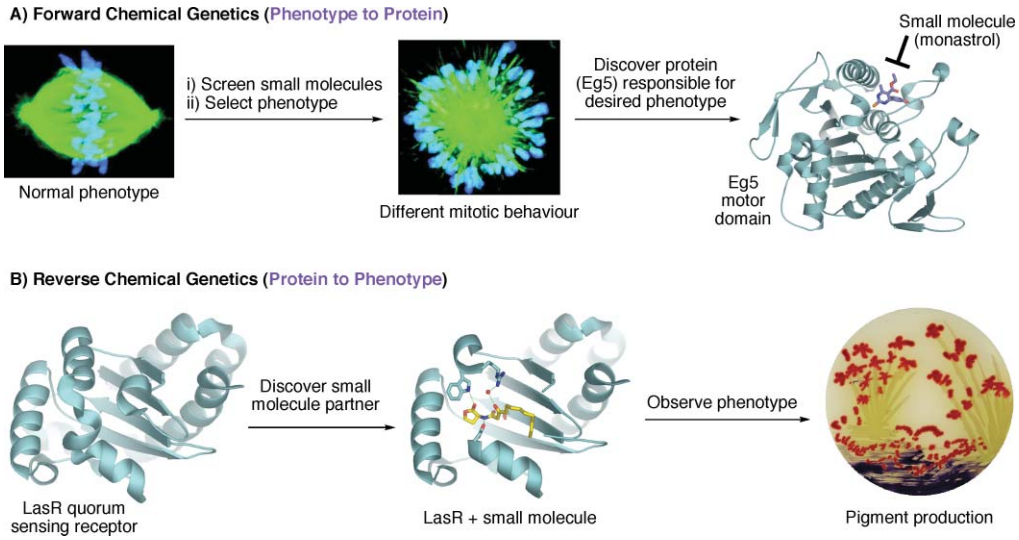


Fig. 1 **(A)** In forward chemical genetics a small molecule eliciting a desired phenotype (abnormal mitotic behaviour in this example; in the cell images tubulin is shown in green, DNA in blue) is identified; its protein partner is discovered subsequently. **(B)** In reverse chemical genetics, the phenotype resulting from protein modulation by its small molecule partner is observed. In this example, binding of an agonist (*e.g.* *N*-(3-oxododecanyl)-L-homoserine lactone) to a LuxR-type protein (*e.g.* LasR in *Pseudomonas aeruginosa*) in Gram-negative bacteria, activates transcription of a diverse range of processes (*e.g.* pigment production). In the image, pigments are produced under quorum sensing control by the bacteria *Serratia marsescens* (red), *Chromobacterium violaceum* (purple), and *P. aeruginosa* (light green).[7]

be identified by screening collections of structurally diverse small molecules. These collections can originate from nature or combinatorial chemistry campaigns (commercially available and proprietary). These sources do, however, have their short-comings (see below) and an alternative is *de novo* library preparation using diversity-oriented synthesis.[1,11]

Another major challenge for the chemical intervention method in biology is that of selectivity. Apart from identifying compounds with the desired activity, it is equally important that promising lead candidates do not exhibit too much promiscuity against other targets.[12] Here, especially in the case of a protein target with a large number of homolog members (such as kinases, proteases or phosphatases), compound selectivity is of major concern. Diversity-oriented synthesis is a tool to explore new areas of chemical space efficiently. It can be argued that more diverse starting points for lead optimization are more likely to lead to scaffolds showing a superior selectivity profile, compared to narrowly defined chemical libraries.

Small molecules and chemical space

Chemical genetics may benefit from access to collections of small molecules that are both structurally complex and diverse. Although there is debate in the literature,[13,14] it has been argued that structural complexity aids specificity in the interactions of chemical modulators with proteins.[10] Incorporating structural diversity into a collection may increase the chances of identifying novel lead compounds. This viewpoint is supported by evidence showing a direct correlation between the chemical space occupied by a collection of compounds and its functional (biological) diversity.[15] Methods of analysing and describing chemical space, therefore, may be useful in assessing the quality of a compound collection.[16,17]

Using computer algorithms, an abstract representation of a molecule can be constructed based on an analysis of its associated *chemical descriptors*.[16,18–20] These descriptors contain information regarding either the bulk properties of the compound[21] or its topological features.[22] Each molecule, therefore, resides at a discrete point in chemical space (more correctly known as multidimensional descriptor space), with the whole of chemical space being defined by the total descriptor space available to all molecules.[23] Thus, the more chemical space interrogated by a compound collection, the more structurally diverse the library.

In order to represent chemical space visually, and to aid the assessment of structural diversity, principle component analysis (PCA) can be used to condense a high-dimensional descriptor space into a representation that is accessible to human interpretation.[15] Recently, this concept has been used to asses the structural diversity of compound collections.[15,16,23,24] Interestingly, performing this type of analysis on any class of bioactive molecules demonstrates that they are not clustered in a discrete region of the chemical space occupied by known pharmacologically active compounds (MDL Drug Data Repository). For example, inhibitors of the cyclooxygenase-1 enzyme are shown in chemical space with the MDL Drug Data Repository in Fig. 2. Thus, the nature of the chemical space coverage displayed by these inhibitors supports the argument for screening skeletally diverse compound collections.

Although graphical representations generated by computational analysis are gratifying visually, they can be misleading. For example, analysing a collection of amides synthesized hypothetically from diverse commercially available amines and carboxylic acids can, in our experience, appear diverse when examined using certain chemical descriptors and PCA. The diversity generated is the result of the different building blocks used and not by virtue of the amide bond forming reaction. A more powerful

 This journal is © The Royal Society of Chemistry 2008

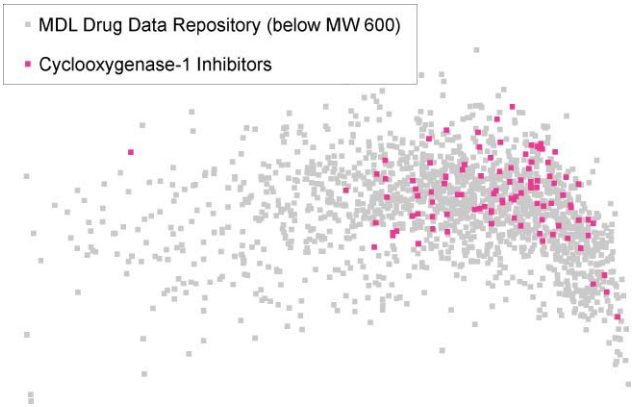


Fig. 2 Visual representation of the diversity of different chemical collections in chemical space. Cyclooxygenase-1 inhibitors (pink squares) are shown on a background of MDL Drug Data Repository compounds (grey squares).

technique to achieve diversity would be to combine this method of combinatorial building block variation with methods that also incorporate skeletal diversity.

This amide synthesis example underlines the difficulties of programming human intuition (here diversity assessment) into a computer; it is often difficult to make explicit what really constitutes diversity, and, furthermore, how to calculate it. Therefore, although useful, it should be remembered that diversity assessment is a very subjective process and depends on which chemical descriptors are used; it is the numerical values of those descriptors that are analysed and not the structures of the compounds themselves. More importantly, in the search for biological probes it is the functional diversity and not the structural diversity of the compound collection that is its measure of success. If a collection does not yield hits, no matter how structurally diverse, an experimenter will deem their research efforts as less than successful.

Although structural diversity is rarely the 'end-game' in a synthesis project, it is, nevertheless, an important consideration when the target molecule is unknown, as in forward chemical genetics. In these instances, libraries that interrogate larger areas of chemical space are useful, since a greater sample of the bioactive chemical universe increases the chance of identifying a compound with desired properties.[15]

Sources of small molecules

There are a number of potential sources of small molecule collections. Traditionally, nature has been a rich source of molecules that effect biological systems, many of which act on specific protein targets. Natural products, which are indeed complex and diverse in structure, have been used for centuries as medicines and have had a profound impact on human lives, but such compounds do have their disadvantages. For example, natural products may be isolated in low quantities and, due to the difficulties associated with purification and characterization, are sometimes screened as mixtures. Furthermore, the structural complexity of natural products makes chemical derivatization, a process especially relevant to drug discovery, extremely challenging.[25,26]

Commercially available combinatorial libraries and pharmaceutical proprietary compound collections are both alternative sources of small molecules. Although a traditional combinatorial library may offer complexity, it may show limited structural diversity by virtue of the 'one-synthesis/one-skeleton' approach in general use.[11] However, by combining many of these libraries together, a certain degree of chemical diversity (and complexity) can be achieved in practice, such as in the compound archives of large pharmaceutical companies, which typically comprise 1 million to 5 million compounds from different sources. Perhaps one drawback of pharmaceutical companies' compound collections is that they tend to be biased by the requirements of previous focused drug discovery programmes[27] or by meeting certain predefined criteria, *e.g.* the Lipinski rule of 5 (RO5).[28] Although these 'rules' are useful, there has been debate about restricting chemists to synthesising RO5 compounds,[29,30] and, furthermore, these rules are less applicable to biological probes than to drugs.[10] Extensive analysis of properties of natural products *vs.* drugs *vs.* combinatorial chemistry products has been performed, for example mean values for molecular weight (414 : 340 : 393), number of chiral centres (6.2 : 3.3 : 0.4), and number of rings (4.1 : 2.6 : 3.2) have been calculated from various databases.[31] These data highlight differences, but also surprising similarities.

Natural products and currently available compound collections occupy only a small proportion of bioactive chemical space.[10,23] So what do we do if we would like to exploit compounds from the unexplored areas of chemical space? The first step is to appreciate that the total number of possible 'drug-like' molecules is astronomic and unobtainable.[1] The second step is to think about how to synthesize molecules efficiently to interrogate wide areas of chemical space simultaneously. This is the aim of diversity-oriented synthesis.

Diversity-oriented synthesis and chemical space

DOS collections of small molecules interrogate larger areas of chemical space, by virtue of their structural diversity, compared with libraries produced using more traditional combinatorial chemistry. As a result, the functional (biological) diversity is greater, also.[15]

In contrast to *target oriented synthesis* (TOS), preparing a collection of compounds using DOS requires the development of a 'forward planning' algorithm to enable simple starting materials to be converted into products.[11] DOS, therefore, differs from TOS as, in the latter case, retrosynthetic analysis is used to plan a synthesis from a complex product to structurally simple building blocks (Fig. 3). As a result of their differing goals, the diversity generating potential of a DOS algorithm is far greater than traditional approaches using TOS. A TOS aims either to populate a discrete point in chemical space, *e.g.* a total synthesis, or to populate more densely a specific area of interest, *e.g.* a focused library synthesis. Conversely, a DOS aims to achieve a diverse and non-focused coverage of biologically active chemical space (Fig. 3, also see Scheme 2B).[32] However, as is the case in a TOS, a DOS project also requires highly efficient, high yielding and stereoselective reactions to be effective.[11] Since the term DOS is used freely in the literature, it is worth expanding on the concept of diversity.

By definition, when any collection of compounds is synthesized, since they are not identical, a degree of structural diversity

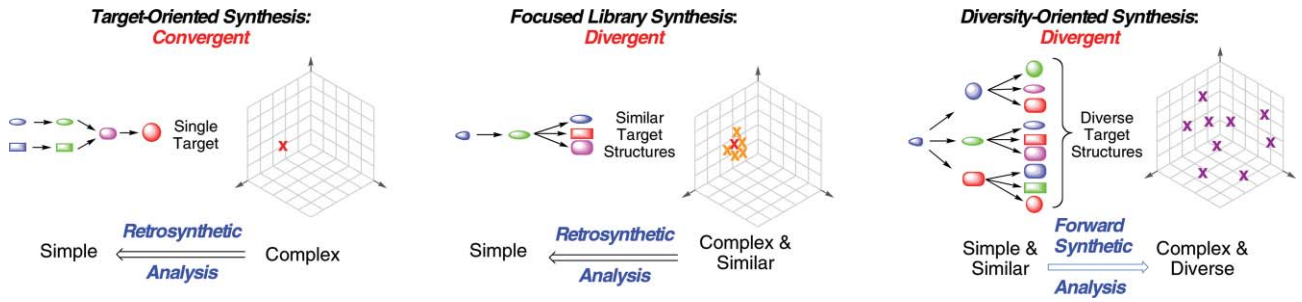


Fig. 3 TOS, focused library synthesis and DOS; a comparison of the planning strategies used (*i.e.* retrosynthetic or forward synthetic analysis and convergent or divergent synthesis) and the chemical space interrogated (*i.e.* a focused point/area or diverse coverage).

is incorporated. As an extension of this, in an extreme case, the racemic synthesis of enantiomers could be classified as a DOS. Clearly this is not our definition or understanding of the term. It may be useful, therefore, to consider structural diversity as a spectrum ranging from, in one extreme, the synthesis of a discrete target compound to, in the other extreme, a situation where maximal chemical space coverage is achieved (Fig. 4).

Although there are caveats associated with diversity assessment and a direct quantitative analysis is sometimes not possible, it should be the goal of a DOS to synthesize, in a qualitative sense, collections as near as possible to the right hand side of the '*molecular diversity spectrum*'. In the future, developments in computational approaches may lead to a better quantification of this spectrum.

Approaches to diversity-oriented synthesis

A successful DOS algorithm must address three principle types of diversity: substitutional (appendage) diversity, stereochemical diversity, and skeletal diversity.[11,33,34] Thus, the products of a DOS should not only be diverse in the appendages they display but also in the three-dimensional orientations of these appendages. The first of these can be achieved by combinatorial variation of building blocks; the second by use of stereocontrolled reactions. The most challenging facet of DOS, and of critical importance to its success, is the ability to incorporate skeletal diversity into a compound collection, *i.e.* the efficient generation of multiple molecular scaffolds from the same starting material.[34] This method is the most effective way of increasing structural diversity.[35] Before returning to consider some recent strategies used, the parallel approach of 'DOS based on privileged scaffolds' will be considered.

DOS around privileged scaffolds and biologically-oriented synthesis

Although there is a need to explore areas of chemical space not occupied by natural products and synthetic drugs,[1,15] basing a synthesis on so-called 'privileged' structures, *i.e.* those structural motifs common to bioactive molecules,[36–41] could be advantageous in some instances. The rationale behind this approach is that evolution over millions of years has made natural products, and hence compounds that resemble them structurally, more likely to exhibit bioactivity.[42,43] This approach has been cited as being distinct from the process of focused combinatorial library synthesis, for example in lead compound optimization, as efforts are directed toward identifying compounds with novel biological properties, discrete from those of the original privileged compound.[42] It has been argued that these strategies of so-called 'rational' diversity oriented synthesis are superior to 'DOS from simple starting materials'[36] in the identification of lead compounds for drug discovery.[43]

This approach is exemplified by the research of Park and co-workers who synthesized 22 unique core skeletons with an embedded privileged benzopyran motif **1**.[38] The discrete skeletons were accessed using two major branching pathways (Path **A** and Path **B**, Scheme 1) down which the substrates **2** and **3** could be channelled. Subjecting **2** and **3** to identical transformations gave access to the 11 core scaffolds **4–14** and hence 22 discrete skeletons (Scheme 1). As predicted, the privileged motif (**1**) conferred 'drug-likeness' to the compounds and a range of IC_{50} values (biological diversity) were reported against a human cancer cell line. More interestingly, the variations in bioactivity were shown to be a function of the molecular skeleton and not the appendages displayed. Further examples of 'natural product-like' libraries and

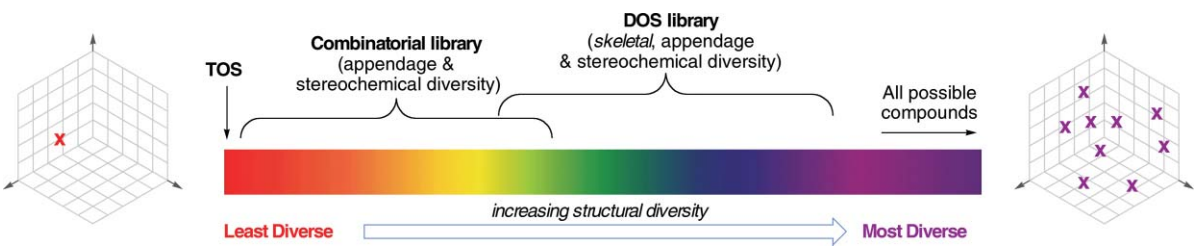
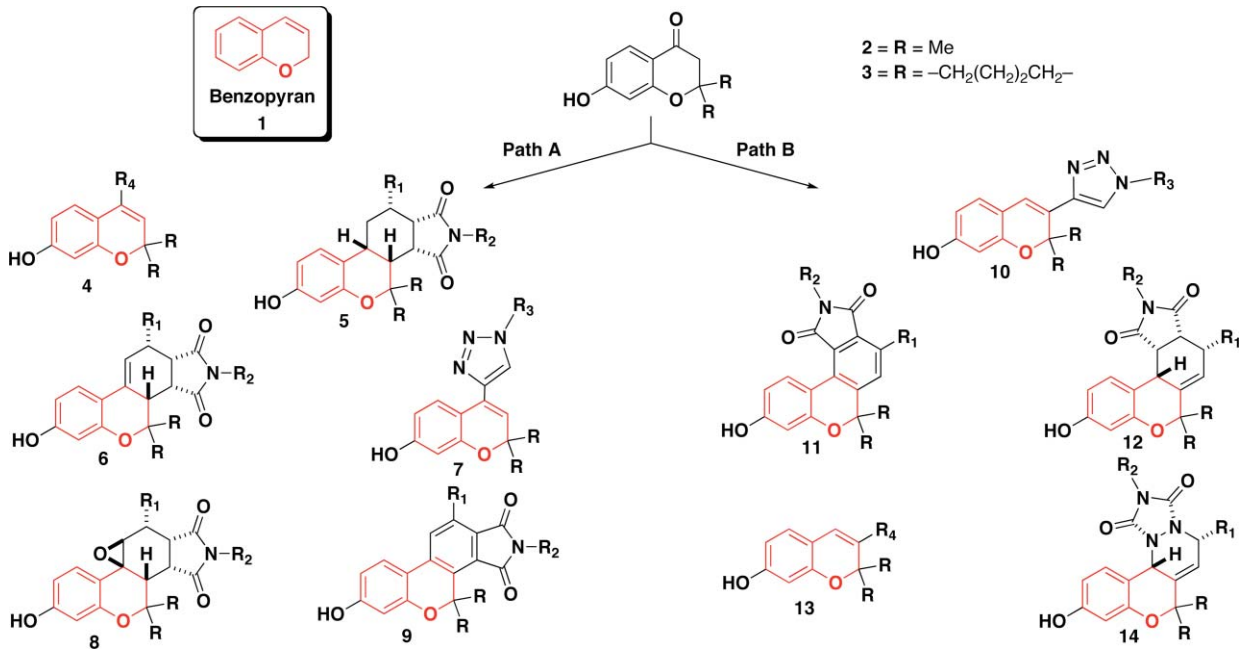


Fig. 4 The '*molecular diversity spectrum*'. In qualitative terms, diversity can be viewed as a spectrum ranging from a TOS to the synthesis of all possible molecular entities. Traditional combinatorial chemistry, where diversity primarily arises from building block variation, and DOS, where skeletal diversity is also incorporated, produce compound collections between these two extremes.

 This journal is © The Royal Society of Chemistry 2008



Scheme 1 Example of DOS around a privileged scaffold by Park and co-workers. By basing a DOS on the privileged benzopyran motif **1** (embedded in the starter units **2** and **3**), 11 distinct scaffolds could be generated using branching pathways. As a result of the variation in the R groups of **2** and **3**, 22 core skeletons were generated. These were found to display a range of IC_{50} values (biological diversity) against a cancer cell line.[38]

DOS around a privileged scaffold can be found in some recently published reviews and articles.[36,40,43]

An extension of the 'DOS around privileged structures' approach to library design uses the concepts of 'protein structure similarity clustering' (PSSC)[44] and the 'structural classification of natural products' (SCONP).[45] The SCONP represents biologically active scaffolds in a hierarchical manner; this hierarchy can be used to identify potential 'pre-validated' scaffolds. PSSC classifies proteins depending on the topology and inhibitory profile of their active-site folds; this can be used to predict the nature of likely small molecule inhibitors to a target protein. Thus, by using information from both of these methods of analysis, compounds have been designed in a hypothesis-driven fashion.[41] This relatively new approach of using the SCONP and PSSC has been termed '*biologically oriented synthesis*' (BIOS) by Waldmann and co-workers and has shown initial success.[41]

Clearly, when a specific protein target is in mind, BIOS and DOS around privileged scaffolds are advantageous approaches. When a less focused approach is required however, such as in a forward chemical genetics experiment, DOS from simple starting materials may be more appropriate.

DOS from simple starting materials

The remainder of this article will focus on DOS strategies used to create diverse libraries from simple starting materials; the process is described by Schreiber in much of his pioneering work in this field.[33,46]

Although there are a number of general strategies that can be used in DOS to incorporate diversity,[32] this article will focus on the use of branching pathways to access distinct molecular scaffolds. Branching pathways are of particular interest in DOS if they also serve to increase structural complexity. Complexity-generating reactions, when used in branching pathways, allow the generation of the complex three-dimensional scaffolds required for the specific interaction with a biomolecule.[10,47] Furthermore, when the product of one complexity-generating reaction is the substrate for the next (a tandem process), structural complexity and diversity are increased efficiently over a short number of steps.[11]

Skeletal diversity

Skeletal diversity can be achieved principally in two ways. The first involves the use of different reagents and a common starting material. This 'reagent-based approach' is also known as a branching pathway. Alternatively, in the 'substrate-based approach', different starting materials, containing suitably pre-encoded skeletal information, are subjected to a common set of conditions leading to different skeletal outcomes (Fig. 5).[34]

A review of the literature suggests successful DOS processes utilize these two approaches in a number of ways by either: (1) the use of a *pluripotent functionality*, where the same part of a molecule is subjected to different transformations induced by different reagents; (2) the use of a *densely functionalized molecule*, where different functionalities in the same molecule are transformed by different reagents (*i.e.* pairing different parts of the same densely functionalized molecule); or, (3) the use of a *folding process*, where different structurally encoding elements (σ), contained in different substrates, are subjected to the same reaction conditions (*i.e.* pairing same parts of different densely functionalized molecules). Examples of these will now be discussed. Strategies 1 and 2 represent reagent-based approaches and Strategy 3 represents the substrate-based approach.

Strategy 1: Pluripotent functional group strategy

In an analogy with stem cell differentiation, branching pathways involve typically the reaction of a pluripotent functionality, *i.e.*

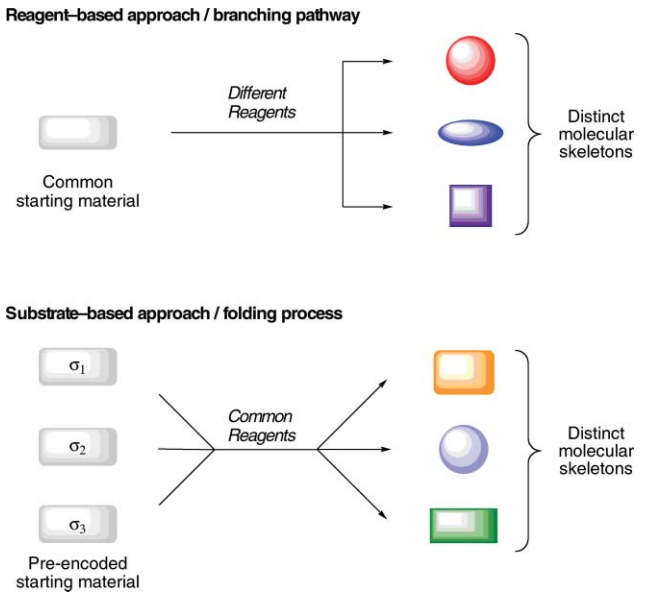


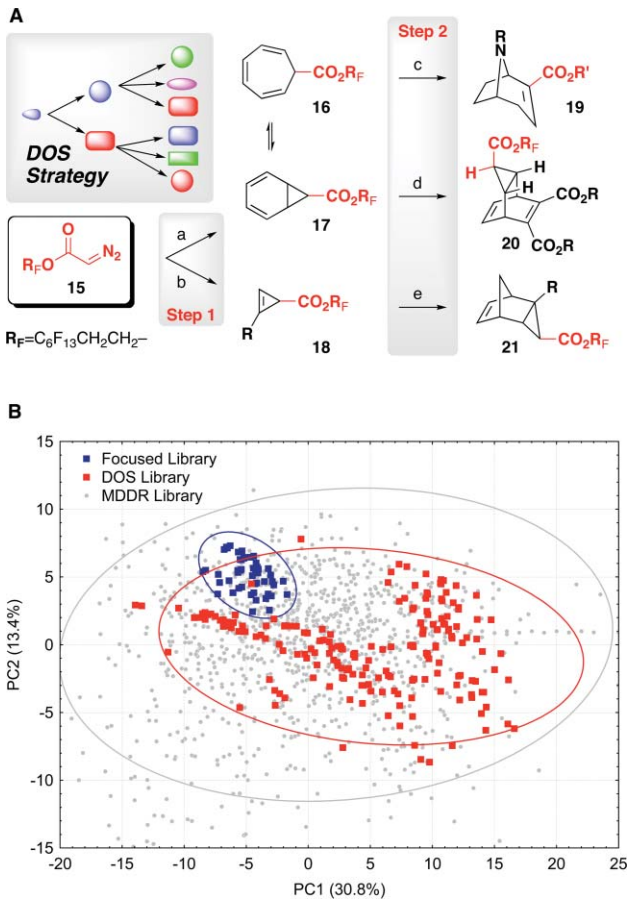
Fig. 5 Generalized methods for achieving skeletal diversity.

one that can participate in a number of different reactions to give discrete molecular scaffolds.[33] An example of this was published by Wyatt *et al.* who exploited the highly reactive and versatile diazo moiety.[35] The fluorous tagged diazoacetate **15** was utilized in divergent reactions to give a wide variety of molecular frameworks, including **16–21** (Scheme 2A). Using this algorithm, a diverse small molecule collection of 223 compounds, with 30 distinct molecular frameworks, was synthesized. The diversity of the library was assessed using molecular descriptors and PCA. The chemical space coverage was then compared with that of a focused library and also with that of known pharmacologically active small molecules (Scheme 2B). In addition to occupying chemical space common to the pharmacologically active compounds, the DOS library covered a larger area of chemical space than the focused library.[48] Normalized diversity on a per-compound basis is about 15 times as diverse for the DOS library[35] than for the focused library, while still only being half as diverse as the drug set derived from the MDDR. Given that the MDDR database comprises compounds from all kinds of sources, such as plant- and marine organism-derived natural products as well as drugs identified *via* HTS and compounds more similar to industry chemicals such as acetylsalicylic acid, it can be understood that it would be hard to mimic this degree of diversity using a single synthetic approach.

Multicomponent coupling reactions (MCR) feature regularly in DOS, since they bring together efficiently three or more building blocks. Unfortunately, they usually produce compounds of identical molecular architecture. This issue can be overcome by using a build–couple–pair sequence,[49] either by: a MCR to produce a densely functionalized molecule that can then be diversified further (Strategy 2);[50] or, by incorporating a 'folding process' into the MCR (Strategy 3).[51]

Strategy 2: Multiple group pairing strategy

Schreiber and co-workers synthesized the highly functionalized β-amino alcohol **22**, *via* the Petasis three-component coupling reaction of **23–25** followed by amine propargylation of the



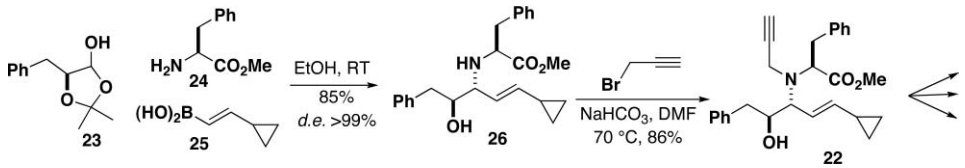
Scheme 2 **(A)** *Conditions:* (a) C_6H_6, $Rh_2(O_2CCF_3)_4$, 70%; (b) RCCH, $Rh_2(OAc)_4$, [BuCCH, 57%]; (c) RNH_2, NaOH then MeOH, H_2SO_4, [$MeNH_2$, 35%]; (d) dienophile [dimethyl acetylenedicarboxylate, 59%]; (e) C_5H_6, 92%; and R_F = $C_6F_{13}CH_2CH_2$– **(B)** Visual representation of the diversity of different chemical collections in physicochemical and topological space using MOE descriptors followed by principal component analysis (PCA). The DOS library is depicted as small red diamonds. For comparison, a focused library (small blue squares) and the MDL Drug Data Repository (small grey dots) are depicted.[35]

resulting compound **26**. Compound **22** was densely functionalized and displayed appendages at specific points; these acted as handles for further diversification (Scheme 3).[50]

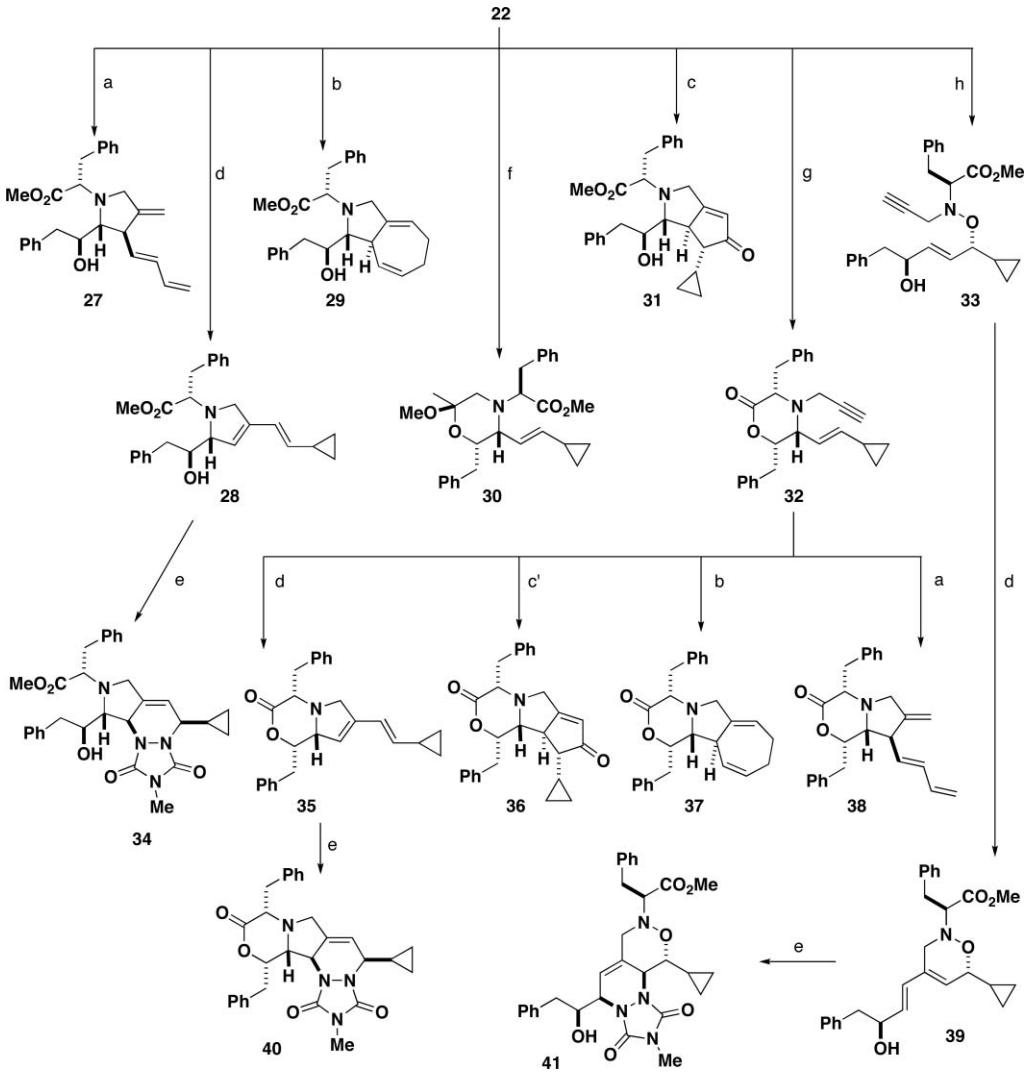
The stereochemical outcome of the Petasis reaction was substrate controlled (by the lactol **23**) and, therefore, allowed potentially the complete matrix of stereoisomers of **22** to be accessed (stereochemical diversity). Four of the functional groups present in **22**, *i.e.* the hydroxyl, the alkene, the alkyne and the cyclopropane moieties, were capable of participating in further complexity and diversity generating reactions (Scheme 4).

Initially, the template **22** was derivatized at each type of functionality affording **27–33**. In some cases the further reaction of selected skeletons furnished a second generation of compounds, *i.e.* **34–39**. For example, using conditions similar to those used in the diversification of **22**, the lactone **32** could be transformed into **35–38**. Also, reactions yielding 1,3-dienes, such as **28**, **35** and **39**, were used in tandem with a Diels–Alder reaction to increase diversity and yield the products **34**, **40** and **41**, respectively. The reaction scheme could be repeated using alternative amine building blocks in the Petasis reaction. In total, 15 different molecular

 This journal is © The Royal Society of Chemistry 2008



Scheme 3 The synthesis of the densely functionalized starting material **22** *via* a Petasis three-component coupling reaction and amine propargylation.[50]

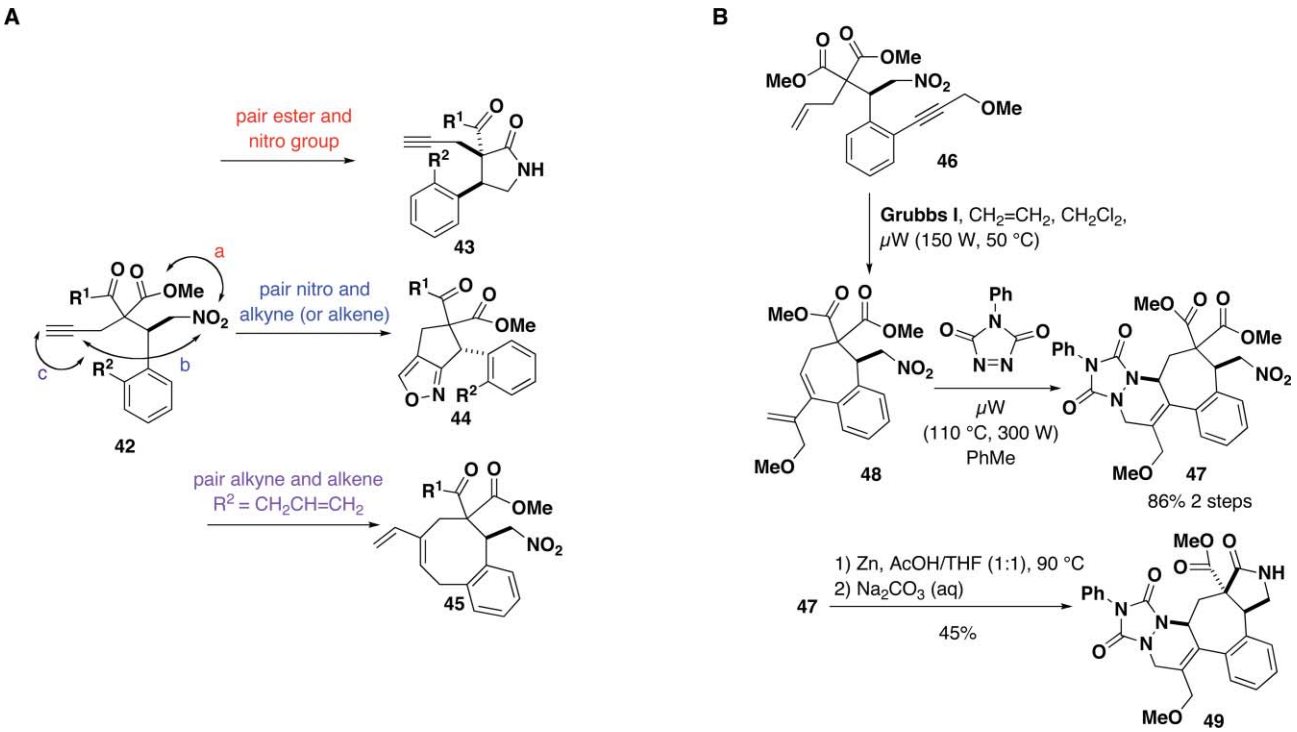


Scheme 4 *Conditions:* (a) $[Pd(PPh_3)_2(OAc)_2]$ (10 mol%), benzene, 80 °C; (b) $[CpRu(CH_3CN)_3PF_6]$ (10 mol%), acetone, RT; (c) $[Co_2(CO)_8]$, trimethylamine *N*-oxide, NH_4Cl, benzene, RT; (c′) $[Co_2(CO)_8]$, trimethylamine *N*-oxide, benzene, RT; (d) Hoveyda–Grubbs second-generation catalyst (10 mol%), CH_2Cl_2, reflux; (e) 4-methyl-1,2,4-triazoline-3,5-dione, CH_2Cl_2, RT; (f) $NaAuCl_4$ (10 mol%), MeOH, RT; (g) NaH, toluene, RT; (h) *m*CPBA, THF, −78 → 0 °C.[50]

skeletons were produced in this elegant example of incorporating complexity, rigidity and diversity (substitutional, stereochemical and skeletal) into a small molecule library.

Another example of the use of a densely functionalized substrate in a divergent branching pathway was reported by Porco and co-workers.[52] In their approach, substrate **42** was synthesized in an enantioselective 1,4-addition of a substituted dicarbonyl to a β-nitrostyrene. Using a similar concept to that reported by Schreiber and co-workers[50] different pairings of the pendant functional groups of **42** allowed for the synthesis of multiple scaffolds **43**–**45** (Scheme 5A). Further diversification was achieved by performing a Diels–Alder reaction with the 1,3-diene substrate **45** to yield another molecular skeleton.

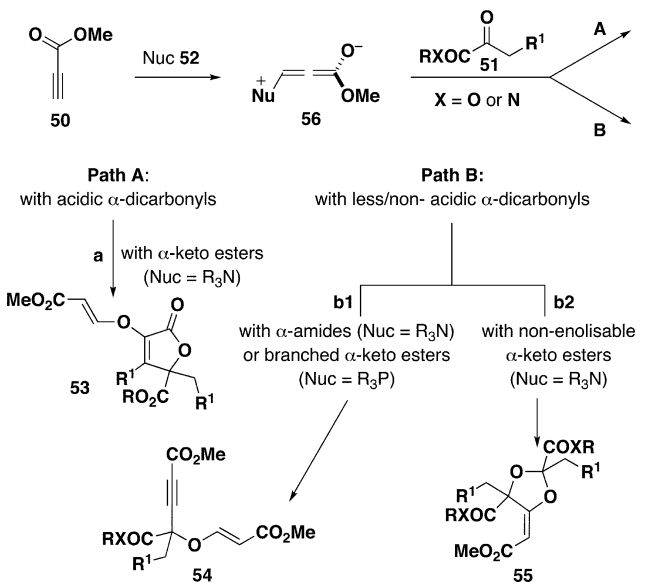
Where the folding reactions left the nitro and ester groups of the densely functionalized starting material untouched (*e.g.* **42** → **45**), their pairing in a later reaction sequence led to the construction of further molecular skeletons. For example, **46** was converted to the Diels–Alder adduct **47**, *via* the diene **48**, in an 82% yield over two steps. A further pairing reaction of the unaltered nitro and ester functionalities then yielded **49** (Scheme 5B).



Scheme 5 **(A)** The pairing reactions of **42** to give the skeletons **43–45**. **(B)** An example of the sequential use of pairing reactions to give **49**.[52]

Strategy 3: Folding pathway strategy

In a synthesis reported by García-Tellado and co-workers, an organocatalyzed ABB′ multicomponent coupling reaction between an alkyne **50** and an α-dicarbonyl compound **51**, in the presence of the nucleophilic catalyst **52**, was reported.[51] Pre-encoded structure determining elements in the substrates (both **50** and **51**), sometimes referred to as σ-factors,[33] determined the skeletal outcome of the folding process and gave the architectures **53–55** (Scheme 6).



Scheme 6 A summary of the ABB′ MCR used to synthesize **53–55**.[51]

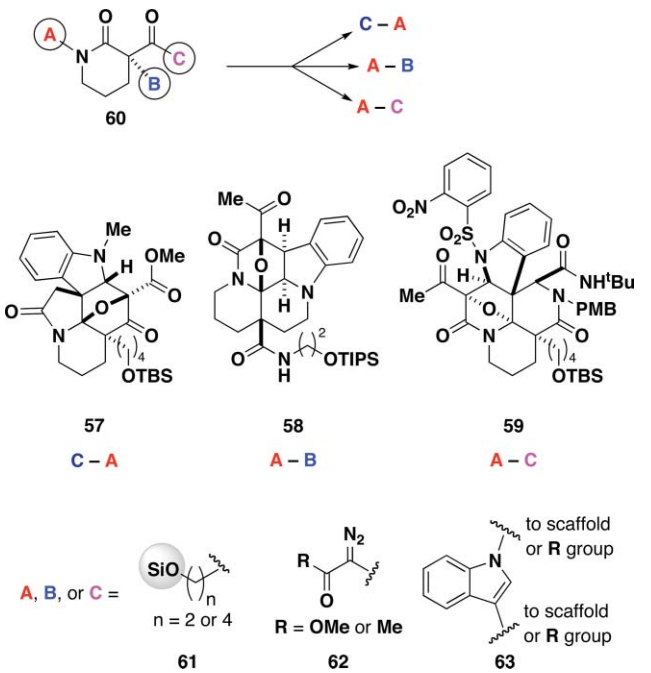
After the formation of the active allenoate species **56**, *via* the reaction of **50** with a nucleophilic catalyst **52**, initial chemodifferentiation resulted from the acidity of **51**. With acidic α-dicarbonyl compounds the reaction proceeded down path **A**, whereas path **B** was accessed when less/non acidic α-dicarbonyl compounds were used. The distinct skeletons **53–55** then resulted from three kinetically controlled reaction pathways, *i.e.* path **a**, path **b1** or path **b2**. In addition to being affected by the acidity of the substrates, the chemodifferentiation in this sequence was also dependent on the nucleophilicity of the catalyst **52** (either R_3N or R_3P) and the electrophilicity of the α-dicarbonyl compound **51**. The use of alternative ABB′ multicomponent coupling reactions has also been reviewed recently.[53]

Folding processes, similar to that described above, have been reported more commonly in DOS pathways as discrete steps.[33,54,55] An excellent example of this was reported by Oguri and Schreiber who were able to synthesize three distinct indole scaffolds **57–59**, under the same reaction conditions, from **60** (Scheme 7).[54] By displaying different combinations of a silyl ether linker **61**, an α-diazo ketocarbonyl group **62** and an indole moiety **63** at the sites A, B, and C of the common scaffold **60**, discrete folding pathways could be accessed. These reaction mechanisms initially involved a Rh(II) induced cyclization of **60**, which led to the formation of the carbonyl ylides **64**. The newly formed ylides **64** could then participate in intramolecular 1,3-dipolar cycloaddition reactions with the pendant indole groups, thus yielding **57–59** (Scheme 8).

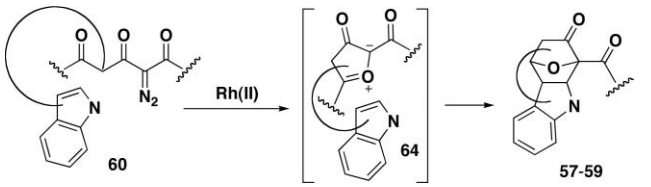
Conclusion

Over the last few years, novel and imaginative strategies have been used to prepare structurally diverse collections of small molecules by DOS. In many cases these collections have been exploited

 This journal is © The Royal Society of Chemistry 2008



Scheme 7 The use of folding pathways to produce the complex and diverse skeleton **57–59** from the starting material **60**. The relative locations of the moieties **61–63** determined the skeletal outcome of the reaction.[54]



Scheme 8 A mechanistic overview of the Rh(II) induced cyclization used to prepare **57–59** from **60**.

successfully to identify modulators for biological systems.[35,38,56–59] The primary aim of this review has been to highlight the different 'forward synthetic planning' strategies being used in DOS to achieve skeletal diversity. Populating diverse regions of chemical space using DOS still represents a significant, and potentially rewarding, challenge for organic chemists.

Acknowledgements

We thank EPSRC, BBSRC, MRC and Augustus and Harry Newman Foundation for financial support.

References

1 D. R. Spring, *Org. Biomol. Chem.*, 2003, **1**, 3867–3870.
2 S. L. Schreiber, *Bioorg. Med. Chem.*, 1998, **6**, 1127–1152.
3 D. R. Spring, *Chem. Soc. Rev.*, 2005, **34**, 472–482.
4 S. L. Schreiber, *Chem. Eng. News*, 2003, **81**, 51–61.
5 D. P. Walsh and Y. T. Chang, *Chem. Rev.*, 2006, **106**, 2476–2530.
6 G. MacBeath, *Genome Biol.*, 2001, **2**, 2005.2001–2005.2006.
7 M. Welch, H. Mikkelsen, J. E. Swatton, D. S. Smith, G. L. Thomas, F. G. Glansdorp and D. R. Spring, *Mol. BioSyst.*, 2005, **1**, 196–202.
8 The number of 324 validated biological targets is derived from a consensus set of small molecule drugs compiled from the FDA Orange Book, as well as biological drugs listed by the website of the Center for Biological Evaluation and Research. Those sources list 1204 distinct small molecule drugs as well as 166 distinct biologics, respectively. For 1065 of the total number of compounds a 'strong evidence of cell-based and/or *in vivo* evidence linking the target (and specific target sub-type) to the effect of the drug [. . .] alongside binding data' could be established, leading to a list of 324 distinct molecular targets with strong evidence of being implicated in the mechanism of action of those drugs. See: J. P. Overington, B. Al-Lazikani and A. L. Hopkins, *Nat. Rev. Drug Discovery*, 2006, **5**, 993–996.
9 A. L. Hopkins and C. R. Groom, *Nat. Rev. Drug Discovery*, 2002, **1**, 727–730.
10 C. Lipinski and A. Hopkins, *Nature*, 2004, **432**, 855–861.
11 M. D. Burke and S. L. Schreiber, *Angew. Chem., Int. Ed.*, 2004, **43**, 46–58.
12 K. Azzaoui, J. Hamon, B. Faller, S. Whitebread, E. Jacoby, A. Bender, J. L. Jenkins and L. Urban, *ChemMedChem*, 2007, **2**, 874–880.
13 M. M. Hann, A. R. Leach and G. Harper, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 856–864.
14 A. Schuffenhauer, N. Brown, P. Selzer, P. Ertl and E. Jacoby, *J. Chem. Inf. Model.*, 2006, **46**, 525–535.
15 S. J. Haggarty, *Curr. Opin. Chem. Biol.*, 2005, **9**, 296–303.
16 S. Fergus, A. Bender and D. R. Spring, *Curr. Opin. Chem. Biol.*, 2005, **9**, 304–309.
17 J. J. Perez, *Chem. Soc. Rev.*, 2005, **34**, 143–152.
18 L. Xue and J. Bajorath, *Comb. Chem. High Throughput Screening*, 2000, **3**, 363–372.
19 A. Bender and R. C. Glen, *Org. Biomol. Chem.*, 2004, **2**, 3204–3218.
20 R. D. Brown and Y. C. Martin, *J. Chem. Inf. Comput. Sci.*, 1996, **36**, 572–584.
21 G. M. Downs, P. Willett and W. Fisanick, *J. Chem. Inf. Comput. Sci.*, 1994, **34**, 1094–1102.
22 E. Estrada and E. Uriarte, *Curr. Med. Chem.*, 2001, **8**, 1573–1588.
23 C. M. Dobson, *Nature*, 2004, **432**, 824–828.
24 S. H. Fitzgerald, M. Sabat and M. Geysen, *J. Comb. Chem.*, 2007, **9**, 724–734.
25 M. S. Butler, *Nat. Prod. Rep.*, 2005, **22**, 162–195.
26 K. S. Lam, *Trends Microbiol.*, 2007, **15**, 279–289.
27 M. J. Valler and D. Green, *Drug Discovery Today*, 2000, **5**, 286–293.
28 C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv. Drug Delivery Rev.*, 1997, **23**, 3–25.
29 T. H. Keller, A. Pichota and Z. Yin, *Curr. Opin. Chem. Biol.*, 2006, **10**, 357–361.
30 C. Abad-Zapatero, *Drug Discovery Today*, 2007, **12**, 995–997.
31 M. Feher and J. M. Schmidt, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 218–227.
32 G. L. Thomas, E. E. Wyatt and D. R. Spring, *Curr. Opin. Drug Discovery Dev.*, 2006, **9**, 700–712.
33 M. D. Burke, E. M. Berger and S. L. Schreiber, *Science*, 2003, **302**, 613–618.
34 M. D. Burke, E. M. Berger and S. L. Schreiber, *J. Am. Chem. Soc.*, 2004, **126**, 14095–14104.
35 E. E. Wyatt, S. Fergus, W. Galloway, A. Bender, D. J. Fox, A. T. Plowright, A. S. Jessiman, M. Welch and D. R. Spring, *Chem. Commun.*, 2006, 3296–3298.
36 A. Reayi and P. Arya, *Curr. Opin. Chem. Biol.*, 2005, **9**, 240–247.
37 J. Clardy and C. Walsh, *Nature*, 2004, **432**, 829–837.
38 S. K. Ko, H. J. Jang, E. Kim and S. B. Park, *Chem. Commun.*, 2006, 2962–2964.
39 C. X. Zhou, A. V. Dubrovsky and R. C. Larock, *J. Org. Chem.*, 2006, **71**, 1626–1632.
40 R. Messer, C. A. Fuhrer, R. H. Iner and R. Haner, *Curr. Opin. Chem. Biol.*, 2005, **9**, 259–265.
41 A. Noren-Muller, I. Reis-Correa, H. Prinz, C. Rosenbaum, K. Saxena, H. J. Schwalbe, D. Vestweber, G. Cagna, S. Schunk, O. Schwarz, H. Schiewe and H. Waldmann, *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 10606–10611.
42 B. C. Goess, R. N. Hannoush, L. K. Chan, T. Kirchhausen and M. D. Shair, *J. Am. Chem. Soc.*, 2006, **128**, 5391–5403.
43 D. S. Tan, *Nat. Chem. Biol.*, 2005, **1**, 74–84.
44 F. J. Dekker, M. A. Koch and H. Waldmann, *Curr. Opin. Chem. Biol.*, 2005, **9**, 232–239.
45 M. A. Koch, A. Schuffenhauer, M. Scheck, S. Wetzel, M. Casaulta, A. Odermatt, P. Ertl and H. Waldmann, *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 17272–17277.
46 S. L. Schreiber, *Science*, 2000, **287**, 1964–1969.
47 N. Kumar, M. Kiuchi, J. A. Tallarico and S. L. Schreiber, *Org. Lett.*, 2005, **7**, 2535–2538.

48 A further example of this strategy has been published recently: G. L. Thomas, R. J. Spandl, F. G. Glansdorp, M. Welch, A. Bender, J. Cockfield, J. A. Lindsay, C. Bryant, D. F. J. Brown, O. Loiseleur, H. Rudyk, M. Ladlow and D. R. Spring, *Angew. Chem., Int. Ed.*, 2008, DOI: 10.1002/anie.200705415.
49 T. E. Nielsen and S. L. Schreiber, *Angew. Chem., Int. Ed.*, 2008, **47**, 48–56.
50 N. Kumagai, G. Muncipinto and S. L. Schreiber, *Angew. Chem., Int. Ed.*, 2006, **45**, 3635–3638.
51 D. Tejedor, A. Sántos-Expósito and F. García-Tellado, *Chem.–Eur. J.*, 2007, **13**, 1201–1209.
52 E. Comer, E. Rohan, L. Deng and J. A. Porco, *Org. Lett.*, 2007, **9**, 2123–2126.
53 D. Tejedor and F. García-Tellado, *Chem. Soc. Rev.*, 2007, **36**, 484–491.
54 H. Oguri and S. L. Schreiber, *Org. Lett.*, 2005, **7**, 47–50.
55 J. K. Sello, P. R. Andreana, D. S. Lee and S. L. Schreiber, *Org. Lett.*, 2003, **5**, 4125–4127.
56 Y. K. Kim, M. A. Arai, T. Arai, J. O. Lamenzo, E. F. Dean, N. Patterson, P. A. Clemons and S. L. Schreiber, *J. Am. Chem. Soc.*, 2004, **126**, 14740–14745.
57 A. N. Koehler, A. F. Shamji and S. L. Schreiber, *J. Am. Chem. Soc.*, 2003, **125**, 8420–8421.
58 D. R. Spring, S. Krishnan, H. E. Blackwell and S. L. Schreiber, *J. Am. Chem. Soc.*, 2002, **124**, 1354–1363.
59 F. G. Kuruvilla, A. F. Shamji, S. M. Sternson, P. J. Hergenrother and S. L. Schreiber, *Nature*, 2002, **416**, 653–657.

This journal is © The Royal Society of Chemistry 2008