Expert Opinion

informa
healthcare

Is synthesis the main hurdle for the generation of diversity in compound libraries for screening?

Warren RJD Galloway & David R Spring†
University of Cambridge, The University Chemical Laboratory, Lensfield Road, Cambridge, CB2 1EW, UK

Background: 'Diversity' is often cited as a crucial consideration when generating compound collections for biological screening. However, what exactly does one mean by 'diversity' and why is it important? *Objective*: How can diversity be incorporated into compound collections and what are the theoretical and technical challenges this poses? In this editorial, we comment on various factors involved in the creation of structurally, and most crucially, functionally diverse compound libraries. *Conclusions*: In particular, we highlight the central role played by organic synthesis and discuss the value of diversity-driven synthetic approaches in the search for new biologically active molecules with potentially exciting and unusual biological properties.

Keywords: diversity, diversity-oriented synthesis, drug discovery, library synthesis

Expert Opin. Drug Discov. (2009) 4(5):467-472

1. Introduction

The screening of compound libraries to identify useful modulators of biological systems is fundamental to the drug discovery process and chemical biology studies in general. However, a crucial consideration is what compounds to use? By the late 1980s, a strong belief had emerged in the pharmaceutical industry that drug discovery was simply a numbers game [1]. The development of combinatorial chemistry strategies allowed companies to generate libraries of hundreds of thousands of different compounds in a rapid manner at comparably low cost [2]. The assumption was that a multitude of drug leads would emerge simply as a consequence of the sheer volume of molecules available for screening. However, the expected surge in productivity has not materialised [3]. This disappointing degree of success is generally attributed to defects in the nature of the libraries produced, which have been described as being intrinsically useless for drug discovery [4]. Indeed, a general consensus has emerged over the past decade that library size is not everything; library diversity, in terms of molecular structure and, more importantly function, is a crucial consideration [5-7].

2. Biological diversity and structural diversity

A functionally diverse library contains compounds displaying a broad range of biological activities. Such collections are especially valuable in 'unbiased' screening processes in which the precise biological target is unknown as it has been argued that a greater sample of the biologically active chemical universe (i.e., of all biologically active molecules) increases the chance of identifying a compound with the desired properties [8-10]. The 'ideal' library in this context is one of such high diversity that, for any given aspect of any biological process, there exists a library compound that can modulate that aspect [5]. The correlation between library functional diversity and the likelihood of identifying small molecule modulators for a broad

10.1517/17460440902916606 © 2009 Informa UK Ltd ISSN 1746-0441

range of biological targets in any screening process [11] is particularly important in modern chemical biology studies. The rapid development of genomics and proteomics approaches to drug discovery is expected to lead to an exponential increase in potential therapeutic targets, creating an ever-burgeoning demand on access to diverse chemical libraries [6].

The biological function of any molecule is intrinsically dependent on its molecular structure. Consequently, the overall functional diversity of a molecule library is directly correlated with its overall structural diversity, which in turn is proportional to the amount of chemical space the library occupies [9].

Although the term 'diversity' is to some degree a subjective one, there are four main components of structural diversity that have been consistently identified in the literature [12]:

1. Appendage diversity: variation in structural moieties around a common skeleton
2. Functional group diversity: variation in the functional groups present
3. Stereochemical diversity: variation in the orientation of potential macromolecule-interacting elements [13] and
4. Skeletal (scaffold) diversity: presence of many distinct molecular skeletons.

Variation in the molecular scaffolds present in the library (so-called scaffold diversity) is crucial, with small multiple scaffold libraries generally regarded as superior to large single-scaffold libraries in terms of bio-relevant diversity [8,14]. In addition to structural diversity, structural complexity is another characteristic that is important in compound libraries for screening as molecules that are structurally complex are more likely to interact with biology in a selective and specific manner [15].

3. Sources of small molecules

Broadly speaking, there are three distinct sources of small molecules for use in biological screens: i) naturally available molecules; ii) commercially available compound collections or iii) new compound collections created by chemical synthesis.

3.1 Natural products and traditional combinatorial chemistry

A multitude of natural products have proven to be useful as drugs or leads [3] and nature still represents a major source of innovative therapeutic agents for infectious diseases [16]. Natural products exhibit enormous structural diversity [2], including scaffold diversity. Unfortunately, there are several problems associated with using natural product compounds in screening experiments including difficulties with purification, biologically active component identification, chemical modification and analogue synthesis. Commercially available compound collections represent an important alternative source of molecules. In the past, such libraries were synthesised in a 'traditional' combinatorial fashion using a 'one-synthesis/one-skeleton' approach and, therefore, tended to show limited structural diversity. However, by combining many of these libraries together, a certain degree of chemical diversity can be achieved, such as in the compound archives of large pharmaceutical companies, which typically comprise several million compounds from different sources. It is worth noting that such corporate compound collections are, historically, heavily biased towards compounds that satisfy certain pre-defined criteria imposed by the confines of traditional medicinal chemistry-lead optimisation campaigns (e.g., Lipinski's 'rule of 5' criteria for orally bioavailable drugs) [15]. Consequently, such collections are typically focused around known biologically active chemical space (that is, the chemical space spanned by known drug molecules and biologically active natural products). Whilst this is, by definition, a fruitful region for the discovery of biologically useful molecules, it does potentially run the risk of omitting a vast number of biologically active small molecules present in unexplored regions of chemical space from any screening process [13]. Another disadvantage associated with the use of commercially available libraries in screening experiments is that such collections are already likely to have been thoroughly panned for biologically active constituents. This is particularly important from a business perspective owing to intellectual property complications that may result.

The problems associated with using natural products and 'traditional' commercially available combinatorial-type libraries in screening experiments have spurred the development of several different synthetic approaches for the *de novo* creation of small molecule collections. Most of these 'modern' library synthesis methods have abandoned the mass synthesis and screening dogma underpinning early combinatorial chemistry and instead seek to either identify and efficiently access areas of chemical space that have an enhanced probability of containing biologically active compounds or efficiently interrogate wide regions of chemical space simultaneously [17]. The former approach is exemplified by methods such as 'biologically-oriented synthesis' [18], 'biology-inspired synthesis' [17] and privileged structure synthesis [19] that seek to generate compound libraries based around the core structures of known biologically active molecules, typically natural product templates. It has been argued that such compound libraries should have a high degree of biological relevance owing to the fact that evolutionary pressure has 'pre-validated' natural products, and thus compounds that are structurally similar, to be able to modulate biological systems [17,20]. However, such biased methods inevitably generate compound collections with a relatively low degree of overall scaffold diversity; thus, only a relatively small region of total chemical space is covered, with a heavy emphasis towards known biologically active regions.

3.2 Diversity-driven approaches towards library synthesis

Biased synthesis methods are particularly relevant when a specific biological target is being considered. However, what if one wishes to access unexplored regions of chemical space? These areas may contain molecules with exciting, novel biological

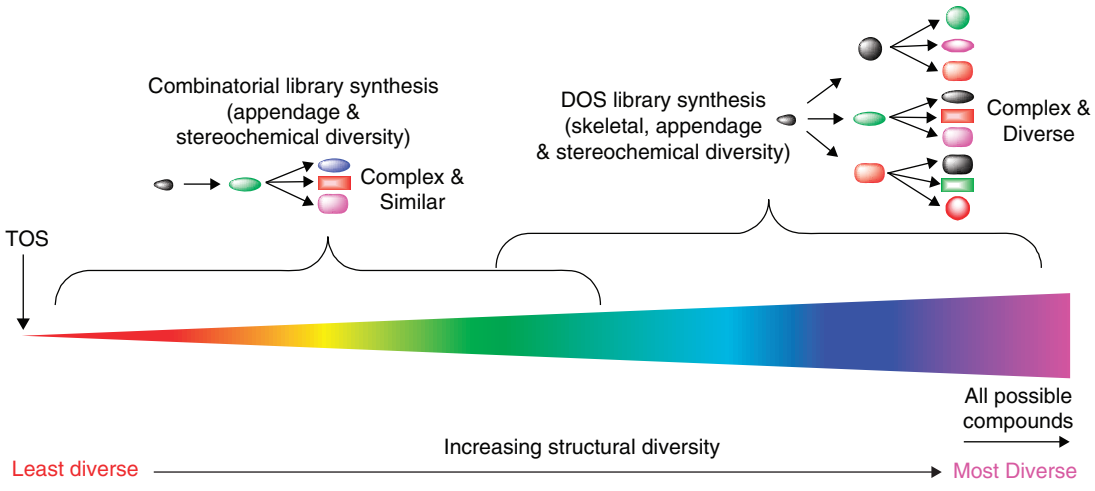

Figure 1. A comparison of the synthetic strategies used in a traditional combinatorial synthesis and a DOS.
DOS: Diversity-oriented synthesis; TOS: Target-oriented synthesis.

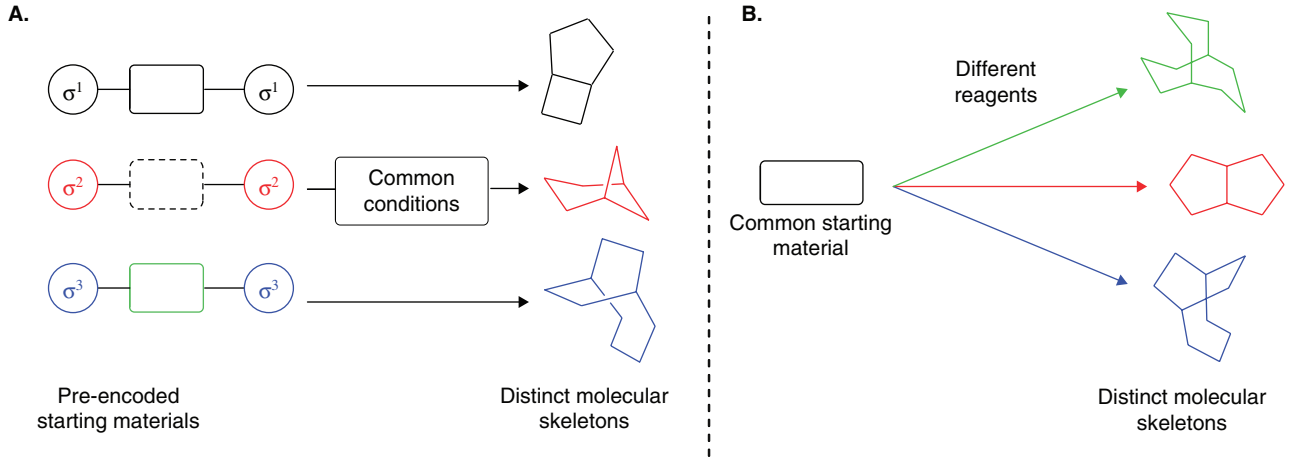

Figure 2. Substrate- and reagent-based approaches. A. An illustration of the substrate-based approach to skeletal diversity. A collection of substrates that are each based around a similar core skeleton but which carry different appendages (the so-called σ-elements, labelled σ^1–σ^3) is shown [13]. Under a common set of reaction conditions, each substrate is converted into a product having a different molecular skeleton, with the skeletal outcome dependent on the nature of the σ–elements present in the substrate; for example, 'red' σ^2 elements react in such a fashion so as to generate the molecular skeleton highlighted in red. Such methods are usually based around intramolecular folding reactions that 'pair' strategically positioned functional groups in the substrates, resulting in compounds with diverse skeletons [26]. **B.** The reagent-based approach involves a short series of divergent, complexity generating reactions from a common starting material to generate a collection of compounds with distinct molecular skeletons [13]. In practise, reagent-based skeletal diversity is achieved through two main methods [13,21]: i) the use of a densely functionalised molecule in which different functionalities in the same molecule are transformed by different reagents (as illustrated) and ii) the use of a pluripotent functionality in which exposure of a given molecule to different reagents results in different reactions occurring at the same part (functional group) of the molecule.

properties, which interact with new target molecules or act through novel modes of action. In this context, a less focused approach is required and the use of non-biased, diversity-driven synthetic approaches, which aim to access a wider range of chemical descriptor space, may be more useful.

The synthesis of a molecular library that achieves this wide coverage of both known and unexplored regions of biologically active chemical space in an efficient fashion presents a formidable challenge to the synthetic chemist. Diversity-oriented synthesis (DOS) is an approach towards library synthesis that seeks to achieve this goal. DOS has been defined as the deliberate, simultaneous and efficient synthesis of more than one target compound in a diversity-driven approach [12]. The overall aim of a DOS is the generation of a library of structurally

complex and structurally diverse small molecules that simultaneously accesses as wide an area of biologically relevant chemical space as possible (**Figure 1**) [12,13,21].

A DOS synthetic pathway is analysed in the forward sense; a single, simple starting material is converted into a collection of structurally diverse small molecules in no more than five synthetic steps. There is a clear distinction between DOS and traditional combinatorial methods: DOS libraries are usually smaller but consist of molecules that are typically structurally more complex, have a greater variety of core structures (skeletons) and possess richer stereochemical variation [22]. The boundary between modern, more considered combinatorial methods and DOS is less clear-cut. Recently, the concept of the molecular diversity spectrum has been introduced as a useful qualitative means for comparing the structural diversity (and thus chemical space coverage) associated with a particular molecular collection [21]. 'Diversity' can be viewed as a spectrum ranging from a target-oriented synthesis of a specific molecule to the synthesis of all possible compounds (i.e., total chemical space coverage), a traditional combinatorial approach and a DOS produce compound collections that sit between these two extremes. A DOS should, therefore, aim to generate, in a qualitative sense, collections of small molecules that are as near as possible to the right hand side of this spectrum [21].

The efficient creation of skeletal diversity represents the most crucial, yet most challenging, facet of a DOS [21]. There are two principle approaches towards this goal, the reagent-based approach and the substrate-based approach (**Figure 2**). Reactions that are capable of rapidly assembling complex molecular skeletons and generating structural complexity, such as pericyclic, cascade and tandem reactions, are particularly valuable in a DOS context. Folding-type DOS processes typically exploit the remarkable utility of ring-closing metathesis to generate complex molecular frameworks from simple starting materials [23]. Diversity-driven approaches such as DOS typically produce libraries that are smaller in size than those resulting from combinatorial-type methods. Nevertheless, compound purification still represents a significant bottleneck in the library generation process. Towards this end, many DOS strategies have used phase-labelling purification techniques (for example, solid-phase synthesis and fluorous-based tags [24]) that provide rapid and generic methods for product isolation.

4. Conclusion

Recent years have witnessed a significant paradigm shift in the selection criteria for compound libraries used in biological screening experiments, with molecular diversity considerations playing an increasingly prominent role. The emergence of diversity-driven synthetic approaches that aim to generate collections of molecules capable of modulating a wide variety of biological processes may provide novel chemical probes for biological research and new drugs for therapeutic interventions, with increased frequency and decreased cost [5]. Indeed, there are numerous examples of novel, biologically useful small molecules that have been discovered through the screening of DOS libraries [25]. However, the challenges associated with the efficient *de novo* creation of functionally diverse compound libraries are formidable and continued advancement in both the design and technical aspects of such methods are required. In particular, future diversity-driven library syntheses should aim to more efficiently couple the generation of molecular complexity, which is required for target specificity, with the structural diversity (principally scaffold diversity) prerequisite for broad biological activity, such that little of the functionality desired in the final compounds needs to be present in the starting substrates.

5. Expert opinion

There is undoubtedly a value in diversity-driven approaches that aim to produce functionally diverse compound collections. The screening of such libraries offers the possibility of discovering novel molecules, with exciting and unusual biological properties, which have thus far escaped the attention of humans and perhaps even nature.

Although recent years have witnessed a surge in the development of strategies that aim to simultaneously access wide areas of chemical space, we feel that the field is still in its infancy. An important consideration is the degree of 'constraint' placed on such methods. Ultimately, the goal of any library screening process is to identify molecules that are capable of interacting with biological macromolecules in a useful fashion. As such, although diversity-driven syntheses typically aim to access a wide area of chemical space, there is still a certain degree of bias associated with the compounds produced; that is, the libraries are designed to access biologically relevant space. Consequently, a diversity-driven synthesis should not simply be a random generation of compounds but rather a carefully considered endeavour that aims to generate molecules with diverse molecular structures that are natural product-like and drug-like in terms of their capability to modulate biological systems. The structural constraints that this consideration imposes on molecules are not precisely known as the true boundaries of biologically relevant chemical space have yet to be defined (if indeed it is ever possible to do so). As such, the degree of 'biologically relevant bias' in a diversity-driven synthesis is dependent on a balance between risks and potential rewards. Making and screening molecules costs, both in terms of time and money. Therefore, given the prevailing economic climate, it is unsurprising that pharmaceutical companies, although recognising the need for incorporating some level of diversity in chemical collections, tend to favour the synthesis of molecules that satisfy existing criteria for being 'pharmaceutically reasonable' and, as such, generate libraries that are biased towards known biologically active chemical space [7]. Although this is (relatively speaking) a low-risk approach, there nevertheless remains the possibility that there exists an untold number of biologically useful compounds in unexplored regions of chemical space.

Furthermore, it is likely that the low-hanging fruit contained with the boundaries of known biologically active chemical space have already been picked and the focused investigation of this area, principally by pharmaceutical companies, means that the crowded intellectual property space is an ever-growing problem. Thus, the exploration of uncharted regions of chemical space may offer great rewards to anyone brave enough to attempt it: a potential source of new biologically active molecules with unusual modes of action acting on unexploited drug targets.

The widespread application of diversity-based approaches will require the development of synthesis strategies that more efficiently and specifically access known and unknown biologically relevant chemical space, rather than chemical space that cannot provide biologically useful molecules. In this sense, whilst synthesis may be considered the most important factor in generating diversity in libraries for screening, it is by no means the only consideration. A better understanding of the boundaries of biologically relevant chemical space, and thus our ability to predict the biological relevance of compound libraries, is also needed. This is by no means an easy task and requires a vastly improved understanding of the relationship between the structural features of molecular libraries and screening outcomes [5].

Declaration of interest

Warren Galloway has previously received PhD funding from the UK Engineering and Physical Sciences Research Council and Pfizer. David Spring receives funding from the UK Engineering and Physical Sciences Research Council, Biotechnology and Biological Sciences Research Council and Medical Research Council, as well as from GlaxoSmithKline, Pfizer, Eli Lilly, UCB, AstraZeneca and Bayer-Schering.

Bibliography

1. Exploiting Chemical Diversity for Drug Discovery. Edited by Bartlett P and Entzeroth M. Royal Society of Chemistry, Cambridge (UK), 2006
•• **Provides a useful discussion of the theoretical and operational considerations involved in exploiting chemical diversity for drug discovery.**
2. Schneider G, Grabowski K. Properties and Architecture of Drugs and Natural Products Revisited. Curr Chemical Biol 2007;1:115-27
3. Newman DJ, Cragg GM. Natural products as sources of new drugs over the last 25 years. J Nat Prod 2007;70:461-77
4. Rouhi AM. Rediscovering natural products. Chem Eng News 2003;81(41):77-8
5. Schreiber SL. Organic chemistry: Molecular diversity by design. Nature 2009;457:153-4
•• **A thought-provoking discussion of the importance of diversity in chemical libraries and how scaffold diversity can be achieved using a metathesis-based folding strategy.**
6. Le GT, Abbenante G, Becker B, et al. Molecular diversity through sugar scaffolds. Drug Discovery Today 2003;8:701-9
7. Lajiness M, Watson I. Dissimilarity-based approaches to compound acquisition. Curr Opin Chem Biol 2008;12:366-71
• **An interesting discussion of how diversity can be incorporated into drug discovery from an industrial perspective.**
8. Schawrz MK, Sauer WHB. Molecular Shape Diversity of Combinatorial Libraries: A Prerequisite for Broad Bioactivity. J Chem Inf Comput Sci 2003;43:987-1003
•• **Provides a detailed discussion of the relationship between structural (principally scaffold) diversity and functional diversity in molecular libraries.**
9. Haggarty SJ. The principle of complementarity: chemical versus biological space. Curr Opin Chem Biol 2005;9:296-303
10. Shelat AA, Guy RK. Scaffold composition and biological relevance of screening libraries. Nat Chem Biol 2007;3:442-6
11. Kennedy JP, Williams L, Bridges TM, et al. Application of Combinatorial Chemistry Science on Modern Drug Discovery. J Comb Chem 2008;10:345-54
12. Spring DR. Diversity-oriented synthesis; a challenge for synthetic chemists. Org Biomol Chem 2003;1:3867-70
13. Burke MD, Schreiber SL. A planning strategy for diversity-oriented synthesis. Angew Chem, Int Ed 2004;43:46-58
•• **A seminal discussion of various planning strategies for diversity-oriented synthesis.**
14. Burke MD, Berger EM, Schreiber SL. Generating diverse skeletons of small molecules combinatorially. Science 2003;302:613-8
15. Lipinski C, Hopkins A. Navigating chemical space for biology and medicine. Nature 2004;432:855-61
• **Provides a useful discussion of the relationship between total chemical space and biologically relevant chemical space.**
16. von Nussbaum F, Brands M, Hinzen B, et al. Antibacterial natural products in medicinal chemistry - Exodus or revival? Angew Chem, Int Ed 2006;45(31):5072-129
17. Kaiser M, Wetzel S, Kumar K, Waldmann H. Biology-inspired synthesis of compound libraries. Cell Mol Life Sci 2008: published online 15th January 2006, doi 10.1007/s00018-007-7492-1
18. Noren-Muller A, Reis-Correa I, Prinz H, et al. Discovery of protein phosphatase inhibitor classes by biology-oriented synthesis. Proc Nat Acad Sci USA 2006;103:10606-11
19. Desimone RW, Currie KS, Mitchell SA, et al. Privileged Structures: Applications in Drug Discovery. Comb Chem High Throughput Screening 2004;7:473-93
20. Koch MA, Schuffenhauer A, Scheck M, et al. Charting biologically relevant chemical space: A structural classification of natural products (SCONP). Proc Nat Acad Sci USA 2005;102:17272-7
21. Spandl RJ, Spring DR. Diversity-oriented synthesis; a spectrum of approaches and results. Org Biomol Chem 2008;6:1149-58
• **An up-to-date and detailed discussion of synthetic strategies used in diversity-oriented synthesis with numerous significant examples.**
22. Borman S. Rescuing Combichem. Chem Eng News: Science and Technology. 2004;82(40):32-40
23. Morton D, Leach S, Cordier C, et al. Synthesis of natural-product-like molecules with over eighty distinct

scaffolds. Angew Chem, Int Ed 2009;48:104-9

- **A library of compounds based around 84 distinct molecular scaffolds is prepared using a metathesis-based strategy; this represented an unprecedented advancement in the capability of synthetic chemists to generate structurally diverse and complex molecules in a rapid manner.**

24. Wyatt EE, Fergus S, Galloway WRJD, et al. Skeletal diversity construction via a branching synthetic strategy. Chem Commun 2006;3296-98

25. Wyatt EE, Galloway WRJD, Thomas GL, et al. Identification of an anti–MRSA dihydrofolate reductase inhibitor from a diversity–oriented synthesis. Chem Commun, 2008;4962-64

26. Nielsen TE, Schreiber SL. Diversity-oriented synthesis - Towards the optimal screening collection: A synthesis strategy. Angew Chem, Int Ed 2008;47:48-56

- **A discussion of some common synthetic elements used in a diversity-oriented synthesis context in which the focus is on the short synthesis of structurally complex and diverse small molecules for use in biological screens.**

Affiliation

Warren RJD Galloway & David R Spring†
†Author for correspondence
University of Cambridge,
The University Chemical Laboratory,
Lensfield Road,
Cambridge, CB2 1EW, UK
Tel: +44 0 1223 336498; Fax: +44 0 1223 336362;
E-mail: drspring@ch.cam.ac.uk