OBC
www.rsc.org/obc

Diversity-oriented synthesis; a challenge for synthetic chemists †

David R. Spring
Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, UK CB2 1EW. E-mail: drspring@ch.cam.ac.uk

Received 3rd September 2003, Accepted 24th September 2003
First published as an Advance Article on the web 13th October 2003

The efficient, simultaneous synthesis of structurally diverse compounds, better known as diversity-oriented synthesis (DOS), is not obvious, and remains a challenge to synthetic chemistry. This personal account details why DOS has such enormous implications for the discovery of small molecules with desired properties, such as catalysts, synthetic reagents, biological probes and new drugs. Also, I describe the evolution behind the current state-of-play of DOS.

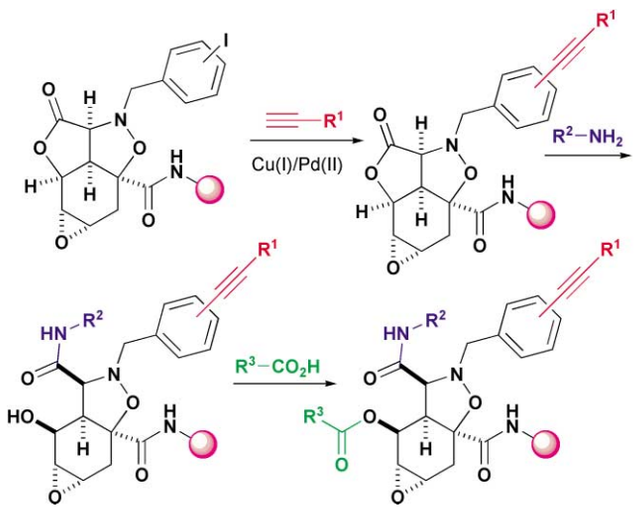
Introduction

Combinatorial chemistry allows for the synthesis of vast numbers of compounds; indeed, millions of compounds are realisable by one chemist on their own in a few weeks using split-pool combinatorial techniques.[1] The problem with combinatorial chemistry so far is that the compounds produced have a limited structural diversity. For example, a collection of over two million compounds was synthesised as shown in Scheme 1,[2] but structurally the compounds all look rather similar. This is because only building block diversity was introduced. The structural diversity of the products was only due to the building blocks and starting scaffold. The resulting molecular framework is the same in every case. In order to achieve the highest levels of structural diversity: (*i*) the building blocks, (*ii*) the stereochemistry, (*iii*) the functional groups and, most importantly, (*iv*) the molecular framework must be varied.

† Electronic supplementary information (ESI) available: Excel file of all the FDA new molecular entities between the years 1998 and July 2003, and new drug approvals between the years 1990 and 2002. See http://www.rsc.org/suppdata/ob/b3/b310752n/

David Spring was born in West Bromwich and attended Oxford University for his undergraduate chemistry degree, graduating in 1995. He stayed at Oxford under the supervision of Sir Jack Baldwin and received his DPhil in 1998 for work on the proposed biosynthesis of the manzamine alkaloids. David then spent two and a half years as a Wellcome Trust postdoctoral fellow and Fulbright scholar at Harvard University with Stuart Schreiber. In 2001 he returned to the UK as a BBSRC David Phillips Fellow and Fellow of Queens' College at Cambridge, where he is starting his own group researching diversity-oriented synthesis and chemical genetics.

David Spring



Scheme 1 Synthesis of over two million compounds.[2] The first structure shown represents three different spacers to the solid support (represented as a shaded sphere), both enantiomers and *ortho*-, *meta*- and *para*-iodo derivatives. The collection was constructed using 30 alkynes, 62 amines and 62 carboxylic acids (and skip codons).

Why do we need to synthesise structurally-diverse collections of compounds? Imagine screening compounds for a desired biological property, although equally you could be looking for a physical property. Compounds that look the same structurally often have a similar biological profile within a few orders of magnitude, although there are exceptions. Moreover, it is no use thinking we can just make everything, since the number of 'drug-like' molecules possible has been estimated to be astronomic (10^{62} to 10^{200}).[3,4] As a comparison there are approximately 10^{51} atoms on earth, so you cannot make every 'drug-like' molecule (let alone ones not considered 'drug-like'); in fact, you cannot come close. You must be selective. Fortunately, there is hope; there is more than one 'answer' to biological 'problems' (*e.g.* the HMG CoA reductase inhibitors: lovastatin, pravastatin, simvastatin, fluvastatin, atorvastatin, cerivastatin. . .) so we don't need to make and screen everything. In fact, biological activity is not a rare chemical property; the reality is that all small molecules are active biologically in some way or another, even ethanol. In terms of lead generation it is quality (structural diversity) and, but not just, quantity (number of compounds) that counts.

How do we synthesise structurally-diverse collections of small molecules? It is not obvious. Whereas, the synthesis of small molecules focussed around a lead structure (the target molecule) is relatively easy: diversify a scaffold with different building blocks. The efficient synthesis of structurally-diverse small molecules has been distinguished from target-

DOI: 10.1039/b310752n

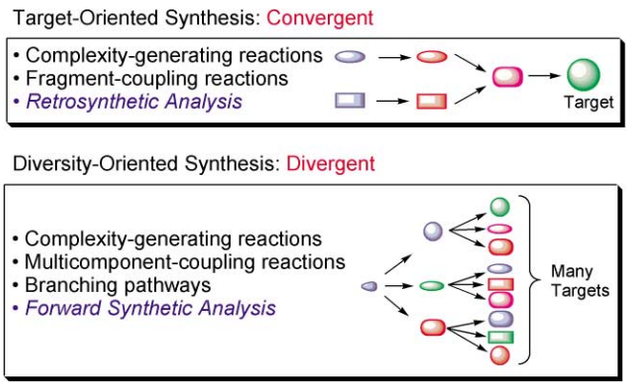


Fig. 1 Comparison of target-oriented synthesis (TOS) *versus* diversity-oriented synthesis (DOS). Note there is no necessity for a solid-support (*e.g.* on polystyrene beads) to perform diversity-oriented synthesis; however, solid-supported synthesis has the advantages of generic purification (filter and wash) and synthetic efficiency using split-pool strategies.[6] No specific meaning is implied by the colours or shapes except that each unit represents a different compound.

oriented synthesis (*e.g.* natural product synthesis and focussed 'library' synthesis) and termed *diversity-oriented synthesis* (Fig. 1).[5]

Diversity-Oriented Synthesis (DOS)

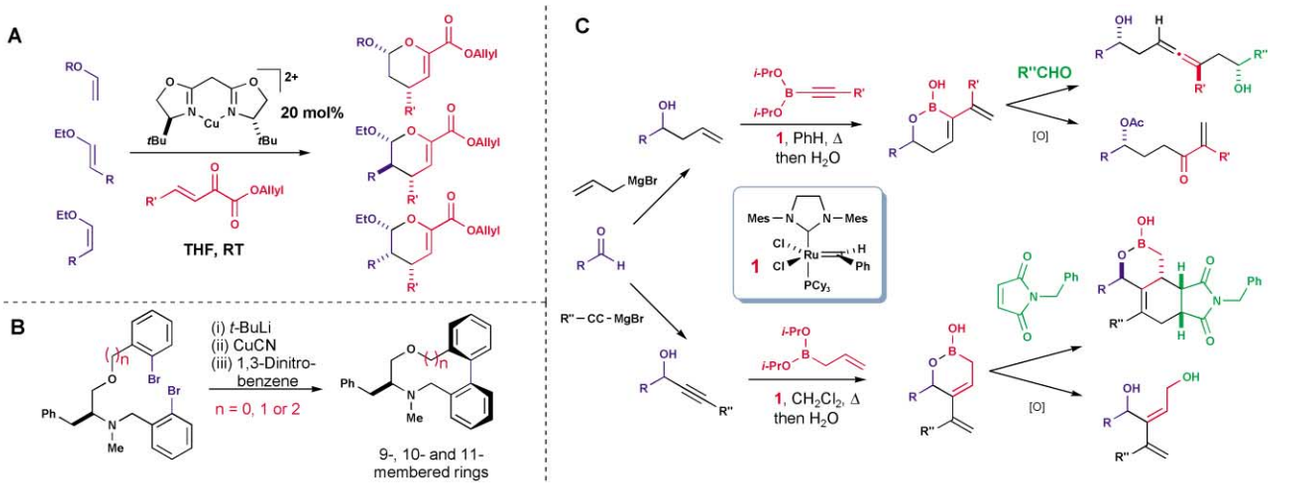
I have been fortunate enough to witness the evolution of DOS in the Schreiber group around the start of 2000. Chemistry subgroup meetings became brainstorming sessions as Stuart Schreiber encouraged us to help formulate forward planning algorithms for designing diversity-oriented small molecule collections, since retrosynthetic analysis was not judged directly relevant. Soon after the meetings, Stuart was seen ascending upstairs to discuss these ideas with the founder of retrosynthetic analysis E. J. Corey (who has the office above Stuart's). The present ideas for the emerging area of DOS are an accumulation of inputs from many people mostly associated with Harvard University ICCB (Institute of Chemistry and Cell Biology). Harvard now offers a full course on DOS, the first of its kind.[7]

There have been several definitions of DOS suggested, but in order to facilitate the present discussion the following definition will be adopted. "*Diversity-oriented synthesis involves the deliberate, simultaneous and efficient synthesis of more than one target compound in a diversity-driven approach to answer a complex problem.*" Complex problems in this context include binding, catalysis, phenotypic effects, *etc*. Most organic chemists are familiar with the problem of catalyst design to achieve enantioselectivity, and know the complications in predicting the most effective chiral ligands. In a hypothesis-based approach ligands would be improved by suggesting important steric and electronic effects, and new ligands would be synthesised and tested. Iterations would be undertaken modifying the ligand structure until results were satisfactory. In a diversity-driven approach (also known as a systematic-based approach) structurally diverse ligands would be screened and then promising results would be investigated as to why they were so successful. Often, both approaches are utilised when a good lead candidate has been found; however, before any ligand is known the likelihood of finding good enantioselectivity (an answer to the complex problem) is dependent either on the initial premise in a hypothesis-driven approach, or on the structural diversity in a diversity-driven approach. In practice, the boundaries are blurred and there is a spectrum between two extremes, because if the diversity-driven approach was favoured it would be folly to disregard a hypothesis such as the requirement for the ligands to be chiral!

When the problem becomes even more complex fewer hypotheses become relevant. Lead generation in drug discovery is a good example of this situation. Consider the problem of finding a new antibacterial drug candidate that has a novel mode of action. Certainly, there is a dire need for new antibacterials due to the ever-increasing problem of resistance to clinically used antibiotics. Which small molecule will work, what functional groups will be important? Often, drug companies impose hypotheses, such as avoiding poor absorption, by following the "Lipinski rule of five".[8] When looking for an antibacterial, which could be used topically or intravenously, such hypotheses become less relevant and a purely diversity-based approach becomes more judicious. The same is true in chemical genetics (biological investigation by modulating protein function with small molecules) where the aim is to understand biology.[9] Effective small molecules need not be potent particularly or have customized pharmacokinetic properties, as they are administered to cell culture rather than *in vivo*, and so the chance of finding a small molecule answer to the complex biological problem increases markedly. For example, just 1100 structurally diverse compounds screened against developing zebrafish identified small molecules that reproduced specifically a heart contractility phenotype and another blocked otolith formation.[10]

With a complex problem in mind how can a diversity-oriented synthesis be designed? In target-oriented synthesis retrosynthetic analysis is employed to find an efficient and convergent route using complexity-generating reactions (Fig. 1),[11] which construct efficiently structural complexity such as the Diels–Alder reaction, where two C–C bonds are made regioselectively (Alder rule), stereospecifically *syn*, stereoselectively (*endo vs. exo*) and enantioselectively (if a chiral mediator is exploited). DOS requires a planning algorithm to deliver an efficient but divergent route. Complexity-generating reactions are again important for efficiency (multicomponent-coupling, cascade and tandem complexity-generating reactions are the most valuable); however, pathways need to be identified that give structurally diverse targets. In order to design a synthetic pathway leading to a collection of compounds with different scaffolds requires the use of branch points, where a common substrate is used in different reactions that give different atomic skeletons. For example, nature takes acetyl CoA and makes terpenes, steroids, polyketides, *etc*., by branching pathways leading to each structural class. The synthesis of structurally diverse and complex collections of small molecules remains a major challenge to synthetic chemists.

As an example of how to design a diversity-oriented synthesis we need to bring complexity-generating reactions together with branching pathways. Ideally, we would like to make chiral compounds; hence, inclusion of stereochemistry is vital (Scheme 2A). Catalytic asymmetric reactions are most useful since the stereochemical outcome of the reaction is determined by the enantiomer of the catalyst added, whereas chiral auxilaries require two substrates to give both enantiomers. Cyclic, bicyclic and polycyclic compounds are often relatively rigid (*e.g.* steroids), which can minimise loss of conformational entropy on binding to a protein/reagent/substrate; however, cyclisation strategies need to be considered carefully, especially with medium and large ring sizes (Scheme 2B). Target-oriented synthesis has shown us that medium and large ring formation can be unpredictable, with subtle changes in substrate substituents, solvent and other conditions being important for reaction success. In diversity-oriented synthesis, methodology has to be used or developed that will work on a wide variety of substrates and be compatible with a wide range of functional groups. Thus, methodology development for DOS is more demanding than for just a total synthesis, for example, where often the method has to work with only one substrate. With methods for controlling stereochemistry and efficient, reliable, general synthetic methodology already



Scheme 2 Diversity-oriented synthesis strategies. **A**: Example of enantioselective catalysis in DOS. The copper bis(oxazoline) Lewis acid catalyses the inverse electron demand heterocycloaddition of a broad range of vinyl ethers and β,γ-unsaturated ketoesters with outstanding efficiency and selectivity.[12] **B**: Example of ring formation in DOS. A wide range of substituted acyclic precursors could be cyclised to give biaryl-containing medium rings efficiently and atropdiastereoselectively.[13] **C**: Example of branching pathways in DOS. Structurally-complex and diverse products are synthesised elegantly by annulation reactions of alcohols and boronic esters (transesterification then ring-closing ene-yne metathesis), followed by divisional, complexity-generating steps.[14]

determined, branching pathways are then conceived. Branch points are devised by choosing reactions that take the same substrate functional group to furnish different functionalities, stereochemistry and molecular frameworks (Scheme 2C). Building blocks are then chosen that contribute best to the structural diversity of products. For instance, if six aldehyde building blocks are required, then structurally diverse ones are chosen, *e.g.* acetaldehyde (small alkyl), trimethylacetaldehyde (large alkyl), benzaldehyde (aromatic), furfural (heteroaromatic), glucose (hydrophilic), and dodecanal (hydrophobic). Within just a few steps a single substrate can be modified into structurally-complex and structurally-diverse outcomes. The key to the structural complexity is the complexity-generating reactions, the key to the structural diversity is the branch points and building blocks.

Drug discovery and DOS

One important area that should benefit significantly from DOS is drug discovery. It is an exciting time since the human genome sequence has allowed the pharmaceutical industry to work on more than 3500 potential drug targets; previous to this only about 400 targets were pursued.[4] However, the pharmaceutical industry is also currently facing a major challenge. Two-thirds of the prescription drugs approved by the US Food and Drug Administration (FDA) between 1989 and 2002 were modified versions of existing medicines or identical to drugs already on the market. † Over the last five years the number of 'new molecular entities' (NME) approved for use as drugs by the FDA has steadily decreased (Fig. 2). In 2002, only 17 NME were approved, and of those only 7 were classified as being significant improvements over existing products.[15]

Applying the principles of DOS to the lead generation step in the drug discovery process should facilitate the discovery of NME. Lead generation can be approached broadly in two ways. The first approach looks for small molecules (drugs are most often organic, small molecules) that give you a desired phenotype (physiological outcome), such as toxicity to bacteria. This approach has been incredibly successful historically, identifying the clinical antibiotics vancomycin, penicillins, streptomycin, tetracyclines, erythromycin, sulfonamides, *etc*. The second approach starts with a target such as a protein, and looks for small molecules that modulate or attenuate its function. Both approaches require the availability of collections of small molecules. Small molecule collections come from four

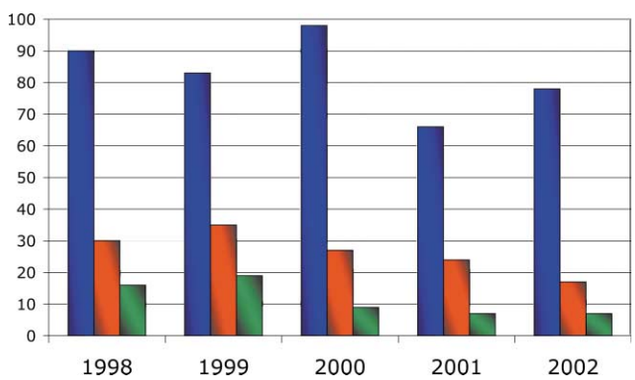


Fig. 2 Number of US Food and Drug Administration (FDA) new drug approvals (NDA, blue), new molecular entities (NME, red) and NME with significant improvements over existing products (green), between the years 1998 and 2002.

sources: existing in-house compounds,[16] commercial collections,[17] nature[18] or synthesis. The ideal collection would contain discrete (pure), structurally-diverse and structurally-complex small molecules (structural complexity should give improved selectivity). DOS has the potential to deliver this ideal collection of small molecules efficiently. Moreover, structure–activity relationships of identified 'hit' structures can be rapidly obtained with different building blocks, facilitating the 'hits to leads' process.

Future developments

What does the future hold for diversity-oriented synthesis? Since it is an emerging area future developments are hard to predict accurately; however, both developments in methodology and forward synthetic analysis will feature undoubtedly. Both will be illustrated by the syntheses of structurally-diverse compound collections. The increased availability of these collections should allow their exploitation in lead generation screening at the chemistry/biology and chemistry/materials interfaces; and this is where the scientific outcomes get really exciting. Novel reagents, reaction catalysts and smart devices should be available to discovery. Chemical genetics should increase our understanding of biological systems and move towards chemical genomics. Even new medicines should be discovered. In fact, the first drug discovery company to exploit this emerging field has been created already: Infinity Pharma-

ceuticals in Cambridge, Massachusetts.[19] More drug discovery companies are sure to follow and, soon, FDA approved new molecular entities that originated from diversity-oriented synthesis lead generation will be a reality.

Acknowledgements

M. D. Shair, C. Abell and M. J. Gaunt are thanked for their insightful comment and suggestions on this article, and I am grateful for financial support from the BBSRC.

References and notes

1 For a well described and cost effective introduction to combinatorial chemistry see: N. K. Terrett, *Combinatorial Chemistry*, Oxford University Press, 1998.
2 D. S. Tan, M. A. Foley, M. D. Shair and S. L. Schreiber, *J. Am. Chem. Soc.*, 1998, **120**, 8565–8566.
3 R. S. Bohacek, C. McMartin and W. C. Guida, *Med. Res. Rev.*, 1996, **16**, 3–50.
4 M. J. Owen, *Biotech Advantage*, 2002, **6**, 17. http://www.biotecplatform. gov.uk/newsletter/.
5 S. L. Schreiber, *Science*, 2000, **287**, 1964–1969.
6 A. Furka, F. Sebestye'n, M. Asgedom and G. Dibo', *Int. J. Pept. Protein Res.*, 1991, **37**, 487–493; K. S. Lam, S. E. Salmon, M. Hersh, E. J. Hruby, W. M. Kazmierski and R. J. Knapp, *Nature*, 1991, **354**, 82–84; R. A. Houghten, C. Pinilla, S. E. Blondelle, J. R. Appel, C. T. Dooley and J. H. Cuervo, *Nature*, 1991, **354**, 84–86.
7 http://www.courses.fas.harvard.edu/%7Echem117/index.html.
8 C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv. Drug Deliv. Rev.*, 1997, **23**, 3–25.
9 S. L. Schreiber, *Chem. Eng. News*, 2003, **81**, 51–61; R. S. Lokey, *Curr. Opin. Chem. Biol.*, 2003, **7**, 91–96; G. E. Ward, K. L. Carey and N. J. Westwood, *Cell. Microbiol.*, 2002, **4**, 471–482; G. MacBeath, *Genome Biol.*, 2001, **2**, 2005.1–2005.6; S. L. Schreiber, *Bioorg. Med. Chem.*, 1998, **6**, 1127–1152; T. J. Mitchison, *Chem. Biol.*, 1994, **1**, 3–6. http://www-schreiber.chem.harvard.edu; http://iccb.med.harvard.edu/.
10 R. T. Peterson, B. A. Link, J. E. Dowling and S. L. Schreiber, *Proc. Natl. Acad. Sci. U.S.A.,* **97**, 12965–12969.
11 E. J. Corey and X.-M. Cheng, *The Logic of Chemical Synthesis*, Wiley, New York, 1989.
12 R. A. Stavenger and S. L. Schreiber, *Angew. Chem., Int. Ed.*, 2001, **40**, 3417–3421.
13 D. R. Spring, S. Krishnan and S. L. Schreiber, *J. Am. Chem. Soc.*, 2000, **122**, 5656–5657; D. R. Spring, S. Krishnan, H. E. Blackwell and S. L. Schreiber, *J. Am. Chem. Soc.*, 2002, **124**, 1354–1363.
14 G. C. Micalizio and S. L. Schreiber, *Angew. Chem., Int. Ed.*, 2002, **41**, 3272–3276; G. C. Micalizio and S. L. Schreiber, *Angew. Chem., Int. Ed.*, 2002, **41**, 152–154.
15 There are many reasons for this trend, for example, the drug discovery process has become much more expensive ($231 million in 1990 to $897 million in 2000; a 388% increase); and although investment in pharmaceutical research and development has also increased dramatically, it has not increased as much as drug discovery costs ($10 billion in 1990 to >$25 billion in 2000; a 250% increase). The Tufts Center for the Study of Drug Development, News Release: 13 May 2003: http://csdd.tufts.edu/NewsEvents/ $231 million in 1990 is equivalent to $318 million in year 2000 dollars.
16 Large pharmaceutical companies have proprietary compound collections that they use for lead generation; many of the small molecules originate from previous medicinal chemistry projects and contain a high proportion of certain compound classes, sometimes called "privileged structures", such as benzodiazepines; thus, maximal structural diversity is inhibited.
17 Several companies (http://www.warr.com/ombichem.html#1) now sell discrete small molecule collections, and, although the molecules are structurally diverse, they tend to be low molecular weight (around 350 Da) and they rarely contain stereocentres.
18 Natural products are small molecules from the environment, synthesised by biosynthetic pathways and tend not to utilise as much chemical diversity available to a chemist, *e.g.* silicon atoms. Most natural products identified to date were isolated by assay-linked purification, singling out the most biologically active compounds. Natural products do show a high degree of structural diversity, but they are often mixtures (making it difficult to identify the active constituent), they are often isolated in low abundance, they are very costly to develop and they are often so structurally complex that chemical derivitisation is challenging synthetically.
19 See: http://www.ipi.com/.