ELSEVIER

Assessment of structural diversity in combinatorial synthesis

Suzanne Fergus, Andreas Bender and David R Spring

This article covers the combinatorial synthesis of small molecules with maximal structural diversity to generate a collection of pure compounds that are attractive for lead generation in a phenotypic, high-throughput screening approach. Nature synthesises diverse small molecules, but there are disadvantages with using natural product sources. The efficient chemical synthesis of structural diversity (and complexity) is the aim of diversity-oriented synthesis, and recent progress is reviewed. Specific highlights include a discussion of strategies to obtain structural diversity and an analysis of molecular descriptors used to classify compounds. The assessment of how successful one synthesis is versus another is subjective, therefore we test-drive software to assess structural diversity in combinatorial synthesis, which is freely available via a web interface.

Addresses
Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, CB2 1EW, UK

Corresponding author: Spring, David R (drspring@ch.cam.ac.uk)

Current Opinion in Chemical Biology 2005, **9**:304–309

This review comes from a themed issue on
Combinatorial chemistry
Edited by Hans-Jörg Roth and Prabhat Arya

Available online 8th April 2005

1367-5931/$ – see front matter

DOI 10.1016/j.cbpa.2005.03.004

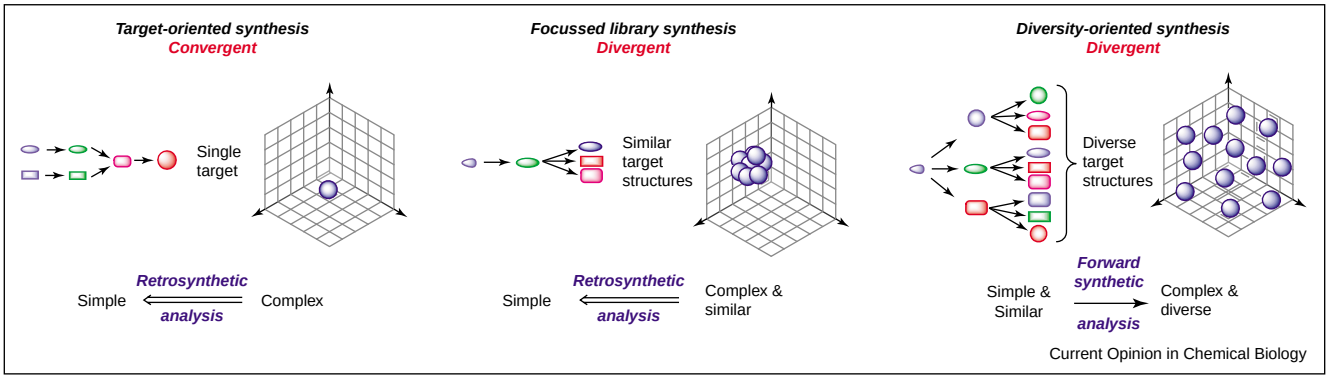
Introduction

There are occasions in research when a structurally diverse compound collection is just what is required. For instance, imagine being faced with the challenge of looking for a new antibacterial drug that selectively kills bacteria but not mammalian cells. The traditional approach would be to start with natural product extracts and screen them in antibacterial assays. Nature produces an astonishing array of structural diversity in secondary metabolites and, moreover, they are structurally complex too. Complex structures are likely to interact with biology more selectively than flat, simple molecules. Therefore, structural complexity is highly desirable because it is easy to kill cells unselectively. Unfortunately, there are disadvantages with using natural product extracts. Firstly, nature does not make secondary metabolites in a pure form; therefore, the extracts are usually screened as mixtures of many compounds, leaving the problem of purifying and identifying the active component(s). Secondly, the natural product extract may come from a limited source, leaving a supply problem. Thirdly, the active natural product may be so complex structurally, such as erythromycin, that making analogues to optimise activity is a formidable synthetic challenge. Fourthly, the chemistry space encompassed by natural products is unlikely to be the only region useful for discovering physical or biological properties of interest and, moreover, may not be the most productive region [1••]. These complications have led organic chemists to take the complementary approach of *synthesizing* structurally diverse and complex (natural-product-like) small molecules directly (Figure 1), an approach known as diversity-oriented synthesis (DOS) [1••,2,3]. Pure compounds can be synthesized for screening in any quantity and structure–activity relationship studies are inherently simpler to conduct, because DOS can be adapted easily to a focussed combinatorial synthesis of analogues [4,5•].

Drug-discovery companies often bias the small molecules that they make to fall within certain defined physical properties to increase the chances that their lead compounds will be orally bio-available [6]; however, DOS is also useful outside of drug discovery programmes. A lot can be leant from the effects of small molecules on biological systems, and this more academic application of DOS has been termed chemical genetics [7–9]. Small molecules have the potential to selectively modulate every function of every protein and this ultimate aim is known as chemical genomics [10]. Since the only requirement for screening small molecules in chemical genetic screens is that they must be soluble in the assay media, the structural diversity of useful compounds is vastly increased.

So there is a clear justification for the synthesis of structural diversity, but structural diversity is subjective. DOS is in itself a subjective expression; since, in an extreme example, the racemic synthesis of enantiomers could be said to be a diversity-oriented synthesis (although we do not endorse this) as the two products are not identical. By definition, the synthesis of a collection of compounds must incorporate structural diversity; therefore, DOS has been freely used in the literature. DOS is the deliberate, simultaneous and efficient synthesis of more than one target compound in a diversity-driven approach (i.e. a hypothesis-*generating* approach rather than a purely hypothesis-*based* approach) [3]. But how do you assess the *degree* of structural diversity created? Intuition? It is clear that a less subjective method of assessment is required.

Figure 1



Target-oriented synthesis (TOS) versus focussed library synthesis versus diversity-oriented synthesis (DOS) [32••]. DOS concerns the efficient synthesis of structurally diverse (and complex) small molecules (i.e. where the molecules differ in their (i) attached groups, (ii) stereochemistry, (iii) functional groups and (iv) molecular frameworks). TOS aims to synthesize a single target molecule, whereas a focussed library generates structurally similar target structures. Synthetic pathways in DOS are branched and divergent and the planning strategy extends simple and similar compounds to more complex and diverse compounds. Retrosynthetic analysis concepts focus on the existence of a defined target structure [33]. In DOS there is no single target structure and therefore retrosynthetic analysis cannot be used directly and a forward synthetic analysis algorithm is required. The three-dimensional grids of molecular descriptors illustrate the product(s) of the syntheses in chemical descriptor space (see later).

This commentary attempts to cover two issues: firstly, to review synthetic strategies used to obtain structural diversity; and, secondly, to assess the degree of structural diversity obtained in DOS libraries computationally. To assess structural diversity computationally we need to use molecular descriptors.

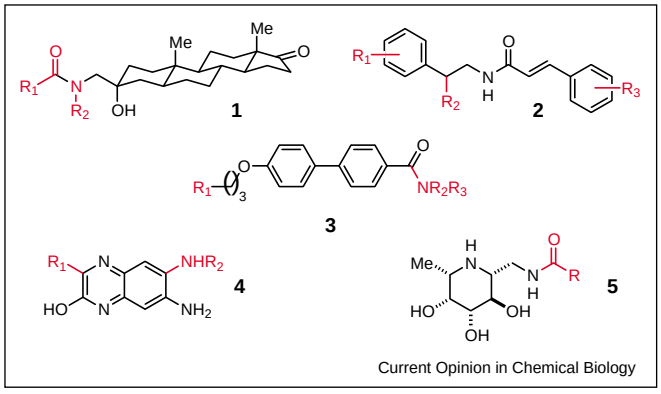
Strategies for structural diversity

Structural diversity is essential for lead generation, as compounds that look the same structurally are likely to share similar physical and biological properties. A collection of compounds with the highest level of structural diversity will consist of molecules that have different building blocks, stereochemistries, functional groups and molecular frameworks [3]. Consider a coupling reaction that involves a substrate, a building block (or more than one building block for multicomponent coupling reactions), and a reagent to give the product. In simple terms, strategies to generate structural diversity would involve varying the building block (appendage decoration), reagent (constitutional isomer generation, stereoisomer generation, divergent reaction pathways) or substrate (divergent folding pathways). The most successful syntheses of structural diversity incorporate multiple strategies.

Appendage decoration is the simplest diversity-generating processes and a central feature in combinatorial chemistry, particularly to improve the biological activity of a lead drug compound; it involves the use of coupling reactions to attach different building blocks to a common molecular framework. Many examples are available from the literature of this approach to combinatorial synthesis. For example, Maltais *et al.* synthesized new 3β-substituted androsterones (**1**, Figure 2), which are potential cancer chemotherapeutics, by varying amine and acid chloride building blocks [11]. Hergenrother and co-workers identified a small molecule that selectively induces apoptosis in cancer cells from a focussed combinatorial library of *N*-acylated aromatic amines (**2**) based on a natural product from *Isodon excisus* [12•]. Researchers from Abbott functionalised a common scaffold with amines at two positions to generate selective histamine H_3 receptor antagonists (**3**) [13]. Zhang *et al.* synthesized the first 2-quinoxalinols library (**4**) and tested the compounds for their inhibition of mouse macrophage cytokine response [14]. If only appendage decoration is used then all the products will have the same molecular frameworks, which is ideal if a focussed library is required. Nevertheless, if a very diverse range of building blocks is used then, although the scaffold is the same, the overall structural diversity is high. For example, Wu *et al.* have discovered selective α–fucosidase inhibitors by attaching 60 structurally diverse carboxylic acids to **5** [15]. To generate an even greater degree of structural diversity in the scaffold, other strategies need to be incorporated into the synthesis too.

Constitutional isomer generation involves using chemoselective and/or regioselective reactions to synthesize different product isomers. Stereoisomer generation involves using reactions that proceed with diastereoselectivity and/or enantioselectivity. Divergent reaction pathways are a very effective way of generating structural diversity, in particular, diverse molecular frameworks and functional groups. Skeletal diversity is generated by using different reagents to transform a common substrate into a collection of products having varied molecular skeletons.

Figure 2



Combinatorial libraries using the appendage decoration strategy. Libraries of (**1**) 3β-substituted androsterones [11], (**2**) *N*-acylated aromatic amines [12•], (**3**) histamine H_3 receptor antagonists [13], (**4**) 2-quinoxalinols [14] and (**5**) α-fucosidase inhibitors [15] are illustrated.

Divergent folding pathways utilizes substrates with different appendages that pre-encode skeletal information into a collection of products having distinct molecular skeletons using common reaction conditions. Most DOS libraries use several strategies to generate structural diversity.

For example, Schreiber and co-workers cyclized ('folded') amino alcohols functionalised with building blocks (*ortho*-bromobenzaldehydes and *ortho*-bromobenzyl bromides) to give the kinetic atropisomer **6**, the thermodynamic atropisomer **7** could be accessed selectively by heat (Figure 3). In addition, divergent reactions accessed acyclic analogues [16]. Diastereomers are usually accessed by diastereoselective reactions, but Itami *et al.* have illustrated the synthesis of *E*- and *Z*-isomers, along with other analogues of tamoxifen **8**, by altering the order of building block addition [17]. Different stereoisomer scaffold fragments were combined to form a diverse range of macrolactone molecular frameworks [18]. Sello *et al.* have exemplified the use of stereochemistry from building blocks to pre-encode skeletal information. Substrates were synthesized that could undergo either a ROM/RCM or an RCM reaction (where ROM is ring-opening metathesis and RCM is ring-closing metathesis) using the same reagent and conditions. The outcome of the pathway selected for each substrate (**9**) was determined by the stereochemistry of a single substrate substituent R* [19]. Alternatively, Oguri *et al.* have elegantly demonstrated that six structurally diverse indole alkaloid-like frameworks can be generated by shifting the relevant functionality around three points on a starting scaffold. A rhodium-catalysed tandem cyclization-cycloaddition reaction was used to efficiently generate distinct frameworks (**10** and **11**) with complete diastereocontrol [20•]. Divergent reaction pathways have been successfully demonstrated with the dihydroisoquinoline [21] and dicyclopropylamino alkene [22]. Burke *et al.* have elegantly combined divergent reactions of furaldehydes that pre-encoded divergent folding pathways via the Achmatowicz reaction and related transformations [23•,24].

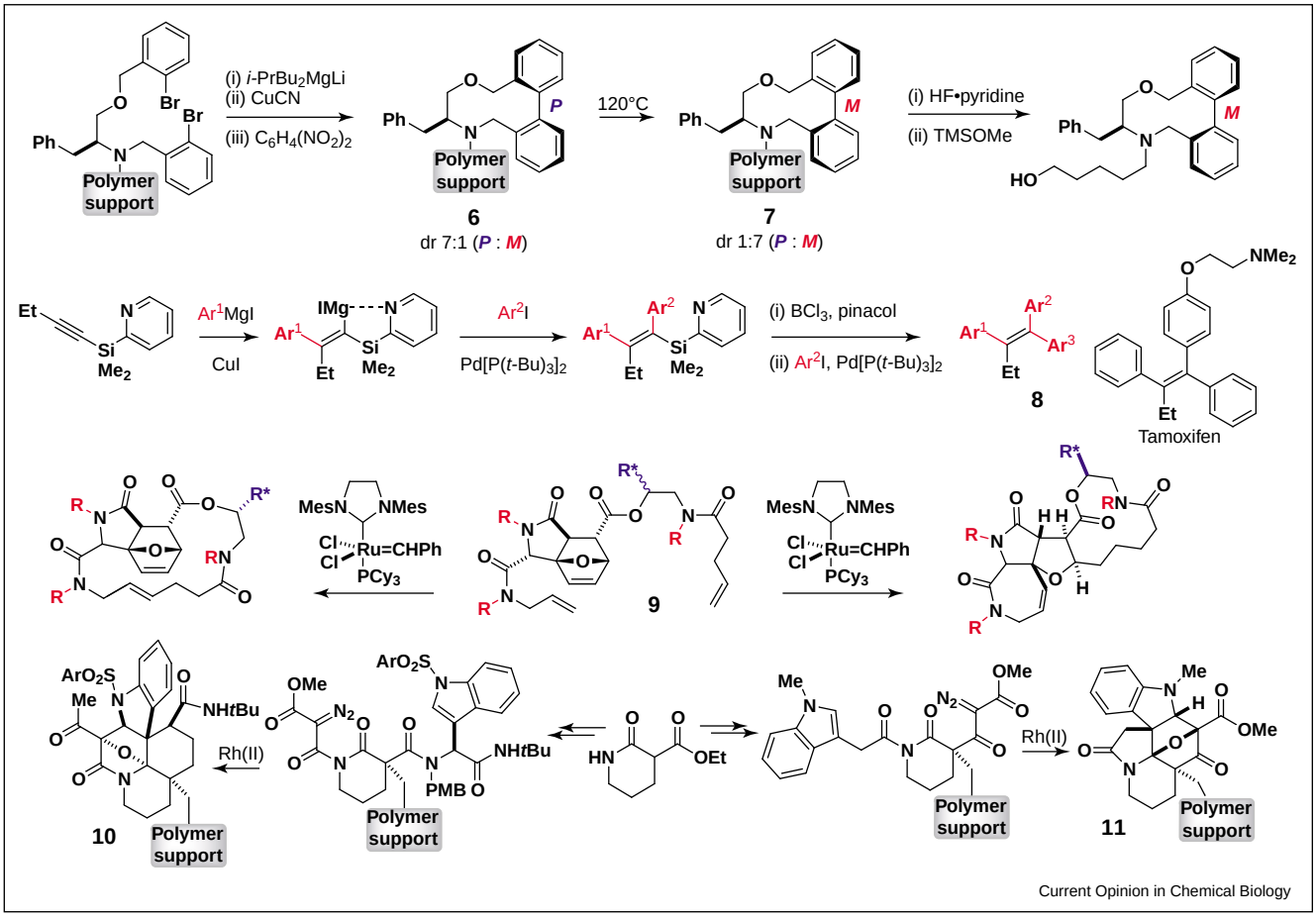
Molecular descriptors: generation and representation

To assess the molecular diversity of a collection of small molecules on a large scale it is necessary to use computer algorithms that, broadly speaking, consist of two operations. Firstly, the structures are put into 'chemical descriptor space' using molecular descriptors; secondly, diversity in chemical descriptor space is calculated [25]. The calculation of descriptors creates an abstract representation of the molecule [26,27]. Representations of molecules can be classified according to their dimensionality [28]:

1. One-dimensional (1D) with bulk properties such as volume, molecular weight and log P [29].
2. Two-dimensional descriptors (2D) are derived from the connectivity table [30] of a molecular structure.
3. Three-dimensional descriptors (3D) use geometrical information from points in 3D space.

Since binding of a ligand to a target is an event in space, the geometry of the ligand in relation to that of the binding pocket is crucial. Is it still advisable to use a 2D method over a 3D method in certain situations? Molecules are not rigid entities, they are conformationally flexible, especially if many single bonds are present in a molecule, leading to a 'curse of dimensionality' when dealing with 3D information. In addition, since the active (binding) conformation of a structure is usually unknown, most of the conformations cannot be excluded. Dealing with the complete conformational ensemble results in an increase in noise, since virtually every spatial arrangement can be fulfilled by the ligand. 2D methods, on the

Figure 3



Strategies to give structurally diverse molecular frameworks. Skeletal diversity can be generated by constitutional isomer and stereoisomer generation, divergent reaction pathways and divergent folding pathways.

other hand, do not explicitly capture shape; however, at least locally, shape is implicitly contained in the connectivity table. Therefore the information one has to deal with is greatly reduced, eliminating noise. This leads to a much faster generation of results while often retaining their validity. Atom environment descriptors are employed as a molecular representation [31], as shown in Figure 4. For diversity assessment, we calculate the average number of atom environments per molecule. The absolute number of features necessarily increases if non-identical structures are added, but here we are interested in a diversity measure relative to the size of the library. This software is freely

Figure 4

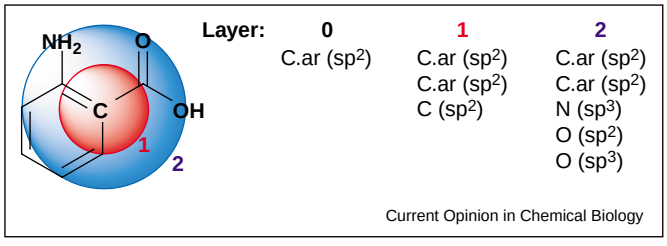


Illustration of descriptor generation step, applied to an aromatic carbon atom. The distance ('layers') from the central atom is shown in brackets. Every heavy atom in the hydrogen-depleted structure of the molecule is assigned its Sybyl atom types [34]. Sybyl atom types and are used to classify atoms according to the element type and hybridization state. An individual atom fingerprint is calculated for each heavy atom in the molecule capturing its local environment at a distance of n bonds. Frequencies of atom types at a given distance (n = 0, 1, 2) are recorded.

Table 1

Diversity value of nine collections of compounds.

Library	Reference	Number of molecules	Number of features	Features per molecule[a]	Diversity value[b]
1	[11]	168	38	0.226	2
2	[12]	88	30	0.341	3
3	[13]	49	50	1.02	10
4	[14]	62	101	1.63	16
5	[18]	24	56	2.33	23
6	[17]	15	46	3.07	30
7	[16]	28	72	4.00	39
8	[15]	60	288	4.80	46
9	[19]	10	55	5.50	53
Ideal[c]		40	414	10.4	100

Increasing molecular diversity

The diversity value is calculated on a scale from 0 to 100 incorporating the number of features per molecule. [a]To 3 significant figures. [b] Nearest integer value. [c]The 'ideal diverse library' consists of acetic acid, alliin, ampicillin, bee pheromone, benzene, bergenin, beta carotene, blebbistatin, caffeine, catechin, cinnamic acid, ciprofloxacin, cocaine, cortisone, cyclosporin, cysteine, D-glucose, dopamine, erythromycin, fluzanim, fumiquinazoline G, genistein isoflavonoid, glucosamine, L-DOPA, methane, methanol, morphine, nandrolone, omega-6 fatty acid, phenylalanine, quinine, rapamycin, serotonin, streptomycin, sucrose, Taxol, testosterone, vitamin A, vitamin E and vitamin K.

available via a web interface at www.cheminformatics.org/diversity.

To test drive this computational assessment of structural diversity a range of combinatorial libraries was chosen from the literature referenced above, and an 'ideal diverse library' consisting of 40 diverse natural products. The diversity values of each library are shown in Table 1.

The diverse libraries generally result in a higher value of diversity than the focused libraries; however, certain limitations require highlighting when evaluating the diversity of a library. The diversity value is dependent on the number of library members; therefore, very small libraries (library members < 10) give illogical results that should be utilized with caution. Also, since the program compares library compounds using two factors — the hybridization and the variation of heavy atoms — a focused library using a common scaffold with varying appendages that contain a wide variety of elements and different degrees of hybridization will give a higher value than perhaps expected. This program is a useful tool in assessing the diversity of a library; however, it should be employed with due care upon understanding some of its limitations as outlined above. A more thorough investigation will be published in due course.

Conclusions

This article has attempted to cover the combinatorial synthesis of structural diversity, known as diversity-oriented synthesis. Progress has been made over recent years in the efficient synthesis of small molecules differing in their building blocks, stereochemistry, functional groups and molecular framework. However, without a means to assess the success of a diversity-oriented synthesis it is difficult to assess the success of the field in general. We have exploited a free computer programme using fragment-based molecular descriptors to quantify the structural diversity of collections of small molecules. The results of using this software are in line with common sense; however, the software should be used with caution as library size and substituent groups influence the results.

We predict that likely future directions of DOS will be, firstly, improvements in library design, synthetic methodology and the computational assessment of structural diversity, and, secondly, the use of the compound collections to discover small molecules with desired physical or biological properties. The applications of small molecules in so many different aspects of science and life in general will guarantee that the emerging area of DOS has an exciting future.

Update

A recent review by Perez on molecular diversity analysis has been published [35•].

Acknowledgements

We would like to thank the Gates Cambridge Trust and BBSRC for funding.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. •• Burke MD, Schreiber SL: **A planning strategy for diversity-oriented synthesis**. *Angew Chem Int Ed Engl* 2004, **43**:46-58.
This review describes the design of structural diversity and complexity in the context of chemical genetics and is highly recommended as an excellent introduction to the field.

2. Schreiber SL: **Target-oriented and diversity-oriented organic synthesis in drug discovery**. *Science* 2000, **287**:1964-1969.

3. Spring DR: **Diversity-oriented synthesis; a challenge for synthetic chemists**. *Org Biomol Chem* 2003, **1**:3867-3870.

4. Ortholand JY, Ganesan A: **Natural products and combinatorial chemistry: back to the future**. *Curr Opin Chem Biol* 2004, **8**:271-280.

5. • Borman S: **Rescuing Combichem diversity-oriented synthesis aims to pick up where traditional combinatorial chemistry left off**. *Chem Eng News* 2004, **82**:32-40.
This recent article highlights the advent of DOS to counter the negative press that combinatorial chemistry had received. Comments from many big players in the field are included.

6. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ: **Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings**. *Adv Drug Deliv Rev* 2001, **46**:3-26.

7. Schreiber SL: **Chemical genetics resulting from a passion for synthetic organic chemistry**. *Bioorg Med Chem* 1998, **6**:1127-1152.

8. Schreiber SL: **The small molecule approach to biology**. *Chem Eng News* 2003, **81**:51-61.

9. Alaimo PJ, Shogren-Knaak MA, Shokat KM: **Chemical genetic approaches for the elucidation of signalling pathways**. *Curr Opin Chem Biol* 2001, **5**:360-367.

10. MacBeath G: **Chemical genomics: what will it take and who gets to play?** *Genome Biol* 2001, **2**:2005.1-2005.6.

11. Maltais R, Luu-The V, Poirier D: **Synthesis and optimization of a new family of type 3 17β-hydroxysteroid dehydrogenase inhibitors by parallel liquid-phase chemistry**. *J Med Chem* 2002, **45**:640-653.

12. • Nesterenko V, Putt KS, Hergenrother PJ: **Identification from a combinatorial library of a small molecule that selectively induces apoptosis in cancer cells**. *J Am Chem Soc* 2003, **125**:14672-14673.
An elegant use of combinatorial chemistry to generate small molecules to probe biology.

13. Faghih R, Dwight W, Bao Pan J, Fox GB, Krueger KM, Esbenshade TA, McVey JM, Marsh K, Bennani YL, Hancock AA: **Synthesis and SAR of aminoalkoxy-biaryl-4-carboxamides: novel and selective histamine H_3 receptor antagonists**. *Bioorg Med Chem Lett* 2003, **13**:1325-1328.

14. Zhang L, Liu G, Zhang S-D, Yang H-Z, Li L, Wu X-H, Yu J, Kou B-B, Xu S, Li J *et al.*: **Parallel approach for solution-phase synthesis of 2-quinoxalinol analogues and their inhibition of LPS-induced TNF-α release on mouse macrophages *in vitro***. *J Comb Chem* 2004, **6**:431-436.

15. Wu CY, Chang CF, Chen JSY, Wong CH, Lin CH: **Rapid diversity-oriented synthesis in microtitre plates for *in situ* screening: discovery of potent and selective α-fucosidase inhibitors**. *Angew Chem Int Ed Engl* 2003, **42**:4661-4664.

16. Spring DR, Krishnan S, Blackwell HE, Schreiber SL: **Diversity-oriented synthesis of biaryl-containing medium rings using a one bead/one stock solution platform**. *J Am Chem Soc* 2002, **124**:1354-1363.

17. Itami K, Kamei T, Yoshida J: **Diversity-oriented synthesis of tamoxifen-type tetrasubstituted olefins**. *J Am Chem Soc* 2003, **125**:14670-14671.

18. Schmidt DR, Kwon O, Schreiber SL: **Macrolactones in diversity-oriented synthesis: preparation of a pilot library and exploration of factors controlling macrocyclization**. *J Comb Chem* 2004, **6**:286-292.

19. Sello JK, Andreana PR, Lee D, Schreiber SL: **Stereochemical control of skeletal diversity**. *Org Lett* 2003, **5**:4125-4127.

20. • Oguri H, Schreiber SL: **Skeletal diversity via a folding pathway: synthesis of indole alkaloid-like skeletons**. *Org Lett* 2005, **7**:47-50.
Inspired by the skeletal diversity of indole alkaloids, the authors exploited Padwa intramolecular cyclizations in three out of six possible modes. The paper illustrates a conceptually simple and general approach to folding substrates to give different molecular frameworks.

21. Taylor SJ, Taylor AM, Schreiber SL: **Synthetic strategy toward skeletal diversity via solid-supported, otherwise unstable reactive intermediates**. *Angew Chem Int Ed Engl* 2004, **43**:1681-1685.

22. Wipf P, Stephenson CRJ, Walczak MAA: **Diversity-oriented synthesis of azaspirocycles**. *Org Lett* 2004, **6**:3009-3012.

23. • Burke MD, Berger EM, Schreiber SL: **Generating diverse skeletons of small molecules combinatorially**. *Science* 2003, **302**:613-618.
The authors outline the approach of using appendages to pre-encode substrates to undergo different folding pathways to different molecular frameworks using the same reaction conditions. This is the first paper to describe this approach.

24. Burke MD, Berger EM, Schreiber SL: **A synthesis strategy yielding skeletally diverse small molecules combinatorially**. *J Am Chem Soc* 2004, **126**:14095-14104.

25. Xue L, Bajorath J: **Molecular descriptors in chemoinformatics, computational combinatorial chemistry and virtual screening**. *Comb Chem High Throughput Screen* 2000, **3**:363-372.

26. Bender A, Glen RC: **Molecular similarity: a key technique in molecular informatics**. *Org Biomol Chem* 2004, **2**:3204-3218.

27. Brown RD, Martin YC: **Use of structure-activity data to compare structure-based clustering methods and descriptors for use in compound selection**. *J Chem Inform Comput Sci* 1996, **36**:572-584.

28. Willett P, Bernard JM, Downs GM: **Chemical similarity searching**. *J Chem Inf Comput Sci* 1998, **38**:983-996.

29. Downs GM, Willett P, Fisanick W: **Similarity searching and clustering of chemical-structure databases using molecular property data**. *J Chem Inf Comput Sci* 1994, **34**:1094-1102.

30. Estrada E, Uriarte E: **Recent advances on the role of topological indices in drug discovery research**. *Curr Med Chem* 2001, **8**:1573-1588.

31. Bender A, Mussa HY, Glen RC: **Similarity searching of chemical databases using atom environment descriptors (MOLPRINT 2D): evaluation of performance**. *J Chem Inf Comput Sci* 2004, **44**:1708-1718.

32. •• Burke MD, Lalic G: **Teaching target-oriented and diversity-oriented organic synthesis at Harvard University**. *Chem Biol* 2002, **9**:535-541.
This article is a comprehensive, useful guide to the comparison of target-oriented and diversity-oriented synthesis.

33. Corey EJ, Cheng XM: *The Logic of Chemical Synthesis*. Wiley; 1995

34. Clark RD, Cramer RD, Van Opdenbosch N: **Validation of the general purpose Tripos 5.2 force field**. *J Comput Chem* 1989, **10**:982-1012.

35. • Perez JJ: **Managing molecular diversity**. *Chem Soc Rev* 2005, **34**:143-152.
An excellent general introduction to computational molecular diversity has appeared, describing four steps for diverse compound selection. Approaches to describe molecular structures are followed by methods to establish their similarities. Next, partitioning/clustering methods are described for diversity-oriented compound selection, accompanied by their validations.