Diversity-Oriented Synthesis

RICHARD J. SPANDL, MÓNICA DÍAZ-GAVILÁN, KIERON M. G. O'CONNELL, GEMMA L. THOMAS, DAVID R. SPRING
Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, UK

ABSTRACT: *Diversity-oriented synthesis* (DOS), which describes the synthesis of structurally diverse collections of small molecules, was developed, in part, to address combinatorial chemistry's shortfalls. In this paper, we hope to give an indication of what can be achieved using the DOS approach to library generation. We describe some of the most successful strategies utilized in DOS, with special focus on our own area of interest; DOS from simple, pluripotent starting materials. © 2008 The Japan Chemical Journal Forum and Wiley Periodicals, Inc. Chem Rec 8: 129–142; 2008: Published online in Wiley InterScience (www.interscience.wiley.com) DOI 10.1002/tcr.20144

Introduction

The sequencing of the human genome has led to scientists achieving a greater understanding of life, at a molecular level, than ever before.[1] With these new insights have come new and formidable challenges; one of the most significant is the annotation of the human genome. Meeting this challenge would not only lead to significant advances in our appreciation of biology but would also be a step toward developing cures for some currently untreatable or poorly understood genetic conditions and disease states.

The elucidation of human gene function may be aided by access to small molecule modulators of the corresponding gene products (proteins).[2] This use of chemical probes to explore biological systems is known as *chemical genetics* (Fig. 1).[2–5] The ultimate goal of chemical genetics is to identify small molecules that specifically perturb the function of every protein;[6] this realization would allow the systematic use of small molecules to explore biological systems.[4] Depending on the circumstances, chemical genetic experiments can be performed in either a forward sense, i.e., going from phenotype to target protein (forward chemical genetics), or a reverse sense, i.e., going from a protein to a phenotype (reverse chemical genetics).[2]

Although it is estimated that only 10% of the human genome encodes proteins suitable for chemical intervention (the "druggable genome"),[7] only about 1000 of these proteins have known small molecule partners which perturb their function.[8] It is clear, therefore, that access to, and the efficient and reliable screening of, compound collections with diverse biological activities may be advantageous in the identification of novel biological modulators.

To meet the potential need for small molecule collections that exhibit a range of bioactivities, and to address some of the shortfalls of traditional combinatorial chemistry,[9] over the last decade or so, *diversity-oriented synthesis* (DOS) has been developed. DOS aims to efficiently synthesize collections of natural product-like and/or drug-like small molecules with diverse molecular structures.[10–12] It is skeletal variety which is of critical importance in DOS and, furthermore, it is skeletal

---

▸ *Correspondence to*: David Spring; e-mail: drspring@ch.cam.ac.uk

variety which has been reported to confer the desired biological diversity to a compound library.[13]

DOS libraries, which contain molecules that are diverse in both structure and (consequently) function, in addition to being useful in chemical genetics, may have applicability in other situations. For example, the screening of such libraries may aid the identification of novel compounds with a desired physiochemical property or may be useful in lead identification for drug discovery. Although these examples demonstrate the broad potential applicability of DOS, this paper will describe the DOS approach in the context of discovering biologically active small molecules.

▶ *Richard Spandl was born in 1982 and was brought up near the village of Eye in Suffolk, England. He obtained his M.Chem. in Biological Chemistry from the University of Leicester in 2004 before joining the Spring Group as a Ph.D. student. Richard's Ph.D., which was funded by the BBSRC and Eli Lilly, has focused on the development of strategies for diversity-oriented synthesis (DOS).* ■

▶ *Mónica Díaz-Gavilán is currently an EU Marie Curie Postdoctoral Fellow at the University of Cambridge, where she works with Dr. David R. Spring on the field of DOS. She joined the Spring Group in May 2006, after gaining her Ph.D. at the University of Granada, Spain. There she obtained a degree in pharmacy and worked with Professor Antonio Espinosa-Úbeda on the development of new benzoxazepine derivatives with antitumoral activity.* ■

Photographer: Nathan Pitt

▶ *Kieron O'Connell was born in London, England in 1984 and grew up in Whitstable, Kent. He obtained his M.Chem. in Chemistry with a year in Industry from the University of Reading in 2007. This course included a year working for Syngenta Agrochemicals based at Jealott's Hill, Bracknell, England. Kieron joined the Spring group as a Ph.D. student working in the area of DOS in October 2007; his project is funded by the BBSRC and UCB Celltech.* ■

Dissecting Biology Using Small Molecule Intervention (Chemical Genetics)

As a complementary approach to traditional genetics, chemical genetics uses small molecules to dissect biological systems.[2–5] For chemical intervention to become a general approach to solving biological problems, an enormous number of small molecules, which specifically perturb protein function, are required.[11] These novel chemical modulators could be identified by screening structurally, and possibly biologically, diverse collections of small molecules in suitably designed assays. It is noteworthy that although there are sources of diverse small molecules available to scientists, this "diversity" is not paralleled in the biological screening assays. Currently, the lack of generally applicable and effective high-throughput screening technologies represents a bottleneck in the identification of bioactive small molecules. Advances toward meeting these challenges are, however, being made.[14]

There are a number of potential sources of diverse small molecules; these include natural products, combinatorial libraries, proprietary compound collections, and libraries synthesized using DOS.

Nature has provided a rich supply of skeletally diverse molecular entities that have a modulating effect on proteins and other biomolecules. Unfortunately, however, the isolation and screening of a specific natural product can be a complicated process. In some instances, this can result from a low natural abundance of the desired compound and/or difficulties associated with the natural product's purification and characterization. Furthermore, total synthesis and subsequent chemical derivatization, both of which are especially relevant in "lead optimization," can be extremely difficult and time-consuming processes.[15,16]

Commercially available combinatorial libraries or in-house pharmaceutical proprietary compound collections are alternative sources of small molecules. As a result of the "one-synthesis/one-skeleton" approach often used,[12] although combinatorial libraries may offer complexity, they sometimes show limited structural diversity within a particular library. Furthermore, these libraries may be biased by previous drug discovery

▶ *Gemma Thomas is currently an EPSRC postdoc within the Spring Group having previously obtained her Ph.D. within the group working on the identification of a novel antibiotic using DOS. Before joining the Spring group in 2003, she obtained her M.Chem. in Chemistry from the University of Southampton.* ■

Photographer: Nathan Pitt

▶ *David Spring is currently an EPSRC Advanced Fellow at the University of Cambridge Chemistry Department and Fellow of Trinity College. Previous to this appointment, he spent 2.5 years as a Wellcome Trust Postdoctoral Fellow and Fulbright Scholar at Harvard University with Professor Stuart L. Schreiber. He gained his Ph.D. for work on the proposed biosynthesis of the manzamine alkaloids at Oxford University under the supervision of Professor Sir Jack E. Baldwin, FRS. David's research program is focused on DOS, synthetic methodology, and chemical genetics.* ■
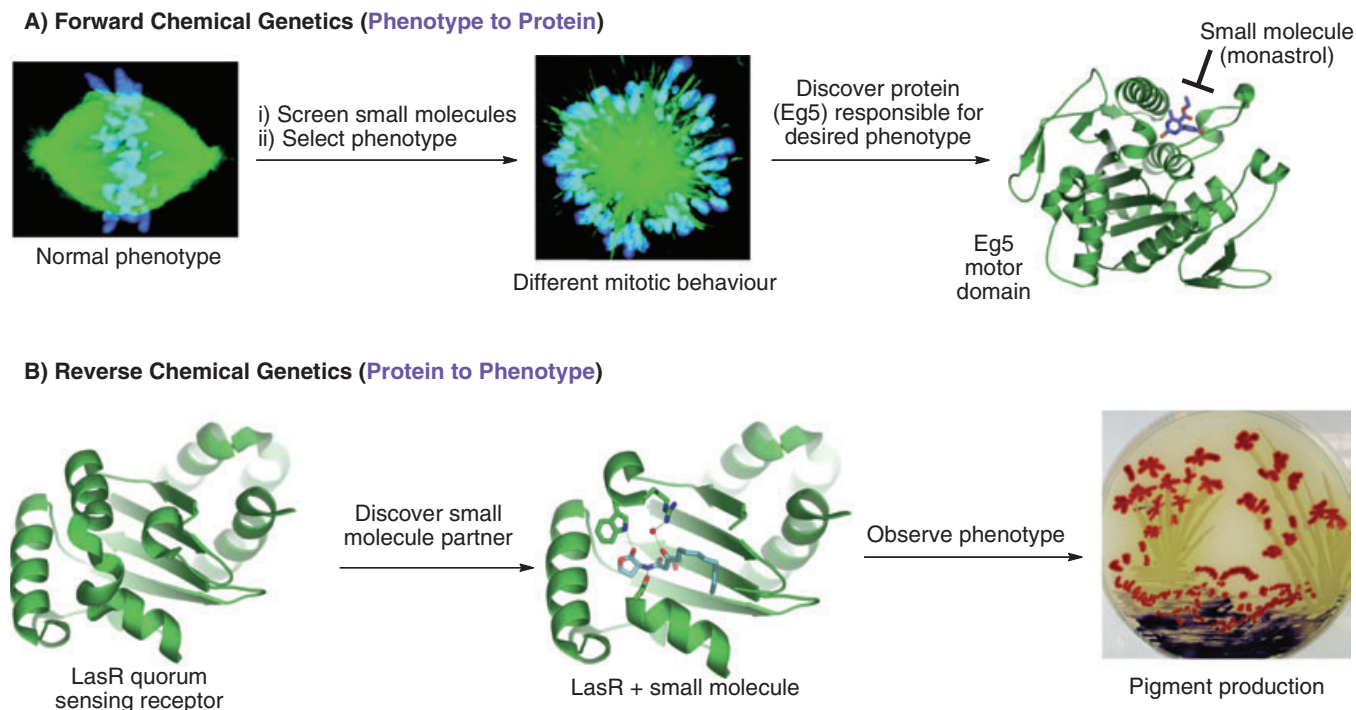
A) Forward Chemical Genetics (Phenotype to Protein)

Normal phenotype

i) Screen small molecules
ii) Select phenotype

Different mitotic behaviour

Discover protein
(Eg5) responsible for
desired phenotype

Small molecule
(monastrol)

Eg5
motor
domain

B) Reverse Chemical Genetics (Protein to Phenotype)

LasR quorum
sensing receptor

Discover small
molecule partner

LasR + small molecule

Observe phenotype

Pigment production

Fig. 1. An overview of chemical genetics; this technique uses chemical modulators to explore biological systems. (A) In forward chemical genetics, small molecules are screened, and those which induce a desired phenotype (different mitotic behavior in this example) are selected. The protein partner, through which these small molecules act, is then identified. (B) Conversely, in reverse chemical genetics, a small molecule modulator, with a known protein partner, is used to perturb the biological system under investigation and the resulting phenotype 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* (red), *Chromobacterium violaceum* (purple), and *P. aeruginosa* (light green).[53–55]

programs and generally contain flat molecules with fewer chiral centers compared with natural products and drugs. This latter point was demonstrated by the results from the computational analysis of various databases. It was found that the number of chiral centers, on average, from combinatorial chemistry, natural products, and drugs was 0.4, 6.2, and 3.3 per molecule, respectively.[17]

Although the compound archives of pharmaceutical companies (which will include many combinatorial libraries) constitute a very diverse collection of molecules, they potentially have their drawbacks for chemical genetics. A possible limitation may be the imposition of certain predefined criteria. One such criteria often evoked is the Lipinski rule of five[18] and, although useful, there has been debate about restricting chemists to synthesizing compounds with these physical properties.[19,20] This limitation is especially relevant in the discovery of chemical probes compared to, for example, orally active drugs since less stringent rules for efficacy apply.[8]

Another potential drawback of using the above small molecule sources is highlighted when the concept of *chemical space* analysis is used. Each molecule, as a function of its associated chemical descriptors,[21–24] resides at a discrete point in chemical space.[25] Natural products and currently available compound libraries occupy only a small proportion of bioactive *chemical space*.[8,25] Therefore, there may be value in screening compounds that have novel molecular architectures. These libraries should interrogate different areas of chemical space within this bioactive region (and possibly also outside of it). A non-focused library, as produced using DOS, may be advantageous in this context. Although structural diversity is incorporated into a DOS library, the molecules are still "drug- and/or natural product-like." As a result, the compounds synthesized are more likely to interact with biomolecules. Thus, since nature "sees" molecules as complex three-dimensional surfaces and produces an amazingly diverse collection of compounds to interact specifically with its biomolecules, DOS aims to mimic this variety. Library construction using

DOS may therefore be a suitable alternative to the above strategies.

The number of possible "drug-like" molecules present in chemical space is astronomic.[10] It would indeed be impossible to prepare molecules to suitably cover even the biologically active region of chemical space. We must therefore be selective and produce libraries that are not only "drug-like" but are also skeletally diverse to ensure maximal chemical space coverage; this should increase the chances of discovering novel bioactive molecules. This is the goal of DOS.[11]

It should be noted that the term "diversity" is very subjective. Although DOS libraries are structurally diverse, they are still designed to have the potential to interact with biomolecules (where chemical probes are sought) and hence are not completely diverse. In short, the term diversity must be taken in context.

How Does DOS Compare to More Traditional Approaches?

DOS describes a process whereby diverse collections of complex small molecules are synthesized in an efficient and deliberate manner.[10,11] A comparison of DOS to *target-oriented synthesis* (TOS) and also to traditional combinatorial chemistry demonstrates this (Fig. 2). Although complexity is not a prerequisite for diversity, it has been proposed to confer specificity in biological interactions.[8] There is, however, debate in the literature about this point.[26,27]

In designing a DOS, analysis is performed in a forward sense and a strategy is developed whereby simple starting materials can be transformed into diverse and complex products. In TOS, retrosynthetic analysis allows a complex product to be deconstructed in a backward sense. There is also a difference in the outcomes (and goals) of both these approaches; whereas TOS aims to synthesize a molecule at a discrete point in chemical space, DOS aims to cover as diffuse an area as possible.[12]

The distinction between DOS and combinatorial chemistry is less clear cut. We would therefore like to suggest that DOS is a more "evolved" version of combinatorial chemistry. Thus, these terms are not mutually exclusive and the technologies overlap. DOS, however, does differ from *traditional* combinatorial chemistry as DOS does not target as selected an area of chemical space. An example of this selected targeting in *traditional* combinatorial chemistry would be in lead optimization for drug discovery. This comparison also serves to highlight another important issue, the subjectivity of diversity.[11]

When a compound collection is synthesized, since the composite molecules are not identical, diversity, to a greater or lesser extent, is incorporated; the racemic synthesis of enantiomers could even be classified as a DOS. As a result of this subjectivity, and the free use of the terms "diversity" and "DOS" in the literature, considering diversity as a spectrum may be useful. In one extreme of the "molecular diversity spectrum" would be where maximal chemical space coverage has been achieved and, in the other extreme, would be a TOS (Fig. 3).[11] It should be the goal of a DOS to synthesize, in a qualitative sense, collections as near to the right-hand side of the "diversity spectrum" as possible.[11]

It should be noted that DOS and TOS are different strategies with different goals. The above serves to compare the diversity achieved using either approach regardless of the eventual aim; it is not to be implied that DOS is better than TOS as it generates more diversity, merely that, to maximize chemical space coverage, skeletal diversity is essential. It is this skeletal diversity that can be incorporated using the DOS (and not the traditional combinatorial chemistry) approach to library synthesis.
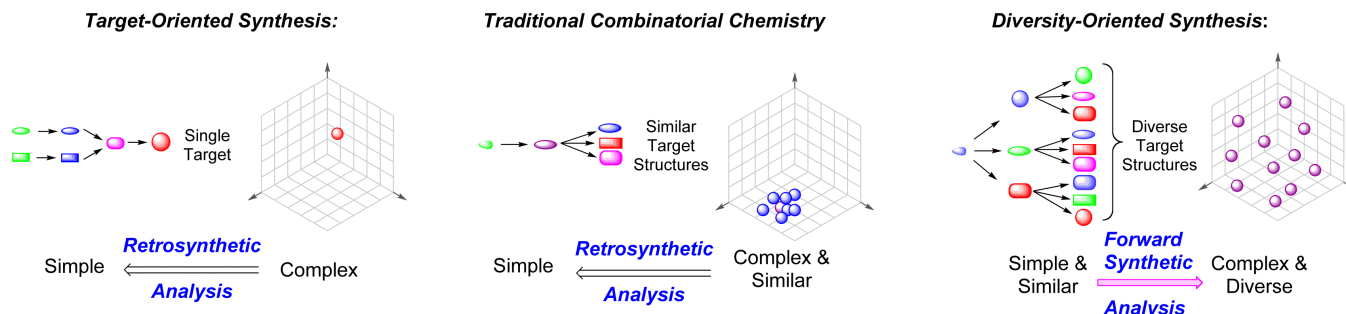

Fig. 2. There is a clear distinction between the strategies and approaches of target-oriented synthesis compared to diversity-oriented synthesis (DOS); the former aims to prepare molecules at discrete points in chemical space, whereas the latter strives to cover as much of chemical space as possible. The partition between DOS and combinatorial chemistry is less clear cut. DOS does, however, differ from traditional combinatorial chemistry (or focused library synthesis). In these latter approaches, a discrete region of chemical space is interrogated, the chemical space around a lead compound, for example.
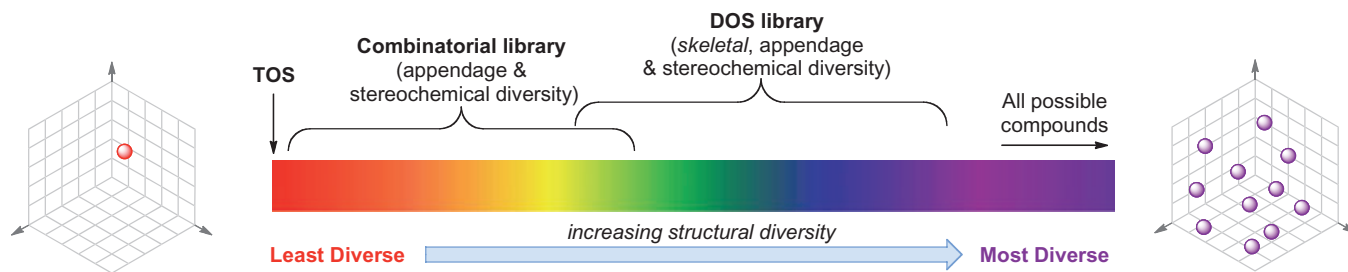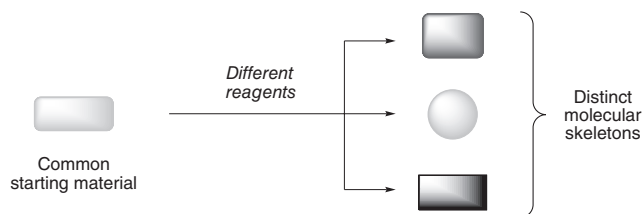
Fig. 3. The "molecular diversity spectrum" which ranges from a target-oriented synthesis (TOS) in one extreme to the synthesis of all possible compounds in the other. It should be the goal of diversity-oriented synthesis (DOS), in a qualitative sense, to synthesize compounds as close to the right-hand side of this spectrum as possible.[11]
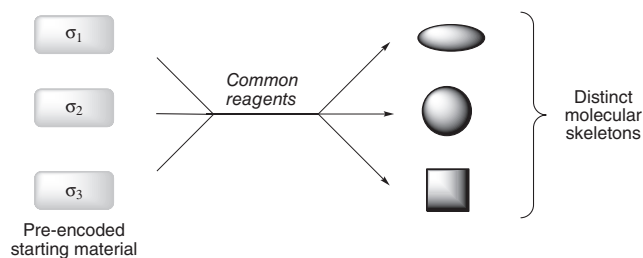

Fig. 4. Access to skeletal diversity can be achieved in two principal ways, either (A) using a common starting material and reacting it with different reagents or (B) taking different starting materials and reacting them with common reagents.

The General Requirements of a DOS

The members of a DOS library should be diverse in both the appendages they display and also in the three-dimensional orientations and locations of these appendages.[12] Thus, in designing a DOS, strategies to incorporate the four key types of diversity must be included.[12,28,29] These four types of diversity are (i) appendage diversity, which can be incorporated by combinatorial variation in the building blocks used; (ii) stereochemical diversity, which can be incorporated using reagent-controlled asymmetric reactions; (iii) functional group diversity, which can be incorporated by chemical manipulations; and, most importantly, (iv) skeletal diversity.[12,28,29]

The most challenging aspect of DOS, and of vital importance to the realizations of its goals, is the ability to generate multiple molecular architectures, i.e., the incorporation of skeletal diversity.[28] Although the main purpose of this paper was to examine some recent strategies to create DOS libraries from simple starting materials, alternative approaches to DOS have been used. These strategies, which will not be considered further, such as "biologically oriented synthesis"[30–32] and "DOS based on privileged scaffolds,"[33–37] have been successful in the identification of novel bioactive molecules. These approaches involve a more discrete targeting of chemical space and may therefore be advantageous when a particular biological target is in

mind.[38,39] However, in forward chemical genetics, when the final "hit" compound is frequently novel in structure and unpredictable prior to experimentation, DOS from simple starting materials may be useful as a result of its less focused nature.

The approach of DOS from simple starting materials utilizes sequential complexity-generating reactions to incorporate skeletal diversity into a compound collection.[40–42] As a result and in relatively few steps, these *branching pathways*, where the product of one reaction is the substrate for the next, act to convert a simple starting material to an array of complex and diverse molecular scaffolds. These sequential transformations have been termed *tandem* reactions.[12]

Skeletal Diversity Using Branching Pathways

There are two principal ways in which skeletal diversity can be achieved. These involve using either branching pathways, as mentioned above, or, alternatively, *folding pathways*. In a branching pathway (also called a "reagent-based approach"), a particular starting material is converted to different architectures using different reagents. Conversely, in a folding pathway (also called a "substrate-based approach"), different substrates containing pre-encoded skeletal information are transformed into different scaffolds using the same reagents (Fig. 4).[28]

A review of the literature suggests that reagent- and substrate-based approaches are utilized in a variety of ways in successful DOS projects.[40] In general, these strategies involve one of the following: (i) the use of a densely functionalized molecule where different functionalities in the same molecule are transformed by different reagents; (ii) the use of a folding process where different structurally encoding elements, contained in different substrates, are subjected to the same reaction conditions; or, (iii) the use of a pluripotent functionality where the same part of a molecule is subjected to different transformations induced by different reagents.

Recently, Schreiber and coworkers have also described a strategy which combines these approaches.[43] This "build/couple/pair" strategy involves firstly "building" the required chiral starting units, possibly from chiral pool reagents. Next, a densely functionalized molecule is synthesized by coupling these starter units, possibly using multicomponent coupling reactions. Finally, pairing different parts of the densely functionalized molecule, in functional group specific reactions, generates different molecular skeletons.

Strategy 1. Densely Functionalized Molecules

Schreiber and coworkers demonstrated the utility of using a densely functionalized β-amino alcohol **1** to generate multiple scaffolds **2–8** via pairing reactions (Scheme 1).[44] The substrate **1** could be synthesized using the Petasis three-component coupling reaction, the stereochemical outcome of which was controllable. A second generation of compounds **9–13** could be produced by the subsequent pairing of functional groups that were not affected by the first round of chemical transformations. In addition, when a 1,3-diene was produced, a tandem Diels–Alder reaction was used to increase the diversity and complexity of the compound collection; this yielded skeletons **14–16**.

A similar strategy has recently been reported by Porco and coworkers. An asymmetric conjugate addition to an α,β-unsaturated nitro compound furnished the enantiopure starting material **17**.[45] Taking advantage of the densely functionalized nature of **17**, distinct molecular scaffolds could be synthesized by subjecting **17** to different reagents and conditions. Thus, these transformations served to pair the functional groups of **17**, in a specific and controllable fashion, and yielded scaffolds **18–20**. Further diversity could be incorporated by using tandem reactions (enyne metathesis–Diels–Alder) and sequential pairing processes. This is exemplified in the synthesis of skeleton **24**. Skeleton **24** was accessed via a pairing reaction of the nitro and ester group of **23**. These functionalities remained unchanged in the formation of this Diels–Alder adduct **23**. The adduct **23** could be accessed via the 1,3-diene **22** which was itself synthesized from the initial alkene–alkyne pairing reaction of these groups in **21** (Scheme 2).

Although a library of compounds was not produced in either of the above cases, these elegant examples, reported by the Schreiber and Porco groups, respectively, demonstrate how multiple scaffolds can be produced efficiently from densely functionalized starting units.

Strategy 2. Folding Pathways

An excellent example of the use of folding pathways in a DOS was reported by Oguri and Schreiber.[46] Under the same conditions, i.e., a Rh(II)-induced cyclization, three distinct indole architectures **26–28** could be produced from different starting materials based around the scaffold **25** (Scheme 3). The starting materials were synthesized by elaborating the basic scaffold **25** at its three reactive sites, i.e., A, B, and C. Attached to these sites was either a linker unit and silyl protecting group/solid support **29**, an α-diazo carbonyl moiety **30**, or an indole group **31**. The relative locations of the α-diazo carbonyl and the indole groups in **25** served to determine the skeletal outcome of the reaction.

### Strategy 3. Using Branching Pathways to Incorporate Diversity

To demonstrate the use of pluripotent functional groups as an effective branching strategy in DOS, we highlight some of our own research efforts.[41,42] For this strategy to be effective, a number of factors must be considered. The starting material should not only undergo a rich variety of chemical transformations, but its synthesis should potentially allow for building block diversity to be incorporated. Furthermore, the novel scaffolds generated in the initial branching process should have the potential to be diversified further, preferably in subsequent branching reactions.

To this end, the high reactivity and mechanistic flexibility of the diazoacetate moiety **32** was exploited in the generation of a small molecule library by Wyatt et al.[42] In addition to participating in complexity-generating C-C bond-forming reactions, the diazoacetate **32** acted as either an electrophile or a nucleophile depending on the reaction conditions employed. The starting material **32** also contained a polyfluorocarbon tag; this allowed for generic purification protocols to be utilized. Three branching pathways were initially employed (step 1), and the starting material **32** was converted to scaffolds **33–39** using either three-membered ring-forming reactions (yielding **33** and **34**), α-deprotonation followed by electrophilic quenching reactions (yielding **36** and **37**), or 1,3-dipolar cycloaddition reactions (yielding **38** and **39**) (Scheme 4).

Following these initial syntheses, a second generation of compounds **40–46** was produced by exploiting complexity-generating reactions (step 2). Compounds **33** and
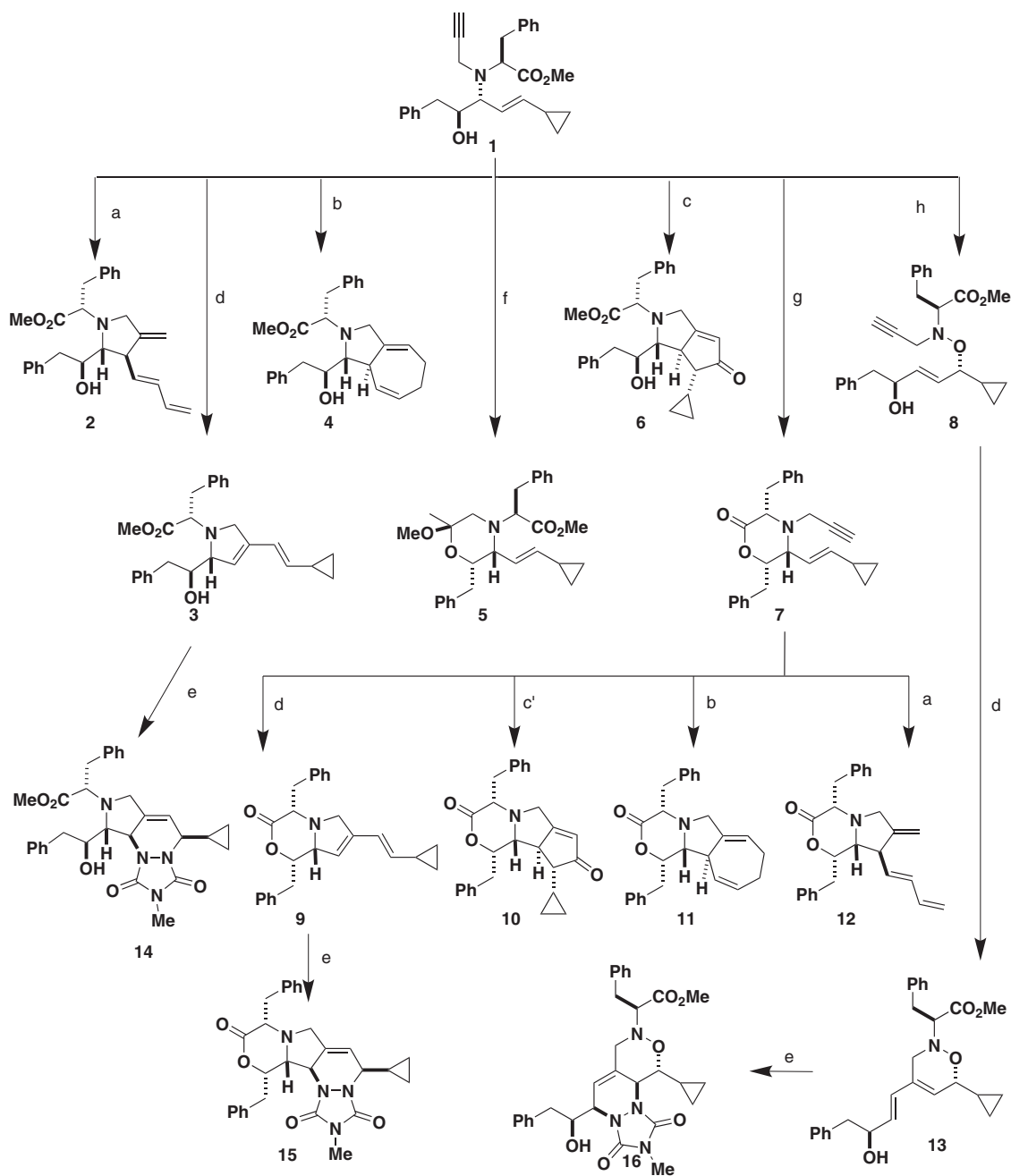
Scheme 1. An example of the use of densely functionalized molecules in a diversity-oriented synthesis. Conditions: (a) [Pd(PPh₃)₂(OAc)₂] (10 mol%), benzene, 80°C; (b) [CpRu(CH₃CN)₃PF₆] (10 mol%), acetone, RT; (c) [Co₂(CO)₈], trimethylamine *N*-oxide, NH₄Cl, benzene, RT; (c′) [Co₂(CO)₈], trimethylamine *N*-oxide, benzene, RT; (d) Hoveyda–Grubbs second-generation catalyst (10 mol%), CH₂Cl₂, reflux; (e) 4-methyl-1,2,4-triazoline-3,5-dione, CH₂Cl₂, RT; (f) NaAuCl₄ (10 mol%), MeOH, RT; (g) NaH, toluene, RT; (h) *m*CPBA, THF, −78 → 0°C. *m*CPBA = *m*-chloroperbenzoic acid.[44]
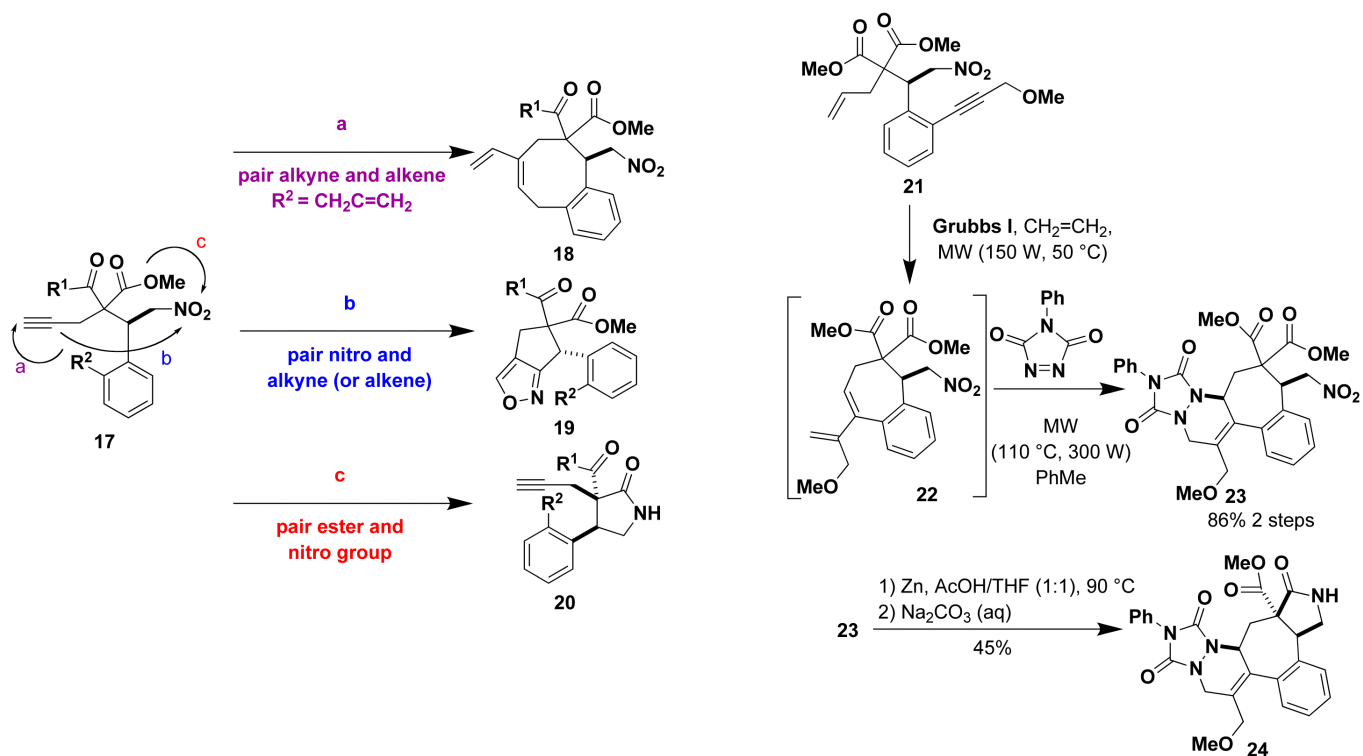
Scheme 2. A further example of the use of a densely functionalized molecule in diversity-oriented synthesis. Different pairing reactions of the densely functionalized molecule **17** gave access to distinct molecular scaffolds **18–20**. Sequential pairing processes yielded alternative architectures.[45] MW = microwave irradiation.
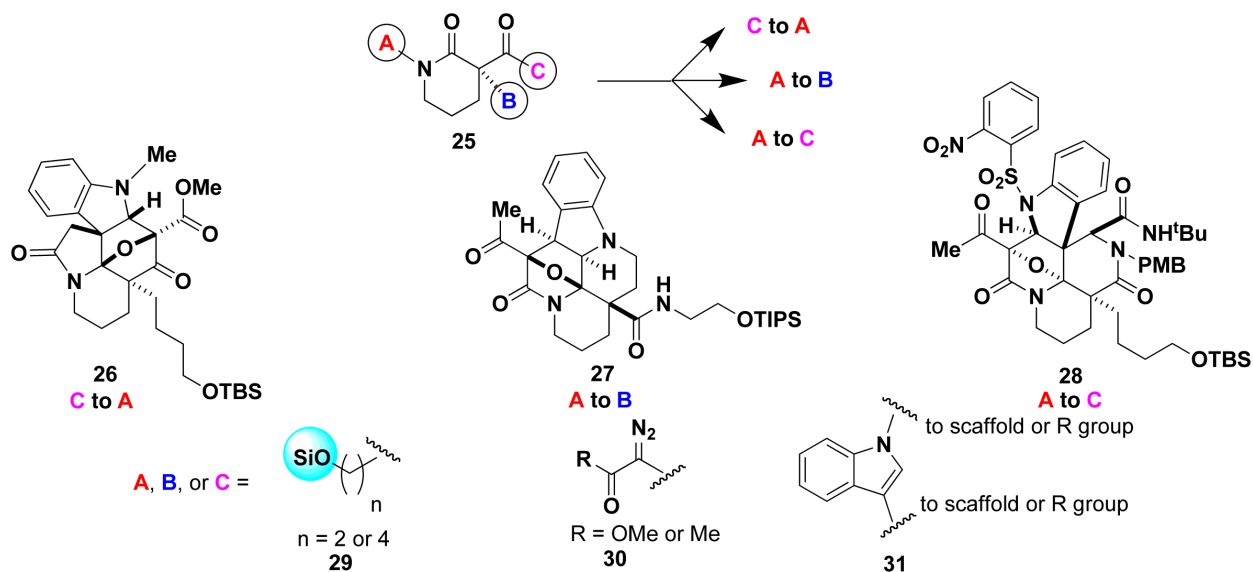
Scheme 3. An example of a folding pathway in a diversity-oriented synthesis. By constructing starting materials with various combinations of **29–31** at the reactive sites A, B, and C of scaffold **25**, the distinct indole-based architectures **26–28** could be accessed.[46] TIPS = triisopropylsilyl; TBS = *tert*-butyldimethylsilyl.
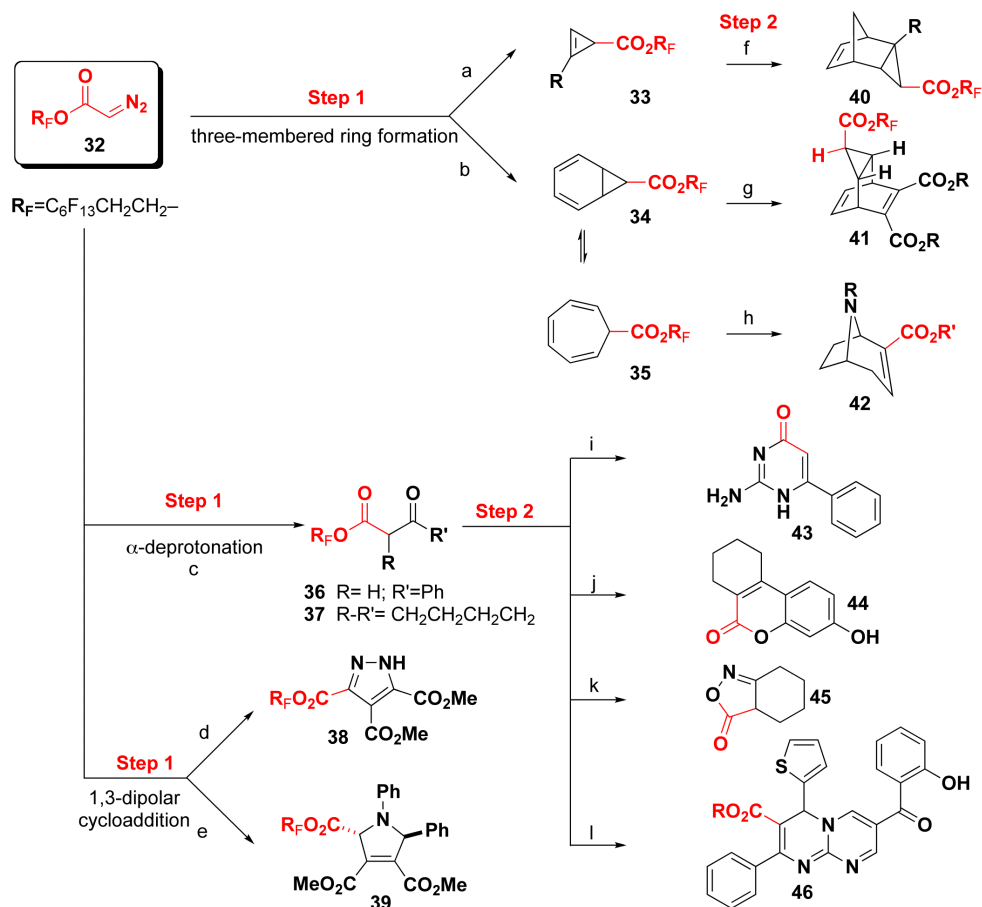
Scheme 4. Diversity-oriented synthesis of 223 small molecules based on 30 discrete frameworks. Step 1: (a) RCCH, Rh$_2$(OAc)$_4$, [BuCCH, 57%]; (b) C$_6$H$_6$, Rh$_2$(O$_2$CCF$_3$)$_4$, 70%; (c) LDA, RCOR' then Rh$_2$(OAc)$_4$; **36**: 49% (90%); **37**: 68% (97%); (d) DMAD, 84% (88%); (e) PhCHO, PhNH$_2$, then DMAD, Rh$_2$(OAc)$_4$, *d.r.* = 20 : 1, 51% (80%). Step 2: (f) C$_5$H$_6$, 92%; (g) dienophile [dimethyl acetylenedicarboxylate, 59%]; (h) RNH$_2$, NaOH then MeOH, H$_2$SO$_4$, [MeNH$_2$, 35%]; (i) guanidine carbonate 62% (96%); (j) resorcinol, H$_2$SO$_4$, 74% (95%); (k) NH$_2$OH, 77% (89%); (l) thiophene-2-carboxaldehyde, guanidine carbonate, then 3-formylchromone, 43% (98%). Yields and purity (in brackets) of the product example following generic purification using (reverse) fluorous solid phase extraction or precipitation shown. Purity determined by HPLC, LCMS, or ^{1}H NMR. DMAD = dimethyl acetylenedicarboxylate.[42]

34, synthesized via the three-membered ring-forming reaction, were converted to the Diels–Alder adducts **40** and **41**. In addition, compound **34** was also converted to the ecgonine scaffold **42** via the ring-expanded molecule **35**, in a reaction with primary amines. A diverse array of heterocyclic scaffolds **43**–**46** was accessed via the β-keto compounds **36** and **37**; these β-keto compounds were synthesized via the α-deprotonation reaction route previously.

A collection 223 small molecules, based around 20 discrete molecular frameworks, was synthesized from **32** and its diversity assessed using computational analysis.* Using the same *in silico* methods, the diversity of a focused library and known pharmaceutically active compounds [MDL Drug Data Repository (MDDR)] was also assessed and the results compared. The DOS library, within the same region of chemical space as the MDDR compounds although occupying less area, was found to interrogate a more diffuse region of chemical space than the focused library (Fig. 5).

As an alternative starting unit, we have also used the versatile solid-supported phosphonate **47** in library synthesis (Scheme 5).[41] This library, synthesized by Thomas et al., exploited the key properties of **48** which could be readily synthesized from **47** in an *E*-selective Horner–Wadsworth–Emmons reaction (step 1). The pluripotent nature of **48** allowed access to the three primary branching pathways via either a Diels–Alder reaction, a dihydroxylation reaction, or a 1,3-dipolar cycloaddition reaction (step 2). Subsequent transforma-

---

*Principal component analysis (PCA) was used to assess the diversity of the compound collection. Each molecular entity was described by a series of physiochemical descriptors.[21] The descriptors were then condensed and analyzed on a per molecule basis.[42]
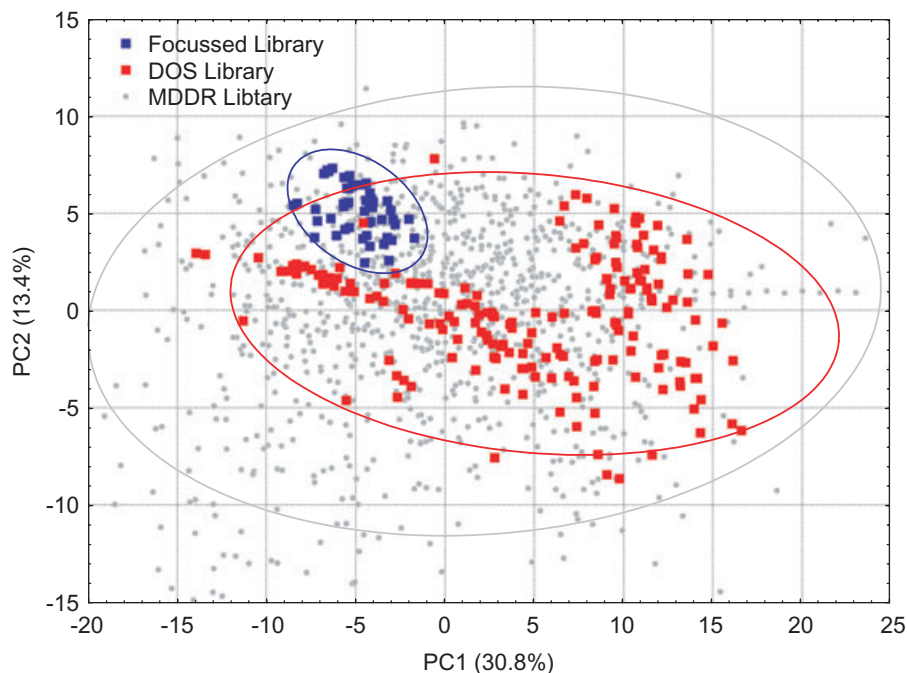
Fig. 5. The diversity (assessed using principal component analysis) of the diversity-oriented synthesis (DOS) library (red squares) was found to occupy a greater area of chemical space than a focused library (blue squares). The chemical space covered by the DOS library was within the same region as the MDL Drug Data Repository (MDDR) compounds (grey squares) that were also assessed.[42] PC = principal component.

tions enabled a second and third generation of compounds to be synthesized; these processes allowed both the appendage and the skeletal diversity to be increased (steps 3 and 4).

In addition to the dihydroxylation and 1,3-dipolar cycloaddition reactions being performed catalytically and asymmetrically, the nature of **48** allowed two-point chiral catalyst binding. As a result, the Diels–Alder reactions were also performed asymmetrically. The imidazolidinone portion of **48** allowed the attachment of the appropriate chemical entity to a novel solid support resin developed previously in the group;[47] this allowed for high-throughput synthesis.

The diversity of the 242 small molecules synthesized from **47** was analyzed using defined chemical descriptors and PCA. The diversity of our library could then be compared to (i) MDDR compounds (molecular weight cutoff 650), (ii) the 3762 compounds marked as "antibacterial" in the MDDR database, and (iii) a focused library (synthesized using traditional combinatorial chemistry). Using the data set chosen, the DOS library, numerically, was even more diverse than the MDDR library; i.e., 22 (relative) diversity units for the DOS library, 19 for MDDR, 13 for the antibacterials, and 0.6 for the focused library (Fig. 6).

This compound collection, based around 18 distinct molecular frameworks, was also screened for antibacterial activity against two UK epidemic strains of methicillin-resistant *Staphylococcus aureus* (MRSA), EMRSA 15 and EMRSA 16. In addition to these MRSA strains being resistant to erythromycin and penicillins, they are responsible for the majority of MRSA infections in the UK.[48] The most active compound **49**, called gemmacin, was investigated further and showed a broad range of activity against Gram-positive bacteria (Table 1). Target identification suggested that gemmacin **49** acted as a cell membrane disruptor; this unique scaffold, identified using DOS, could potentially be useful in antibacterial development.

In conclusion, although progress has been made in the synthesis of DOS libraries and the subsequent identification of biological modulators,[49–52] accessing diverse regions of chemical space still represents a significant but potentially rewarding challenge for organic chemists.
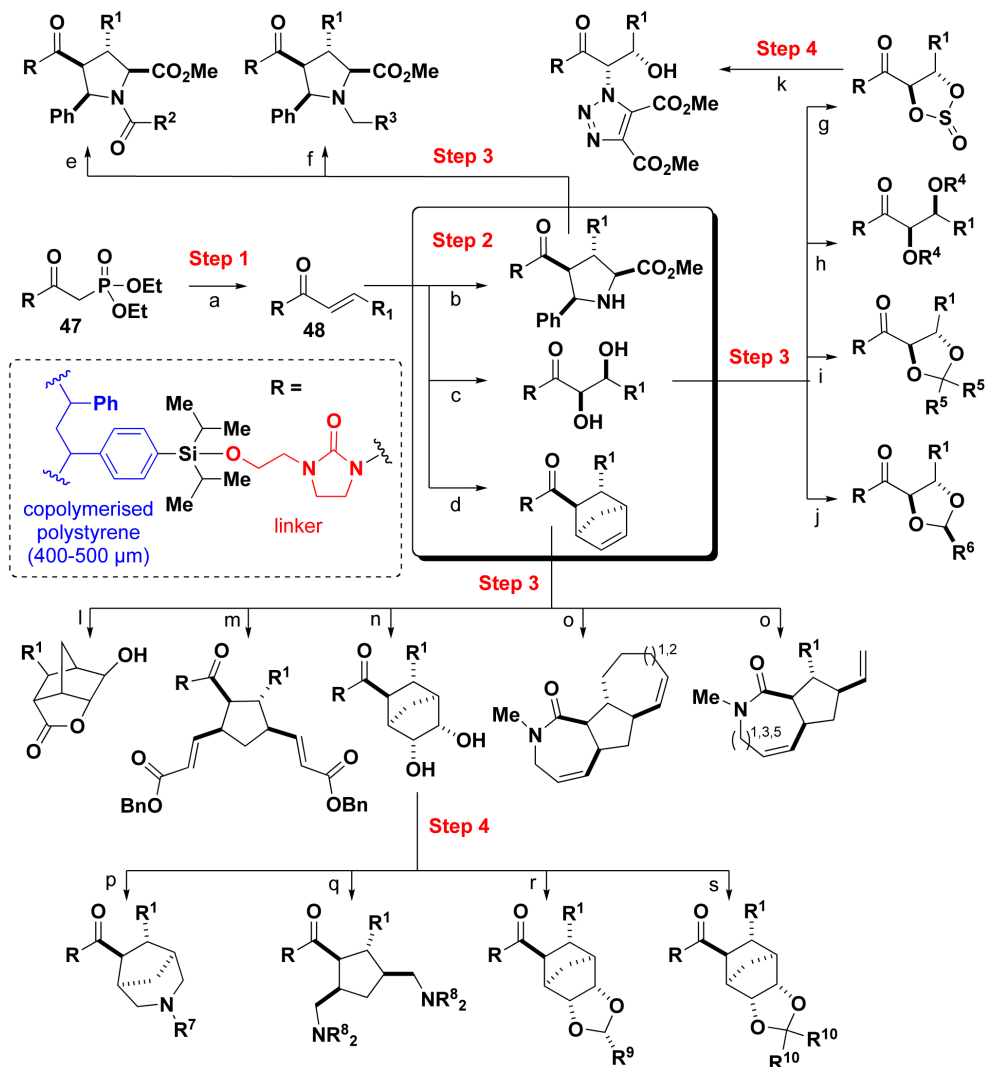
Scheme 5. Diversity-oriented synthesis of 242 compounds based on 18 discrete molecular frameworks. Conditions: (a) LiBr, 1,8-diazabicyclo[5.4.0]undec-7-ene, R^1CHO, MeCN; (b) (*R*)-QUINAP, AgOAc, i-Pr_2NEt, THF, $-78 \rightarrow$ 25°C; (c) AD-mix-β, (DHQD)PHAL, THF:H_2O (1:1); (d) chiral bis(oxazoline), Cu(OTf)$_2$, 3Å MS, CH_2Cl_2, C_5H_6; (e) R^2COCl, DMAP, pyridine, CH_2Cl_2; (f) R^3CHO, $BH_3 \cdot$ pyridine, MeOH; (g) $SOCl_2$, pyridine, CH_2Cl_2, 40°C; (h) R^4Br, Ag_2O, CH_2Cl_2, 40°C; (i) $R^5C(O)R^5$, TsOH, DMF, 65°C; (j) R^6CHO, TsOH, DMF, 65°C; (k) NaN_3, DMF, 100°C then dimethyl acetylenedicarboxylate, PhMe, 65°C; (l) *m*CPBA, CH_2Cl_2 then MeOH, 65°C; (m) CH_2=$CHCO_2$Bn, PhMe, 120°C, Grubbs II, CH_2=CH_2; (n) OsO_4, NMO, $CH_3C(O)CH_3$:H_2O (10:1); (o) RNH_2, Me_2AlCl, PhMe 120°C; then NaH, R^{11}X, DMF, THF; then PhMe, 120°C, Grubbs II, CH_2=CH_2; (p) $NaIO_4$, THF:H_2O (1:1); then R^7NH_2, $NaB(OAc)_3H$, CH_2Cl_2; (q) $NaIO_4$, THF:H_2O (1:1); then R^8NHR^8, $NaB(OAc)_3H$, CH_2Cl_2; (r) R^9CHO, DMF, TsOH, 60°C; (s) $R^{10}C(O)R^{10}$, DMF, TsOH, 60°C. DMF = *N,N*-dimethylformamide; THF = tetrahydrofuran; DMAP = *N,N*-dimethylaminopyridine; (DHQD)PHAL = hydroquinidine 1,4-phthalazinediyl diether; *m*CPBA = meta-chloroperbenzoic acid; Ts = para-toluenesulfonyl; Grubbs II = 1,3-[bis(mesityl)-2-imidazolidinylidene] dichloro (phenylmethylene) (tricyclohexylphosphine) ruthenium; NMO = 4-methylmorpholine-N-oxide.[41]
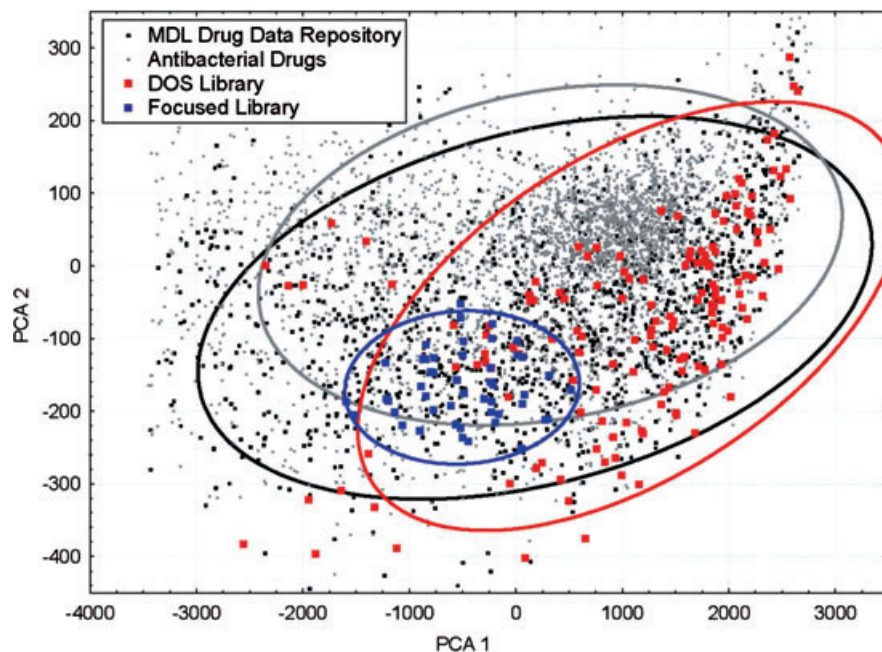
Library (MW <650)	STDDEV (PC1)	STDDEV (PC2)	STDDEV (PC3)	Average 'Chemical Space' occupied per compound
MDDR	1472.11	122.53	104.56	18.86
Antibacterials	1276.50	109.01	91.92	12.79
DOS	1176.97	135.69	139.78	22.32
Focused	473.57	43.91	31.00	0.64

Fig. 6. Visual representation of the diversity of different chemical collections in physicochemical and topological space using 184 molecular operating environment (MOE) descriptors followed by principal component analysis (PCA). The diversity-oriented synthesis (DOS) library is depicted by red squares. For comparison, a focused library (blue squares), the MDL Drug Data Repository (black dots), and antibacterial drugs (grey dots) are depicted.[41] PC = principal component; MW = molecular weight.

Table 1. Structure and activity (MIC_{50}) of gemmacin 49 with growth inhibitory activity against three strains of *Staphylococcus aureus*.
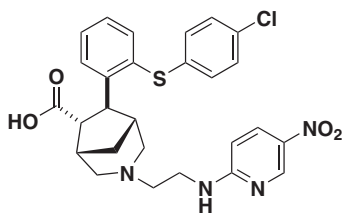
(−)-Gemmacin	MIC_{50} ($\mu g\ ml^{-1}$)		
	MSSA	EMRSA 15	EMRSA 16
(±)-Gemmacin	2	16	32
(−)-Gemmacin	**ND**	**8**	**16**
(+)-Gemmacin	ND	16	32
Erythromycin	0.5	>64	>64
Oxacillin	0.5	>32	>32

ND = not determined; MSSA = methicillin-susceptible *S. aureus*;[41] MIC_{50} = minimum inhibitory concentration for 50% of bacterial cells.

REFERENCES

[1] Collins, F. S.; Lander, E. S.; Rogers, J.; Waterston, R. H. Nature 2004, 431, 931.

[2] Spring, D. R. Chem Soc Rev 2005, 34, 472.

[3] Schreiber, S. L. Bioorg Med Chem 1998, 6, 1127.

[4] Schreiber, S. L. Chem Eng News 2003, 81, 51.

[5] Walsh, D. P.; Chang, Y. T. Chem Rev 2006, 106, 2476.

[6] MacBeath, G. Genome Biol 2001, 2, comment 2005.1.

[7] Hopkins, A. L.; Groom, C. R. Nat Rev Drug Discovery 2002, 1, 727.

[8] Lipinski, C.; Hopkins, A. Nature 2004, 432, 855.

[9] Borman, S. Chem Eng News 2004, 82, 32.

[10] Spring, D. R. Org Biomol Chem 2003, 1, 3867.

[11] Spandl, R. J.; Spring, D. R.; Bender, A. Org Biomol Chem 2008, 6, 1149.

[12] Burke, M. D.; Schreiber, S. L. Angew Chem Int Ed 2004, 43, 46.

[13] Haggarty, S. J. Curr Opin Chem Biol 2005, 9, 296.

[14] Nicholson, R. L.; Welch, M.; Ladlow, M.; Spring, D. R. ACS Chem Biol 2007, 2, 24.

[15] Butler, M. S. Nat Prod Rep 2008, 22, 162.

[16] Lam, K. S. Trends Microbiol 2007, 15, 279.

[17] Feher, M.; Schmidt, J. M. J Chem Inf Comput Sci 2003, 43, 218.

[18] Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Adv Drug Deliver Rev 1997, 23, 3.

[19] Keller, T. H.; Pichota, A.; Yin, Z. Curr Opin Chem Biol 2006, 10, 357.

[20] Abad-Zapatero, C. Drug Discovery Today 2007, 12, 995.

[21] Fergus, S.; Bender, A.; Spring, D. R. Curr Opin Chem Biol 2005, 9, 304.

[22] Xue, L.; Bajorath, J. Comb Chem High Throughput Screening 2000, 3, 363.

[23] Bender, A.; Glen, R. C. Org Biomol Chem 2004, 2, 3204.

[24] Brown, R. D.; Martin, Y. C. J Chem Inf Comput Sci 1996, 36, 572.

[25] Dobson, C. M. Nature 2004, 432, 824.

[26] Hann, M. M.; Leach, A. R.; Harper, G. J Chem Inf Comput Sci 2001, 41, 856.

[27] Schuffenhauer, A.; Brown, N.; Selzer, P.; Ertl, P.; Jacoby, E. J Chem Inf Mod 2006, 46, 525.

[28] Burke, M. D.; Berger, E. M.; Schreiber, S. L. J Am Chem Soc 2004, 126, 14095.

[29] Burke, M. D.; Berger, E. M.; Schreiber, S. L. Science 2003, 302, 613.

[30] Noren-Muller, A.; Reis-Correa, I.; Prinz, H.; Rosenbaum, C.; Saxena, K.; Schwalbe, H. J.; Vestweber, D.; Cagna, G.; Schunk, S.; Schwarz, O.; Schiewe, H.; Waldmann, H. Proc Nat Acad Sci USA 2006, 103, 10606.

[31] Dekker, F. J.; Koch, M. A.; Waldmann, H. Curr Opin Chem Biol 2005, 9, 232.

[32] Koch, M. A.; Schuffenhauer, A.; Scheck, M.; Wetzel, S.; Casaulta, M.; Odermatt, A.; Ertl, P.; Waldmann, H. Proc Nat Acad Sci USA 2005, 102, 17272.

[33] Reayi, A.; Arya, P. Curr Opin Chem Biol 2005, 9, 240.

[34] Clardy, J.; Walsh, C. Nature 2004, 432, 829.

[35] Ko, S. K.; Jang, H. J.; Kim, E.; Park, S. B. Chem Commun 2006, 2962.

[36] Zhou, C. X.; Dubrovsky, A. V.; Larock, R. C. J Org Chem 2006, 71, 1626.

[37] Messer, R.; Fuhrer, C. A.; Iner, R. H.; Haner, R. Curr Opin Chem Biol 2005, 9, 259.

[38] Goess, B. C.; Hannoush, R. N.; Chan, L. K.; Kirchhausen, T.; Shair, M. D. J Am Chem Soc 2006, 128, 5391.

[39] Tan, D. S. Nat Chem Biol 2005, 1, 74.

[40] Thomas, G. L.; Wyatt, E. E.; Spring, D. R. Curr Opin Drug Discovery Dev 2006, 9, 700.

[41] Thomas, G. L.; Spandl, R. J.; Glansdorp, F. G.; Welch, M.; Bender, A.; Cockfield, J.; Lindsey, J. A.; Bryant, C.; Brown, D. F. J.; Loiseleur, O.; Rudyk, H.; Ladlow, M.; Spring, D. R. Angew Chem Int Ed 2008, 47, 2808.

[42] Wyatt, E. E.; Fergus, S.; Galloway, W.; Bender, A.; Fox, D. J.; Plowright, A. T.; Jessiman, A. S.; Welch, M.; Spring, D. R. Chem Commun 2006, 3296.

[43] Nielsen, T. E.; Schreiber, S. L. Angew Chem Int Ed 2008, 47, 48.

[44] Kumagai, N.; Muncipinto, G.; Schreiber, S. L. Angew Chem Int Ed 2006, 45, 3635.

[45] Comer, E.; Rohan, E.; Deng, L.; Porco, J. A. Org Lett 2007, 9, 2123.

[46] Oguri, H.; Schreiber, S. L. Org Lett 2005, 7, 47.

[47] Thomas, G. L.; Ladlow, M.; Spring, D. R. Org Biomol Chem 2004, 2, 1679.

[48] Johnson, A. P.; Aucken, H. M.; Cavendish, S.; Ganner, M.; Wale, M. C.; Warner, M.; Livermore, D. M.; Cookson, B. D. J Antimicrob Chemother 2001, 48, 143.

[49] Kim, Y. K.; Arai, M. A.; Arai, T.; Lamenzo, J. O.; Dean, E. F.; Patterson, N.; Clemons, P. A.; Schreiber, S. L. J Am Chem Soc 2004, 126, 14740.

[50] Koehler, A. N.; Shamji, A. F.; Schreiber, S. L. J Am Chem Soc 2003, 125, 8420.

[51] Spring, D. R.; Krishnan, S.; Blackwell, H. E.; Schreiber, S. L. J Am Chem Soc 2002, 124, 1354.

[52] Kuruvilla, F. G.; Shamji, A. F.; Sternson, S. M.; Hergenrother, P. J.; Schreiber, S. L. Nature 2002, 416, 653.