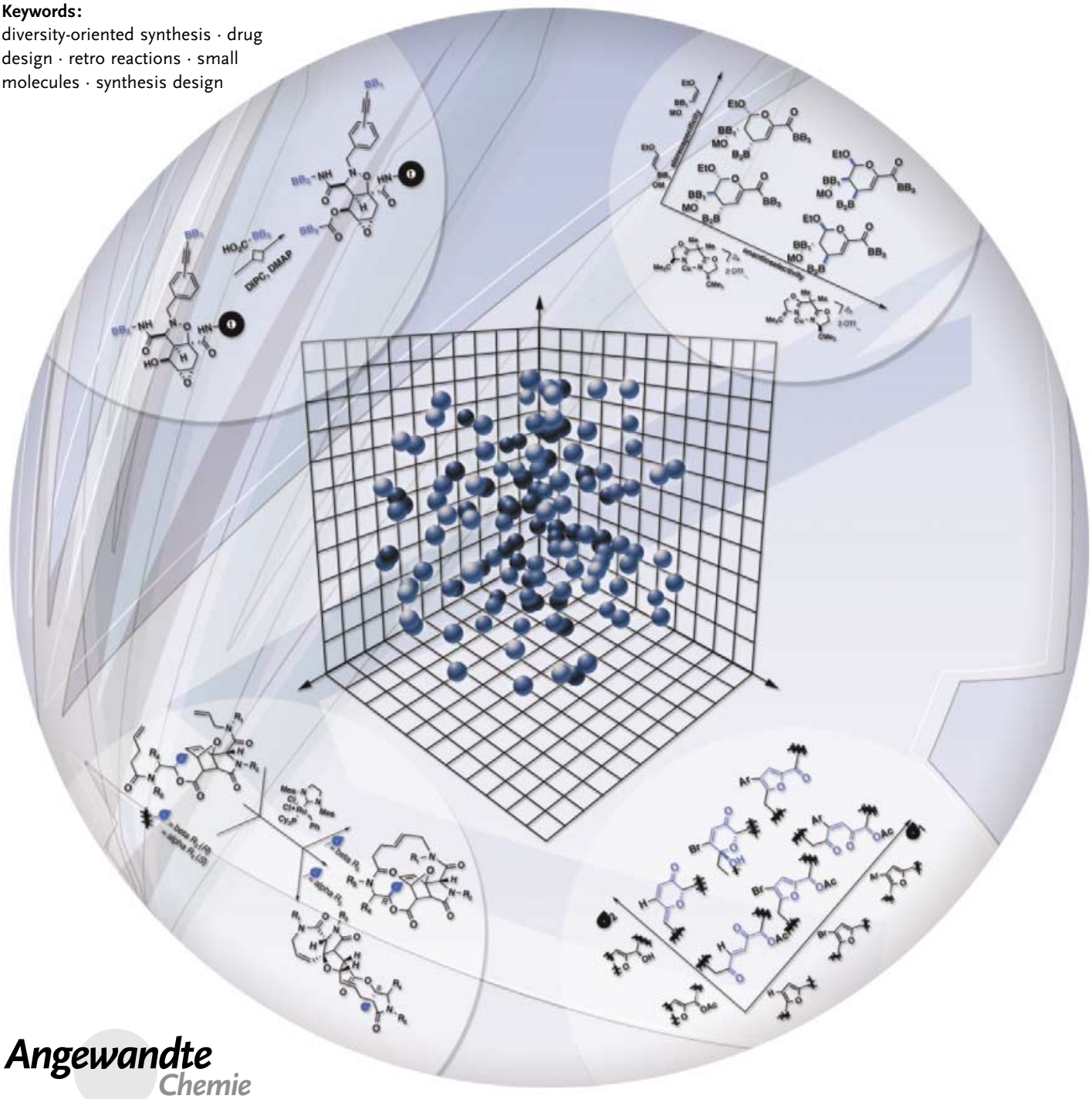
Discovery of Small Molecules

A Planning Strategy for Diversity-Oriented Synthesis

*Martin D. Burke and Stuart L. Schreiber**

Keywords:
diversity-oriented synthesis · drug design · retro reactions · small molecules · synthesis design



DOI: 10.1002/anie.200300626

In contrast to target-oriented synthesis (TOS) and medicinal or combinatorial chemistry, which aim to access precise or dense regions of chemistry space, diversity-oriented synthesis (DOS) populates chemical space broadly with small-molecules having diverse structures. The goals of DOS include the development of pathways leading to the efficient (three- to five-step) synthesis of collections of small molecules having skeletal and stereochemical diversity with defined coordinates in chemical space. Ideally, these pathways also yield compounds having the potential to attach appendages site- and stereoselectively to a variety of attachment sites during a post-screening, maturation stage. The diverse skeletons and stereochemistries ensure that the appendages can be positioned in multiple orientations about the surface of the molecules. TOS as well as medicinal and combinatorial chemistries have been advanced by the development of retrosynthetic analysis. Although the distinct goals of DOS do not permit the application of retrosynthetic concepts and thinking, these foundations are being built on, by using parallel logic, to develop a complementary procedure known as forward-synthetic analysis. This analysis facilitates synthetic planning, communication, and teaching in this evolving discipline.

From the Contents

1. Introduction

Small-molecules can exert powerful effects on the functions of macromolecules that comprise living systems. This remarkable ability makes them useful, both as research tools for understanding life processes and as pharmacologic agents for promoting and restoring health. Synthetic organic chemists aim to gain access to these compounds using three general approaches.

The first approach uses target-oriented synthesis (TOS) and relies primarily on nature to discover small-molecules with useful, macromolecule-perturbing properties. Natural compounds can be identified in screens of extract mixtures, isolated, and then structurally characterized by using a variety of spectroscopic techniques. Once such a structure has been identified, it can become a target for chemical synthesis. The aim of the synthesis effort in TOS is to access a precise region of chemical space,[1] which is often defined by a complex natural product known to have a useful function (Figure 1 A).

The second approach uses either medicinal chemistry or combinatorial chemistry and aims to explore a dense region of chemistry space in proximity to a precise region known to have useful properties (Figure 1 B). The source of the starting or lead compounds can vary and may include a natural product, a known drug, or a rationally designed structure developed from a mechanistic hypothesis and/or a crystal structure of a macromolecule of interest.

Synthetic chemists' ability to access precise or dense regions of chemistry space defined by natural products or known drugs have led to major advances in the chemical and life sciences. Nevertheless, the following question remains unanswered: Are the regions of chemistry space defined by natural products and known drugs, which have been so intensely scrutinized to date, the best or most fertile regions for discovering small-molecules that modulate macromolecular function in useful ways? Given the extraordinary potential for such small molecules to promote the understanding and betterment of human health, it is urgent that organic chemists begin to answer this basic question. One aim of diversity-oriented synthesis[2] (DOS) is to meet this challenge.

The synthesis effort in DOS aims to create a broad distribution of compounds in chemistry space, including currently poorly populated (or even vacuous) space, and in the future, in space found empirically to correlate best with desired properties (Figure 1 C). The first step toward achieving this aim is to recognize that the problem of accessing broad regions of chemistry space is different than the problem of accessing precise or dense regions. These problems are different and, therefore, present distinct challenges and require distinct solutions.

[*] Dr. M. D. Burke, Prof. S. L. Schreiber
Department of Chemistry and Chemical Biology
Howard Hughes Medical Institute
ICCB, Harvard University
12 Oxford Street, Cambridge, MA 02138 (USA)
Fax: (+1) 617-495-0751
E-mail: sls@slsiris.harvard.edu

 DOI: 10.1002/anie.200300626

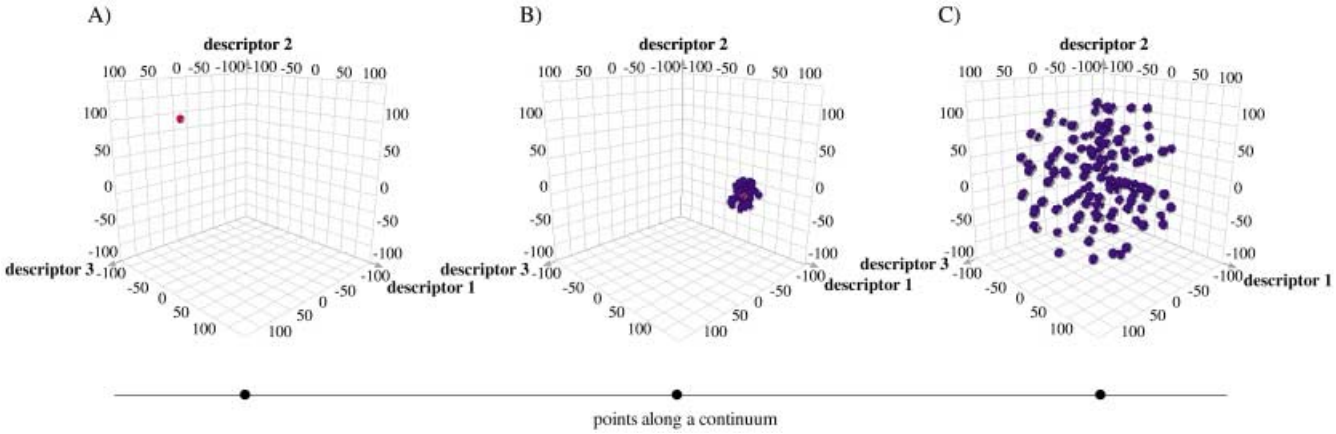


Figure 1. Comparison of TOS (A), medicinal and combinatorial chemistry (B), and DOS (C). Each three-dimensional plot is meant to represent the chemical product or collection of products derived from a single synthesis pathway. Each axis plots a calculable or measurable property of a small molecule (for example, molecular weight, solubility). A) The aim in TOS is to synthesize a single target structure having known or predicted properties (red sphere). B) The goal in medicinal and combinatorial chemistry is to synthesize a collection of analogues (blue spheres) of a target structure having known or predicted properties (red sphere). C) The aim in DOS is to populate chemistry space broadly with complex and diverse structures having unknown properties (blue spheres) as a first step in the small molecule discovery process. In some ways, these three approaches to synthesizing small-molecules represent points along a continuum.

TOS as well as medicinal and combinatorial chemistries have been advanced by the development (beginning over 40 years ago) of a general planning strategy known as retrosynthetic analysis,[3] in which a complex target is transformed into a sequence of progressively simpler structures by formally performing chemical reactions in the reverse-synthetic direction. Prior to this, strategic solutions to the problems of synthesizing different target structures were developed on a case-by-case basis. The introduction of a general planning strategy had a revolutionizing impact on these fields in at least three ways: by assisting chemists in planning efficient synthesis pathways that access complex target structures, by creating a language and defining concepts to facilitate communication between colleagues, and by providing a framework for teaching this field to new generations of organic chemistry students.

Retrosynthetic concepts and thinking depend on the existence of a defined target structure. Retrosynthetic analysis cannot be effectively applied in DOS because there is no single target structure. However, the foundations of retrosynthetic analysis are being used (by applying parallel logic) to develop a complementary strategy to facilitate synthetic planning, communication, and teaching within the realm of DOS.

2. Retrosynthetic Analysis and Forward-Synthetic Analysis

Synthesis pathways in TOS are linear and convergent, and they are planned in the reverse-synthetic direction by using retrosynthetic planning, which aims to move in the direction of complex→simple. In contrast, in DOS, where the structural complexity of the individual compounds and the structural diversity of the overall collection are maximized, synthesis pathways are branched and divergent, and they are planned in the forward-synthetic direction[4] by using forward-synthetic analysis. Forward-synthetic planning aims to move in the direction of simple and similar→complex and diverse. (Maximizing diversity and correlating structure with activity are inherently computational challenges. This article will not focus on these elements of DOS despite their critical nature;

Stuart L. Schreiber was born in Virginia, USA in 1956. He received his BA at the University of Virginia in 1977, completed his PhD at Harvard University under the supervision of R. B. Woodward and Y. Kishi, and joined the faculty at Yale University in 1981. In 1988 he returned to Harvard and is now an Investigator at the Howard Hughes Medical Institute, and Morris Loeb Professor and Chair of the Department of Chemistry and Chemical Biology. His research interests focus on developing systematic ways to explore biology and medicine using small molecules.

Martin D. Burke was born in 1976 in Maryland, USA. He graduated in chemistry from Johns Hopkins University in 1998 and recently completed his PhD thesis in synthetic organic chemistry under the supervision of Prof. Schreiber at Harvard University.

here we simply note the recent launching of ChemBank on the Internet[5] as a first step towards providing the requisite tools to the chemical community.)

The basic subunit of retrosynthetic planning is the transform, that is, the theoretical transformation of a product into a substrate by formally performing a chemical reaction in the reverse-synthetic direction. To make use of a transform in retrosynthetic analysis one must first identify the corresponding "retron", that is, the enabling structural subunit ("keying element") that permits its application, in the chemical target. In contrast, the basic subunit of forward-synthetic planning is the process, namely, the transformation of a collection of substrates into a collection of products by performing a number of chemical reactions together in the forward-synthetic direction. The key element for implementing a process is common reactivity, that is, the inherent chemical reactivity common to a collection of compounds that makes them all potential substrates for the same reaction(s). To plan efficient DOS pathways containing iterative processes, it is critical to identify products-equals-substrates relationships, such that the products of one process have some common inherent chemical reactivity that makes them all potential substrates for another process.

TOS and DOS share the aim of accessing complex structures efficiently. Structurally simplifying transforms are critical in TOS when devising a retrosynthesis for a complex target structure, and iterative application of these transforms can lead to a plan for an efficient synthesis.[6] In contrast, when planning a diversity-oriented synthesis in the forward direction, complexity-generating reactions are most valuable for accessing complexity in an efficient manner. Moreover, identification of pairwise relationships, where the product of one complexity-generating reaction is the substrate for another, can lead to highly complex products with just a few synthetic steps.

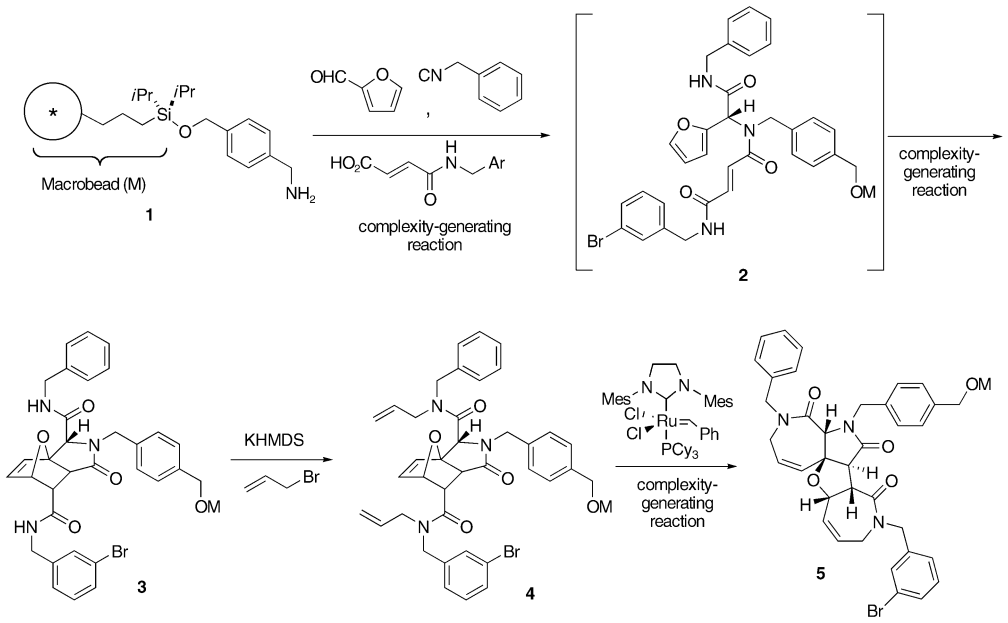
TOS does not share the aim of accessing diversity. The aim in medicinal and combinatorial chemistry is to access diversity to some degree, and usually involves synthesizing analogues of a given target structure. This can be accomplished efficiently using solid-phase synthesis[7] to append different sets of building blocks to a common molecular skeleton.[8] Retrosynthetic planning is used in this context to devise pathways to a target structure that permit the addition of diverse sets of building blocks during the actual synthesis. If this common skeleton contains multiple reactive sites with potential for orthogonal functionalization, the powerful technique of split-pool synthesis[9] can be used to access all possible combinations of building blocks (namely, the complete matrix) efficiently. In DOS, where there is no one target structure, the problem of diversity is subdivided into three diversity elements: appendages (for example, building blocks and "σ elements"; see below), stereochemistry, and molecular skeletons. Forward-synthetic planning aimed at accessing these diversity elements relies on the use of diversity-generating processes, which is defined as the transformation of a collection of relatively similar substrates into a collection of more diverse products. In an ideal DOS pathway all of the products of one diversity-generating process are substrates for another, thus making it possible to use split-pool synthesis to access combinatorially matrices of building blocks, stereochemical isomers, and even molecular skeletons.

3. Complexity-Generating Reactions and their Use in DOS To Generate Complex Products Efficiently (Simple→Complex)

The structures and functions of natural products suggest that structural complexity may be positively correlated with macromolecule-perturbing function and specificity of action. This correlation is particularly striking in small molecules known to disrupt protein–protein interactions. Therefore, it is a goal of DOS to access small molecules with complex molecular skeletons,[10] and forward-synthetic planning aims to proceed in the direction of simple→complex. Moreover, in contrast to the relatively flat molecular skeletons often used in medicinal and combinatorial chemistry that tend to project appendages outward along the perimeter of a circle, the aim in DOS is to access more globular or spherical molecular skeletons to which substituents can be potentially appended along the surface of a sphere during a post-screening, optimization stage.

To maximize efficiency and, in the case of researchers affiliated with ICCB,[11] to be compatible with one-bead/one-stock solution technology platforms, synthesis pathways in DOS should be no more than three to five steps (which leaves no room for protective-group manipulations). Therefore, to achieve skeletal complexity in DOS it is critical to identify and to implement complexity-generating reactions that rapidly assemble complex molecular skeletons. Moreover, the identification in the forward direction of pairwise relationships, where the product of one complexity-generating reaction is the substrate for another,[12] can lead to high levels of molecular complexity in a very efficient manner.

For example, as shown in Scheme 1, the Ugi four-component coupling reaction[13] can be used to assemble a complex product from simple starting materials in a single step. If those simple starting materials are selected to include both a diene and a dienophile, then the product of this first complexity-generating reaction **2** is a substrate for another, namely an intramolecular Diels–Alder reaction.[14] The identification of this pairwise relationship made it possible to generate complex molecular skeleton **3** from simple starting materials in a single synthetic step. Moreover, it was recognized that the product **3** is almost a substrate for an additional complexity-generating reaction, namely a ring-opening/ring-closing metathesis,[15] and can be transformed into such a substrate (**4**) by bisallylation with KHMDS and allyl bromide. (This type of forward-synthetic planning is analogous to the identification of partial retrons and the use of functional-group manipulations to make the application of structurally simplifying transforms possible in retrosynthetic analysis.) Treatment of **4** with the Grubbs catalyst[16] resulted in a complexity-generating ring-opening/ring-closing metathesis reaction to generate product **5**, which has a highly complex 7-5-5-7 polycyclic molecular skeleton.



Scheme 1. Three-step synthesis of a complex 7-5-5-7 polycyclic ring system using complexity-generating reactions having the product-equals-substrate relationship. KHMDS = potassium 1,1,1,3,3,3-hexamethyldisilazine, Mes = 2,4,6-trimethylphenyl, Cy = cyclohexyl.

4. Diversity-Generating Processes and their Use in DOS To Generate Diverse Products Efficiently (Similar→Diverse)

DOS pathways aim to proceed in the direction of similar structures→diverse structures to gain access to broad regions of chemistry space efficiently. To achieve this requires planning (in the forward direction) a series of products-equals-substrates relationships, that is, the products of one diversity-generating process should share some common inherent chemical reactivity. This common reactivity serves as a keying element that makes the products collective substrates for a subsequent diversity-generating process. The goal of achieving diversity can be simplified by considering three distinct diversity elements: appendages, stereochemistry, and skeletons.

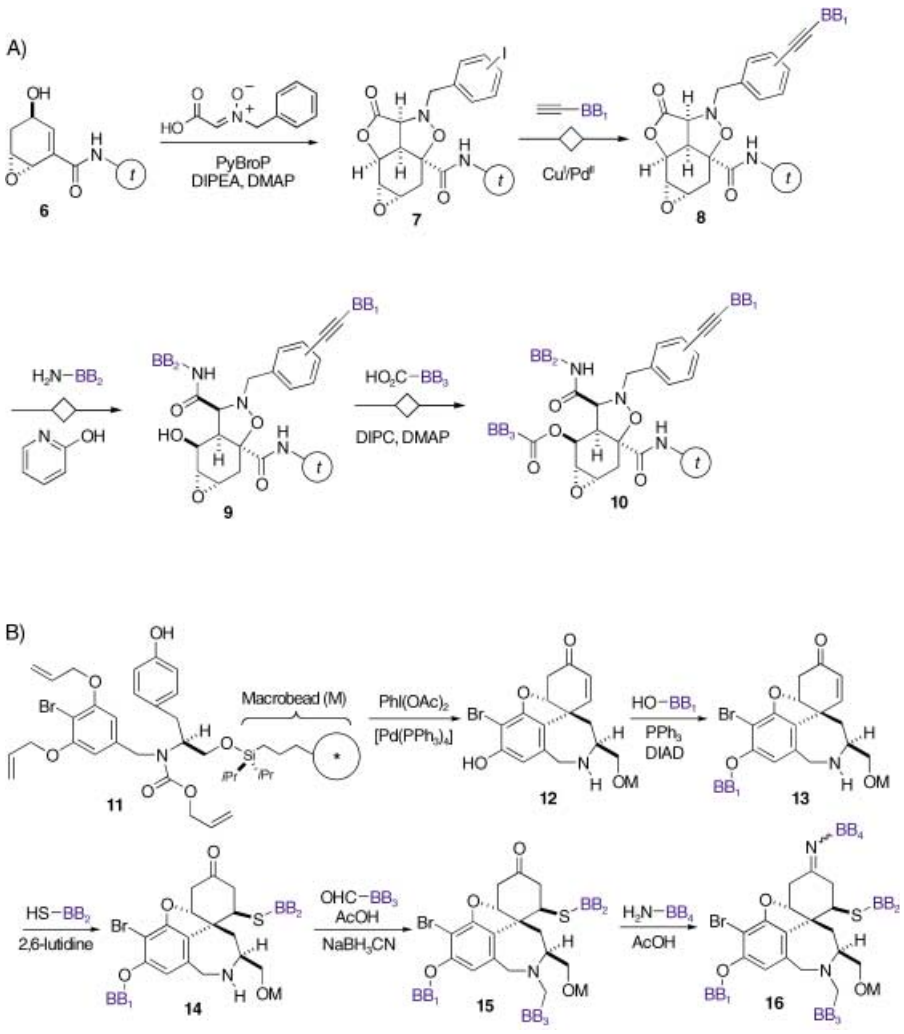
4.1. Appendage Diversity

The simplest diversity-generating process is the central feature of combinatorial chemistry and involves the use of coupling reactions to attach different appendages to a common molecular skeleton. In forward-synthetic analysis these are referred to as appending processes. If a molecular skeleton has multiple reactive sites with potential for orthogonal functionalization, then the technique of split-pool synthesis can be used to harness the power of combinatorics (a multiplicative increase in the number of products with an additive increase in the number of reaction conditions), and thereby generate all possible combinations of appendages (that is, the complete matrix) efficiently.

The origins of DOS were combinatorial chemistry efforts that simply used increasingly sophisticated organic transformations. These efforts began with a complexity-generating reaction to yield a single, complex molecular skeleton having several attachment points followed by a series of diversity-generating appending processes (potentially in split-pool format) to attach all possible combinations of building blocks to this common skeleton. This one-synthesis/one-skeleton approach has proven to be highly general and capable of generating hundreds, thousands, or even millions of distinct small molecules in just three to five steps.[8,17,18]

For example, a complexity-generating, consecutive transesterification–cycloaddition reaction was used to generate, in one step, the tetracyclic skeleton **7** with potential for functionalization through a series of diversity-generating appending processes (Scheme 2A).[17] A Sonogashira coupling reaction was first used to append a diverse collection of alkyne building blocks (BB_1) to the iodoaryl moiety of **7** and thereby to generate the collection of more diverse products **8**. Although these products differ in the identity of BB_1, they all have a common electrophilic lactone moiety. This common reactivity makes this collection of products **8** of the first diversity-generating appending process a collection of potential substrates for another appending process, namely, an amine-mediated lactone-opening reaction that generated a collection of new products **9**. Similarly, while members of this new collection of products **9** differ in the identity of both BB_1 and BB_2, they all share a common nucleophilic secondary hydroxy group, thus making them all substrates for a third appending process that resulted in their coupling with a collection of carboxylic acid building blocks (BB_3). This series of products-equals-substrates relationships made it possible to carry out this four-step synthetic pathway by using split-pool synthesis and thereby generate the complete matrix of building blocks **10** in a highly efficient manner.

A second example that vividly illustrates the power of a complexity-generating reaction is shown in Scheme 2B. In this case, a biomimetic, complexity-generating oxidative



Scheme 2. Building-block diversity generated combinatorially. See text for details. A) Split-pool synthesis of compounds derived from Shikimic acid. The encircled "*t*" represents a solid support of tentagel. The diamond-filled arrow is used to represent an appending process carried out in split-pool format. B) Diversity-oriented synthesis of galanthamine-related compounds. PyBroP = bromotris(pyrrolidino)phosphonium hexafluorophosphate, DIPEA = *N*,*N*-diisopropylethylamine, DMAP = 4-dimethylaminopyridine, DIPC = diisopropylcarbodiimide, DIAD = diisopropylazodicarboxylate.

cyclization reaction was used to transform acyclic precursor **11** into the rigid skeleton **12**, which has four sites of potential reactivity (two nucleophilic and two electrophilic) that can each be orthogonally functionalized by a series of diversity-generating appending processes.[18] In the first of these appending process, a Mitsunobu reaction was used to couple a diverse collection of building blocks derived from primary alcohols to the phenolic alcohol of **12**. The products **13** of this process share a common cyclic enone functionality, which was selectively functionalized by the conjugate addition of a collection of thiols (BB_2). While the products of this second appending process differ in terms of the identities of BB_1 and BB_2, they all share a common nucleophilic secondary amine, which makes them all substrates for a third appending process involving coupling to a diverse collection of aldehyde, acid chloride, and isocyanate building blocks (BB_3, only the aldehydes are shown). The resulting collection of products **15** represents all possible combinations of a three-dimensional matrix of building blocks, yet they all share a common electrophilic ketone. This common moiety imparts common reactivity to all the members of this collection, and made it possible to carry out a final appending process with a diverse collection of hydrazine and hydroxylamine building blocks (BB_4). In this example a consecutive series of four products-equals-substrates relationships enabled the efficient generation of a four-dimensional, combinatorial matrix of building block diversity elements appended to a complex molecular skeleton **16**.

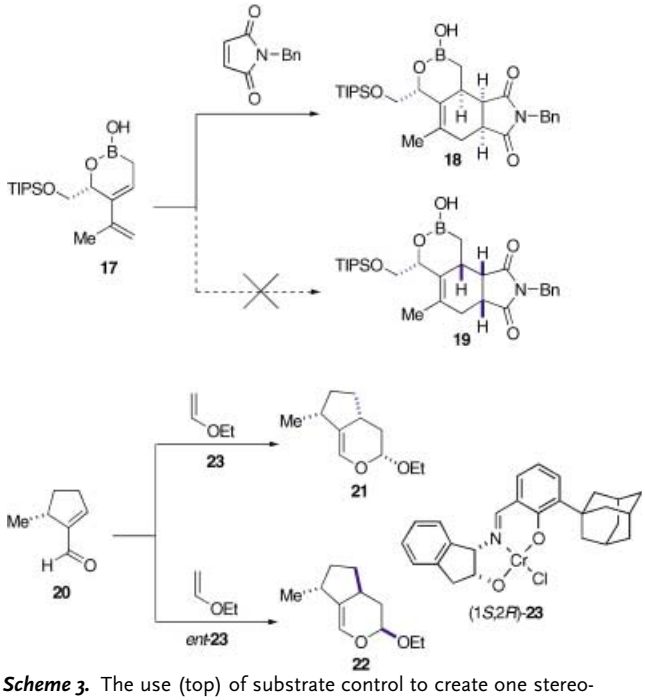
This one-synthesis/one-skeleton approach has proven to be general and highly efficient;[8] however, its impact in the academic and pharmaceutical realms has thus far been limited.[19] This may be because compounds having a common molecular skeleton display chemical information similarly in three-dimensional space, thus limiting the pool of potential binding partners to only those macromolecules with a complementary three-dimensional binding surface. Thus, an important (and intellectually challenging) aim in DOS is to develop efficient synthesis pathways that yield products that represent many diverse displays of chemical information in three-dimensional space. To achieve this goal it is necessary to gain efficient access to stereochemical and skeletal diversity.

4.2. Stereochemical Diversity

Stereochemical diversity increases the number of relative orientations of potential macromolecule-interacting elements in small molecules. It can best be achieved by using stereospecific reactions that proceed with enantio- or diastereoselectivity. The corresponding transforms for these types of processes are well-known in the context of retrosynthetic planning. Since diversity-generating processes involve the transformation of a collection of substrates into a collection of products, it is critical that the processes used to generate new stereogenic centers are both selective and general.[20] The collective transformation of chiral substrates into products having increased stereochemical diversity (namely, diaster-

eoselective diversity-generating processes) requires powerful reagents that can override substrate bias and deliver diastereomeric products with very high selectivity.[21]

For example, a diastereoselective intermolecular Diels–Alder reaction was used to transform the chiral dialkenylboronic acid **17** into the cycloadduct **18**, with the selective formation of three new stereogenic centers (Scheme 3).[22]



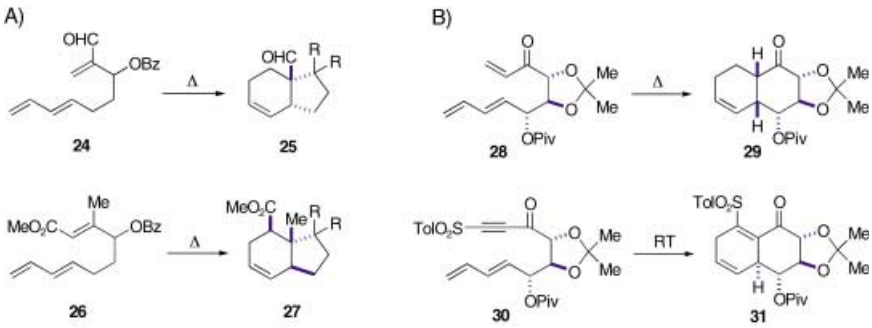
Scheme 3. The use (top) of substrate control to create one stereoisomer selectively and (bottom) of a chiral reagent **23** to override the stereochemical bias of a chiral substrate—a possible solution to the challenge of stereochemical diversity in DOS. TIPS = triisopropylsilyl, Bn = benzyl.

This reaction and its relatives represent a promising development for application to DOS pathways. It also illustrates the double-edged sword of highly stereoselective reactions. Since the diastereoselectivity of this transformation is under a powerful substrate control (steric interactions with the TIPS-protected hydroxymethyl group direct cycloaddition to the less sterically hindered face of the diene) it may prove challenging to generate the opposite diastereomeric product **19**.

Clues for progress come from advances made in double (more generally, multiple) diastereoselection reactions.[23] For example, Jacobsen and co-workers have demonstrated the use of chiral catalyst **23**, which can override the stereochemical bias of a chiral substrate and generate diastereomeric products with high selectivity.[24,25] Since transformations with such a catalyst are reagent-controlled, it is possible to use both enantiomers of **23** in a diastereoselective stereochemical diversity-generating process and thereby transform a common, chiral substrate into a collection of products having increased stereochemical diversity. For example, (1*S*,2*R*)-**23** was used to transform chiral enal **20** into dihydropyran **21** through a catalyst-controlled inverse-electron-demand diastereoselective hetero-Diels–Alder reaction.[25] Alternatively, it was possible to override the stereochemical bias of the chiral substrate and generate the diastereomeric dihydropyran **22** using the enantiomer of this catalyst, (1*R*,2*S*)-**23**. The discovery of these types of powerful reagents is critical to achieving stereochemical diversity in DOS. While catalysts such as **23** are capable of controlling the face selectivity of one coupling partner (in this case the chiral enal substrate), the development of double-diastereoselective reagents that can override the face selectivity of both coupling partners, for example, to achieve *exo* versus *endo* selectivity in the Diels–Alder reaction, would be highly valuable. For example, chemists have succeeded in discovering highly effective catalysts to yield the *endo*-Diels–Alder product of cyclopentadiene and acrolein enantioselectively. It remains as a formidable challenge, however, to develop an effective catalyst for the same reaction that yields the *exo* product enantioselectively.

While certain types of stereochemical flexibility (for example, the ability to achieve both *exo* and *endo* relative face selectivity in a Diels–Alder reaction process) are sometimes difficult to achieve in intermolecular reactions, there are a number of examples of this type of stereochemical control in intramolecular transformations. In the Diels–Alder example, Alder's "*endo* rule" is often not obeyed when the reaction is performed in the intramolecular mode.[26] In this case, subtle changes in the structure of substrates can exert powerful effects on the stereochemical outcome of intramolecular reactions that produce new stereogenic centers.[26–28] Such controlling elements may prove to be valuable for achieving stereochemical diversity in DOS.

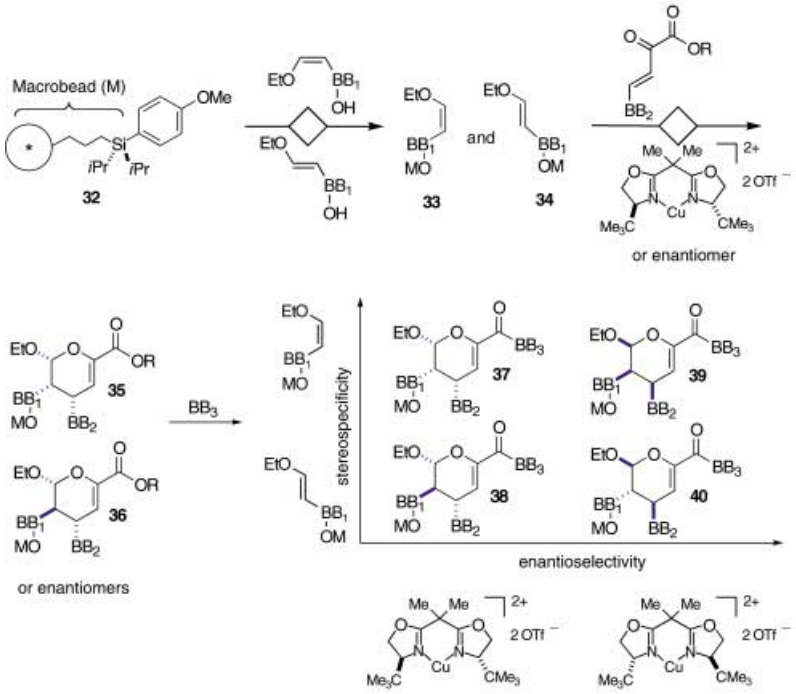
For example, Roush and co-workers found that the position of the activating carbonyl group of the dienophile in substrates such as **24** and **26** can control *exo* versus *endo* selectivity for intramolecular Diels–Alder reactions (Scheme 4 A).[26] Substrate **24**, with an activating aldehyde group on the internal position of the dienophile, yields predominantly the *cis*-fused perhydroindan ring system (*exo* product). Alter-



Scheme 4. Subtle changes in substrate structure can dictate distinct stereochemical outcomes for intramolecular reactions that generate new stereogenic centers. Bz = benzoyl, Tol = tolyl, Piv = trimethylacetyl.

natively, the terminally activated substrate **26** undergoes an intramolecular Diels–Alder reaction to yield the *trans*-fused (*endo*) cycloadduct as the major product. Sulikowski and coworkers reported that intramolecular Diels–Alder cycloaddition with vinyl ketone **28** and alkynone **30** led to cycloaddition adducts with opposite diastereoselectivity (Scheme 4B).[28] It is conceivable that these types of stereochemical diversity-generating transformations could be carried out with spatially segregated, pooled substrates under common reaction conditions.

Forward-synthetic planning that incorporates multiple stereochemical diversity-generating processes into a single pathway should also make it possible to generate stereochemical diversity in a combinatorial fashion, analogous to the ability of appending processes to generate building-block diversity in a combinatorial manner. An early example of this is shown in the DOS pathway in Scheme 5, in which both stereospecific and enantioselective stereochemical diversity-generating processes were used to generate a combinatorial matrix of four stereoisomeric products.[29]



Scheme 5. Combinatorial stereochemical diversity: both stereospecific and enantioselective stereochemical diversity-generating processes were used to generate a combinatorial matrix of four stereoisomeric products. Tf = trifluoromethanesulfonyl.

4.3. Skeletal Diversity

DOS pathways that yield collections of products with many distinct molecular skeletons are particularly effective at achieving a diverse display of chemical functionality in three-dimensional space. There are, at present, two different strategies for planning DOS pathways that generate skeletal diversity.

The first strategy involves using different reagents to transform a common substrate with the potential for diverse reactivity into a collection of products having distinct molecular skeletons (Figure 2A).[2,30] This approach is analogous to the natural process of cell differentiation in which a

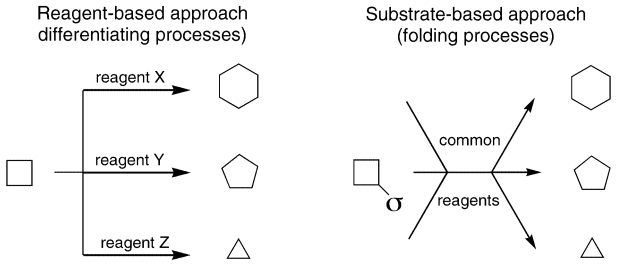
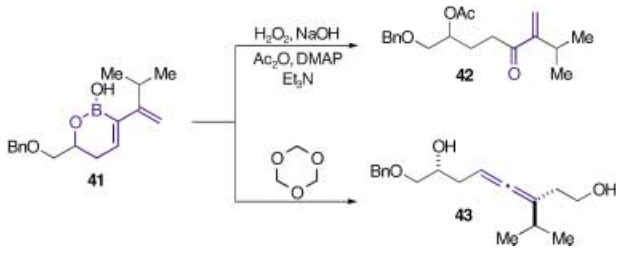


Figure 2. Two general approaches for planning synthesis pathways that generate skeletal diversity. See text for details.

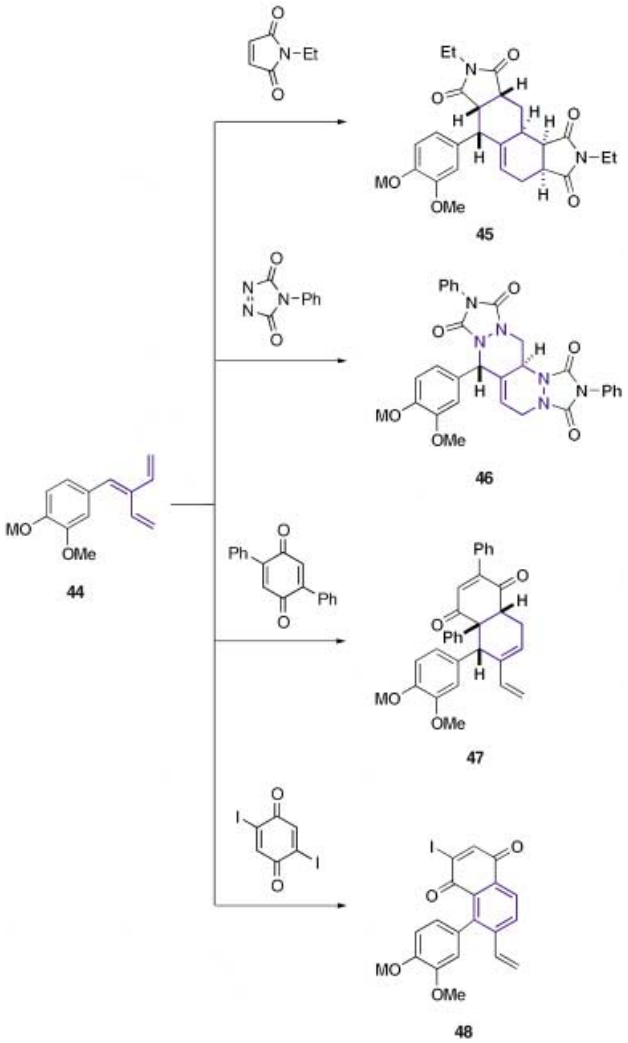
pluripotent stem cell is transformed into different cell types on exposure to distinct differentiation factors. These reagent-based skeletal diversity-generating transformations are, therefore, also referred to as differentiating processes. For example, the unsaturated, cyclic dialkenylboronic ester **41** has potential for diverse reactivity, and thus, different reagents can be used to transform this common, pluripotent substrate into different products (Scheme 6; two are shown), each having a distinct molecular skeleton.[31] Treatment of **41** with hydrogen peroxide and sodium hydroxide effects an oxidation which leads to enone **42**. Alternatively, treatment of the same substrate with 1,3,5-trioxane effects transformation into the trisubstituted allene **43**.

Another example of this reagent-based approach for generating skeletal diversity is shown in Scheme 7. It was determined that the Fallis-type[32] triene **44** is another pluripotent substrate that can be transformed into a collection of products with distinct molecular skeletons by the actions of different reagents.[33] For example, treatment of **44** with highly reactive, cyclic disubstituted dienophiles such as ethyl maleimide led to double cycloaddition reactions and yielded unsaturated decalin skeletons functionalized with maleimide-derived building blocks (for example, **45**). Treatment of the same substrate **44** with a different reagent, specifically a substituted triazol-3,5-dione, produced an unsaturated tetraazadecalin skeleton **46** through a hetero-Diels–Alder reaction. Treatment of **44** with less-reactive tri- and tetrasubstituted dienophiles resulted in single cycloaddition reactions and yielded functionalized cyclohexene derivatives such as **47**. Alternatively, treatment of **44** with halogenated quinones resulted in cycloaddition followed by spontaneous dehydrohalogenation and aromatization to yield benzene derivatives such as **48**.

In contrast to appending processes, these differentiating processes have not (as of yet) been used to generate skeletal



Scheme 6. A differentiating process: the use of different reagents to transform a common, pluripotent substrate into a collection of products having distinct molecular skeletons.



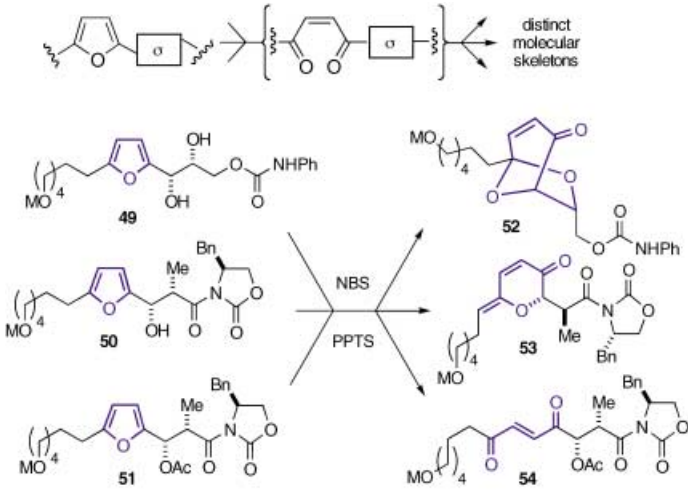
Scheme 7. The transformation of a common, pluripotent substrate into products having distinct skeletons by the actions of different reagents.

diversity combinatorially. Doing so will require the identification of differentiating processes having the products-equals-substrates relationship, that is, all of the skeletally distinct products of one differentiating process must share a common chemical reactivity that makes them all potential substrates for another differentiating process. This type of forward-synthetic planning is challenging, and will require nonmutually exclusive approaches to the two, potentially conflicting, goals of maximizing structural diversity and maintaining common reactivity.

An alternative synthesis strategy circumvents this potential conflict. In this case, diverse skeletons of small molecules can be accessed combinatorially by transforming a collection of substrates having different appendages that pre-encode skeletal information (called σ elements) into a collection of products having distinct molecular skeletons using common reaction conditions (Figure 2B).[34] This strategy is analogous to the natural process of protein folding,[35] in which different structural information pre-encoded in primary amino acid sequences is transformed into structurally diverse macromolecules using a common folding buffer. Thus, these substrate-based skeletal diversity-generating transformations are referred to as folding processes in forward-synthetic analysis. An advantage of this approach is that sets of σ elements can be identified that act in combination, that is, a matrix of σ elements can pre-encode all combinations of distinct skeletal outcomes.

These folding processes can be planned by first identifying a relatively unreactive core structure that can be transformed into a more reactive intermediate upon treatment with mild reagents. Distinct skeletal outcomes can then be pre-encoded into a collection of substrates by attaching to this common core different appendages (σ elements) having complementary reactivity with the latent, reactive intermediate. Mild conditions can then be used to liberate the reactive intermediate and to realize the pre-encoded, complementary reactivity, thus resulting in the formation of different skeletons.

The aromatic furan ring, for example, is a relatively unreactive core structure that, upon treatment with a mild oxidant, can be transformed into a more reactive, electrophilic *cis*-enedione intermediate.[36] As shown in Scheme 8, by appending three distinct two-carbon side chains containing



Scheme 8. A skeletal diversity-generating folding process: the transformation of substrates having different σ elements (that is, appendages that pre-encode skeletal information) into products having different skeletons under a common set of reaction conditions.

two, one, or zero nucleophilic hydroxy groups to a common furan core, it was possible to transform three structurally similar substrates into three products having distinct molecular skeletons through a common set of oxidative and acidic reaction conditions (NBS and PPTS, respectively). Furan derivative **49**, with a side chain containing two nucleophilic hydroxy groups, underwent NBS-mediated oxidative ring expansion and subsequent ketalization[37] to yield the [3.2.1]bicyclic ketal **52.** Alternatively, the Evans aldol product **50**, with a single hydroxy group on its side chain, underwent oxidative ring expansion and acid-catalyzed dehydration to yield the alkylidene pyran-3-one **53**. Finally, treatment of furan derivative **51**, with no nucleophilic hydroxy groups on its two-carbon side chain, under the same reaction conditions resulted in oxidative opening of the furan ring followed by olefin isomerization[38] to yield the *trans*-enedione **54**.

The use of this substrate-based approach to generate skeletal diversity combinatorially (namely, achieving a multiplicative increase in skeletons with an additive increase in appendages) requires at least two sets of σ elements that can be appended at different sites and function in combination to pre-encode a matrix of distinct skeletal outcomes. For example, it was determined that different appendages at the 4-position of the furan core can also pre-encode distinct molecular skeletons. Moreover, a combinatorial matrix of these two different σ elements (H, Br, or aryl at the 4-position of furan combined with OH or OAc on the α carbon atom; a 3×2 matrix) can pre-encode a complete, combinatorial matrix of six distinct skeletal outcomes that were realized in a one-pot reaction under common conditions (Scheme 9).

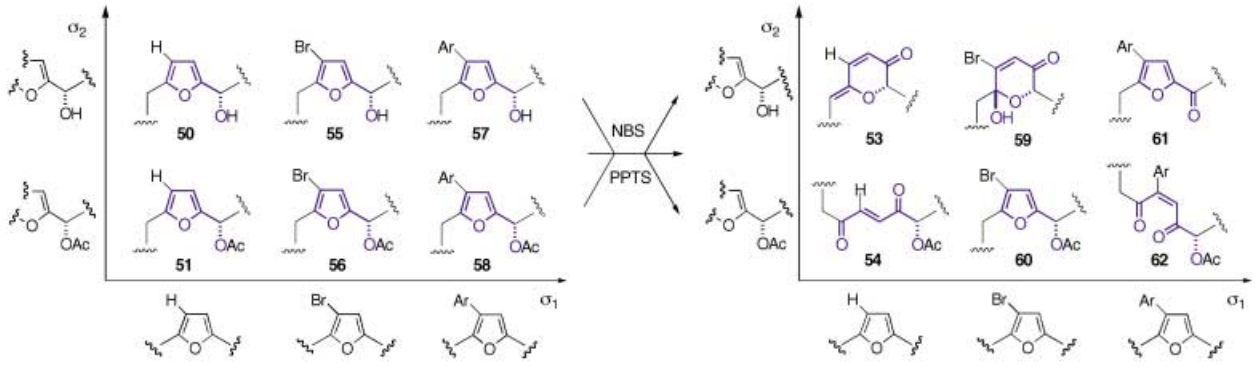
In contrast to the one-synthesis/one-skeleton approach (which typically involves forming a single molecular skeleton early in a synthesis), a folding process can be used to generate new skeletons at the end of a synthesis pathway. This approach facilitates the generation of functionalized skeletons that might otherwise be difficult to access, such as those having building blocks coupled through carbon–carbon bonds at stereogenic quaternary carbon centers (for example, **59**) and/or potentially unstable structural elements (for example, enediones **54** and **62**). Additionally, σ elements can be attached to a common molecular skeleton by using appending processes (similar to the way building blocks are appended in the one-synthesis/one-skeleton approach). The maintenance of structural similarity and, therefore, common reactivity until late in the synthesis pathway facilitates the realization of this approach using the split-pool technique. These potential advantages were realized in the context of a five-step, fully encoded, split-pool synthesis which yielded a collection of products representing overlapping, combinatorial matrices of molecular skeletons and appended building blocks in both enantiomeric and diastereomeric forms (Scheme 10).

5. Integrated Forward-Synthetic Analysis for Generating Both Complexity and Diversity (Simple and Similar→Complex and Diverse)

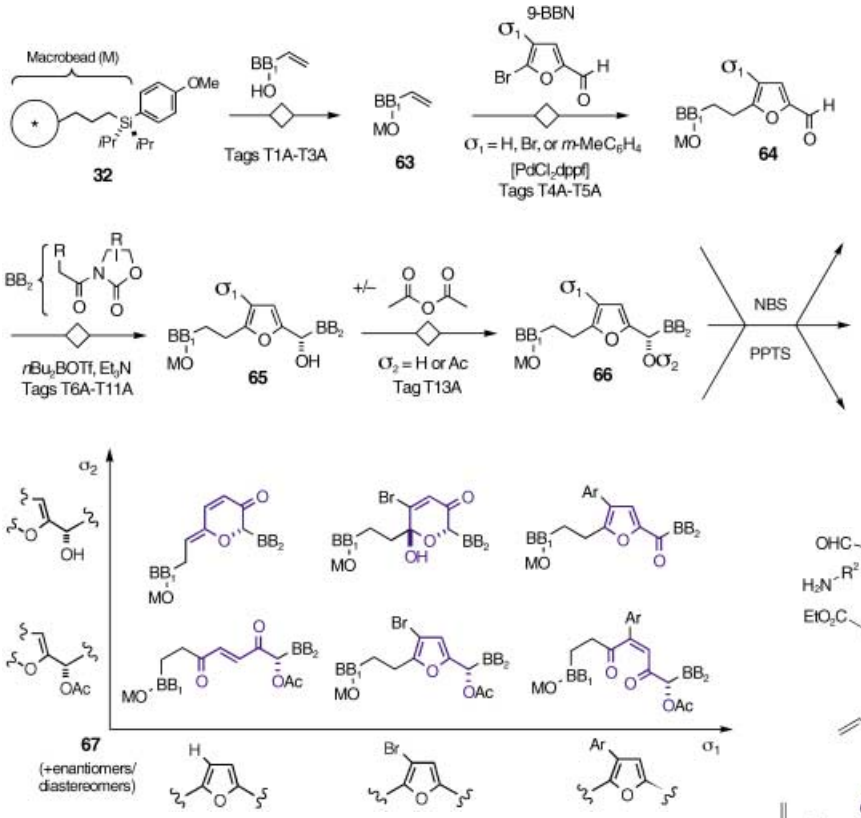
As described in the previous sections, two goals of DOS (namely, generating structural complexity and structural diversity in an efficient manner) can be considered independently, and different strategies have been developed to address each of these distinct challenges. However, achieving high levels of both complexity and diversity in the context of a single DOS pathway will require integrated forward-synthetic planning. One logical approach is to incorporate complexity-generating reactions into stereochemical and skeletal diversity-generating processes. This is another challenging frontier—however, some recent progress suggests that this approach can be effective (Scheme 11).[39]

6. A Challenge for Synthesis in the Future.

DOS, as it has evolved in 2004, entails the development of pathways leading to the efficient (3–5 step) synthesis of collections of small molecules having rich skeletal and stereochemical diversity and, we propose here, the potential to attach appendages (during a post-screening, maturation stage) site- and stereoselectively to several attachment sites (Scheme 12). In contrast to earlier efforts in DOS, as well as to past and present efforts in medicinal and combinatorial chemistry, we suggest that appending processes involving building blocks (as distinct from σ elements) should be of less importance in the original synthesis. Incorporating untapped

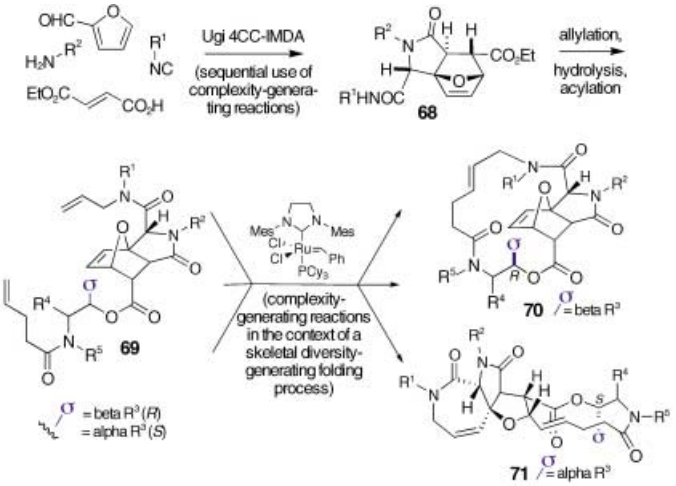


Scheme 9. Combinatorial skeletal diversity: the transformation of a collection of substrates having a combinatorial matrix of σ elements appended to a common molecular skeleton into a collection of products that represents a complete, combinatorial matrix of distinct skeletal possibilities. NBS = *N*-bromosuccinimide, PPTS = pyridinium *p*-toluenesulfonate.



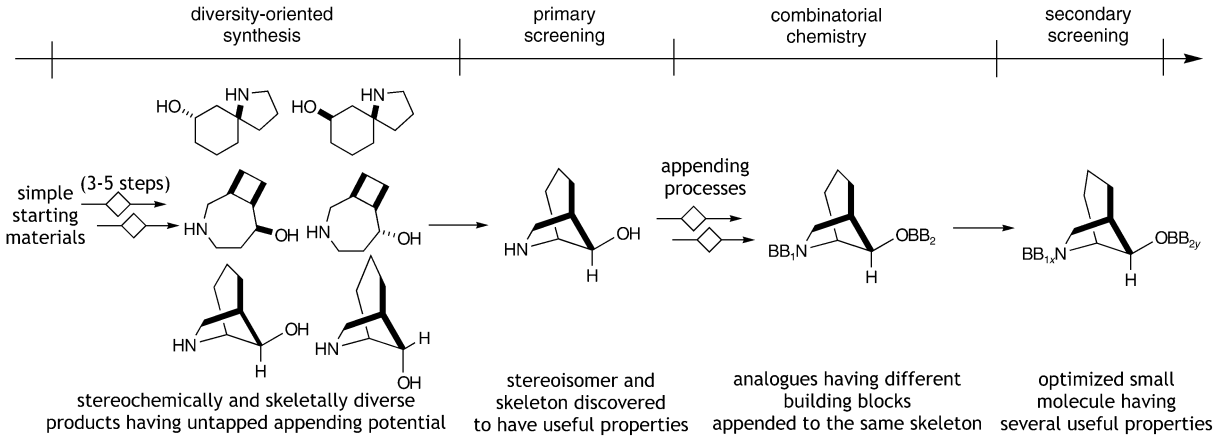
Scheme 10. Split-pool synthesis of a collection of compounds representing all possible combinations of building block, stereochemical, and skeletal diversity elements. 9-BBN = 9-borabicyclo[3.3.1]nonane, dppf = 1,1′-bis(diphenylphosphanyl)ferrocene.

synthetic chemistry. By including this consideration into future DOS pathways, the possibility for an overall discovery process illustrated in Scheme 12 is made possible, one that creates synergistic links between the strengths of DOS and combinatorial chemistry. Finally, subsets are selected from the large collection of potential products from any given DOS pathway for synthesis with guidance from computations of molecular descriptors, analyses of these descriptors relative to reference small-molecules previously annotated, and the application of filters to minimize undesired properties, especially poor solubility.



Scheme 11. The generation of both skeletal complexity and skeletal diversity through the incorporation of complexity-generating reactions into a skeletal diversity-generating folding process. It was determined that a single stereocenter in otherwise similar substrates **69** pre-encodes the formation of highly complex products **70** and **71**, which have very different molecular skeletons, under a common set of reaction conditions. This example also illustrates the potential of using folding processes to link stereochemical diversity to skeletal diversity.

appending potential into products of DOS pathways ensures the possibility for facile, post-screening appending processes that can ease the optimization of properties not examined in primary, "discovery" screens. In this way, synthetic planning anticipates the likely need for optimization and provides a unique and general solution for it—something not yet addressed, to our knowledge, in other applications of



Scheme 12. A potentially general approach for discovering small molecules with useful properties that begins with the efficient synthesis of collections of small molecules having structural complexity, stereochemical and skeletal diversity, and untapped appending potential.

7. Summary

Achieving the goals of DOS requires new advances in strategic thinking. Although the logic of diversity-oriented synthesis is still evolving, some guiding principles have emerged that provide the basis for a forward-synthetic analysis. Structural complexity can be most efficiently accessed using complexity-generating reactions, ideally in series, where the product of one complexity-generating reaction is the substrate for another. Structural diversity can be accessed using diversity-generating processes, and planning efficient DOS pathways depends on the identification of such processes where the products of one diversity-generating process are the substrates for another. It is necessary to gain efficient access to both stereochemical and skeletal diversity to achieve a diverse display of chemical information in three-dimensional space. Stereochemical diversity can be generated by using stereospecific and stereoselective diversity-generating processes, with the latter relying heavily on the development of powerful reagents that can override substrate bias to generate, ideally, all possible diastereomeric products with a high degree of selectivity. Skeletal diversity can be achieved by using both reagent-based (differentiating) and substrate-based (folding) strategies, with the latter having a demonstrated potential for generating skeletal diversity combinatorially. Achieving both complexity and diversity in an efficient manner requires integrated forward-synthetic planning, for example, the incorporation of complexity-generating reactions into stereochemical and skeletal diversity-generating processes.

The answer to the question: "Are the regions of chemistry space defined by natural products and known drugs, which have been so intensely scrutinized to date, the best or most fertile regions for discovering small-molecules that modulate macromolecular function in useful ways?" is not known. However, we believe that the answer to this question is likely to be "no", that is, the vast, previously unexplored regions of chemistry space likely contain small molecules having extraordinary properties that can contribute in unprecedented ways to the understanding and betterment of human health. Fortunately, this hypothesis can be tested with experiment, although doing so will require synthetic organic chemists to gain broad access to these as-of-yet unexplored regions of chemistry space in a highly efficient manner,[40] and thereby bring these extraordinary small molecules into existence. DOS aims to achieve this objective, and, although the challenge is daunting, synthetic organic chemists have a history of rising to such challenges.

Glossary of terms

Chemical space: n-dimensional space defined by the value of *n* descriptors; these descriptors can be of a chemical or biological nature and are either computed or measured.

Molecular skeleton: the combination of rigidifying elements (covalent bonds, non-covalent bonds, and non-bonding interactions) that define a molecule's overall three-dimensional shape; a complex molecular skeleton is one that is defined by a large number and/or variety of rigidifying elements.

Target-oriented synthesis
Retrosynthetic analysis: A problem-solving technique for transforming the structure of a synthetic target molecule to a sequence of progressively simpler structures along a pathway which ultimately leads to simple or commercially available starting materials for a chemical synthesis.
Transform: The exact reverse of a synthetic reaction.
Retron: The enabling structural subunit that permits the application of a transform.

Diversity-oriented synthesis
Forward-synthetic analysis: A problem-solving technique for transforming a collection of simple and similar starting materials into a collection of more complex and diverse products.
Process: The transformation of a collection of substrates into a collection of products.

Colleagues at Harvard University's Department of Chemistry and Chemical Biology and ICCB, especially M. D. Shair, E. J. Corey, D. Brittain, and G. Lalic are gratefully acknowledged for helpful discussions. M.D.B. is an HHMI predoctoral fellow. We thank Eric Keller for creating the frontispice for this Review.

Received: August 19, 2003 [A626]

[1] Chemistry space is multidimensional, with each dimension defined by a value of a descriptor; descriptors can be of a biological or chemical nature and are either computed or measured.
[2] S. L. Schreiber, *Science* **2000**, *287*, 1964; S. L. Schreiber, *Chem. Eng. News* **2003**, *81*(9), 51.
[3] E. J. Corey, X.-M. Cheng, *The Logic of Chemical Synthesis*, Wiley, New York, **1995**; E. J. Corey, *Angew. Chem.* **1991**, *103*, 469; *Angew. Chem. Int. Ed. Engl.* **1991**, *30*, 455 *(Nobel Lecture)*; E. J. Corey, *Pure. Appl. Chem.* **1967**, *14*, 19; E. J. Corey, W. T. Wipke, *Science* **1969**, *166*, 178; E. J. Corey, *Q. Rev. Chem. Soc.* **1971**, *25*, 455.
[4] M. R. Spaller, M. T. Burger, M. Fardis, P. A. Bartlett, *Curr. Opin. Chem. Biol.* **1997**, *1*, 47.
[5] http://chembank.med.harvard.edu/.
[6] For reviews, see "Cascade Reactions": *Tetrahedron Symposium in Print*, No. 62 (Ed.: R. Grigg), Elsevier, New York, **1996**; F. E. Ziegler in *Comprehensive Organic Synthesis*, *Combining C-C π-Bonds*, *Vol. 5* (Ed.: L. A. Paquette), Permagon, Oxford, **1991**, chap. 7.3; T.-L. Ho, *Tandem Organic Reactions*, Wiley, New York, **1992**.
[7] R. B. Merrifield, *J. Am. Chem. Soc.* **1963**, *85*, 2149; C. C. Leznoff, J. Y. Wong, *Can. J. Chem.* **1972**, *50*, 2892; F. Camps, J. Castells, M. J. Ferrando, J. Font, *Tetrahedron Lett.* **1971**, *12*, 1713; F. Camps, J. Castells, J. Pi, *An. Quim.* **1974**, *70*, 848; A. Patchornik, M. A. Kraus, *J. Am. Chem. Soc.* **1970**, *92*, 7857; J. I. Crowley, H. Rapoport, *J. Am. Chem. Soc.* **1970**, *92*, 6363; V. Yedida, C. C. Lexnoff, *Can. J. Chem.* **1980**, *58*, 1140; M. J. Farrall, J. M. J. Fréchet, *J. Org. Chem.* **1976**, *41*, 3877; J. M. J. Fréchet, *Tetrahedron* **1981**, *37*, 663.
[8] B. A. Bunin, J. A. Ellman, *J. Am. Chem. Soc.* **1992**, *114*, 10997; R. J. Simon, K. S. Kania, R. N. Zuckermann, V. D. Huebner,

D. A. Jewell, S. Banville, S. Ng, L. Wang, S. Rosenberg, C. K. Marlowe, D. C. Spellmeyer, R. Tan, A. D. Frankel, D. V. Santi, F. E. Cohen, P. A. Bartlett, *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 9367; S. H. Dewitt, J. S. Kiely, C. J. Stankovic, M. C. Schroeder, D. M. R. Cody, M. R. Pavia, *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6909; P. H. H. Hermkens, H. C. J. Ottenheijm, D. Rees, *Tetrahedron* **1996**, *52*, 4527; P. H. H. Hermkens, H. C. J. Ottenheijm, D. Rees, *Tetrahedron* **1997**, *53*, 5643; R. E. Dolle, K. H. Nelson, Jr., *J. Comb. Chem.* **1999**, *1*, 235; R. E. Dolle, *J. Comb. Chem.* **2000**, *2*, 383; R. E. Dolle, *J. Comb. Chem.* **2001**, *3*, 477; R. E. Dolle, *J. Comb. Chem.* **2002**, *4*, 369.

[9] A. Furka, F. Sebestyén, M. Asgedom, G. Dibó, in *Highlights of Modern Biochemistry, Proceedings of the 14th International Congress of Biochemistry, Prague, Czechoslovakia* (VSP, Utrecht, Netherlands), **1988**, *13*, 47; A. Furka, F. Sebestyén, M. Asgedom, G. Dibó, *Int. J. Pept. Protein Res.* **1991**, *37*, 487; R. A. Houghten, C. Pinilla, S. E. Blondelle, J. R. Appel, C. T. Dooley, J. H. Cuervo, *Nature* **1991**, *354*, 84; K. S. Lam, S. E. Salmon, E. M. Hersh, V. J. Hruby, W. M. Kazmierski, R. J. Knapp, *Nature* **1991**, *354*, 82.

[10] It is useful to describe the molecular skeleton of a small molecule as the combination of rigidifying elements (covalent bonds, noncovalent bonds, and nonbonding interactions) that define its overall three-dimensional shape. In that vein, a complex molecular skeleton is one that is defined by a large number and/or variety of rigidifying elements.

[11] http://iccb.med.harvard.edu/.

[12] For a review of such transformations, see L. F. Tietze, U. A. Beifuss, *Angew. Chem.* **1993**, *105*, 137; *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 131.

[13] I. Ugi, A. Dömling, W. Hörl, *Endeavour* **1994**, *18*, 115.

[14] K. Paulvannan, *Tetrahedron Lett.* **1999**, *40*, 1851; D. Lee, J. K. Sello, S. L. Schreiber, *Org. Lett.* **2000**, *2*, 709.

[15] W. J. Zuercher, M. Hashimoto, R. H. Grubbs, *J. Am. Chem. Soc.* **1996**, *118*, 6634; W. J. Zuercher, M. Scholl, R. H. Grubbs, *J. Org. Chem.* **1998**, *63*, 4291; R. Stragies, S. Blechert, *Tetrahedron* **1999**, *55*, 8179; O. Arjona, A. G. Csákÿ, M. C. Murcia, J. Plumet, *J. Org. Chem.* **1999**, *64*, 9739.

[16] M. Scholl, S. Ding, C. W. Lee, R. H. Grubbs, *Org. Lett.* **1999**, *1*, 953; A. K. Chatterjee, R. H. Grubbs, *Org. Lett.* **1999**, *1*, 1751.

[17] D. S. Tan, M. A. Foley, M. D. Shair, S. L. Schreiber, *J. Am. Chem. Soc.* **1998**, *120*, 8565.

[18] H. E. Pelish, N. J. Westwood, Y. Feng, T. Kirchhausen, M. D. Shair, *J. Am. Chem. Soc.* **2001**, *123*, 6740.

[19] R. Breinbauer, I. R. Vetter, H. Waldmann, *Angew. Chem.* **2002**, *114*, 3002; *Angew. Chem. Int. Ed.* **2002**, *41*, 2878.

[20] W. S. Knowles, M. J. Sabacky, *Chem. Commun.* **1968**, 1445; T. P. Dang, H. B. Kagan, *Chem. Commun.* **1971**, 481; W. S. Knowles, M. J. Sabacky, B. D. Vineyard, D. J. Weinkauff, *J. Am. Chem. Soc.* **1975**, *97*, 2567.

[21] K. B. Sharpless, *Chem. Scr.* **1985**, *25*, 71; S. Masamune, W. Choy, J. S. Petersen, L. R. Sita, *Angew. Chem.* **1985**, *97*, 1; *Angew. Chem. Int. Ed. Engl.* **1985**, *24*, 1; S. Masamune, *Heterocycles* **1984**, *21*, 107; S. Y. Ko, A. W. M. Lee, S. Masamune, L. A. Reed III, K. B. Sharpless, F. J. Walker, *Science* **1983**, *220*, 949; B. A. Harisson, G. L. Verdine, *Org. Lett.* **2001**, *3*, 2157; T. M. Gierasch, Z. Shi, G. L. Verdine, *Org. Lett.* **2003**, *5*, 621.

[22] G. C. Micalizio, S. L. Schreiber, *Angew. Chem.* **2002**, *114*, 160; *Angew. Chem. Int. Ed.* **2002**, *41*, 152.

[23] O. I. Kolodiazhnyi, *Tetrahedron* **2003**, *59*, 5953.

[24] A. G. Dossetter, T. F. Jamison, E. N. Jacobsen, *Angew. Chem.* **1999**, *111*, 2549; *Angew. Chem. Int. Ed.* **1999**, *38*, 2398; G. D. Joly, E. N. Jacobsen, *Org. Lett.* **2002**, *4*, 1795; K. Gademann, D. E. Chavez, E. N. Jacobsen, *Angew. Chem.* **2002**, *114*, 3185; *Angew. Chem. Int. Ed.* **2002**, *41*, 3059.

[25] D. E. Chavez, E. N. Jacobsen, *Org. Lett.* **2003**, *5*, 2563.

[26] W. R. Roush, S. M. Peseckis, *J. Am. Chem. Soc.* **1981**, *103*, 6696.

[27] J. W. Coe, W. R. Roush, *J. Org. Chem.* **1989**, *54*, 915; F. K. Brown, K. N. Houk, *Tetrahedron Lett.* **1985**, *26*, 2297; J. D. White, B. G. Sheldon, *J. Org. Chem.* **1981**, *46*, 2273; D. F. Taber, C. Campbell, B. P. Gunn, C. I. Chiu, *Tetrahedron Lett.* **1981**, *22*, 5141; for a review on diastereofacial selection in ring-closing reactions, see T. Tokoroyama, *Chem. Eur. J.* **1998**, *4*, 2397.

[28] K. Kim, U. S. M. Maharoof, J. Raushel, G. A. Sulikowski, *Org. Lett.* **2003**, *5*, 2777.

[29] R. A. Stavenger, S. L. Schreiber, *Angew. Chem.* **2001**, *113*, 3525; *Angew. Chem. Int. Ed.* **2001**, *40*, 3417.

[30] D. Lee, J. Sello, S. L. Schreiber, *J. Am. Chem. Soc.* **1999**, *121*, 10648.

[31] G. C. Micalizio, S. L. Schreiber, *Angew. Chem.* **2002**, *114*, 3406; *Angew. Chem. Int. Ed.* **2002**, *41*, 3272.

[32] S. Woo, N. Squires, A. G. Fallis, *Org. Lett.* **1999**, *1*, 573.

[33] O. Kwon, S. B. Park, S. L. Schreiber, *J. Am. Chem. Soc.* **2002**, *124*, 13402.

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

[35] C. B. Anfinsen, *Science* **1973**, *181*, 223.

[36] N. Clauson-Kaas, P. Dietrich, J. T. Nielson, *Acta Chem. Scand.* **1953**, *7*, 845; N. Elming, *Adv. Org. Chem.* **1960**, *2*, 67.

[37] O. Achmatowicz, Jr., P. Bukowski, B. Szechner, Z. Zwierzchowska, A. Zamojski, *Tetrahedron* **1971**, *27*, 1973; J. M. Harris, G. A. O'Doherty, *Tetrahedron Lett.* **2000**, *41*, 183.

[38] Y. Kobayashi, M. Nakano, G. B. Kumar, K. Kishihara, *J. Org. Chem.* **1998**, *63*, 7505.

[39] J. K. Sello, P. R. Andreana, D. Lee, S. L. Schreiber, *Org. Lett.* **2003**, *5*, 4125.

[40] Realizing the goals of DOS will require organic chemists to adopt computational tools from computer and information science, since the complexities of multidimensional (chemical and biological) descriptor space can only be tackled using these tools.

