**Towards drugging the ‘undruggable’: enhancing the scaffold diversity of synthetic small molecule screening collections using diversity-oriented synthesis**

**Abstract**

Medicinal chemistry research has traditionally focused upon a limited set of biological targets. Many other human disease-related targets have been termed ‘undruggable’ as they have proved largely impervious to modulation by small molecules. However, it is becoming increasingly evident that such targets can indeed be modulated; they are simply being challenged with the wrong types of molecules. Traditionally, screening libraries were composed of large numbers of structurally similar compounds. However, library size is not everything; the structural diversity of the library, which is largely dictated by the range of molecular scaffolds present, is crucial. Diversity-oriented synthesis (DOS) generates small molecule libraries with high levels of scaffold, and thus structural, diversity. Such collections should provide hits against a broad range of targets with high frequency, including ‘undruggable’ targets. Examples in the area of scaffold diversity generation taken from the author’s laboratories are given.

**Keywords**

Diversity-oriented synthesis • Undruggable targets • Antibacterials

© Versita Sp. z o.o.
molecular structure in this context is the molecular scaffold, that is the core rigidifying feature of the molecule. It is the molecular scaffold that has the most influence upon how a molecule displays chemical information in 3D space; indeed, the shape-space coverage of any compound set stems mainly from the nature and 3D geometries of the central scaffolds, with the peripheral substituents being of minor importance [1,11,12]. Traditional combinatorial libraries typically possessed low levels of scaffold diversity, and thus a limited overall structural, and consequently functional, diversity [6]. This explains their poor performance in many biological screens, particularly those where the precise nature of the biological target was unknown or ill-defined (e.g. phenotypic screening). A lack of scaffold diversity is still apparent in many of the synthetic screening sets employed in more recent years, despite a growing appreciation of the importance of this feature [13]. Indeed, deficiencies in current compound collections are evidenced by the continuing decline in drug discovery successes by pharmaceutical companies [1,6].

It can be argued that the lack of scaffold, and thus functional, diversity found in typical small molecule screening sets is a direct consequence of the nature of drug discovery over the course of the last few decades. Medicinal chemistry research has traditionally focused upon a limited set of biological targets; indeed there are only approximately 500 distinct targets of the current pharmacopoeia [1,10]. Nearly 36% of all drugs act upon G-protein coupled receptors (GPCRs) and approximately 29% are enzyme modulators [2]. Unsurprisingly, therefore, pharmaceutical proprietary collections and commercially available (combinatorial) libraries are typically heavily biased towards compounds that satisfy predefined criteria for successful modulation of such targets (e.g. the Lipinski ‘rule of 5’ criteria for oral bioavailability) [1,6,14]. However, it is widely acknowledged that the targets of the current pharmacopoeia represent only a tiny subset of potential targets that could impact on disease [1,2,10,15]. The term ‘undruggable’ has been coined to describe those biological targets and processes that bear little resemblance to the molecular drug targets exploited in present-day drug therapy. These have historically been thought of as difficult, if not impossible, to modulate with small molecules (examples include protein-protein interactions, transcription factors, regulatory RNAs and protein-DNA interactions) [1,10,16]. Many of these are now validated targets for the treatment of a range of human diseases [1,10,16]. It is clear that there are a myriad of biological targets and processes that are potentially extremely useful as drug-targets, which the pharmaceutical industry has ignored or been unable to exploit. Clearly small molecule modulators of these targets offer exciting therapeutic potential.

There is a widespread consensus that increasing the scaffold diversity in a small molecule library is one of the most effective ways of increasing its overall shape and structural diversity [1,9,17]. Libraries which are small in size but which contain compounds based around multiple scaffolds are generally regarded to have better biorelevant diversity than libraries of large numbers of compounds based around single scaffolds [1,11,12]. In principle, the screening of collections with higher levels of scaffold diversity should provide hits against a broad range of biological targets with increased frequency and decreased cost [1,10]. This includes so-called ‘undruggable’ targets and processes. It has been argued that one of the reasons why these are traditionally viewed to be intractable to small molecule modulation is because of deficiencies in existing compound collections. That is, the molecules comprising such collections seem to be well suited to modulating ‘traditional’ medicinal chemistry targets, but lack the necessary structural features required to modulate other processes [1,10,16]. Therefore, in order to identify molecules capable of modulating these more challenging targets there is a clear need for new small molecule collections with increased levels of structural (scaffold) diversity that span regions of bioactive chemical space not accessed by traditional compound libraries [1]. When considering the generation of scaffold diversity, it is important to note that the known universe of organic chemistry is dominated by a remarkably small number of molecular scaffolds [18]. For example, in a recent study of known cyclic molecules, 0.25% of the molecular scaffolds were found in 50% of known compounds [16]. That is, a large percentage of the compounds were represented by only a small percentage of scaffolds [13]. As another illustration of this point, there are around 25,000 possible mono- and bicyclic heteroaromatic compounds, yet only ~ 1,700 have been synthesised and reported to date [19]. Furthermore, there are many examples of molecular scaffolds with proven biological relevance that are under-represented in current small molecule collections. In the majority of such cases, this relative paucity of compounds can be attributed to synthetic intractability. Thus, there is a need for the development of new and efficient methods of broad utility for the synthesis of both ‘atypical’ and novel molecular scaffolds so that the biological usefulness of these structures can be investigated and exploited further [1]. The enrichment of screening libraries with molecules based on such frameworks will allow the sampling of previously untapped regions of chemical space [1]. Compounds in such areas may have exciting, novel biological properties, including the ability to modulate more challenging underexploited drug targets.

Diversity-oriented synthesis (DOS) is a synthetic approach that seeks to efficiently generate structurally diverse compound collections [1]. Schreiber and co-workers first described the concept of DOS in two seminal publications in 2000 [20,21]. These reports introduced the idea that the forward analysis of synthetic pathways could be used to devise strategies to convert small numbers of structurally simple building blocks into larger collections of structurally complex and diverse compounds that efficiently interrogate wide regions of chemical space simultaneously. Since these landmark works, the field of DOS has evolved rapidly. Alternative DOS planning strategies have been developed which are more “focused” in nature, seeking to efficiently explore areas of chemical space that are thought to have an enhanced probability of containing bioactive compounds.
Towards drugging the ‘undruggable’: enhancing the scaffold diversity of synthetic small molecule screening collections using diversity-oriented synthesis

A synthetic strategy commonly employed in DOS is the so-called reagent-based branching approach \([1,3]\). This involves the use of a common starting material and different reagents. A short series of divergent, complexity-generating reactions are carried out, leading to the formation of a structurally diverse collection of functionalized compounds based around a diverse range of molecular scaffolds \([3]\). Two examples from our group of DOS using a reagent-based approach are shown in Scheme 1.

A key strategic feature in both cases was the incorporation of a phase-tag in the starting materials, which greatly simplified purification during library synthesis. In the first example a structurally simple fluorous-tagged diazaoacetate \([1]\) was converted to a library of 223 structurally diverse compounds based around 30 different core molecular scaffolds \([25]\). At the time this represented the highest number of different molecular scaffolds [22,23] and DOS around privileged structures [24].

The Spring group has a strong interest in the development of innovative DOS strategies, with an emphasis placed upon the efficient incorporation of scaffold diversity. There are two related strands running through this work: (1) the DOS of scaffold, and therefore structurally, diverse compound collections and (2) the development of novel methodologies and strategies that allow for the efficient synthesis of under-explored and novel molecular scaffolds, and the application of these methodologies in a DOS context. In this minireview some selected representative examples of work taken from the author’s laboratories in both of these branches of research are described, with a particular focus upon new methods and strategies for the synthesis of unusual molecular scaffolds.

A synthetic strategy commonly employed in DOS is the so-called reagent-based branching approach \([1,3]\). This involves the use of a common starting material and different reagents. A short series of divergent, complexity-generating reactions are carried out, leading to the formation of a structurally diverse collection of functionalized compounds based around a diverse range of molecular scaffolds \([3]\). Two examples from our group of DOS using a reagent-based approach are shown in Scheme 1.

A key strategic feature in both cases was the incorporation of a phase-tag in the starting materials, which greatly simplified purification during library synthesis. In the first example a structurally simple fluorous-tagged diazaoacetate \([1]\) was converted to a library of 223 structurally diverse compounds based around 30 different core molecular scaffolds \([25]\). At the time this represented the highest number of different molecular scaffolds [22,23] and DOS around privileged structures [24].

The Spring group has a strong interest in the development of innovative DOS strategies, with an emphasis placed upon the efficient incorporation of scaffold diversity. There are two related strands running through this work: (1) the DOS of scaffold, and therefore structurally, diverse compound collections and (2) the development of novel methodologies and strategies that allow for the efficient synthesis of under-explored and novel molecular scaffolds, and the application of these methodologies in a DOS context. In this minireview some selected representative examples of work taken from the author’s laboratories in both of these branches of research are described, with a particular focus upon new methods and strategies for the synthesis of unusual molecular scaffolds.

Scheme 1. Two examples of DOS using a reagent-based branching synthetic strategy. A) DOS from a fluorous tagged diazaoacetate \([25]\); B) DOS using a solid-supported phosphonate \([26]\).
scaffolds incorporated into a synthetically-derived small molecule library. The second example involved the generation of a library of small molecules starting from a solid-supported phosphonate 2 [26]. Overall 242 small molecules based around 18 molecular scaffolds were generated. Many of these scaffolds are rare or have no known representation in nature (e.g. the cis-trans-fused 7-5-7-tricycle of the form 3), highlighting the capability of this DOS approach to generate compounds that populate new, unexplored regions of chemical space [4,26]. Biological screening of these compound sets identified a range of novel small molecules, based around a diverse range of scaffolds, which were capable of inhibiting the proliferation of a range of bacterial strains. Significantly these included two UK epidemic methicillin-resistant Staphylococcus aureus strains, EMRSA 15 and EMRSA 16, which are responsible for the majority of infections with MRSA in the UK [4,27].

One class of scaffolds that are arguably underrepresented in current synthetic small molecule screening collections are macrocycles. Compounds containing macrocyclic ring structures (a ring size of 12 atoms or above) are capable of potent biological activity and specificity, as exemplified by the success of over 100 marketed macrocycle drugs derived from natural products [28,29]. Despite the proven biological relevance of macrocycles, there is a widely thought to be a lack of synthetic macrocycles in drug discovery. This is mainly attributed to concerns about synthetic intractability [30,31]. Towards this end recent years have witnessed considerable interest in the development of new synthetic strategies to efficiently access macrocyclic structures, and several DOS campaigns targeted at this structural type have been reported [32-34].

Within our own group we have developed two DOS strategies towards complex macrocyclic structures. The first is based around a so-called ‘two-directional’ approach (Scheme 2) [28]. This involved the initial two-directional synthesis of symmetrical ‘linear’ precursors containing reactive functionality at two positions. Linear precursors containing mutually complementary functionality could then be reacted together in a two-directional macrocyclisation process. In initial proof-of-principle work we specifically focused upon two types of linear precursors: bis-enyne amides 4 and conjugated dienes 5 (which could be readily accessed from 4 by ring-closing enyne metathesis). The complementary reacting partners were bis-maleimides 6 and bis-azides 7. Diels-Alder reaction between bis-dienes 5 and bis-maleimides 6 then gave macrocycles of type 8. Copper-catalysed azide-alkyne cycloaddition between bis-enyne amides 4 and bis-azides 7 gave macrocycles of type 9. An example of a final compound, 10, synthesized using this approach is shown in Scheme 2. The linker units between the reactive groups can be regarded as ‘scaffold elements’ in the sense that different combinations of appropriately functionalized scaffold elements will lead directly to distinct macrocycle scaffolds (with two distinct scaffolds accessible for every combination of reactive linear precursors, depending on the macrocyclisation process employed) [28]. In principle therefore the combinatorial variation of scaffold elements could be used to rapidly and efficiently access a large range of unique scaffolds. In a proof-of-principle study this approach was used to synthesize 14 macrocyclic compounds, including examples of nine different ring sizes, in no more than five synthetic steps from commercially available materials.

We have also developed another DOS strategy for generating macrocyclic scaffolds, more specifically a subset of macrocycles termed macrocyclic peptidomimetics [35]. Peptidomimetics are compounds whose essential structural
elements mimic a natural protein or peptide chain in 3D and which are often utilized in an attempt to circumvent some of the problems associated with the application of natural peptides in a therapeutic context (e.g. stability against proteolysis and poor bioavailability) [36]. Numerous biologically active molecules contain cyclic peptide and peptidomimetic structural units including diketopiperazines (DKPs, the smallest possible cyclic peptides) and many compounds incorporating macrocyclic peptides and peptidomimetics are also known to be capable of modulating biological systems [29,35,37-41].

Despite such valuable properties, macrocyclic peptides (as with macrocyclic compounds in general) are generally viewed as a poorly explored structural class within drug discovery [29,35]. We have described a strategy for the DOS of a library of structurally unique and diverse macrocyclic peptidomimetics based around two general structural types (A and B, Scheme 3) from simple, readily available amino acid starting materials [35]. Each structural type contains a triazole ring in place of an amide bond, and structure B also incorporates a DKP in the macrocycle. Our strategy for the synthesis was based around a so-called build/couple/pair three-phase approach described by Nielsen and Schreiber [42]. In our case the build step involves the preparation of two types of chiral building blocks: (1) ‘azido-amine’ building blocks 11 which contain a free amine and an azide and (2) ‘alkyne-acid’ building blocks 12 which contain a carboxylic acid and an alkyne. Coupling of three of these building blocks via amide bond formation furnishes a range of tripeptide derivatives 13; this process provides the basis for stereochemical diversity. The subsequent pair phase provides the basis for scaffold diversity and is comprised of two cyclization steps. First, a ‘click’-type 1,3-dipolar cycloaddition, catalyzed by either ruthenium or copper, to selectively combine the azide and alkyne functionalities of these tripeptides, thus generating the desired macrocyclic peptidomimetic architectures 14 or 15 respectively. Intramolecular cyclization reactions between amine and carbonyl moieties would then introduce the DKP motif into the macrocyclic frameworks to form 16 and 17. In a proof-of-concept study a small library of 14 compounds based around four different macrocyclic scaffolds was generated (namely cis-DKPs, trans-DKPs, 1,4-, and 1,5-triazoles).

Cheminformatic analysis of both macrocycle proof-of-principle libraries suggested that they access a relatively large area of chemical space given the small number of compounds synthesized [28,35]. This included regions not explored by molecules from ‘traditional’ medicinal chemistry. Both these macrocyclisation strategies are modular in nature. Thus we anticipate that they hold significant potential for the DOS of larger libraries of macrocycles with greater levels of scaffold, and thus overall structural diversity. In preliminary biological screens a number of compounds from both proof-of-principle libraries showed interesting biological properties including antibacterial effects and antiproliferative activity against cancer cell lines.

Scheme 3. DOS of macrocyclic peptidomimetics.
Further biological screening is ongoing, as are studies aimed at applying both of these synthetic strategies to the DOS of larger libraries of macrocyclic compounds.

Another class of scaffolds that are arguably underrepresented in current synthetic small molecule screening collections are nitrogen-linked medium ring biaryls [43]. Numerous compounds based around seven membered N-linked biaryl scaffolds have been found to demonstrate extraordinary biological properties [43]. Eight-membered-ring derivatives have also been reported, but are comparatively scarce (Scheme 4). In general, N-linked medium-ring biaryl compounds are a very rare class of compounds. This can primarily be attributed to synthetically intractability; there are well-know difficulties associated with medium-ring synthesis in general [44] and medium-ring biaryl scaffolds in particular are known to be especially challenging synthetic targets. We have developed a novel strategy for the synthesis of seven-, eight- and nine-membered N-linked biaryl ring systems of the general form 22 from acyclic precursors 23. The method is based on the premise of generating a high active intramolecular copper species 24 that can facilitate C-(aryl)-N bond formation and thus affect ring closure (Scheme 4). The process proceeds under relatively mild conditions, is technically simple, displays a broad substrate scope and forms biologically valuable products that are difficult to synthesize by other methods.

In addition to targeting the synthesis of scaffolds of proven biological relevance that are underrepresented in small molecule screening collections, our research group also has an interest in devising efficient routes to more unusual (including novel) complex molecular scaffolds whose biological properties have yet to be explored or exploited. Towards this end we have recently developed a folding-type process for the generation of skeletally diverse small molecules [47]. Substituted norbornene derivatives 26 were used in a tandem domino enyne metathesis-Diels Alder sequence that allows the generation of complex polycyclic scaffolds 27 in a highly efficient and atom-economical manner (Scheme 5) [47]. Variation in the norbornene starting material and the dienophile used allowed access to a diverse range of complex molecular skeletons. For example, cis-norbornene derivative 28 could be converted directly into the complex adduct 29 in a high-yielding, one-pot procedure. It should be noted that a detailed mechanistic understanding of this process was essential for the development of a high-yielding procedure. Extensive optimisation studies established that the choice of Grubbs’ metathesis catalyst affected the reaction pathway, the product distribution, and the yield. Optimum results were obtained when Grubb’s 1 and 2 were used in a stepwise fashion, to facilitate ROM of the norbornene before the RCM reactions (both olefin and enyne) occurred. The reactions had to be run in an ethylene saturated solvent in order to minimise the occurrence of ring opening metathesis polymerization. In addition, it was determined that the addition of a polar isocyanate to the reaction mixture was required (Step 3 in the conversion of 28 to 29, Scheme 5) in order to minimise product decomposition before the Diels-Alder step.

**Scheme 4.** A) Some examples of biologically active compounds that are based around N-linked medium-ring biaryl scaffolds. Oxcarbazepine (18) is employed in the treatment of epilepsy [45]; compound 19 has been reported to have anti-inflammatory effects and imipramine (20) and desipramine (21) are employed as anti-depressants [46,47]. B) Mechanistic blueprint for the copper(I)-catalysed N-linked biaryl cyclisation process. The mechanism is proposed to begin with co-ordination of the substrate 23 to the copper catalyst followed by deprotonation of the aniline, giving rise to a 6-membered chelate 24. Oxidative addition of the copper(I) species across the aryl C-Br bond forms a copper(III)-based bicyclo[4.3]metal-cyclic intermediate 25. Reductive elimination yields the cyclised product 22 and regenerates the copper(I) catalyst. The auxiliary nitrogen atom is expected to allow internal chelation to form the active intramolecular copper species and its geometry could possibly facilitate ring closure. Possible ancillary ligands omitted for clarity [45].
Towards drugging the ‘undruggable’: enhancing the scaffold diversity of synthetic small molecule screening collections using diversity-oriented synthesis

In recent years; in 2006 the highest number of distinct molecular scaffolds incorporated into a synthetically derived small molecule library was 30 [25], whereas that figure now stands at 84 [48,49]. However, the construction of large libraries with hundreds of scaffolds remains an as-yet unmet challenge. In the context of generating new small molecule libraries for biological testing it is also vital that chemists broaden their horizons and look beyond the relatively limited set of molecular scaffolds that dominate typical screening sets. The enhancement of screening collections with molecules based around underrepresented or novel molecular scaffolds would allow the exploration of uncharted regions of chemical space. Compounds in such areas may have exciting and unusual biological properties that have thus far escaped the attention of humans and perhaps even nature [1, 6]. There is therefore a clear need for the continued development of new synthetic approaches to efficiently access such underexploited molecular scaffolds [13].

In summary, there is an undoubted need for synthetic small molecule screening sets with higher levels of scaffold diversity in order to address challenging ‘undruggable’ biological targets which have proven largely intractable to small molecule modulation using traditional compound libraries. The efficient de novo generation of such collections presents a formidable challenge to the synthetic chemist; over the course of the last decade DOS has established itself as a powerful tool in this regard [1]. The screening of DOS libraries has led to the identification of numerous biologically active small molecules, including modulators of a range of ‘undruggable’ targets and process, thus validating the usefulness of DOS [1,10]. However, significant improvements in library synthesis and screening are required in order for the full potential of the DOS-approach to be realized [2]. From a synthetic perspective, the efficiency of scaffold diversity generation represents the most important issue. Some progress has undoubtedly been made in this area in recent years; in 2006 the highest number of distinct molecular scaffolds incorporated into a synthetically derived small molecule library was 30 [25], whereas that figure now stands at 84 [48,49]. However, the construction of large libraries with hundreds of scaffolds remains an as-yet unmet challenge. In the context of generating new small molecule libraries for biological testing it is also vital that chemists broaden their horizons and look beyond the relatively limited set of molecular scaffolds that dominate typical screening sets. The enhancement of screening collections with molecules based around underrepresented or novel molecular scaffolds would allow the exploration of uncharted regions of chemical space. Compounds in such areas may have exciting and unusual biological properties that have thus far escaped the attention of humans and perhaps even nature [1,6]. There is therefore a clear need for the continued development of new synthetic approaches to efficiently access such underexploited molecular scaffolds [13].

References
