A unified lead-oriented synthesis of over fifty molecular scaffolds†

Richard G. Doveston,a Paolo Tosatti,a Mark Dow,a Daniel J. Foley,a Ho Yin Li,a Amanda J. Campbell,b David House,b Ian Churcher,b Stephen P. Marsden*a and Adam Nelson*a,c

Controlling the properties of lead molecules is critical in drug discovery, but sourcing large numbers of lead-like compounds for screening collections is a major challenge. A unified synthetic approach is described that enabled the synthesis of 52 diverse lead-like molecular scaffolds from a minimal set of 13 precursors. The divergent approach exploited a suite of robust, functional group-tolerant transformations. Crucially, after derivatisation, these scaffolds would target significant lead-like chemical space, and complement commercially-available compounds.

Introduction

Control of molecular properties is crucial in drug discovery, and experience has provided sets of guidelines that can steer medicinal chemists towards drug-like chemical space. For example, Lipinski’s rule-of-five was formulated to predict oral bioavailability,1,2 and, in the case of central nervous system (CNS) drug discovery, more restricted guidelines can assist the design of compounds able to penetrate the blood-brain barrier.3,4 Conversely, relaxation of molecular property guidelines is generally required to enable the discovery of small molecule inhibitors of protein–protein interactions.5,6 The molecular properties of clinical candidates strongly influence the probability of successful progression towards marketed drugs.7–11 Key parameters that correlate strongly with success in drug discovery include molecular weight,7,8 lipophilicity8–10 and the fraction of sp3-hybridised carbons,11 some of which have been captured within a single metric that estimates drug-likeness.12

The identification of high quality lead compounds is a key challenge in early stage drug discovery. Lead optimisation tends to increase both molecular weight and lipophilicity, as well as molecular complexity.13–16 It is advisable, therefore, to control the properties of lead compounds to facilitate the emergence of clinical candidates with desirable drug-like characteristics. As a result, approaches have been developed to define chemical space that is populated by molecules with good lead-like properties typically by considering factors such as lipophilicity (e.g.17 −1 < clog P < 3), molecular size (e.g.17 14 ≤ heavy atoms ≤ 26) and removal of undesirable substructures.17

High-throughput screening provides a significant source of starting points for drug discovery.14,18 Control of the molecular properties of screening compounds can both aid subsequent development into high-quality lead molecules, and increase the efficiency of the exploration of chemical space.19 However, sourcing large numbers of diverse compounds with appropriate molecular properties has been identified as a major challenge.17 In a recent analysis, just 2.6% of 4.9 M commercially-available compounds were found to survive filtering by molecular weight (200 ≤ Mw ≤ 360), lipophilicity (−1 < clog P < 3) and various structural features known to be undesirable in drug candidates.17 In addition, the ability of emerging synthetic methods to deliver lead-like compounds has been very poorly demonstrated17 despite significant development of diversity-oriented synthetic approaches.20–22 The problem is in fact exacerbated when diversity is also considered because chemists have historically explored chemical space unsystematically and in an uneven manner;13 indeed, metrics have been developed to capture the scaffold diversity of screening collections.24 The realisation of lead-oriented synthesis has recently been framed as a major academic and practical challenge17 which has prompted the development of approaches25,26 to the synthesis of specific classes of lead-like molecules.27–30

We have established a programme focussed on developing general synthetic approaches to diverse and novel lead-like molecular scaffolds i.e. scaffolds with the potential, following

---

†It has been suggested that new approaches may be required to translate such molecules into marketed drugs (see ref. 5).

aSchool of Chemistry, University of Leeds, Leeds, LS2 9JT, UK.
E-mail: s.p.marsden@leeds.ac.uk, a.s.nelson@leeds.ac.uk
bGlaxoSmithKline Medicines Research Centre, Stevenage, SG1 2NY, UK
cAstbury Centre for Structural Molecular Biology, University of Leeds, Leeds, LS2 9JT, UK.
†Electronic supplementary information (ESI) available. See DOI: 10.1039/c4ob02287d
diversification, to yield large numbers of compounds with lead-like molecular properties. In this case, our approach to exemplifying the strategy exploited a single connective reaction in combination with just six distinct cyclisations (Scheme 1). Ir-catalysed amination\(^3\) was selected as the connective reaction because we had previously retooled it for compatibility with polar, sp\(^3\)-rich, functionalised substrates\(^3\) that would have particular value for targeting lead-like chemical space. Thus, reaction between allylic carbonates \(1\) and amines \(2\) would provide cyclisation precursors \(3\). Importantly, the building blocks would be armed with functional groups to enable subsequent cyclisation or functionalisation (filled circles), third building blocks (red) and new bonds (bold) are indicated.

Results

Synthesis of cyclisation precursors

We selected a number of densely functionalised cyclisation precursors that could facilitate the synthesis of a wide range of novel and diverse lead-like molecular scaffolds. Thirteen cyclisation precursors were prepared from combinations of an amine (10 alternatives) and an allylic carbonate (2 alternatives) building block in good to excellent yield, and with high enantio- or diastereoselectivity (Scheme 2). The building blocks were either commercially available or were prepared on a multi-gram scale using well established methods. In most cases, our previously reported protocol, which enables efficient coupling of unprotected polar amines, was used.\(^3\) Thus, the active catalyst was generated from 2 mol% [Ir(dbcot)Cl]\(_2\), 4 mol% chiral ligand [either \((R,R,R)-\) or \((S,S,S)-10\)] and 4 mol% BuNH\(_2\) in DMSO at 55 °C before the building blocks were

---

**Scheme 1** Overview of the unified approach to the synthesis of lead-like molecular scaffolds. Reaction between an allylic carbonate (1; blue) and an amine (2; green) would yield a cyclisation precursor 3. Reactive sites that may enable cyclisation or functionalisation (filled circles), third building blocks (red) and new bonds (bold) are indicated.

**Scheme 2** Synthesis of cyclisation precursors by Ir-catalysed reaction between amine and allylic carbonate building blocks. a \((S,S,a)-10\) was used. b PrNH\(_2\) and THF were used. c The amine HCl salt and 1.3 eq. K\(_3\)PO\(_4\) were used. dbcot = dibenzo[a]cyclooctatetraene; TBDPS = tert-butyldiphenylsilyl; Ns = 2-nitrobenzenesulfonyl.

---
added. In some cases, the use of DMSO was not required, and THF was used as the reaction solvent.\textsuperscript{31}

**Synthesis of lead-like molecular scaffolds**

A toolkit of six cyclisation reactions was then exploited to convert the thirteen cyclisation precursors (11–23) into lead-like molecular scaffolds. Selected syntheses of scaffolds are presented in Schemes 3 and 4 (see ESI\textsuperscript{†} for full details). In order to maximise the number of scaffolds prepared, a divergent synthetic approach was adopted that exploited each of the six methods in the first cyclisation step (Methods A–F; Scheme 3).

For example, Pd-catalysed aminoarylation (Method A),\textsuperscript{37,38} if necessary after protection of the secondary amine, enabled conversion of the cyclisation precursors 11, 14 and 15 into the pyrrolidines 24–28. In each case, the reaction proceeded

---

**Scheme 3** Selected syntheses of lead-like scaffolds prepared using a single cyclisation step (indicated by colour: A, yellow, Pd-catalysed aminoarylation; B, purple, iodocyclisation/displacement; C, pink, reaction with CDI; D, green, reaction with α-halo acetyl halide; E, light blue, ring-closing metathesis; F, peach, lactamisation). Typical methods (see ESI\textsuperscript{†} for full details): A: Aryl bromide (1.2 eq.), 5 mol% Pd(OAc)\textsubscript{2}, 10 mol% DPE-Phos, Cs\textsubscript{2}CO\textsubscript{3} (2.5 eq.), 1,4-dioxane, 105 °C; B: (i) NsCl (1.2 eq.), NEt\textsubscript{3} (2.0 eq.), DMAP (0.1 eq.), rt, then TBAF (1.2 eq.), AcOH (1.2 eq.), THF, rt; (ii) NIS (1.5 eq.), MeCN, 65 °C; (iii) ArSH (1.5 eq.), DBU (2.5 eq.), MeCN, rt; (iv) mCPBA (4.0 eq.), CH\textsubscript{2}Cl\textsubscript{2}, rt; (v) PhSH (1.2 eq.), DBU (1.5 eq.), MeCN, rt; C1: CH\textsubscript{2}Cl\textsubscript{2}–TFA, 0 °C – rt, then CDI (1.5 eq.), DBU (4.0 eq.), THF, 50 °C; C2: CDI (4.5 eq.), DMF, 110 °C; C3: CDI (1.5 eq.), DBU (2.5 eq.), THF, 50 °C; D1: Chloroacetyl chloride (1.5 eq.), NEt\textsubscript{3} (5.0 eq.), CH\textsubscript{2}Cl\textsubscript{2}, 0 °C – rt, then NaH (2.0 eq.), NaI (1.0 eq.), THF, rt; D2: (i) TMSCl (1.1 eq.), NEt\textsubscript{3} (3.0 eq.), CH\textsubscript{2}Cl\textsubscript{2}, 0 °C – rt, then bromoacetyl bromide (1.5 eq.), then 20% AcOH (aq.), rt; (ii) 35% NaOH (aq.) (5.0 eq.), Bu\textsubscript{4}NSO\textsubscript{4} (0.5 eq.), CH\textsubscript{2}Cl\textsubscript{2}, 0 °C – rt; E1: 5 mol% Grubbs II, CH\textsubscript{2}Cl\textsubscript{2}, reflux; F1: CH\textsubscript{2}Cl\textsubscript{2}–TFA, 0 °C – rt, then K\textsubscript{2}CO\textsubscript{3} (6.0 eq.), CH\textsubscript{2}Cl\textsubscript{2}, H\textsubscript{2}O, rt; F2: CH\textsubscript{2}Cl\textsubscript{2}–TFA, 0 °C – rt, then NaOEtBu (2.0 eq.), THF, reflux; TBDPS = tert-butyldiphenylsilyl; Ts = 2- or 4-nitrobenzenesulfonyl; DMAP = 4-dimethylaminopyridine; TBAF = tetra-n-butyrammonium fluoride; DBU = 1,8-diazabicycloundec-7-ene; mCPBA = m-chloroperbenzoic acid; DPE-Phos = bis-[2-[(diphenylphosphino)phenyl]ether; TFA = trifluoroacetic acid; CDI = carbonyl diimidazole.
efficiently and with high diastereoselectivity. § Iodocyclisation 39 (Method B) enabled the cyclisation precursor 11 to be converted, after protecting group manipulation, into the corresponding morpholine (dr 56:44); subsequent treatment with 4-methyl-4\(H\)-1,2,4-triazole-3-thiol and DBU gave separable sulfides which were oxidised and then deprotected to give the morpholine 29. Methods A and B both exploit a potentially variable building block – a (het)aryl bromide or a thiol respectively – and may therefore enable variation of the specific scaffold prepared. The application of alternative electron-deficient (het)aryl bromides in an initial cyclisation by aminoarylation was exemplified in the synthesis of nine scaffolds, five of which are shown in Scheme 3.

Two cyclisation reactions (Methods C and D) exploited simple bifunctional reagents that enabled the formation of a range of five-, six- and seven-membered ring systems. Thus, reaction with CDI (Methods C), if necessary after a telescoped protecting group removal, gave either cyclic ureas (e.g. 30 and 31) or oxazolidinones (e.g. 32). Similarly, reaction with either chloroacetyl chloride or bromoacetyl bromide, followed by alkylation, (Methods D) enabled the synthesis of ketopiperazines (e.g. 33) and ketomorpholines (e.g. 34). In addition,

§ Substrates bearing a remote o-nitrosulfonyl-protected amine did not undergo clean aminoarylation (see ESI‡).
2-keto-1,4-diazepane formation was also possible from the homologous substrates, although generally telescoped with a second cyclisation step (see below). Ring-closing metathesis (Methods E) was often used as a single cyclisation to form tetrahydropyridines (e.g. 35) or dihydropyroles (e.g. 36) but could also be telescoped with a second cyclisation step (see below). Finally, lactamisation (Methods F) enabled the formation of either six- (e.g. 37) or seven- (e.g. 38) membered lactams.

In many cases, a second cyclisation step was possible using one of the complementary cyclisation reactions (Scheme 4). The robustness of Pd-catalysed aminoaarylation (Method A) enabled the formation of pyrrolidines from a wide range of substrates. For example, the substrates 31–34 were converted into the pyrrolidines 39–43 with consistently high diastereoselectivity. Again, the ability to vary the scaffold prepared through careful choice of (het)aryl bromide was widely demonstrated. Method C also had utility as a second cyclisation: Boc deprotection of 29, followed by reaction with CDI, enabled the synthesis of the bicyclic scaffold 44.

As already noted, telescoped procedures incorporating a second cyclisation step were often used to great effect (Scheme 4; structures shown in blue). For example, ring-closing metathesis (Methods E) followed by urea formation (Methods C) enabled the synthesis of bicyclic compounds 45 and 46 directly from acyclic cyclisation precursors (19 and 17). Alternatively, telescoped diketopiperazine or 2-keto-1,4-diazepane formation (Methods D) and ring-closing metathesis procedures enabled the facile synthesis of compounds 47–49.

Lactamisation (Method F) had particular value for substrates prepared by aminoaarylation with 2-methoxycarbonylphenyl bromide in the first cyclisation. For example, deprotection of 24, 26 and 27 triggered cyclisation to yield either six- (e.g. 50, 52 or 53) or seven- (e.g. 51) membered lactams. In the case of pyrrolidine 26, the selective formation of two distinct scaffolds (50 and 51) was possible using orthogonal deprotection reactions.

Discussion

To assess the value of the 52 scaffolds prepared, a virtual library of functionalised compounds was enumerated (see ESI† for full details). The deprotected scaffolds were combined with 59 exemplar medicinal chemistry capping groups. For most scaffolds, up to two capping groups were used. However, for scaffolds whose synthesis had already involved a variable reagent (e.g. those prepared by aminoaarylation), only one capping group was exploited. The resulting virtual library comprised 19 530 likely synthetically-accessible small molecules. To confirm the validity of this analysis, we demonstrated experimentally that scaffold decoration was possible to yield exemplar lead-like compounds from the virtual library (see ESI†).

First, the lead-likeness of the members of the virtual library was assessed (Fig. 1, Panel A). Compounds were successively filtered by molecular size (14 ≤ number of heavy atoms ≤ 26), lipophilicity (−1 ≤ Alog P ≤ 3) and undesirable structural features (see ESI† for specific structural filters). About 59% of the compounds in the virtual library had lead-like molecular

†† The properties of these compounds are representative of the entire ZINC database of ~9 M commercially-available compounds (see ESI†).

The properties of these compounds are representative of the entire ZINC database of ~9 M commercially-available compounds (see ESI†).
properties, and the majority of the outlying compounds only narrowly failed the molecular property filters (heavy atoms: \( \mu = 23.8, \sigma = 4.0; \text{AlogP: } \mu = 0.3, \sigma = 1.3 \)). By comparison, with these specific filters, just 23% of \( \sim 9 \) M commercially-available compounds from the entire ZINC database\(^{40} \) were lead-like, with the majority of compounds lying well outside lead-like chemical space (heavy atoms: \( \mu = 25.9, \sigma = 5.4; \text{AlogP: } \mu = 1.7, \sigma = 2.9 \)) (Fig. 1, Panel B).\(^{11} \) Remarkably, we also showed that, with this set of capping groups, each one of the 52 scaffolds allowed significant regions within lead-like chemical space to be targeted (see ESI†). Our unified synthetic approach thus specifically targeted lead-like chemical space.

Second, we determined the fraction of sp\(^3\)-hybridised carbon atoms (Fsp\(^3\)) in the virtual compounds (Fig. 1, Panel C). It has previously been shown that Fsp\(^3\) correlates strongly with success because compounds in the discovery phase (mean Fsp\(^3\): 0.36) have lower Fsp\(^3\) than marketed drugs (mean Fsp\(^3\): 0.47).\(^{11} \) It has thus been stated that accessing more three-dimensional lead compounds is a desirable goal.\(^{11,17} \)

The mean Fsp\(^3\) of the virtual compounds (0.58) compared very favourably with that of the random sample of compounds from the ZINC database (0.33). Thus, our synthetic approach can yield compounds with significantly greater sp\(^3\) character than most commercially-available compounds, thereby expanding the range of molecular architectures available within lead-like chemical space and offering more flexibility in lead optimisation.

Third, the novelty and diversity of the 52 scaffolds was assessed. A substructure search was performed in which the ZINC database (9 046 036 compounds) was interrogated with each of the deprotected scaffolds. In 43 cases (82%), the deprotected scaffold was not found as a substructure in any compound in the database. Even in the remaining 9 cases, the deprotected compound was not known in the CAS registry. The diversity of, and relationship between, the scaffolds was assessed using an hierarchical analysis (see ESI†).\(^{42} \) It was found that 42 frameworks were represented at the graph-node-bond level, which were related hierarchically to 13 “parent” frameworks. There is significant scaffold diversity at each level of hierarchy, meaning that the scaffolds are not simply closely related derivatives.

In total, 52 diverse molecular scaffolds were prepared from just thirteen different cyclisation precursors. Initially, pairs of building blocks were combined using a single connective reaction – Ir-catalysed allylic amination – before a divergent synthetic approach was used to convert these cyclisation precursors into 52 molecular scaffolds. This approach exploited a toolkit of just six cyclisation reactions, and required on average just 3.0 operations|| for the synthesis of the scaffolds from the constituent building blocks. Furthermore, the unified and modular nature of the strategy means that it has the potential to deliver many additional scaffolds through expansion of the range of building blocks used (e.g. by use of homologated, and stereo- or regioisomerically substituted variants).

Conclusions

Our unified synthetic approach yielded molecular scaffolds that were novel, diverse and lead-like. It was shown that functionalisation of the scaffolds would allow significant lead-like chemical space to be targeted that complements that occupied by commercially-available molecules. A key challenge in lead-oriented synthesis is the identification of complementary and robust reactions with broad functional group compatibility, particularly convergent reactions that may be used to link building blocks.**\(^{41} \) The success of our unified lead-oriented synthetic approach was founded in the identification of a robust convergent reaction that, together with a range of cyclisation reactions, could be exploited in the synthesis of a wide range of novel, diverse and lead-like molecular scaffolds. An increased armoury of such robust convergent reactions would crucially expand the relevant chemical space accessible to drug discovery programmes, and may help to address the grand challenge of increasing productivity in the pharmaceutical sector.\(^{44,45} \)

Acknowledgements

We thank EPSRC (EP/J00894X/1) and GSK for funding, Stephen Pickett for computational analyses, George Burslem, Stuart Warriner and Alan Nadin for useful discussions.

Notes and references


|| For the purposes of this analysis, a synthetic operation is defined as a process conducted in a single reaction vessel.

** An approach to identify such highly robust reactions has recently been reported (ref. 43).
Organic & Biomolecular Chemistry