This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the **Information for Authors**.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard **Terms & Conditions** and the **Ethical guidelines** still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

Journal Name

COMMUNICATION

A Systematic Approach to Diverse, Lead-Like Scaffolds from α,α-Disubstituted Amino Acids†

Daniel J. Foley,[a] Richard G. Doveston,[a] Ian Churcher,[b] Adam Nelson*[a,c] and Stephen P. Marsden*[a]

A powerful strategy for the efficient lead-oriented synthesis of novel molecular scaffolds is demonstrated. Twenty two scaffolds were prepared from just four α-amino acid-derived building blocks and a toolkit of six connective reactions. Importantly, each individual scaffold has the ability to specifically target lead-like chemical space.

There is an intrinsic link between the molecular properties of initial screening compounds and the likelihood of their successful progression towards drug candidates.[1,2] Since both molecular weight and lipophilicity tend to increase during the lead optimisation phase,[3] controlling the molecular properties of lead molecules is crucial if the final compounds are to remain in drug-like space. However, a recent survey concluded that just 2.6% of 4.9M commercially-available molecules were lead-like,[4] even before compound diversity[5] was considered. It has also been shown that compounds with higher fractions of sp^3-hybridised atoms have lower attrition rates than less saturated compounds,[2] and thus may serve as better leads. Obtaining large numbers of diverse, highly three-dimensional, lead-like small molecules is thus a major challenge in maintaining high-quality screening collections. The concept of lead-oriented synthesis has been proposed[4] to frame the challenge of delivering diverse molecules with specific molecular properties of direct relevance to the drug discovery and development process.[4,6,7] Lead-oriented syntheses should, inter alia,[4] focus on the diversity of structures generated, synthetic efficiency, tolerance to polar functionality, and not be susceptible to 'LogP drift', the phenomenon whereby higher failure rates of more polar compounds in a given array synthesis leads to an increase in mean and median Log P for final libraries versus that designed. The significant academic challenge laid down by such demands has recently begun to prompt the syntheses of specific classes of lead-like molecules.[8]

Making diverse compound libraries is not necessarily difficult *per se*, but without careful planning comes at the expense of synthetic efficiency and, until recently,[8a] has lacked a lead-oriented focus. In order to realise efficient lead-oriented synthesis, our approach is to exploit small building blocks that contain multiple chemically-orthogonal functional groups; systematic exploitation of a minimal toolkit of reactions for the pairwise coupling of the functional groups present would rapidly deliver diverse scaffolds that, on decoration of the newly formed or residual functional groups, should target broad regions of lead-like chemical space.
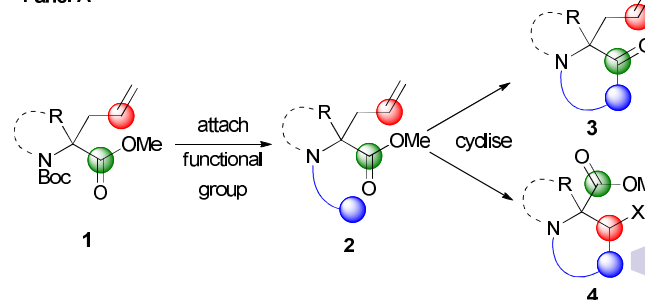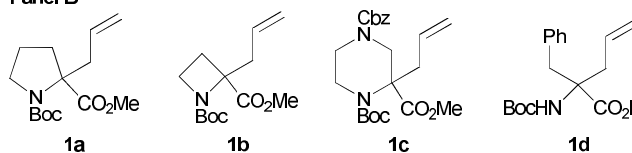
Scheme 1: Overview of the modular synthetic approach.
Panel A: Attachment of alternative functional groups (blue) to α-allyl α-amino acid building blocks 1 yields intermediates 2 which may cyclise to yield alternative scaffold topologies 3 or 4. Panel B: Structures of the specific building blocks used.

To validate our strategy, we chose readily available[9] α-allyl α-amino acid derivatives **1** as the substrates, the amino functionality of which would then be armed with a functional group (blue) to yield a trifunctional intermediate **2**, tuning the precursor for cyclisation with either the adjacent ester (green; → **3**) or alkene (red; → **4**) (Scheme 1). Variation of the resulting scaffold should be possible by exploiting acyclic or cyclic amino acid derivatives (e.g. **1a**-**d**) as substrates; by varying the appended functionality (blue); and by varying the cyclisation reaction.

a. School of Chemistry, University of Leeds, Leeds, LS2 9JT, UK.
b. GlaxoSmithKline Medicines Research Centre, Stevenage, SG1 2NY, UK.
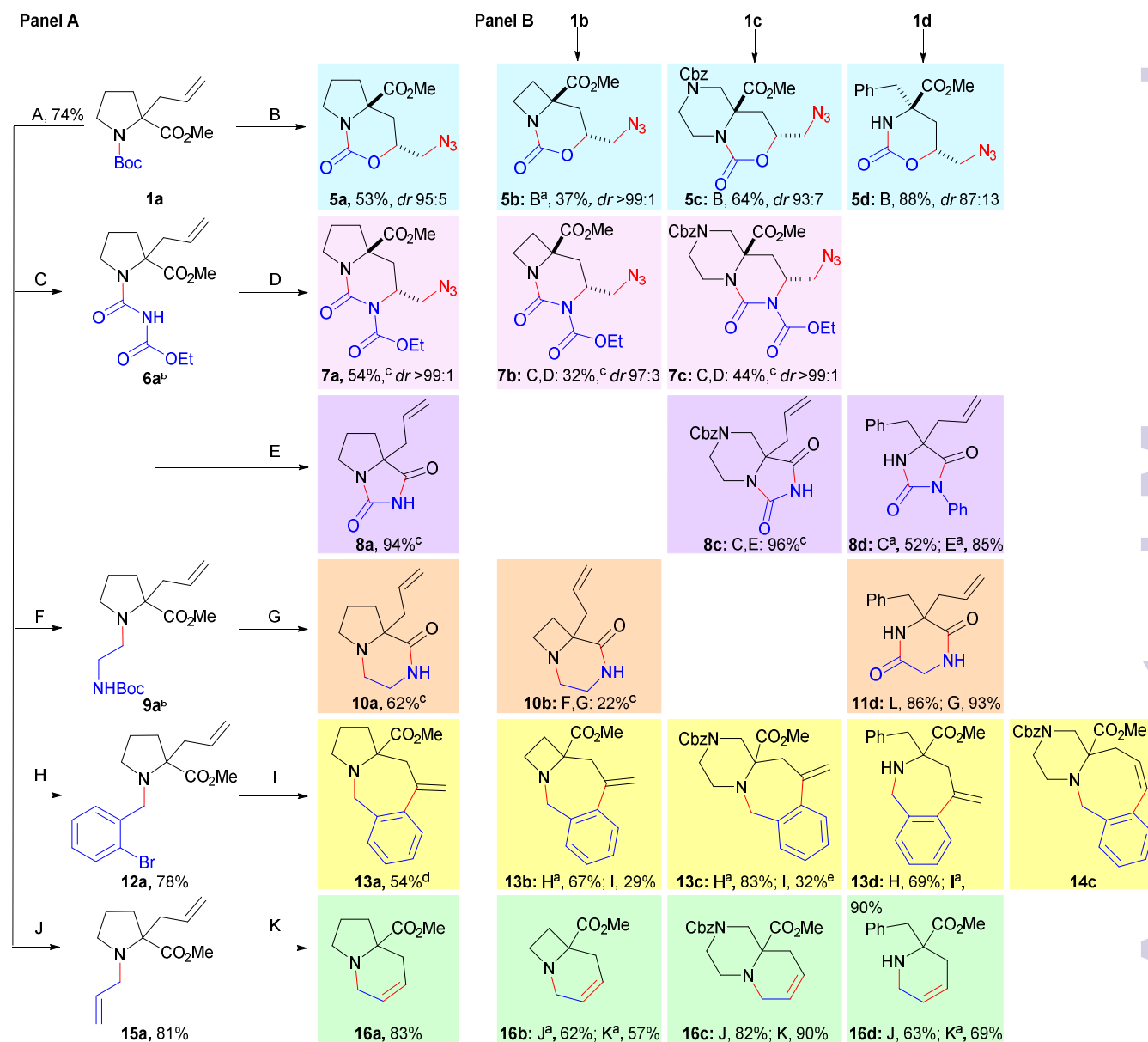c. Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, LS2 9JT, UK.

COMMUNICATION

Panel A

Panel B 1b 1c 1d

1a

5a, 53%, dr 95:5

5b: B[a], 37%, dr >99:1

5c: B, 64%, dr 93:7

5d: B, 88%, dr 87:13

6a[b]

7a, 54%,[c] dr >99:1

7b: C,D: 32%,[c] dr 97:3

7c: C,D: 44%,[c] dr >99:1

8a, 94%[c]

8c: C,E: 96%[c]

8d: C[a], 52%; E[a], 85%

9a[b]

10a, 62%[c]

10b: F,G: 22%[c]

11d: L, 86%; G, 93%

12a, 78%

13a, 54%[d]

13b: H[a], 67%; I, 29%

13c: H[a], 83%; I, 32%[e]

13d: H, 69%; I[a], 90%

14c

15a, 81%

16a, 83%

16b: J[a], 62%; K[a], 57%

16c: J, 82%; K, 90%

16d: J, 63%; K[a], 69%

Scheme 2: Synthesis of diverse lead-like molecular scaffolds. Coloured boxes correspond to different cyclisation reactions, new bonds/groups are shown in red. Panel A: Synthesis of scaffolds from **1a**. Panel B: Extension to alternative scaffolds using building blocks **1b–d**. Standard methods: A: TFA, CH$_2$Cl$_2$; B: (i) I$_2$, THF–H$_2$O; (ii) NaN$_3$, DMF; C: (EtO$_2$C)NCO, CH$_2$Cl$_2$; D: (i) I$_2$, Li[Al(O^{t}Bu)$_4$], THF, −5 °C; (ii) NaN$_3$, DMF; E: NaOMe, 85:15 MeOH–PhMe, 65 °C; F: N-Boc-glycinal, NaBH(OAc)$_3$, CH$_2$Cl$_2$, 4 Å MS; G: (i) TFA, CH$_2$Cl$_2$; (ii) Cs$_2$CO$_3$, DMF, Δ; H: 2-Br-C$_6$H$_4$CHO, NaBH(OAc)$_3$, CH$_2$Cl$_2$, 4 Å MS; I: 5 mol% Pd(PPh$_3$)$_4$, Et$_3$N, MeCN, μw, 125 °C; J: allyl bromide, K$_2$CO$_3$, DMF; K: 2.5 mol% Grubbs II, TsOH, PhMe, Δ; L: EDCI, N-Boc-glycine, Et$_3$N, CH$_2$Cl$_2$. [a]See Supplementary Information for variation from standard methods. [b]Product not purified before subsequent reaction. [c]Yield over 2 steps. [d]92:8 mixture with the regioisomeric azocane. [e]The regioisomeric azocane **14c** also isolated (31% yield).

Firstly, allylation of commercial N-Boc-protected amino acid esters furnished the trifunctional building blocks **1**.[9] Additional intermediates bearing alternative N-substituents were prepared to tune the precursors **1** for cyclisation (Scheme 2, Panel A). Removal of the Boc-protecting group was followed by formation of acyl ureas **6** (treatment with ethyl isocyanatoformate) or by alkylation to give protected 1,2-diamines **9**, o-bromobenzylamines **12** and allylamines **15**. We then used these substrates to synthesise cyclic molecular scaffolds, by exploiting the use of a toolkit of three broad reaction types, all targeting medicinally relevant motifs.

1,2-Amino alcohols, 1,2-diamines and their functionalised derivatives are prevalent in many bioactive molecules.[10] We therefore explored introduction of these motifs by pairwise coupling of the π-nucleophilic allyl group with the nucleophilic N-acyl substituents of **1** and **6** (carbamate[11] or urea[12] respectively).

upon activation with iodine; subsequent treatment with NaN_3 gave cyclic carbamates **5a**-**5d** and the biologically relevant[13] ureas **7a**-**c**.

Pairwise coupling of the electrophilic carboxylate ester with nitrogen-centred nucleophiles led to hydantoins **8a,c**-**d** (a motif known for anticonvulsant biological activity, e.g. phenytoin) from acyl ureas **6** and lactams **10a**-**b**/diketopiperazine **11d** from protected amines **9**. Finally, ring closure between the *N*-substituent and the C2-allyl group could be facilitated through transition metal catalysis: Pd-catalysed Heck cyclisations of substrates **12** gave the azepanes **13a**-**d** and the azocane **14c**. Similarly, Ru-catalysed ring-

Figure 1: Analysis of molecular properties and shape diversity. A virtual library of 1110 compounds derived from the 22 scaffolds (orange) compared with 90911 randomly-selected compounds from the ZINC database (blue). Panel A: Distribution of molecular properties (black rectangle: guide properties used to delineate lead like space). Panel B: Normalised principal moment of inertia plot.

closing metathesis allowed the synthesis of the indolizidine alkaloid-like tetrahydropyridines **16a** and the congeners **16b**-**d**. In this way we prepared 22 distinct scaffolds from the four building blocks **1a**-**d** in a total of just 49 synthetic operations[†] using a toolkit consisting of only six reaction methodologies and one *N*-capping event.

To assess the novelty of these scaffolds, substructure searches against the ZINC and CAS databases were performed. A total of 37 variants of the scaffolds (with protecting groups removed and allowing for modifications available through standard functional group interconversions; see Supporting Information for full details) were screened against the ZINC database of 9,039,756 commercially available compounds. This returned only 28 hits, all of which were derived from just two of the scaffolds (**10a/16a**). A further analysis of the Murcko assemblies of all scaffolds (with alpha-attachments) against a random 5% of the ZINC database

again only returned hits (4) against the same two scaffolds **10a/16a**. Thus, it is clear that the scaffolds display a very high degree of novelty relative to commercial collections. Finally, the scaffolds were also assessed for skeletal diversity: using an hierarchical approach[15] it was found that each scaffold comprised a unique molecular framework at the GNB level.[‡] Thus, the scaffolds are not simple derivatives of each other, but represent a skeletally diverse collection.

The potential of the scaffolds to provide access to lead-like screening compounds was demonstrated by enumeration of a virtual library of 1110 compounds from the deprotected scaffolds and 80 typical medicinal chemistry capping groups (see Supplementary Information). Assessment of the lead-likeness of the virtual library (Figure 1, Panel A) showed that 66% of compounds survived filtering by molecular size ($14 \leq$ heavy atom count ≤ 26; $\mu = 22.8$, $\sigma = 3.57$), lipophilicity ($-1 < AlogP < 3$; $\mu = 0.38$, $\sigma = 1.38$) and structural filters (see Supplementary Information). By comparison, just 23% of the ~9M compounds from the ZINC database[15] survived this filtering process, with most compounds lying well outside lead-like chemical space (heavy atoms: $\mu = 25.9$, $\sigma = 5.4$; Alog P: $\mu = 1.7$, $\sigma = 2.9$). Remarkably, using the same set of capping groups, each of the 22 scaffolds allowed significant lead-like chemical space to be targeted. In addition to this *in silico* study, we have shown practically that decoration of exemplar scaffolds is possible (Supplementary Information).
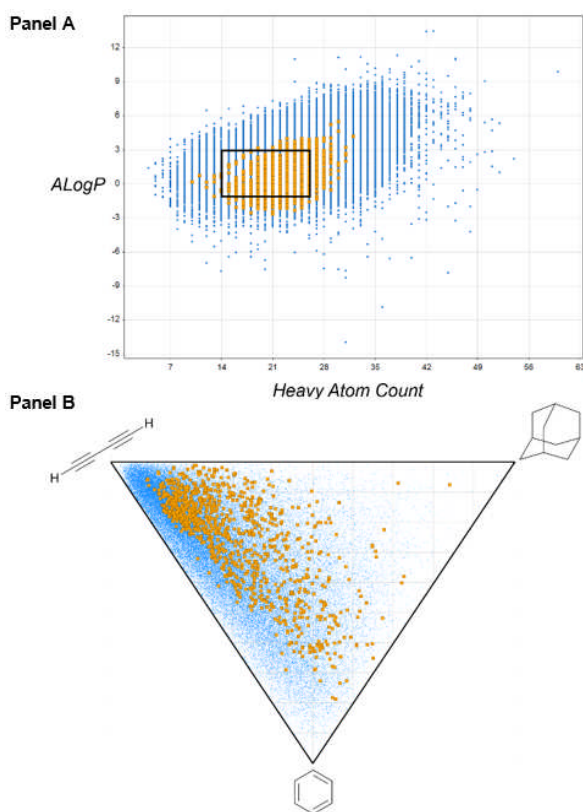
The shape diversity of the virtual library was compared with that of 90911 randomly-selected compounds from the ZINC database (Figure 1, Panel B). For each compound, the three normalised principal moments of inertia values were determined for a low energy conformation.[16] The virtual library was significantly more three-dimensional (Fsp^3: $\mu = 0.57$) than the commercially-available compounds (Fsp^3: $\mu = 0.33$).

In this study we have shown that careful selection of small, polyfunctional substrates facilitates an efficient approach to the synthesis of molecular scaffolds that are novel, diverse, and can specifically target lead-like chemical space. This was achieved using a toolkit of just six cyclisation methodologies, leading to 22 scaffolds in only 49 synthetic operations. In particular, the efficient and diverse coverage of 3D space from such a small set of reagents is notable. This general approach should be applicable to many classes of polyfunctional substrate, enabling the more efficient exploration of lead-like chemical space.

Notes and references

† Defined as a process conducted in a single reaction vessel.

‡ See the scaffold tree in SI for further detail.

1 (a) M. C. Wenlock, R. P. Austin, P. Barton, A. M. Davis and P. D. Leeson, *J. Med. Chem.* 2003, **46**, 1250; (b) P. D. Leeson and B. Springthorpe, *Nat. Rev. Drug. Discov.* 2007, **6**, 881.

2 F. Lovering, J. Bikker and C. Humblet, *J. Med. Chem.* 2009, **52**, 6752.

3 (a) T. I. Oprea, A. M. Davis, S. J. Teague and P. D. Leeson, *J. Chem. Inf. Comput. Sci.* 2001, **41**, 1308; (b) G. M. Keserü and G. M. Makara, *Nat. Rev. Drug. Discov.* 2009, **8**, 203.

4 A. Nadin, C. Hattotuwagama and I. Churcher, *Angew. Chem. Int. Ed.* 2012, **51**, 1114.

5   A. H. Lipkus, Q. Yuan, K. A. Lucas, S. A. Funk, W. F. Bartelt, R. J. Schenck and A. J. Trippe, *J. Org. Chem.* 2008, **73**, 4443.

6   P. MacLellan and A. Nelson, *Chem. Commun.* 2013, **49**, 2383.

7   R. Doveston, S. Marsden and A. Nelson, *Drug Discov. Today* 2014, **19**, 813.

8   (a) R. G. Doveston, P. Tosatti, M. Dow, D. J. Foley, H. Y. Li, A. J. Campbell, D. House, I. Churcher, S. P. Marsden and A. Nelson, *Org. Biomol. Chem.* 2015, **13**, 859. (b) S. V Ryabukhin, D. M. Panov, D. S. Granat, E. N. Ostapchuk, D. V Kryvoruchko and O. O. Grygorenko, *ACS Comb. Sci.* 2014, **16**, 146; (c) A. Borisov, V. Voloshchuk, M. Nechayev and O. Grygorenko, *Synthesi*s 2013, 2413; (d) T. James, I. Simpson, J. A. Grant, V. Sridharan and A. Nelson, *Org. Lett.* 2013, **15**, 6094; (e) T. James, P. MacLellan, G. M. Burslem, I. Simpson, J. A. Grant, S. Warriner, V. Sridharan and A. Nelson, *Org. Biomol. Chem.* 2014, **12**, 2584.

9   P. N. Confalone, E. M. Huie, S. S. Ko and G. M. Cole, *J. Org. Chem.*, 1988, **53**, 482.

10 (a) T. J. Donohoe, C. K. A. Callens, A. Flores, A. R. Lacy and A. H. Rathi, *Chem. Eur. J.*, 2011, **17**, 58. (b) D. Lucet, T. Le Gall and C. Mioskowski, *Angew. Chem. Int. Ed.*, 1998, **37**, 2580.

11 M. Pattarozzi, C. Zonta, Q. B. Broxterman, B. Kaptein, R. De Zorzi, L. Randaccio, P. Scrimin and G. Licini, *Org. Lett.* 2007, **9**, 2365.

12 M. Fujita, O. Kitagawa; T. Suzuki and T. Taguchi, *J. Org. Chem.* 1997, **62**, 7330.

13 E. De Clercq, *Biochim. Biophys. Acta* 2002, **1587**, 258.

14 A. Schuffenhauer, P. Ertl, S. Roggo, S. Wetzel, M. A. Koch and H. Waldmann, *J. Chem. Inf. Model.*, 2007, **47**, 47.

15 J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad and R. G. Coleman, *J. Chem. Inf. Model.*, 2012, **52**, 1757.

16 W. H. Sauer and M. K. Schwarz, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 987.

COMMUNICATION

Graphical abstract:

A powerful strategy for the efficient lead-oriented synthesis of novel molecular scaffolds is demonstrated. Twenty two distinct scaffolds were prepared from four quaternary α-amino acid building blocks in a total of only 49 synthetic operations, using a toolkit of six connective reactions. The ability of each individual scaffold to specifically target lead-like chemical space was demonstrated computationally.