Article:
Morgentin, R, Dow, M, Aimon, A et al. (5 more authors) (2018) Translation of Innovative Chemistry into Screening Libraries: An Exemplar Partnership from the European Lead Factory. Drug Discovery Today, 23 (9). pp. 1578-1583. ISSN 1359-6446

https://doi.org/10.1016/j.drudis.2018.05.007

Translation of Innovative Chemistry into Screening Libraries: An Exemplar Partnership from the European Lead Factory

Remy Morgentin,[a] Mark Dow,[b,c] Anthony Aimon,[b,c] George Karageorgis,[b,c] Tuomo Kalliokoski,[d] Didier Roche,[a] Stephen Marsden[b] and Adam Nelson[b,c]

[a]Edelris, 115 Avenue Lacassagne, F-69003 Lyon, France

[b]School of Chemistry, University of Leeds, Leeds LS2 9JT, UK

[c]Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, UK

[d]Lead Discovery Center GmbH, Otto-Hahn-Straße 15, 44227 Dortmund, Germany

[e]Current address: Orion Pharma, Orionintie 1A, 02101 Espoo, Finland

Corresponding author email address: a.s.nelson@leeds.ac.uk

Abstract

The identification of high-quality starting points for drug discovery is an enduring challenge in medicinal chemistry. Yet, the chemical space explored in discovery programmes tends be limited by the narrow toolkit of robust methods that are exploited in discovery workflows. The European Lead Factory (ELF) was established in 2013 to boost early-stage drug discovery within Europe. In this Feature, we describe an exemplar partnership from that has led to the addition of 21 119 distinctive screening compounds to the ELF Joint European Compound Library. The partnership may serve as a blueprint for the translation of innovative academic chemistry into discovery programmes.

Introduction

The identification of high-quality starting points for drug discovery is an enduring challenge in medicinal chemistry. To facilitate the discovery of high-quality lead molecules, lead-likeness guidelines have been formulated[1] that account for the tendency for molecular complexity, molecular size and lipophilicity to increase during optimisation.[2] Yet, chemical innovation in molecular discovery is still limited by the historically uneven and unsystematic exploration of chemical space.[3] Indeed, medicinal chemists tend to harness a remarkably small number of robust methods[4-6] that may be easily accommodated within discovery workflows. As a result, chemists are often driven away from optimal property space,[7] increasing their focus on flatter and more lipophilic compounds. Recently, a range of initiatives – both within companies and collaboratively – has been established to broaden the chemical space that is available in early-stage drug discovery.

The European Lead Factory (ELF) is an example of a collaborative platform that was established in 2013 to boost early-stage drug discovery within Europe. At the heart of ELF is its Joint European Compound Library (JECL) that was initiated by the contribution of 321 000 screening compounds by seven pharmaceutical company partners. This initial set has now being complemented through the addition of ~200 000 new compounds: the public compound collection (PCC).[8] The PCC was designed to populate biologically-relevant chemical space that is not explored in existing compound collections.

In this Feature, we describe our experience of translating innovative chemistry into screening libraries as part of the ELF initiative. The libraries described here were proposed by staff at the University of Leeds (Leeds, UK); reviewed and selected by the ELF Library Selection Committee; and validated experimentally in Leeds.[9] For each library proposal, the aims of the validation work were to develop a practical and scalable (10s of grams) route to the required scaffolds; and to demonstrate the ready decoration into exemplar screening compounds. After successful validation, a library of diverse screening compounds was designed and nominated by Lead Discovery Center (Dortmund, DE) using capping groups selected using a previously described process.[8] The screening compounds were produced by Edelris (Lyon, FR) as racemic mixtures of single diastereoisomers.This partnership yielded a total of 21 119 screening compounds that were based on 34 distinct library proposals (19 924 on the >5 mg scale that was targeted). The molecular properties and novelty of many of these compounds has already been described together with the details of the underpinning synthetic chemistry (for details for specific libraries, see the references in Figure 1). Around 20 000 additional compounds were also produced by Edelris based on other library proposals. The novelty (compared to other screening collections and ChEMBL) and molecular properties of the first ~50 000 compounds in the PCC has been described.[8] Crucially, it was demonstrated that produced compounds were diverse and distinctive, and have appropriate molecular properties. Our experience of successful collaboration may be of value in the construction of other partnerships that are focused on the translation of innovative chemistry for exploitation in industrial practice.

Overview of collaborative approach

Before the formal start of the ELF initiative, the University of Leeds and Edelris had already identified a common interest in sp^3-rich and natural product-like compounds. Crucially, both partners were aware of – and were excited by – the challenges that would likely be associated with the production of large libraries of such compounds. Dedicated project managers were appointed at each partner who subsequently liaised closely and oversaw local activities throughout the project. To assist the translation of validated chemistry into library production, protocols for routine scaffold decoration and compound purification were shared at the outset.

The libraries described in this Feature were proposed by academic staff, postdoctoral researchers and PhD students at the University of Leeds. To enable rapid progress at the start of the ELF, early library proposals were adapted from synthetic approaches that had already been developed at Leeds. Scaffold syntheses were designed to exploit combinations of (i) known synthetic approaches to natural product and natural product-like scaffolds; (ii) recently-disclosed transformations that had not been widely applied to natural products; and (iii) reactions recently developed in Leeds (see below for details of underpinning chemistry). Judicious choice of substrates enabled the synthesis of scaffolds with generally 2-3 sites for decoration. Before submission of library proposals to the ELF Library Selection Committee, the project manager at Edelris had the opportunity to provide feedback on problems that might be encountered during library production: for example, safety issues, the cost of chemicals, and the length of proposed synthetic routes. In most cases, this feedback informed the aims of the experimental validation work; and, occasionally, it was decided not to submit a library proposal to ELF.

The library production (at Edelris) was assessed using tools that had been developed to monitor progress. In particular, the success rate for the synthesis and purification of final screening compounds was carefully monitored throughout the project (and increased from ~72% to ~85% during this time). In addition to monthly teleconferences, annual face-to-face meetings brought the project teams together. Moreover, the teams met to celebrate their shared success when key milestones were met (e.g. following the synthesis of the 10 000th screening compound by Edelris). Throughout the ELF project, a key challenge to maintain a stocked pipeline of library proposals at all stages from the initial proposal through library production.

Overview of the libraries produced

The 34 library proposals that were translated into screening compounds are summarised in Figure 1 (average library size: 586). In each case, a feasible synthesis of the scaffold(s) was developed at the University of Leeds, and was typically been executed on a 5-10 g scale. Notably, the molecular complexity of the scaffolds tended to increase as the partnership developed (for the approximate chronological order, see Figure 1). To demonstrate the scope for scaffold decoration, 10-20 final compounds were typically prepared at Leeds to investigate a range of different capping chemistries. To ensure ready translation, Edelris shared both standard capping reaction conditions and purification protocols, which were implemented at Leeds. Based on the demonstrated scope, and liaison with Edelris, libraries were then nominated by Lead Discovery Center for production.

The ease of production of screening compounds varied widely between libraries (see Figure 1 for an assessment of the ease of the final production step). In three cases, the difficulties encountered meant that compound production was stopped prematurely: specific

difficulties encountered with libraries **9**, **12** and **33** (see below) resulted in relatively few compounds being produced (35, 50 and 37 compounds respectively). In other cases, more general problems were encountered. First, the diversity of the produced compounds was often limited by the actual scope of the chemistry (e.g. libraries **3**, **19** and **24**), the availability and/or price of reactants (e.g. libraries **3** and **6**) and by the need to control molecular weight and lipophilicity (e.g. library **19**). Second, the properties of some final compounds (e.g. more polar analogues in library **3** and more lipophilic compounds in library **19**) prompted the use of a focused solvent gradient for HPLC in conjunction with mass-triggered purification.[10] Finally, in some cases, specific final compounds were not stable, for example under the conditions used during purification (e.g. libraries **3**, **12**, **15**, **24**). Nonetheless, for many libraries, high synthetic feasibility and diversity potential meant that large numbers of diverse compounds could be produced (>800 compounds for **13**, **14**, **17**, **22**, **25**, **26** and **28**). Furthermore, the diversity of five libraries (**17**, **18**, **21**, **25** and **26**) was substantially increased by preparing sub-libraries in which the final decoration step (which tended to introduce more alternative substituents) was changed.

A wide range of reaction classes was harnessed, many of which are rarely[4-6] exploited to drive medicinal chemistry programmes. The reaction classes exploited in the key cyclisation step(s) are summarised in the Table. Notably, key reactions that had been harnessed successfully were often exploited again subsequently in the synthesis of additional scaffolds. For example, cycloadditions ([4+2], [3+2], [2+2], [5+2] and [4+3]) were exploited in the synthesis of many scaffolds. In addition, the range of transition metal-catalysed cyclisations employed is also notable: for example, Pd-catalysed aminoarylation;[11] Ru-catalysed ring-closing metathesis; Au- or Pt-catalysed activation of alkynes;[12, 13] and Rh-catalysed C-H activation chemistry.[14] Pd-catalysed aminoarylation had particular value since it enabled both the scaffold synthesis and its diversification.

Examples of successes and challenges encountered in validation and production

Some of the successes and challenges encountered during the translation process are illustrated by five library proposals that we have not previously described elsewhere (Figure 2). The synthesis of the bridged bicyclic scaffold **37** was straightforward, and exploited a Pt-catalysed reaction sequence that was optimised during validation work (Library **10**; Panel A). Here, a key finding was that, with this substrate, $PtCl_4$ was superior to other reported[13] catalyst systems involving gold or other platinum salts. The syntheses of the fused bicyclic scaffolds **39** and **40** were also straightforward (Library **24**; Panel B): reaction of the Diels–Alder reaction adduct **38** with $LiAlH_4$ resulted in reduction of the imide and reductive ring-opening of the bridged ether. After diversification of the secondary alcohol ($\rightarrow$ **40**), highly diastereoselective dihydroxylation was possible. Several problems did arise, however, during library production when specific *O*-decorating groups were exploited. With an *O*-2-thiazolyl group, catalyst poisoning prevented efficient *N*-debenzylation; whilst with

an *O*-benzoxazolyl group, *N*-debenzylation was followed by *O*→*N* migration of the hetaryl group.

Significant problems were encountered during both the validation and the production of the library **9** (Panel C). During validation work, the Lewis acid-catalysed reaction of the cyclic allylsiloxane **42** with aldehydes[19] was found to be limited in scope: although good yields and diastereoselectivity were observed with simple aliphatic aldehydes (e.g. → **42a** and **42b**), low yields (at best) were obtained with aromatic substrates. Oxidative cleavage of the alkene of **42a**, and reductive amination, yielded either the bridged lactam **44** (with primary amines e.g. *p*-chlorobenzylamine) or the tetrahydrofuran **45** (with secondary amines e.g. morpholine). On the basis of this validation work, a library was nominated for production. However, on a larger scale, significant epimerisation was observed during both reductive amination (e.g. of **43**) and subsequent saponification (e.g. of **45**). For these reasons, and because of the length of the synthesis, production was abandoned after a small library (35 final compounds) had been prepared; because of the premature termination of the project, no final compounds were ultimately based on the bicyclic lactam scaffold.

With two proposed libraries, successful validation was not possible despite the availability of an efficient scaffold syntheses. The doubly spirocyclic scaffold **49** was prepared in a single step via three-component reaction of the isatin **46**, sarcosine **47** and the α,β-unsaturated lactam **48** (80%; dias.: >98:<2). However, the scaffold **49** and its derivatives were extremely insoluble, which prevented efficient decoration. In a similar vein, oxidation of the phenol **51** with PIFA was followed by intramolecular Diels–Alder reaction to give the complex scaffold **52** in 80% yield and >98:<2 diastereoselectivity. Once more, efficient decoration was not possible, in part due to solubility problems. In both cases, a library was not nominated for production.

Scaffold level analysis

A wide range of distinctive and sp^3-rich scaffolds was exploited in the screening compounds prepared. In 31 of the 34 cases, the Murcko frameworks[20] of scaffolds (both with and without α atoms) are not found as substructures in a random 2% of compounds in the ZINC[21] database of commercially-available compounds. In contrast, the Murcko frameworks of three monocyclic frameworks are found as substructures of compounds in the ZINC database (without α atoms: **8**, found in 0.84% of this random selection of compounds from the ZINC database; **9**, 2.3% of compounds; **24**: 0.16% of compounds). In these cases, novelty is markedly increased when α atoms are included (**8**, found in 0.01% of this random selection of compounds from the ZINC database; **9**, 0.12% of compounds; **24**: 0% of compounds), reflecting the unusual *C*-substitution of these rings.

The shape diversity of the scaffolds was also assessed using the computational tool LLAMA[22] (Figure 3). Many of the scaffolds are highly three-dimensional, highlighting the potential for unusual vectors to be explored through decoration. Thus, our libraries are nicely distributed within novel, lead-like chemical space. In addition, many of the scaffolds have high natural product likeness scores[23] that may increase the biological relevance of the resulting compound libraries.

Concluding remarks

An efficient partnership was established to harness innovative chemistry from academia in the production of screening compounds. The produced screening compounds were generally based on highly distinctive molecular scaffolds that were sp^3-rich and often natural product-like. It is envisaged that the partnership could serve as a blueprint for the translation of innovative academic chemistry into compounds that align with the needs of discovery-based industries. We recognise that the full impact of the produced compounds within the PCC can only be realised following screening against a wide range of targets, and the overall performance of this screening set will be described in due course.

Acknowledgements

We acknowledge support from the Innovative Medicines Initiative Joint Undertaking under grant number 115489, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (F97/2008–2013), EFPIA companies' in kind contribution. We thank Rachel Guilleux, Martin Ohsten, Nicolas Ruiz-Llamas, Claude Lardy, Peter Nussbaumer, Philip Craven and George Heath for their contributions, and Shiao Chow and Sam Liver for assistance with figure preparation.

References

[1] Nadin A, Hattotuwagama C, Churcher I. Lead-Oriented Synthesis: A New Opportunity for Synthetic Chemistry. Angew. Chem. Int. Ed. 2012; 51, 1114-1122.

[2] Oprea TI, Davis AM, Teague SJ, Leeson PD. Is there a difference between leads and drugs? A historical perspective. J. Chem. Inf. Comput. Sci. 2001; 41, 1308-1315.

[3] Lipkus AH, Yuan Q, Lucas KA, Funk SA, Bartelt III WF, Schenck RJ, Trippe AJ. Structural Diversity of Organic Chemistry. A Scaffold Analysis of the CAS Registry. J. Org. Chem. 2008; 73, 4443-4451.

[4] Roughley SD, Jordon AM. The Medicinal Chemist's Toolbox: An Analysis of Reactions Used in the Pursuit of Drug Candidates. J. Med. Chem. 2011; 54, 3451-3479.

[5] Cooper TWJ, Campbell IB, Macdonald SJF. Factors determining the selection of organic reactions by medicinal chemists and the use of these reactions in arrays (small focused libraries). Angew. Chem., Int. Ed. 2010; 49, 8082-8091.

[6] Brown DG, Boström J. Analysis of Past and Present Synthetic Methodologies on Medicinal Chemistry: Where Have All the New Reactions Gone? J. Med. Chem. 2016; 59, 4443-4458.

[7] Walters WP, Green J, Weiss JR, Murcko MA. What Do Medicinal Chemists Actually Make? A 50-Year Retrospective. J. Med. Chem. 2011; 54, 6405-6416.

[8] (a) Karawajczyk A, Giordanetto F, Benningshof J, Hamza D, Kalliokoski T, Pouwer K, Morgentin R, Nelson A, Müller G, Piechot A, Tzalis D. Expansion of chemical space for collaborative lead generation and drug discovery: The European Lead Factory Perspective. Drug Discov. Today 2015; 20, 1301-1316; (b) Kalliokoski T., Combinatorial Library Optimization Process In The European Lead Factory Project. 10th International Conference on Chemical Structures & 10th German Conference on Chemoinformatics, Noordwijkerhout, Netherlands, 2014. http://www.int-conf-chem-structures.org/fileadmin/user_upload/ICCS_2014/posters/P22-Kalliokoski.pdf.

[9] Giordanetto F, Jones P, Nelson A, Benningshof J, Müller G, Pannifer A, van Boeckel S, Tzalis D. European Lead Factory. in Comprehensive Medicinal Chemistry III 2017, 505-519. [10] Koza P, Gonnot V, Pelleter J, Right-First-Time Isocratic Preparative Liquid Chromatography-Mass Spectrometry Purification. ACS Comb. Sci. 2012; 14, 273-279.

[11] Bertrand MB, Leathen ML, Wolfe JP. Mild Conditions for the Synthesis of Functionalized Pyrrolidines via Pd-Catalyzed Carboamination Reactions. Org. Lett. 2007; 9, 457-460.

[12] Firth JD, Zhang R, Morgentin R, Guilleux R, Kalliokoski T, Warriner S, Foster R, Marsden SP, Nelson A. Exploitation of the Ugi–Joullié Reaction in the Synthesis of Libraries of Drug-Like Bicyclic Hydantoins. Synthesis 2015; 47, 2391-2406.

[13] Barluenga J, Fernández A, Satrústegui A, Diéguez A, Rodríguez F, Fañanás FJ. Tandem Intramolecular Hydroalkoxylation–Hydroarylation Reactions: Synthesis of Enantiopure Benzofused Cyclic Ethers from the Chiral Pool. Chem Eur. J. 2008; 14, 4153-4156.

[14] (a) Rakshit S, Grohmann C, Besset T, Glorius F. Rh(III)-Catalyzed Directed C−H Olefination Using an Oxidizing Directing Group: Mild, Efficient, and Versatile. J. Am.

Chem. Soc. 2011, 133, 2350-2353; (b) Davis TA, Hyster TK, Rovis T. Rhodium(III)-Catalyzed Intramolecular Hydroarylation, Amidoarylation, and Heck-type Reaction: Three Distinct Pathways Determined by an Amide Directing Group. Angew. Chem. Int. Ed., 2013, 52, 14181-14185; (c) Guimond N, Gorelsky SI, Fagnou K. Rhodium(III)-Catalyzed Heterocycle Synthesis Using an Internal Oxidant: Improved Reactivity and Mechanistic Studies. J. Am. Chem. Soc. 2011; 133, 6449–6457.

[15] Craven P, Aimon A, Dow M, Fleury-Bregeot N, Guilleux R, Morgentin R, Roche D, Kalliokoski T, Foster R, Marsden SP, Nelson A. Design, synthesis and decoration of molecular scaffolds for exploitation in the production of alkaloid-like libraries. Bioorg. Med. Chem. 2015; 23, 2629-2635.

[16] Colomer I, Adeniji O, Burslem GM, Craven P, Rasmussen MO, Willaume A, Kalliokoski T, Foster R, Marsden SP, Nelson A. Aminomethylhydroxylation of alkenes: Exploitation in the synthesis of scaffolds for small molecule libraries. Bioorg. Med. Chem. 2015; 23, 2736-2740.

[17] Firth JD, Craven PGE, Lilburn M, Pahl A, Marsden SP, Nelson A. A biosynthesis-inspired approach to over twenty diverse natural product-like scaffolds. Chem. Commun. 2016; 52, 9837-9840.

[18] Foley DJ, Craven PGE, Collins PM, Doveston RG, Aimon A, Talon R, Churcher I, von Delft F, Marsden SP, Nelson A. Synthesis and Demonstration of the Biological Relevance of sp^3-rich Scaffolds Distantly Related to Natural Product Frameworks. Chem Eur. J. 2017; 23, 15227-15232.

[19] (a) Cassidy JH, Marsden SP, Stemp G. Stereoselective Synthesis of 2,3,5-Trisubstituted Tetrahydrofurans by an Allyl Silane Metathesis - Nucleophilic Addition Sequence. Synlett 1997; 1411-1413; (b) Miles SM, Marsden SP, Leatherbarrow RJ, Coates WJ. Reagent-Controlled Stereoselective Synthesis of Lignan-Related Tetrahydrofurans. J. Org. Chem. 2004; 69, 6874-6882.

[20] Bemis GW, Murcko MA. The properties of known drugs. 1. Molecular frameworks. J. Med. Chem. 1996, 19, 2887-2893.

[21] Irwin JJ, Sterling T, Mysinger MM, Bolstad ES, Coleman RG. ZINC: A Free Tool to Discover Chemistry for Biology. J. Chem. Inf. Model. 2012; 52,1757–1768.

[22] Colomer I, Empson CJ, Craven P, Owen Z, Doveston RG, Churcher I, Marsden SP, Nelson A. A divergent synthetic approach to diverse molecular scaffolds: assessment of lead-likeness using LLAMA, an open-access computational tool. Chem. Commun. 2016; 52, 7209-7212.

[23] Ertl P, Roggo S, Schuffenhauer A. Natural product-likeness score and its application for prioritization of compound libraries. J. Chem. Inf. Model. 2008; 48, 68-74. Implemented

in RDKit v2015.09.2 (Greg Landrum; Open Source Cheminformatics; http://www.rdkit.org; last accessed 27-Apr-2016).

Table: Reaction classes exploited in scaffold synthesis

Reaction class	**Scaffold**	**Reference**
Pd-catalysed aminoarylation	**1**, **2**, **14**	[15]
[4+2] cycloaddition	**3**, **8**, **18**, **24**, **25**, **34**	**3**: [15] **8**: [16] **18**, **25**: [17]
Other cycloaddition	**4**, **7**, **15**, **16**, **19**, **20**, **23**, **28**, **31**, **32**	**4**, **7**: [15] **16**, **19**, **20** and **23**: [18], *a*
Ring-closing metathesis	**8**	**8**: [16]
Reaction of alkene with electrophile	**9**, **29**	**9**: *b*
Au- or Pt-catalysed cyclisation	**10**, **11**	**10**: *b* **11**: [12]
Ugi reaction	**11**, **17**	[12]
Other Pd-catalysed reaction	**12**, **22**	*a*
Rh-catalysed C-H activation	**13**, **21**, **26**, **30**	*a*
Oxidative dearomatisation	**27**	*a*
Other C-X bond formation	**5**, **6**, **17**, **25**, **33**	**17**: [12] **25**: [17]

[a]The translation of this chemistry has not been published to date. [b]Described elsewhere in this paper.

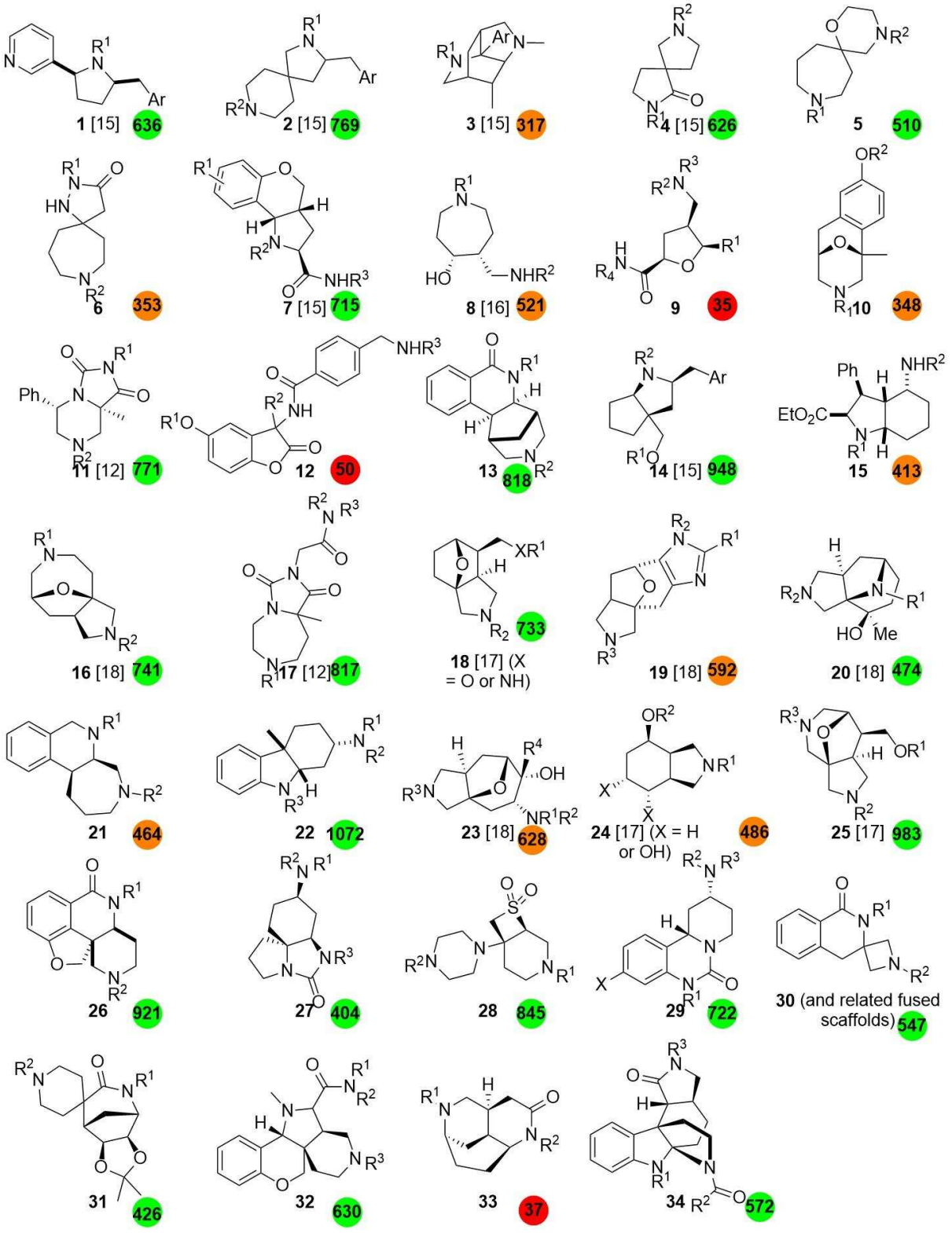

Figure 1: Library proposals that were translated into produced screening compounds (number of final compounds produced on the targeted >5 mg scale in circle) shown in approximate chronological order. In each case, the ease of production of the final compounds is indicated (green: straightforward; orange: challenging; red: difficulties resulted in production being stopped). The synthetic chemistry for library proposals **9**, **10** and **24** is summarised in Figure 2, and other published synthesis cited.

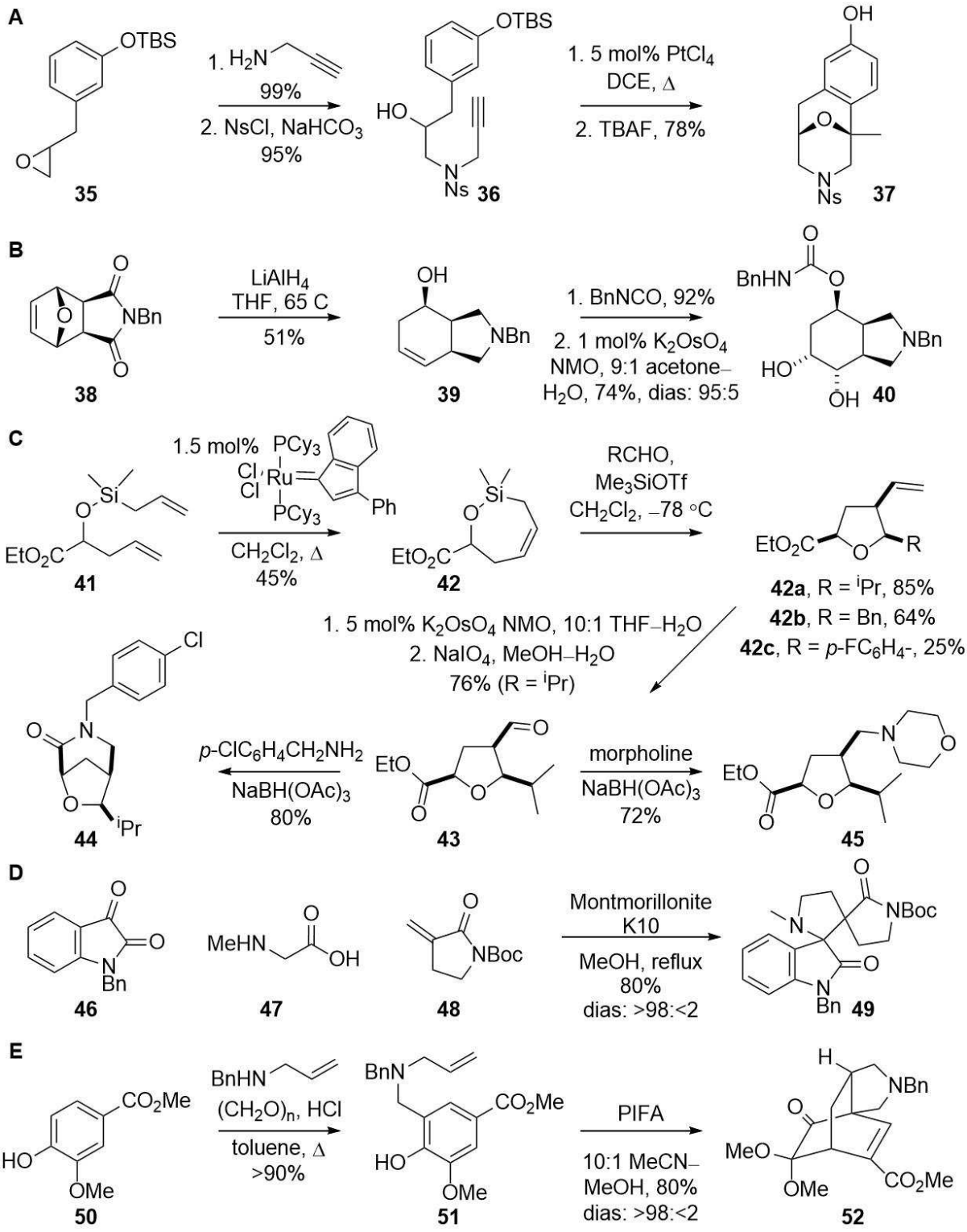

Figure 2: Synthesis of exemplar scaffolds that were explored for exploitation in library proposals: **10** (Panel A), **24** (Panel B), **9** (Panel C) and two proposals that were not validated (Panels D and E).

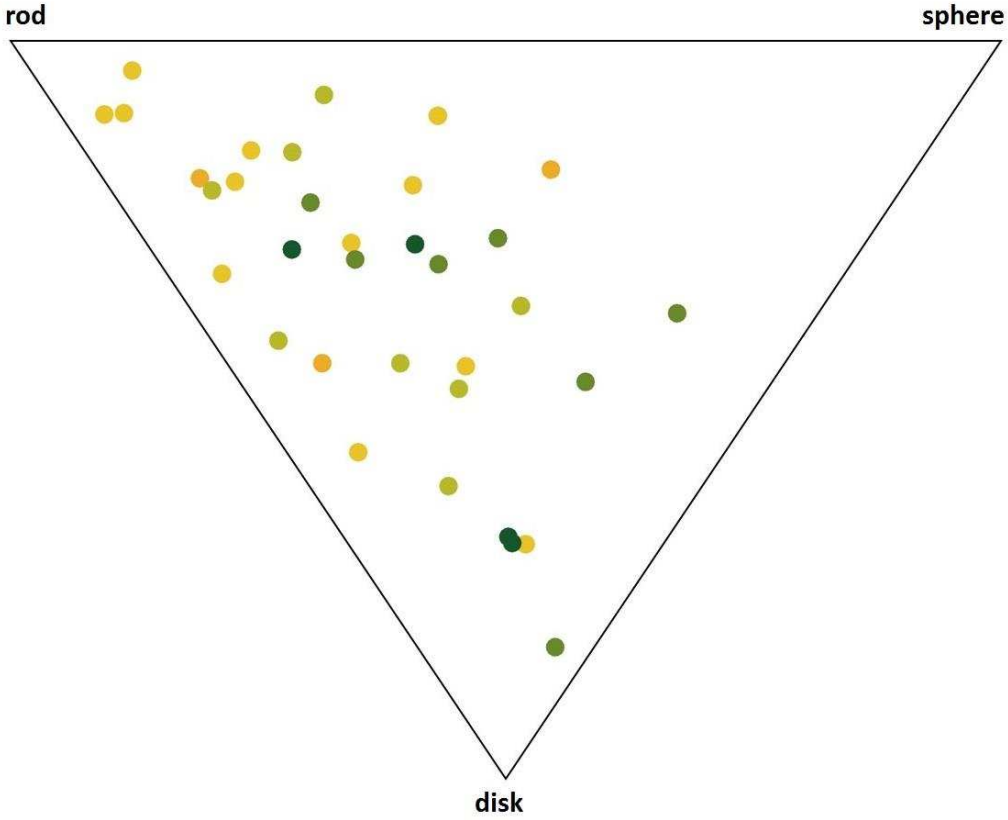

Figure 3: PMI plot to capture the three-dimensionality of the scaffolds **1**-**34** (with R^n = H and a representative Ar group). Natural product-likeness scores are indicated by colour (negative: red; 0: yellow; positive: green).