ARTICLE

Natural Product-informed exploration of chemical space to enable bioactive molecular discovery

Adam Nelson[a,b] and George Karageorgis[a,b*]

The search for new bioactive molecules remains an open challenge limiting our ability to discover new drugs to treat disease and chemical probes to comprehensively study biological processes. The vastness of chemical space renders its exploration unfeasible by synthesis alone. Historically, chemists have tended to explore chemical space unevenly without committing to systematic frameworks for navigation. This Minireview covers a range of approaches that take inspiration from the structure or origin of natural products, and help focus molecular discovery on biologically-relevant regions of chemical space. All these approaches have enabled the discovery of distinctive and novel bioactive small molecules such as useful chemical probes of biological mechanisms. This Minireview comments on how such approaches may be developed into more general frameworks for the systematic identification of currently unexplored regions of biologically-relevant chemical space, a challenge that is central to both chemical biology and medicinal chemistry.

1. Introduction

Small molecules dominate our ability to treat disease[1] and can facilitate our understanding of complex biological mechanisms.[2] The discovery of bioactive small molecules is facilitated by the ability to navigate biologically-relevant chemical space effectively and efficiently, including previously unexplored regions. However, the vastness of chemical space[3] prevents exploration by synthesis alone. Although estimates vary widely,[4–6] extrapolation of the exhaustive fragments database GDB17 has suggested that there are ca. 10^{33} possible drug-like small molecules.[7] The historic exploration of chemical space has been uneven and sparse,[8] which has hampered the discovery of bioactive molecules based on novel molecular scaffolds.[9] This uneven exploration may stem from the over-reliance on a limited palette of established, reliable chemical transformations[10] despite the recent development of many novel synthetic methodologies.[11] Furthermore, in the context of natural products, there has been an historic focus on the target-oriented synthesis of specific complex molecules.[12] To enable diverse chemical space to be explored, approaches such as diversity-oriented synthesis (DOS)[13] and lead-oriented synthesis (LOS)[14] have been developed. Although DOS and LOS can enable diverse and novel regions of chemical space to be explored efficiently, they tend not to be informed by biology (see below). In contrast, fragment-based ligand discovery, which has been reviewed extensively elsewhere,[15,16] starts with fragments (typically with MW<250) that bind weakly, yet efficiently to a target protein. The exploration of chemical space is rendered more tractable by focusing on fragment-like chemical space; in a few cases, fragment sets based on natural

product (NP) substructures have been designed[17,18] and productively exploited.[18]

In this Minireview, we focus on approaches in which the exploration of chemical space is informed by the structures or origin of natural products (see Section 2). In contrast to approaches that enable the optimisation of structure-function relationships, the featured approaches enable new regions of biologically-relevant chemical space to be identified and explored. In each case, the approaches that can drive the discovery of structurally distinctive functional molecules such as drugs and chemical probes.

2. Natural Product-informed Approaches for the Discovery of Functional Molecules

Several approaches to help identify novel biologically-relevant chemical space have been developed that are informed by known bioactive compounds or the evolution of biosyntheses of natural products (NPs). Additionally, ingenious platforms have been developed to enable the discovery of molecules that bind/modulate a specific target.[19] A number of these platforms harness evolved biological mechanisms to enable molecular discovery. These include, but are not limited to, the split-intein circular ligation of peptides and proteins (SICLOPPS) technology[20] for the discovery of bioactive peptidic marcocycles, the flexizyme platform for the expression of peptides consisting of unnatural amino acids,[21] and phage-display technology that has enabled the discovery of protein-protein interaction inhibitors.[22] In this review, we focus on approaches which are underpinned by synthetic chemistry, and have enabled the discovery of novel and distinctive series of bioactive compounds which lie in new regions of biologically-

[a] School of Chemistry, University of Leeds, Woodhouse Lane, LS2 9JT, U.K.
[b] Astbury Centre for Structural and Molecular Biology, University of Leeds, Woodhouse Lane, LS2 9JT, U.K.

relevant chemical space. In each case, the approaches are informed by the structures or the origin of natural products. All of these approaches have enabled useful tools to be developed that may facilitate the investigation of complex biological processes.

2.1. Biology-Oriented Synthesis

Biology-Oriented Synthesis (BIOS) takes inspiration from the structures of NP scaffolds. NPs are inherently biologically relevant as they have evolved to interact dynamically with multiple proteins during their biosynthesis, and bind to evolved small-molecule binding sites. Using the computational algorithm SCONP to enable systematic simplification of NP scaffolds,[23] it is possible to select NP-inspired scaffolds which may retain biological relevance. An interactive computational tool, termed Scaffold Hunter,[24] has been developed to facilitate the identification of NP-inspired scaffolds that are sub-structures of NP scaffolds. In this manner, BIOS identifies and exploits the gaps in the coverage of chemical space by NPs, focusing synthetic effort on simplified, yet unexplored, molecular scaffolds. Compound libraries resulting from BIOS scaffolds can be regarded as more biologically relevant than conventional combinatorial compound libraries.[25]

Inspired by NPs such as sodwanone S[26] (**1**, Fig. 1 Panel A), a compound library based on a bicyclic oxepane scaffold was designed. The development of a multistep, one-pot synthetic sequence was crucial, and enabled the preparation of 91 derivatives. Utilising a reporter gene assay, 50 of these compounds were found to modulate the Wnt pathway.[27] The Wnt signaling pathway is involved in biological processes such as cell migration, renewal, and polarity, and is implicated in the proliferation of cancerous tissue.[28] The number of active compounds allowed structure-activity relationships to be established, and ten modulators with low micromolar activity to be discovered. The synthetic route also enabled the preparation of biotinylated analogues which were used to validate the observed bioactivity of the most active analogue, Wntepane (**2**,

Fig. 1 Panel B), by means of competitive immunoblotting. Additional, immunoenrichment investigations revealed that Wntepane activates the Wnt pathway by binding reversibly to the protein Vangl1,[29] for which no previous small-molecule ligands had been reported.

The Hedgehog (Hh) pathway is another conserved, developmental signaling pathway involved in a number of biological processes such as tissue regeneration and repair, and has been linked with birth defects.[30] Inspired by NPs such as sominone[31] (**3**, Fig. 1, Panel A), a BIOS approach, involving five linear steps and three parallel derivatisations, resulted in the preparation of 30 compounds. Four analogues were identified as Hh pathway inhibitors of which the most potent compound (**4**, Fig. 1, Panel B) was revealed to act through modulation of the protein Smoothened.[32] Novel chemotypes for modulation of the Hh pathway are in high demand as they may potentially have direct clinical applications.[33]

BIOS has yielded small molecule inhibitors for non-mammalian biological targets as well, demonstrating the general applicability of this approach in exploring biologically-relevant chemical space.[34] MptpB is a protein tyrosine phosphatase in *M. Tuberculosis*, which alters host signaling pathways and is a key biological target for the development of new drugs against this pathogen.[35] Brachiation along the branch of the NP yohimbine (**5**, Fig. 1, Panel A), suggested a tetracyclic indoloquinolizidine scaffold. A solid-support synthesis enabled the preparation of a library with 188 compounds. Biological evaluation then revealed that eleven compounds inhibited MptpB, with compound **6** (Fig. 1, Panel B) being active in the low micromolar range. Notably, this activity was not shared by the NP yohimbine itself. Moreover, **6** was selective for MptpB over its isoform MptpA, and other mammalian-derived protein tyrosine phosphatases such as Cdc25A and PTP1B.

Overall, it has been demonstrated that BIOS can afford selective compounds which reside in previously unexplored portions of biologically-relevant chemical space. Crucially, bioactive compounds based on BIOS scaffolds can have functions that are wholly distinct from those of the parent NP.

A. Natural Products

sodwanone S, **1**   sominone, **3**   yohimbine, **5**
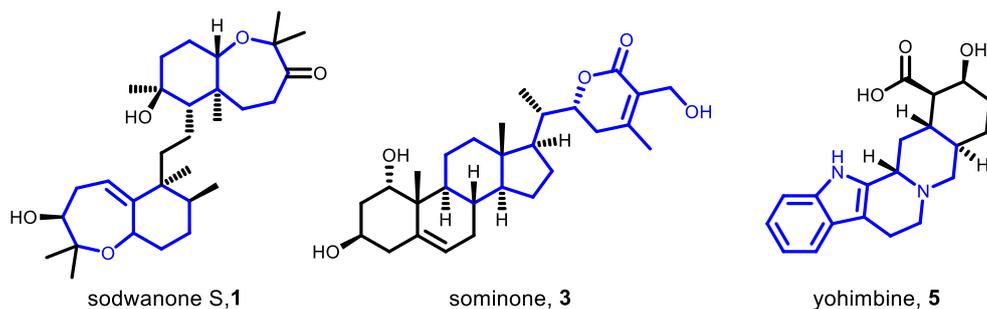
B. NP-inspired, bioactive compounds

Wntepane, **2**   **4**   **6**

Wnt Pathway
Modulator

Hedgehog Pathway
Inhibitor

Selective MptpB
inhibitor
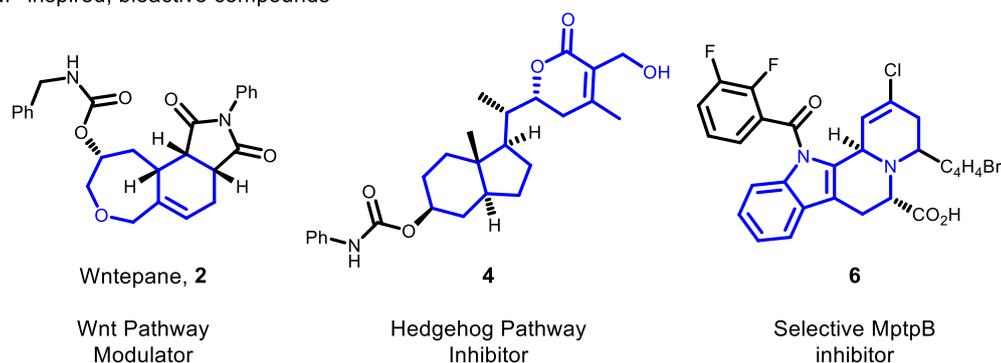
Figure 1: NPs inspire the design of BIOS libraries. BIOS exploits structurally simplified molecular scaffolds derived from natural products and has been effective in the identification of biologically relevant small molecules with diverse bioactivities. Panel A: Structures of Natural Products that have provided inspiration for BIOS. Panel B: Structures of corresponding bioactive compounds which retain conserved portions of the NPs (blue).

2.2.   Complexity to Diversity

In contrast to NP synthesis research, "Complexity-to-Diversity" (CtD) exploits NPs themselves as starting materials for the preparation of diverse scaffolds. The resulting scaffolds are related to the guiding NPs, but are distinct from those generated through scaffold simplification in BIOS. Reactions that retain certain parts of a NP structure, while modifying other portions, lead to unprecedented structures which are likely to retain the biological relevance of NPs. Through chemoselective reactions, the core scaffold of a given NP, can be systematically transformed into new scaffolds, for example through ring expansion, fusion or rearrangement (or combinations of these processes). The resulting compounds may also have favourable key properties for bioactive compound discovery, such as high number of stereogenic centres,[36] number of sp^3 hybridised carbons,[37] and solubility.[38] Additionally, as the structural complexity is already embedded into the resulting scaffold, synthetic effort can be focused on the preparation of diverse analogues.

Novel scaffolds were prepared from the readily-available diterpene gibberellic acid (**7**, Fig. 2, Panel A), the steroid andrenosterone (**8**, Fig. 2, Panel A), and the alkaloid quinine (**9**, Fig. 2, Panel A).[39] Compounds were prepared in three to five synthetic steps using suitably applicable chemoselective ring cleavage or ring expansion reactions. A compound library based on all three starting NPs **7**-**9** was subjected to a cheminformatic analysis, looking at desired molecular properties metrics. This analysis showed that CtD compounds had higher three-dimensionality than compounds in the ChemBridge MicroFormat Library. Further pairwise Tanimoto similarity[40] analysis provided additional evidence for the high diversity of the CtD library.

Demonstrating the general applicability of CtD approach, the preparation of compound libraries using the diterpene abietic acid (**10**, Fig. 2)[41] or the alkaloid yohimbine (**5**)[42] has been reported. In both cases, compounds where prepared by the application of known chemical transformations following a general pattern. Ring-cleavage (*e.g.* → **12**, **13**, Fig. 2) reactions enabled dramatic structural changes in one chemical step and provided new functional groups that can be further diversified. Ring-expansions, (*e.g.* → **12**, **14** Fig. 2) for example *via* the Baeyer–Villiger reaction, led to the formation of novel ring systems, and also preceded ring-cleavage reactions. Ring-fusion reactions provided further diversification by connecting distal groups in the pre-existing scaffold or by merging a new ring to it. Finally, ring-rearrangement (*e.g.* → **11**, **13**, **15**, Fig. 2) reactions, which drastically change the core scaffold, were used on a case-by-case basis. The molecular scaffolds prepared were demonstrated to be structurally diverse between them and have high fraction of sp^3 carbons and number of stereogenic centres. The yohimbine based library was subjected to a range

of phenotypic screens for example related to inflammation, proliferation, and anti-bacterial activity. This comprehensive biological evaluation led to the identification of a compound (**15**, Fig. 2), which demonstrated anti-inflammatory and anticancer activities by modulating the HIF functional pathway.[43] Notably, yohimbine does not share this bioactivity profile, further validating CtD as an approach for exploring novel biologically-relevant chemical space in the pursuit of small molecules with diverse biological functions.
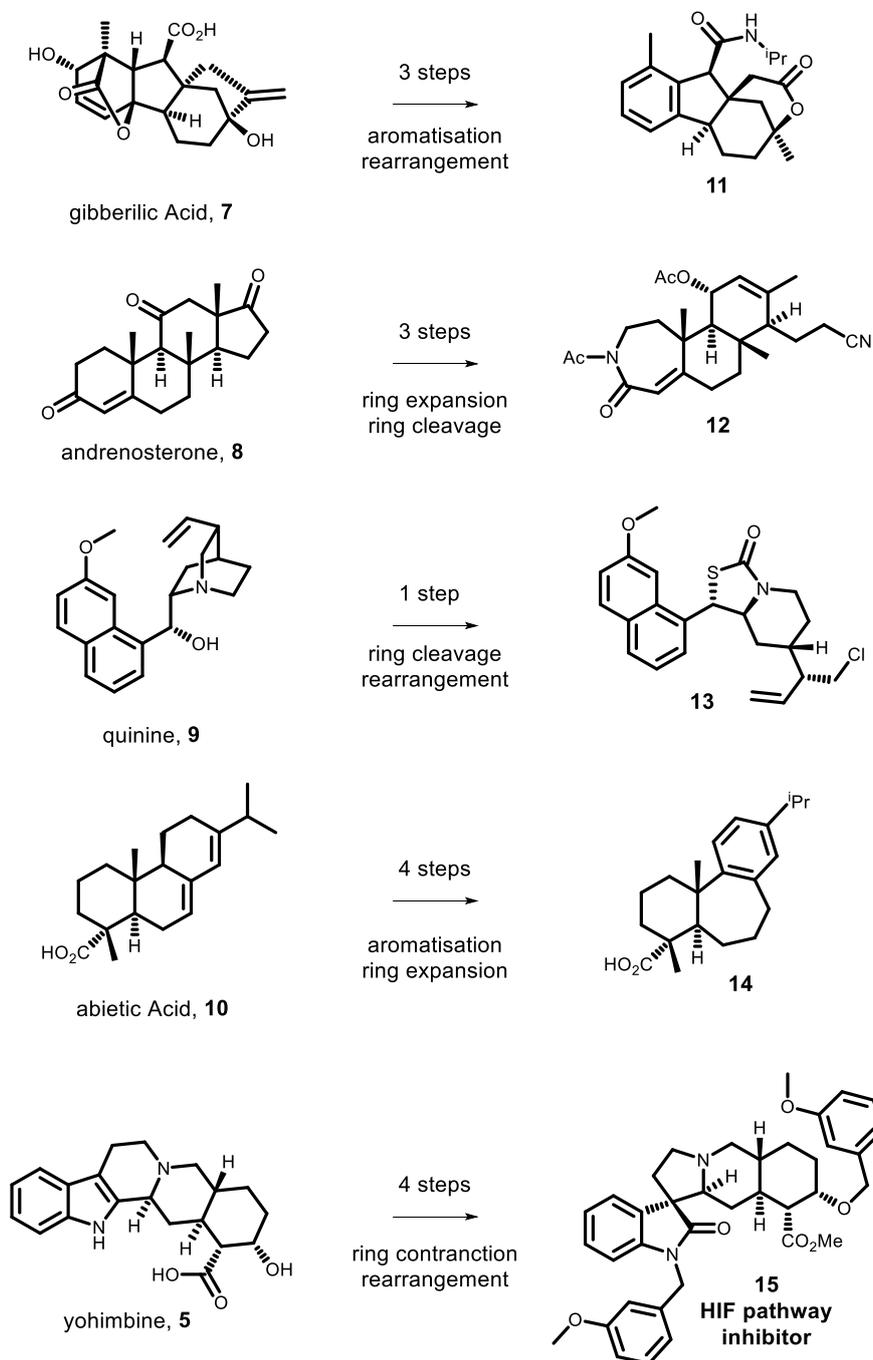
Figure 2: From complex NPs to diverse scaffolds. CtD exploits NPs as starting materials for organic synthesis. Chemoselective transformations, can yield structurally complex molecules which are not accessible by nature. Examples of NPs (left) that have been transformed into novel complex molecular scaffolds (right) are shown. The approach enabled identification of a HIF pathway inhibitor with antiproliferative effects.

ARTICLE

2.3.   Pseudo-Natural Products

Fragment-based ligand discovery[44] (FBLD) is an established approach for biomolecular discovery, enabling the rapid exploration of chemical space. NPs have been exploited directly as fragments,[45] and have inspired sets of NP-derived fragments.[18] However, it is also possible to merge NP-derived fragments to yield unprecedented molecular scaffolds. The resulting compounds have been termed *pseudo-natural products* as they are inspired by NPs and may be inherently biologically relevant. Moreover, these scaffolds lie in portions of chemical space which are not accessible via biosynthesis, and are not simplified versions of NP scaffolds.

To design new pseudo-NPs, it has been proposed to increase the high fraction of stereogenic centres, combine fragments of diverse biological relevance, and fragments with complementary heteroatoms, *e.g.* oxygen and nitrogen. Following these basic principles, a new bridged molecular scaffold based on the chromane and tetrahydropyrimidinone (THP) NP-derived fragments was designed.[46] An efficient synthesis enabled the preparation of 44 chromopynone derivatives (**22**, Fig. 3). A cheminformatic analysis showed that these compounds occupied a different portion of chemical space compared to NPs as judged by their NP-score distribution,[47] and possessed desired molecular properties for bioactive molecular discovery. Biological evaluation in a broad range of phenotypic and cell-based assays monitoring macroscopic effects in cell signalling and metabolic processes revealed that some chromopynones were active glucose uptake inhibitors. Malignant cells are over-reliant on glucose as an energy source[48] and selective modulators of glucose uptake can have clinical applications in cancer therapy.[49] A more in-depth biological investigation revealed that the most potent compound (**22**) was a dual GLUT-1/-3 inhibitor. Additionally, compounds containing either a chromane or a THP fragment alone did not inhibit glucose uptake, indicating that chromopynones display a novel bioactivity profile and are more than just the sum of their constituent parts. This observation demonstrates the potential of this approach to enable the discovery of bioactive small molecules with new functions. Notably, indolomorphane pseudo-NP compounds such as **23** (Fig. 3,), stemming from the combination of the indole and morphan fragments were discovered to display a complementary glucose uptake inhibition profile, engaging primarily with GLUT-3.[50]

Merging biosynthetically-unrelated NP-derived fragments can also afford pseudo-NP scaffolds and can enable the identification of compounds with novel bioactivity profiles. For example, combining the pyridone and dihydropyran fragments resulted in the design of pyrano-furo-pyridones (**24**, Fig. 3) which were readily prepared by the application of a Tsuji-Trost oxa-Michael cascade reaction.[51] Related cheminformatic analysis demonstrated that pyrano-furo-pyridones reside in a different portion of chemical space compared to NPs, and they may have optimal physiochemical properties as judged by the distribution of molecular weight and estimated lipophilicity (ALogP). The pyrano-furo-pyridones were evaluated for bioactivity in multiple bioassays covering a range of biological processes. This broad evaluation led to the identification of compound **24** as a potent inducer of reactive oxygen species (ROS). ROS have been implicated in several diseases[52] and compounds which allow modelling their generation in a controlled manner are of high value. Once more, the observed bioactivity was not shared by compounds containing either the pyridone or pyran fragments, indicating that the biological effect is unique to the pseudo-NP scaffold. Further biological studies showed that **24** induces the production of ROS by inhibiting mitochondrial complex I.

From the above case studies and others,[53] more specific guidelines for the design of pseudo-NPs have been developed.[54] In brief, these guidelines build on connectivity patterns between fragments that are observed in NPs themselves. Designing pseudo-NP scaffolds following these patterns, allows the resulting scaffold to inherit the biological relevance of its constituent NP-derived fragments. Crucially, bioactive molecules based on pseudo-NP scaffolds have tended to have biological functions that are distinct from compounds based on either of the parent fragments.

OH

HO

OH

OH

OH

Catechin, 18

TAN-1057, **19**

Chromopynone-1, 22
Glucose uptake inhibitor
via dual GLUT-1/3 modulation

Ergoline, 16

Morphine, 17

Indolomorphane, 23
Glucose Uptake Inhibitor
mainly *via* GLUT-3 modulation

Camptothecin, 20

Catapol, 21

Pyranofuropyridone, 24
ROS inducer
Mitochondrial
Complex I inhibitor
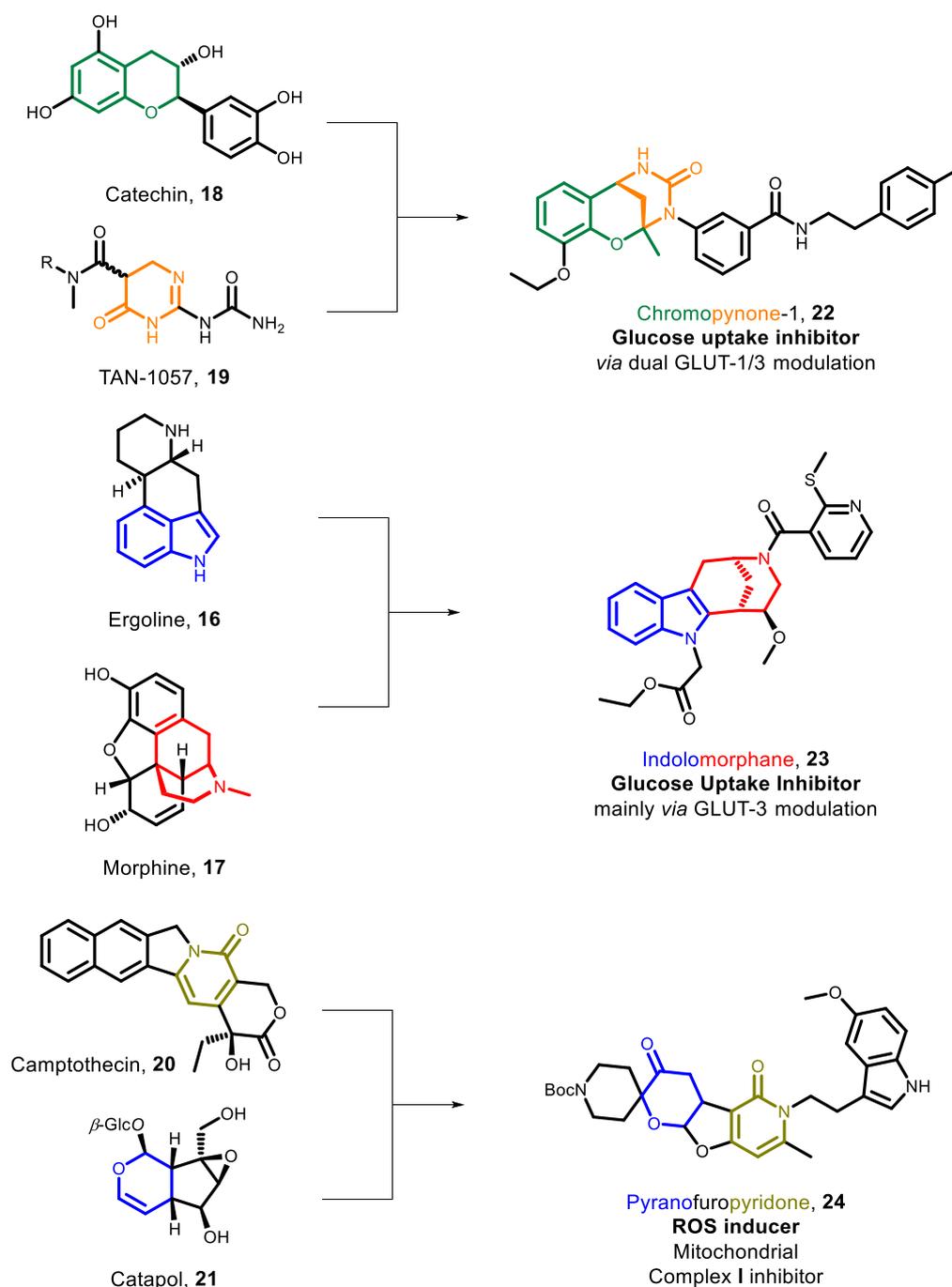
Figure 3: From NPs to Pseudo-NPs. Merging NP-derived fragments yields novel pseudo-NP scaffolds. These compounds retain the biological relevance of NPs, yet reside in chemical space that is not accessible through biosynthesis. Embedded fragments in NPs (left) were fused to yield pseudo-NPs (right). The embedded and merged NP-derived fragments are indicated by colour.

2.4.   Activity-Directed Synthesis

The discovery of non-naturally occurring bioactive small molecules is broadly achieved through iterative rounds of design, synthesis and testing. Within this context, medicinal chemists tend to rely on a narrow ensemble of well-established chemical transformations with predictable outcomes such as metal-catalysed biaryl couplings and heteroatom functionalisations.[10] In stark contrast, NPs have evolved in tandem with their respective biological targets and their associated biosynthetic pathways, in a function-driven manner i.e. to provide evolutionary advantage to the host organism.[55] Activity-Directed Synthesis (ADS) is a function-driven discovery approach that draws inspiration from the evolution of

biosynthetic pathways to NPs (rather than the structures of specific NPs). However, ADS exploits reactions which are not catalyzed by biosynthetic machinery, instead harnessing promiscuous reactions with multiple potential outcomes. Through iterative rounds of reaction arrays, screening and array redesign, ADS focuses attention on active compounds, rather investing resources equally on all compounds within a designed array.

Starting with a structural motif[56] (**25**) found in known ligands, ADS enabled the discovery of androgen receptor modulators (**26** and **27**) that were based on scaffolds with no previously annotated activity for this target (Figure 5).[57,58] Both intra- and intermolecular reactions were exploited in sequential rounds of carbenoid reactions which had many alternative outcomes. Three iterative rounds of screening crude product mixtures and design of subsequent reaction arrays enabled the rapid discovery of reactions that yielded bioactive products. When harnessing intramolecular reactions, a total of 272 microreactions was performed. Initially, an array of 36 reactions was performed in which 12 diazo substrates **25** were treated with 3 catalysts. The crude reaction mixtures were screened, enabling the identification of four hit reactions. In round 2, the most promising substrates were treated with an expanded range of 8 catalysts in 4 different solvents. Based on ten hit reactions from round 2, 4 additional diazo substrates were prepared. Finally, in round 3 the selection was focused on six catalysts and three solvents, leading to the identification of eight promising reactions that were prioritised for scale-up. This

iterative approach enabled the parallel discovery of multiple ligand series. Retrospective analysis showed that ADS enabled parallel optimisation of both the structure of bioactive molecules, and the routes for their syntheses. It was subsequently shown that intermolecular reactions could also be harnessed to drive molecular discovery.[57] Here, using non-exhaustive reaction arrays totalling 326 microreactions, additional structurally-diverse sub-micromolar modulators of the androgen receptor (**30**, Fig. 4, Panel B) were also discovered.

More recently, the applicability of ADS to more challenging targets was demonstrated by the discovery of diverse inhibitors of the p53/*h*DM2 protein-protein interaction (PPI).[59] In this case, the co-substrates and diazo-substrates contained motifs that were intended to mimic p53 hotspot residues.[60] In just two rounds of ADS, a total of 346 microscale reactions was performed leading to the identification of four diverse novel p53/*h*DM2 inhibitors (**31**-**33**, Fig. 4, Panel C). The structures of these inhibitors were shown to be dissimilar to each other, as judged by Tanimoto similarity index values, as well as to over 1000 reported *h*DM2 ligands. This observation demonstrates that ADS can enable the parallel discovery of multiple distinct and novel chemotype. It was noted that ADS had enabled scaffold-hopping: that is, the discovery of ligands in which a common pharmacophore is displayed in the context of different scaffolds. In addition, this study showed that ADS was useful approach for lead generation against a target without an evolved small-molecule binding site.

A.
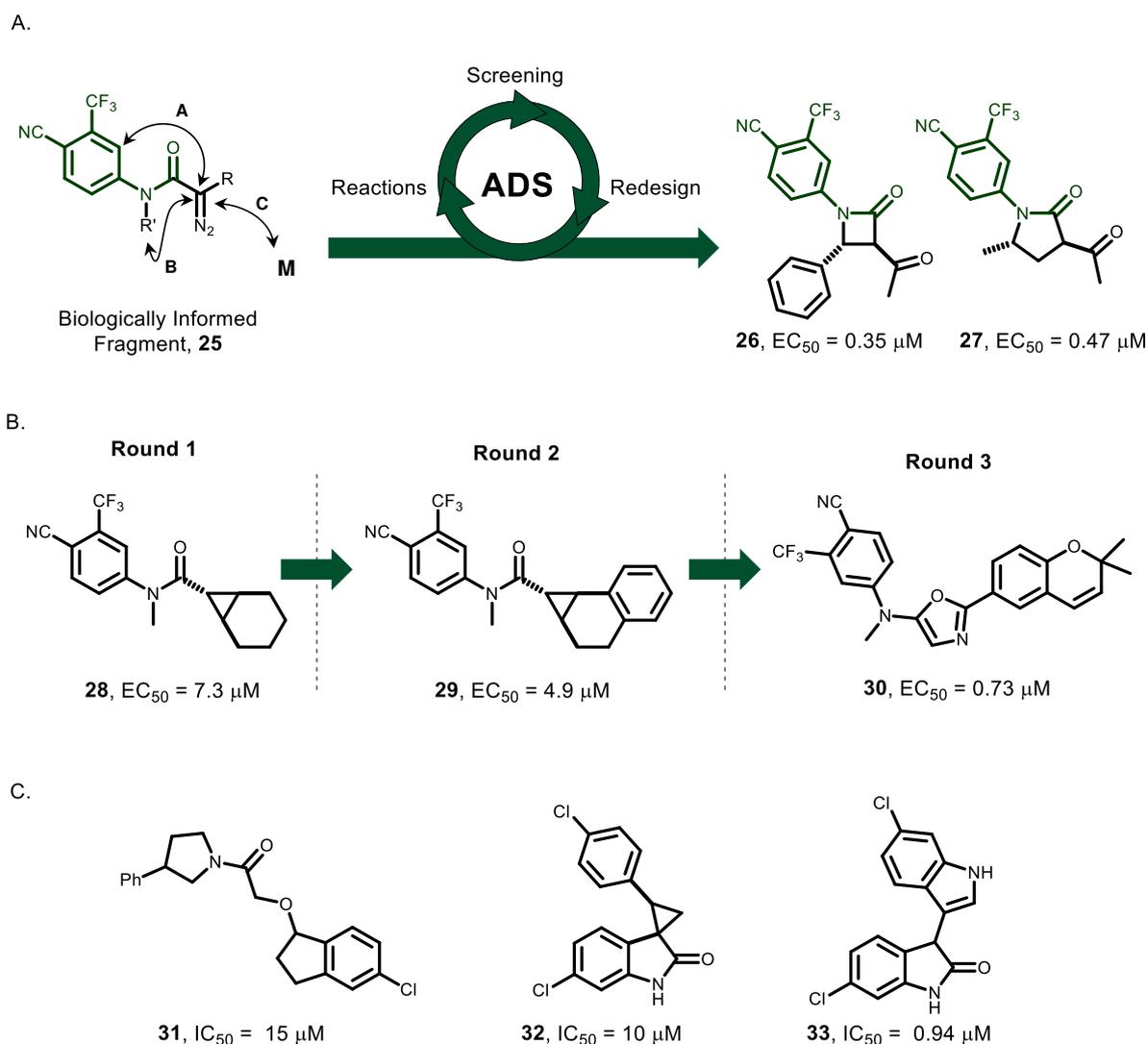
Figure 4: Applications of Activity-Directed Synthesis. ADS is a function-driven approach where bioactive molecules emerge in tandem with their associated synthetic routes, through iterative rounds of activity optimisation. Panel A: ADS was used to discover AR modulators based on scaffolds with no previously annotated AR activity. Panel B: ADS was used to "grow" a known fragment to yield structurally-diverse ligands with increased activity against AR. Panel C: Structures of recently-reported inhibitors of the p53/hMDM2 interaction discovered through ADS. AR: Androgen Receptor.

3. Towards Systematic Frameworks for Exploring Biologically-Relevant Chemical Space.

The chemocentric approaches presented (BIOS, CtD, Pseudo-NPs, ADS) can enable the discovery of functional molecules based on novel small-molecule scaffolds. Although these approaches have been described here in the context of specific case studies, they may also enable more systematic identification and exploration of biologically-relevant chemical space. The progress of each approach highlighted in this Minireview towards enabling the systematic exploration of biologically-relevant chemical space has been summarised in Table 1.

BIOS and pseudo-NPs focus on molecular scaffolds that are informed by specific NP structures. Crucially, both approaches

can identify fertile regions of chemical space for exploration, and can enable discovery of molecules with functions other than those of the parental NPs. For BIOS, the SCONP algorithm can enable systematic navigation of NP-related scaffolds, and can inspire the selection of specific scaffolds for synthesis. This analysis may be facilitated by the computational tool, Scaffold Hunter.[18] Unfortunately, a similar tool is not available to facilitate the generation of pseudo-NP structures by fusion of NP-derived fragments, preventing systematic navigation of pseudo-NP scaffolds. Additionally, both approaches require specific scaffolds to be selected for synthesis, which is currently a human-driven process. We note, however, that computational approaches have been developed to identify large numbers of likely synthetically-accessible structures[61] from which future BIOS and pseudo-NP scaffolds could be selected systematically. Finally, considerable resource is required to address the synthetic challenges tend to arise (and

be solved) on a scaffold-by-scaffold basis. Screening of NP-informed libraries against multiple targets can be achieved through phenotypic, cell-based assays including multi-parametric screening platforms.[62] "Cell-painting" assays combine the high information content that is captured in the form of a phenotypic fingerprint. Cell-painting has enabled the systematic assessment of the functional diversity of specific pseudo-NP classes to be captured.[51] Whilst BIOS and pseudo-NP approaches can identify fertile chemical space for the discovery of bioactive small molecules, it remains an open question whether the guidance from NP scaffolds is more effective than from other classes of bioactive molecules (for example, human-designed FDA drugs). It has been shown that BIOS and pseudo-NP libraries have relatively high hit rates compared to conventional compound libraries used in high-throughput screening campaigns.[63–66] However, rigorous assessment of the productivity of the BIOS and pseudo-NP approaches would require statistical analysis of the performance of many libraries across many assays. It is possible that large-scale screening efforts, such as the European Lead Factory,[67] may enable such analyses to be performed.

CtD focuses on the design of transformations that exploit the functionality within complex NPs as starting materials for synthesis. Thus far, the approach has required the selection of specific NP/reaction combinations by individually. Similar to BIOS and pseudo-NPs, a tool for the systematic selection of NP starting materials, together with transformations that may enable the preparation of diverse scaffolds, have not been yet reported. Again, we note the potential value of computational tools that enable systematic assessment of synthetic feasibility.[61] As with BIOS and Pseudo-NPs, in each example reported, significant resource was invested in the synthesis of each novel scaffold. In addition, learnings from the optimisation of reactions involving specific substrate/transformation combinations are unlikely to be transferrable to other scaffolds.

In contrast, ADS focuses on the exploitation of promiscuous reactions to generate compounds *in situ*, rather than on specifically designed target structures. ADS thereby links actual synthetic availability with biological relevance. The design of subsequent reaction arrays is informed by the function of product mixtures obtained in previous rounds. Although this design has been human-driven to date, the approach may lend itself to being algorithmically-driven. For example, one can envisage computational tools that enable reactant combinations to be selected on the basis of the reaction hits from previous rounds. Furthermore, having designed these arrays, their execution could be performed by a liquid handling robot. Crucially, the experimental stages in the workflow are all performed in parallel and may be readily integrated (and potentially automated). As such ADS raises the prospect of realising fully autonomous molecular discovery. We note that autonomous robotic approaches have been developed to address chemical discovery problems such as the discovery of ligands for supramolecular chemistry,[68] and the identification of novel chemical reactivity.[69]

Finally, novel chemistry has always been at the heart of biomolecular discovery. Innovative chemical transformations can prove crucial in our ability to efficiently explore chemical space in general. Recent advances in late-stage functionalisations[70,11] can be implemented in the context of either of the highlighted discovery approaches to expedite their coverage of chemical space. For example, methods which allow chemoselective ring formations from functionalised fragments[71] or linear precursors,[71] could expand the current scope of CtD and Pseudo-NPs. Additionally, new multicomponent[72] or cascade reactions could also be beneficially as part of the workflow of any of these approaches.[73] Modifying NP-derived fragments or larger structural motifs creates opportunities for the application of biocatalytic methods due to the fact that NPs have evolved to interact with multiple protein binding sites. Together these new reactions may facilitate the productive exploration of biologically-relevant chemical space.

Table 1: Progress towards the systematic exploration of biologically-relevant chemical space.

Approach	How informed by NPs	Scaffold selection	Synthesis
Biology-oriented synthesis	Simplified NP scaffolds	SCONP algorithm and Scaffold Hunter tool help scaffold identification	Tailored synthesis developed for each prioritised scaffold
Complexity-to-diversity	NPs are starting materials for synthesis	Tools not currently available to identify NPs that may be suitable substrates	Synthetic methods developed for each NP starting material
Pseudo-NPs	Scaffolds are merged NP fragments	Algorithms to generate merged structures not currently available	Tailored synthesis developed for each prioritised scaffold
Activity-directed synthesis	Broadly parallels emergence of biosynthetic pathways to NPs	Scaffolds are not designed: functionalised scaffolds emerge on the basis of their biological function	Common protocols used for each reaction class used. Reaction arrays are informed by the activity of products obtained in previous rounds. Array design is currently human-driven.

ARTICLE

4. Summary and Outlook

Small molecules are powerful tools for the study of complex biological processes and treating disease, but their discovery requires exploration of relevant chemical space. Approaches which are informed by biology have met great success in navigating the vast chemical space and identifying unexplored, yet biologically-relevant regions. Genetically- and DNA-encoded libraries exploit biological mechanisms leading to the discovery of functional molecules. However, they are inherently limited in their ability to identify bioactive compounds based on previously unexplored molecular scaffolds. BIOS, CtD, Pseudo-NPs, and ADS exploit natural or non-naturally occurring compounds, building on their biological function by trimming, merging, or expanding the original structures. Each approach has led to the successful identification of bioactive small molecules with novel compounds which, in selected cases, have been used as probes to uncover key biological targets.

Although these approaches have been proven to be successful on a case-by-case basis, they have generally not been implemented in a systematic manner: we have commented in Section 3 how this might be done in the future. In particular, tools which integrate the design of new scaffolds with the assessment of synthetic accessibility may have great value in opening up broader swathes of unexplored chemical space of direct relevance to molecular discovery. The resulting integrated approaches will thereby address central challenges at the chemistry/biology interface. As such they would have a transformative effect on our collective ability to discover novel small-molecule drugs and chemical tools for elucidating complex biological mechanisms.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgements

References

1   A. V Fuentes, M. D. Pineda and K. C. N. Venkata, *Pharm. (Basel, Switzerland)*, 2018, **6**, 43.

2   A. M. Edwards, R. Isserlin, G. D. Bader, S. V Frye, T. M. Willson and F. H. Yu, *Nature*, 2011, **470**, 163.

3   C. M. Dobson, *Nature*, 2004, **432**, 824–828.

4   P. Kirkpatrick and C. Ellis, *Nature*, 2004, **432**, 823.

5   J.-L. Reymond, R. van Deursen, L. C. Blum and L. Ruddigkeit, *Medchemcomm*, 2010, **1**, 30–38.

6   J.-L. Reymond, *Acc. Chem. Res.*, 2015, **48**, 722–730.

7   P. G. Polishchuk, T. I. Madzhidov and A. Varnek, *J. Comput. Aided. Mol. Des.*, 2013, **27**, 675–679.

8   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**, 4443–4451.

9   S. R. Langdon, N. Brown and J. Blagg, *J. Chem. Inf. Model.*, 2011, **51**, 2174–2185.

10   D. G. Brown and J. Boström, *J. Med. Chem.*, 2016, **59**, 4443–4458.

11   J. Boström, D. G. Brown, R. J. Young and G. M. Keserü, *Nat. Rev. Drug Discov.*, 2018, **17**, 709–727.

12   E. J. Corey, *Angew. Chemie Int. Ed. English*, 1991, **30**, 455–465.

13   M. D. Burke and S. L. Schreiber, *Angew. Chem. Int. Ed.*, 2004, **43**, 46–58.

14   R. Doveston, S. Marsden and A. Nelson, *Drug Discov. Today*, 2014, **19**, 813–819.

15   C. W. Murray and D. C. Rees, *Nat Chem*, 2009, **1**, 187–192.

16   C. W. Murray and D. C. Rees, *Angew. Chem. Int. Ed.*, 2016, **55**, 488–492.

17   A. R. Hanby, N. S. Troelsen, T. J. Osberger, S. L. Kidd, K. T. Mortensen and D. R. Spring, *Chem. Commun.*, 2020, **56**, 2280–2283.

18   B. Over, S. Wetzel, C. Grütter, Y. Nakai, S. Renner, D. Rauh and H. Waldmann, *Nat Chem*, 2013, **5**, 21–28.

19   J. S. Ebo, J. C. Saunders, P. W. A. Devine, A. M. Gordon, A. S. Warwick, B. Schiffrin, S. E. Chin, E. England, J. D. Button, C. Lloyd, N. J. Bond, A. E. Ashcroft, S. E. Radford, D. C. Lowe and D. J. Brockwell, *Nat. Commun.*, 2020, **11**, 1816.

20   D. J. Asby, F. Cuda, M. Beyaert, F. D. Houghton, F. R. Cagampang and A. Tavassoli, *Chem. Biol.*, 2015, **22**, 838–848.

21   H. Hirose, C. Tsiamantas, T. Katoh and H. Suga, *Curr. Opin. Biotechnol.*, 2019, **58**, 28–36.

22   D. J. Hughes, C. Tiede, N. Penswick, A. A.-S. Tang, C. H. Trinh, U. Mandal, K. Z. Zajac, T. Gaule, G. Howell, T. A. Edwards, J. Duan, E. Feyfant, M. J. McPherson, D. C. Tomlinson and A. Whitehouse, *Sci. Signal.*, 2017, **10**, eaaj2005.

23   M. A. Koch, A. Schuffenhauer, M. Scheck, S. Wetzel, M. Casaulta, A. Odermatt, P. Ertl and H. Waldmann, *Proc. Natl. Acad. Sci. United States Am.*, 2005, **102**, 17272–17277.

24   S. Wetzel, K. Klein, S. Renner, D. Rauh, T. I. Oprea, P. Mutzel and H. Waldmann, *Nat Chem Biol*, 2009, **5**, 581–583.

25   H. van Hattum and H. Waldmann, *J Am Chem Soc*, 2014, **136**, 11853–11859.

26   J. Dai, J. A. Fishback, Y.-D. Zhou and D. G. Nagle, *J. Nat. Prod.*, 2006, **69**, 1715–1720.

27   S. Basu, B. Ellinger, S. Rizzo, C. Deraeve, M. Schürmann, H. Preut, H.-D. Arndt and H. Waldmann, *Proc. Natl. Acad. Sci.*, 2011, **108**, 6805–6810.

28   J. N. Anastas and R. T. Moon, *Nat Rev Cancer*, 2013, **13**, 11–26.

29   E. Torban, A.-M. Patenaude, S. Leclerc, S. Rakowiecki, S. Gauthier, G. Andelfinger, D. J. Epstein and P. Gros, *Proc. Natl. Acad. Sci.*, 2008, **105**, 3449–3454.

30   J. Briscoe and P. P. Therond, *Nat Rev Mol Cell Biol*, 2013, **14**, 416–429.

31   L.-X. Chen, H. He and F. Qiu, *Nat. Prod. Rep.*, 2011, **28**, 705–740.

32   R. J. Lipinski, J. J. Gipp, J. Zhang, J. D. Doles and W. Bushman, *Exp. Cell Res.*, 2006, **312**, 1925–1938.

33   K. Seifert, A. Büttner, S. Rigol, N. Eilert, E. Wandel and A. Giannis, *Bioorg. Med. Chem.*, 2012, **20**, 6465–6481.

34   A. Nören-Müller, I. Reis-Corrêa, H. Prinz, C. Rosenbaum, K. Saxena, H. J. Schwalbe, D. Vestweber, G. Cagna, S. Schunk, O. Schwarz, H. Schiewe and H. Waldmann, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 10606–10611.

35   A. Koul, T. Herget, B. Klebl and A. Ullrich, *Nat Rev Micro*, 2004, **2**, 189–202.

36   F. Lovering, J. Bikker and C. Humblet, *J. Med. Chem.*, 2009, **52**, 6752–6756.

37   D. F. Veber, S. R. Johnson, H.-Y. Cheng, B. R. Smith, K. W. Ward and K. D. Kopple, *J. Med. Chem.*, 2002, **45**, 2615–2623.

38   C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv. Drug Deliv. Rev.*, 1997, **23**, 3–25.

39   R. W. Huigens III, K. C. Morrison, R. W. Hicklin, T. A. Flood Jr, M. F. Richter and P. J. Hergenrother, *Nat. Chem.*, 2013, **5**, 195–202.

40   D. J. Rogers and T. T. Tanimoto, *Science*, 1960, **132**, 1115–1118.

41   R. J. Rafferty, R. W. Hicklin, K. A. Maloof and P. J. Hergenrother, *Angew. Chem. Int. Ed.*, 2014, **53**, 220–224.

42   N. G. Paciaroni, R. Ratnayake, J. H. Matthews, V. M. Norwood IV, A. C. Arnold, L. H. Dang, H. Luesch and R. W. Huigens III, *Chem. – A Eur. J.*, 2017, **23**, 4327–4335.

43   K. Burkitt, S. Y. Chun, D. T. Dang and L. H. Dang, *Mol. Cancer Ther.*, 2009, **8**, 1148–1156.

44   D. A. Erlanson, S. W. Fesik, R. E. Hubbard, W. Jahnke and H. Jhoti, *Nat. Rev. Drug Discov.*, 2016, **15**, 605–619.

45   H. Vu, L. Pedro, T. Mak, B. McCormick, J. Rowley, M. Liu, A. Di Capua, B. Williams-Noonan, N. B. Pham, R. Pouwer, B. Nguyen, K. T. Andrews, T. Skinner-Adams, J. Kim, W. G. J. Hol, R. Hui, G. J. Crowther, W. C. Van Voorhis and R. J. Quinn, *ACS Infect. Dis.*, 2018, **4**, 431–444.

46   G. Karageorgis, E. S. Reckzeh, J. Ceballos, M. Schwalfenberg, S. Sievers, C. Ostermann, A. Pahl, S. Ziegler

47   K. Vanii Jayaseelan, P. Moreno, A. Truszkowski, P. Ertl and C. Steinbeck, *BMC Bioinformatics*, 2012, **13**, 106–112.

48   R. Scatena, P. Bottoni, A. Pontoglio and B. Giardina, *PROTEOMICS – Clin. Appl.*, 2010, **4**, 143–158.

49   C. C. Barron, P. J. Bilan, T. Tsakiridis and E. Tsiani, *Metabolism*, 2016, **65**, 124–139.

50   J. Ceballos, M. Schwalfenberg, G. Karageorgis, E. S. Reckzeh, S. Sievers, C. Ostermann, A. Pahl, M. Sellstedt, J. Nowacki, M. A. Carnero Corrales, J. Wilke, L. Laraia, K. Tschapalda, M. Metz, D. A. Sehr, S. Brand, K. Winklhofer, P. Janning, S. Ziegler and H. Waldmann, *Angew. Chem. Int. Ed.*, 2019, **58**, 17016–17025.

51   A. Christoforow, J. Wilke, A. Binici, A. Pahl, C. Ostermann, S. Sievers and H. Waldmann, *Angew. Chem. Int. Ed.*, 2019, **58**, 14715–14723.

52   I. Liguori, G. Russo, F. Curcio, G. Bulli, L. Aran, D. Della-Morte, G. Gargiulo, G. Testa, F. Cacciatore, D. Bonaduce and P. Abete, *Clin. Interv. Aging*, 2018, **13**, 757–772.

53   H. Xu, C. Golz, C. Strohmann, A. P. Antonchick and H. Waldmann, *Angew. Chem. Int. Ed.*, 2016, **55**, 7761–7765.

54   G. Karageorgis, D. J. Foley, L. Laraia and H. Waldmann, *Nat. Chem.*, 2020, **12**, 227–235.

55   R. D. Firn and C. G. Jones, *Nat. Prod. Rep.*, 2003, **20**, 382–391.

56   W. Gao, C. E. Bohl and J. T. Dalton, *Chem. Rev.*, 2005, **105**, 3352–3370.

57   G. Karageorgis, M. Dow, A. Aimon, S. Warriner and A. Nelson, *Angew. Chem. Int. Ed.*, 2015, **54**, 13538–113544.

58   G. Karageorgis, S. Warriner and A. Nelson, *Nat. Chem.*, 2014, **6**, 872–876.

59   A. Green, F. Hobor, C. Tinworth, S. Warriner, A. Wilson and A. Nelson, *Chem. – A Eur. J.*, , DOI:10.1002/chem.202002153.

60   Y. Zhao, A. Aguilar, D. Bernard and S. Wang, *J. Med. Chem.*, 2015, **58**, 1038–1052.

61   E. P. Gajewska, S. Szymkuć, P. Dittwald, M. Startek, O. Popik, J. Mlynarski and B. A. Grzybowski, *Chem*, 2020, **6**, 280–293.

62   J. C. Caicedo, S. Cooper, F. Heigwer, S. Warchal, P. Qiu, C. Molnar, A. S. Vasilevich, J. D. Barry, H. S. Bansal, O. Kraus, M. Wawer, L. Paavolainen, M. D. Herrmann, M. Rohban, J. Hung, H. Hennig, J. Concannon, I. Smith, P. A. Clemons, S. Singh, P. Rees, P. Horvath, R. G. Linington and A. E. Carpenter, *Nat. Methods*, 2017, **14**, 849–863.

63   R. Breinbauer, I. R. Vetter and H. Waldmann, *Angew. Chem. Int. Ed.*, 2002, **41**, 2878–2890.

64   D. H. Drewry and R. Macarron, *Curr. Opin. Chem. Biol.*, 2010, **14**, 289–298.

65   S. Wetzel, R. S. Bon, K. Kumar and H. Waldmann, *Angew. Chem. Int. Ed.*, 2011, **50**, 10800–10826.

66   S. C. K. Sukuru, J. L. Jenkins, R. E. J. Beckwith, J. Scheiber, A. Bender, D. Mikhailov, J. W. Davies and M. Glick, *J. Biomol. Screen.* , 2009, **14**, 690–699.

67   A. Karawajczyk, F. Giordanetto, J. Benningshof, D. Hamza, T. Kalliokoski, K. Pouwer, R. Morgentin, A. Nelson, G. Müller, A. Piechot and D. Tzalis, *Drug Discov. Today*, 2015,

20, 1310–1316.

68   L. Porwol, D. J. Kowalski, A. Henson, D.-L. Long, N. L. Bell and L. Cronin, *Angew. Chem. Int. Ed.*, 2020, **59**, 11256–11261.

69   J. M. Granda, L. Donina, V. Dragone, D.-L. Long and L. Cronin, *Nature*, 2018, **559**, 377–381.

70   T. Cernak, K. D. Dykstra, S. Tyagarajan, P. Vachal and S. W. Krska, *Chem. Soc. Rev.*, 2016, **45**, 546–576.

71   E. T. Hennessy and T. A. Betley, *Science*, 2013, **340**, 591–595.

72   P. Slobbe, E. Ruijter and R. V. A. Orru, *Medchemcomm*, 2012, **3**, 1189–1218.

73   P. N. Devine, R. M. Howard, R. Kumar, M. P. Thompson, M. D. Truppo and N. J. Turner, *Nat. Rev. Chem.*, 2018, **2**, 409–421.