
Article:
Chow, S orcid.org/0000-0002-3600-0497, Liver, S and Nelson, AS orcid.org/0000-0003-3886-363X (2018) Streamlining bioactive molecular discovery through integration and automation. Nature Reviews Chemistry, 2 (8). pp. 174-183. ISSN 2397-3358

https://doi.org/10.1038/s41570-018-0025-7

Streamlining bioactive molecular discovery through integration and automation

Shiao Chow,[a,b] Sam Liver[a,b] and Adam Nelson[a,b]

[a]School of Chemistry, University of Leeds, Leeds, LS2 9JT, UK; [b]Astbury Centre for , University of Leeds, Leeds, LS2 9JT, UK. Email: a.s.nelson@leeds.ac.uk

Abstract

The discovery of bioactive small molecules is generally driven via iterative design-make-purify-test cycles. Automation is routinely harnessed at individual stages in order to increase the productivity of drug discovery. We describe recent progress to automate and integrate two or more adjacent stages within discovery workflows. The value of these integrated technologies is illustrated in the context of specific discovery case studies. We note that, to maximise impact on the productivity of discovery, each of the integrated stages would need to have both high and matched throughput. We also consider the longer-term goal of realising the fully autonomous discovery of bioactive small molecules through integration and automation of all stages of discovery.

1. Introduction and Vision

The grand challenge of increasing productivity within the pharmaceutical sector[1] has led to the automation of many parts of the drug discovery process.[2] For example, high-throughput screening of large compound collections is widely exploited to generate series of hit molecules.[3] In addition, arrays of molecules are often synthesised and purified using automated approaches[4] significantly reducing the volume of labour-intensive tasks required. Currently, however, automation within discovery workflows almost always focuses on isolated individual stages of the discovery process (Figure 1).

The quest to improve the productivity of drug discovery, however, is at odds with the aspiration to increase chemical and biological innovation. The targets of the future, such as protein-protein interactions and large macromolecular complexes, are likely to bring new scientific challenges and to require different classes of ligand.[5] The challenge of discovering such ligands is, however, heightened by the vastness of chemical space: although estimates vary extremely widely, extrapolation from GDB-17 (a database of enumerated molecules with up to 17 heavy atoms) led to an estimation of ≈10^{33} possible molecules with up to 36 heavy atoms.[6] Yet, chemists' historical exploration of chemical space has already been highly uneven and unsystematic,[7] and this lack of scaffold diversity is also reflected in exemplified medicinal chemistry space[8] and drugs.[9] Furthermore, despite overwhelming recognition of the importance of molecular properties in drug discovery,[10,11] medicinal chemists are actually

increasing their attention on more lipophilic and flat compounds [12,13] This behaviour is driven, in large part, by current discovery workflows that are underpinned by a remarkably narrow (though gradually increasing[13]) toolkit of robust reaction types.[14-17] The introduction of new ultra-high throughput chemical technologies, such as DNA-encoded libraries,[18] can greatly increase the number of compounds explored, albeit generally with a more limited palette of underpinning reaction classes. How then can the productivity of drug discovery be improved, whilst enabling currently challenging targets to be addressed?

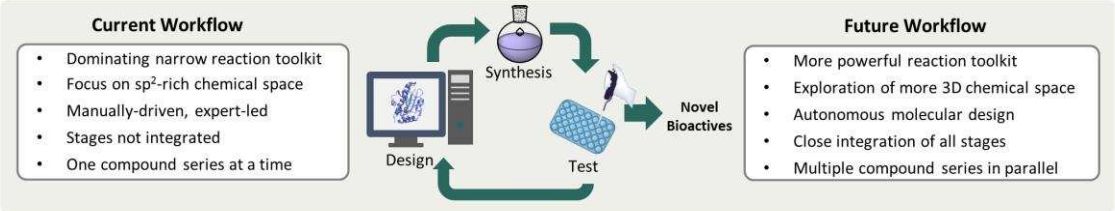

Figure 1 | Current features and future aspirations in drug discovery workflows

In this Perspective, we describe progress towards the automation and integration of adjacent stages within discovery workflows. We have deliberately focused on integrated approaches that have already been exploited in discovery, rather than on isolated methods that have future potential to be integrated into discovery workflows. To provide context, the value of the integrated approaches is illustrated using specific case studies.

2. Integrated Platforms to Drive Bioactive Small Molecule Discovery

The automation of chemical synthesis[4,19] can significantly increase the efficiency and reproducibility of synthetic chemistry within discovery workflows. Although the adoption of parallel synthesis is widespread, compound purification is almost always performed in series. This bottleneck is imposed by the technologies that are widely adopted (such as mass-directed HPLC), and the limited scope of current parallel purification approaches (which include scavenging[20] and extraction[21]). Many reaction classes have been automated, including multi-phase reactions; and the automation of the dominant reactions within the medicinal chemistry toolkit – such as heteroatom functionalisations and metal-catalysed cross-couplings – is widespread. Furthermore, automated multi-step synthesis is possible, for example, in the synthesis β- and γ-amino acid derivatives using several different reaction classes.[22] Although the broader context of automated synthesis is relevant to this Perspective, we focus here on approaches in which the automation of synthesis has been integrated with adjacent stages within discovery workflows.

2.1 Integration of Reaction Optimisation and Synthesis

The molecules that are designed within medicinal chemistry programmes generally need to be prepared from polar and functionalised substrates.[11,23] Such reactants tend to perform systematically poorly in chemical reactions, resulting in a significant drift in logP between designed and produced arrays.[23] Several teams[24-28] have therefore developed high-throughput approaches for the optimisation of reactions within medicinal chemistry workflows. The ability to explore many different reaction parameters[29] rapidly, using small amounts of material, can significantly increase the value of specific reaction classes in discovery applications. Although published examples are essentially limited to the standard medicinal chemistry toolkit (Table 1), these approaches should ultimately broaden the range of underpinning reactions that can be reliably harnessed for molecular discovery. For example, C-H functionalisations at sp^3-hybridised carbon offer great promise[17] for, but are largely untapped in, drug discovery.[30, 31] Notably, high-throughput experimentation can also enable the initial discovery of new catalytic reactions[32] and the identification of transformations that are particularly functional group-tolerant.[33]

Table 1. Case studies based on platforms that can integrate reaction optimisation with small molecule synthesis

Reaction Type	Format	Optimisation demonstrated?	Quantification/Purification	Ref.
Suzuki-Miyaura	Oscillator Flow (series)	Partial	HPLC purification	24
Suzuki-Miyaura	Flow (series)	Yes	HPLC purification	25
S_NAr	Batch (vials) (parallel)	Yes	HPLC, UV & MS analysis	26
C–C, C–O and C–N Pd-catalysed couplings	Batch (plates) (parallel)	Yes	LC-MS (For product detection)	27
C–C, C–O and C–N Pd-catalysed couplings	Flow (series)	Yes	LC-MS (For product detection)	28

HPLC, high-performance liquid chromatography; UV, ultraviolet; MS, mass spectrometry; LC-MS, liquid chromatography-mass spectrometry.

The high-throughput optimisation of palladium-catalysed cross-couplings leading to highly functionalised products has been conducted on a nanomole scale (Figure 2, Panel A).[27] Metal-catalysed couplings, including Suzuki-Miyaura and Buchwald-Hartwig reactions, have received particular attention since significant optimisation is often needed for each pair of substrates. The study focused on palladium-catalysed C–C, C–O and C–N bond-forming reactions involving combinations of 12 polar nucleophiles and 8 complex electrophiles. Initial attempts to synthesise the 96 possible products in 96-well format on a 500 μmol scale had resulted in 54 successful reactions (Panel A1).

For the optimisation study, arrays of reactions were set up in plastic 1536-well plates using a nanolitre liquid-handling robot. The reactions were performed on a ~100 nmol scale in 1.0 μL DMSO at room temperature without agitation, and the outcomes were determined using mass spectrometry. Attention was focussed on 32 nucleophile/electrophile combinations that had failed during the initial synthesis (see above). These 32 combinations

were each subjected to 48 different conditions (6 catalysts x 8 organic superbases), leading to a total of 1536 test reactions. For 21 of the combinations, the required product mass was detected, and 16 of the products were successfully isolated after subsequent scale up (to 500 μmol). For six of the remaining failed reactions, 48 combinations of reagent stoichiometry and catalyst loading were then explored, and better conditions were identified in five cases. Thus, from these two high-throughput experiments, 21 complex products that had previously eluded synthesis could be prepared (see Panel A2 for an example). Finally, to demonstrate that nanomole reaction optimisation can inform the development of practical gram-scale syntheses, it was shown that conditions identified on a 0.02 mg scale did enable efficient coupling of the chloride **4** and the amine **5** on 25 mg and 1 g scales (Panel A3).

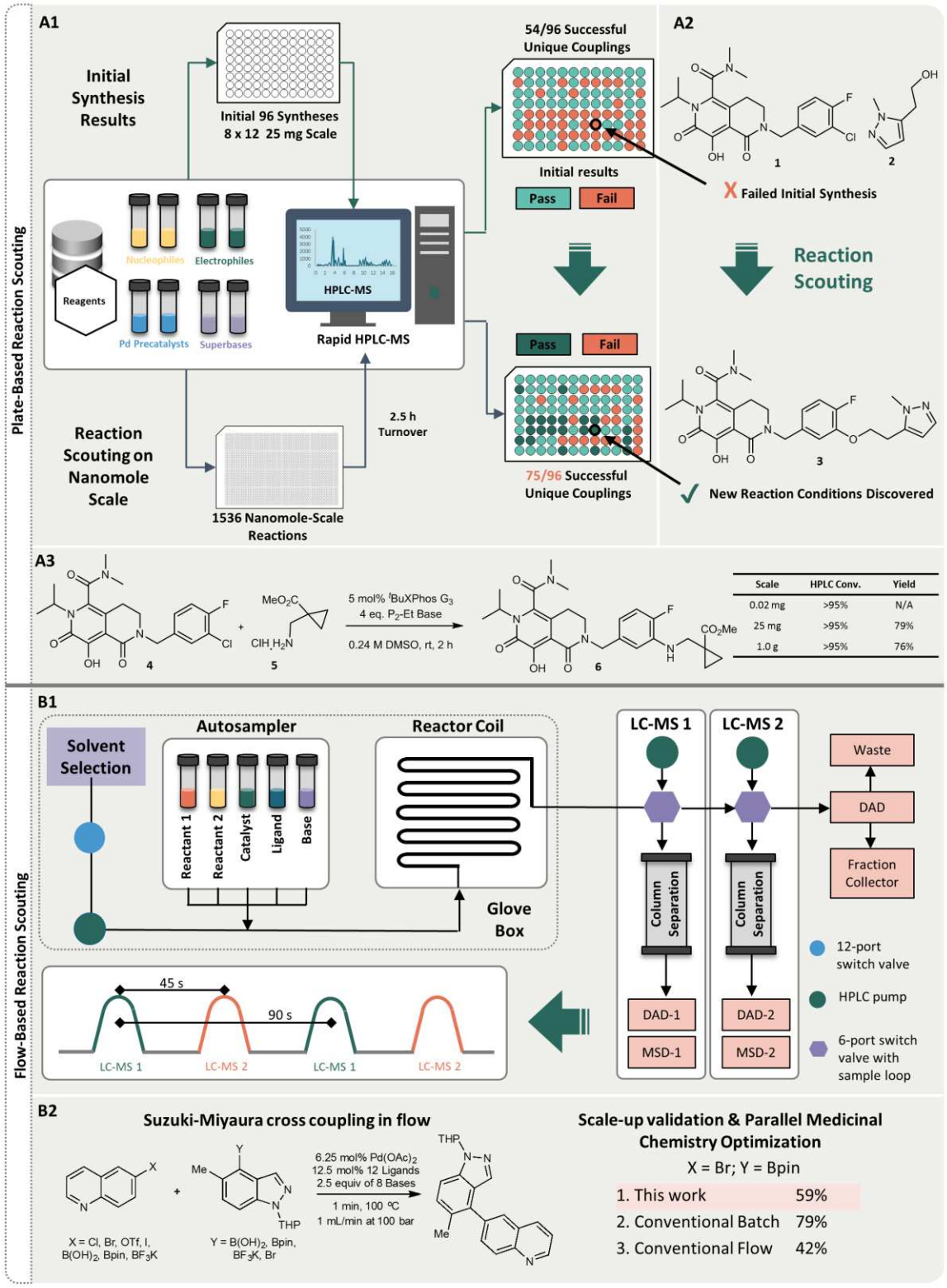

Figure 2 | Platforms for high-throughput reaction optimisation to yield highly-functionalised products. (A) Batch-based nanomole-scale optimisation. A1: Overview of platform. A2: Identification of conditions for the reaction of electrophile **1** and nucleophile **2** to give the product **3**. A3: Successful reaction of the electrophile **4** and the nucleophile **5** on 0.02 mg, 25 mg and 1.0 g scales to generate the product **6**. (B) Flow-based nanomole-scale reaction optimisation. B1: Overview of platform. B2: Range of substrates, ligands and bases explored in the optimisation of a Suzuki-Miyaura coupling, and translation of the optimised reaction to conventional batch- and flow-based approaches. DAD, diode array detector; MSD, mass-selective detector.

In another reaction optimization study, an automated flow-based platform was harnessed that enables both nanomole-scale reaction screening and micromole-scale synthesis (Figure 2, Panel B).[28] As an exemplar, a Suzuki–Miyaura coupling was optimised through exploration of 5760 combinations of 15 pairs of substrates, 12 ligands, 8 bases and 4 solvents. Although the approach is inherently serial in nature, high-throughput operation (>1500 per day) was possible. The 5760 reaction conditions were explored by injection of the appropriate components into the same flow segment, and analysis of reaction outcome by LC-MS (Panel B1). Furthermore, it was demonstrated that the required product could be prepared on a micromole scale on the platform through repetition of the optimised reaction; and on 50-200 mg scale in standard flow and batch modes (Panel B2).

The integration of high-throughput reaction optimisation into discovery workflows is particularly valuable when bespoke conditions are required for each specific combination of reactants. High throughput is possible either by operation in parallel (in batch mode) or in series (e.g. in flow mode with a fast cycle time). To achieve full impact, it will be necessary to exploit these platforms to harness a wider range of underpinning reaction classes within discovery workflows.

2.2 Integration of Synthesis and Biological Evaluation

Several platforms that integrate automated synthesis and biological evaluation have been developed and exemplified in bioactive molecule discovery (Table 2).[34-42] These platforms generally have chemical reactor module(s), off-line compound processing unit(s) and a biological testing module. Inherently modular systems can enable flexible configuration to meet the needs of specific discovery projects.

Throughput depends on both the timescale of, and the parallel or serial nature of, the activities that are performed. For example, in a programme focused on the polycomb protein EED,[36] batch synthesis in vials (in parallel), mass-directed HPLC purification (in series), and a plate-based biochemical assay (in parallel) enabled the synthesis and evaluation of 22 compounds in ~24 hr. The throughput of a programme focused on the peptidase DPP4 was similar (~2 hr / compound), and was driven by synthesis in a microfluidic reactor (in series), HPLC purification (in series) and a plate-based biochemical assay (in parallel).[37]

In-line continuous flow technologies can enable multiple stages within drug discovery workflows to be integrated. For example, affinity chromatography has been exploited to determine immediately the affinity of synthesised compounds for a target protein.[38,39] The approach is particularly valuable when in-line purification is possible, as has been demonstrated in the discovery of ligands for human serum albumin.[39] Here, it was also

demonstrated that manual operations could be omitted through dilution of reaction products, and direct injection onto the affinity column.[39]

Integrated flow technologies have been exploited in the discovery of inhibitors of the protease BACE1 (Figure 3, Panel A).[40] The on-chip assay technology integrated a thin capillary (to generate a ligand concentration gradient over several orders of magnitude) and a glass chip that enabled activity assessment in flow after combining ligand, enzyme and substrate streams. Measurement of fluorescence at different locations on the chip enabled dose-dependent activity determination at 100 different ligand concentrations. The use of flow technologies at all stages enabled each compound to be prepared, purified and evaluated in series in ~1 hr. The approach enabled the SAR of a series of inhibitors (e.g. **7** and **8**) to be defined.

A platform that integrated reaction optimisation, library synthesis and biological evaluation was exploited in the discovery of CHK1 kinase inhibitors (Figure 3, Panel B).[41] An array of 384 target compounds was designed that would be accessible via coupling of the 2-bromothiazole **9** with 384 different nucleophiles: 48 thiols (for C–S cross-couplings); 32 alkynes and 80 boronates (for C–C cross-couplings); 64 alcohols (for C–O cross-couplings); and 32 amides, 32 sulfonamides and 96 amines (for C–N cross-couplings). Initially, four productive reaction conditions were identified for each class of nucleophile. The 384 cross-couplings were then each performed on a ~100 nmol scale under the four reaction conditions. By exploiting the best of the four alternative conditions, 345 of the 384 targeted products could be prepared (compared with just 158 products using a single reaction condition). Pools of 12-24 reaction products were then incubated with the target protein and eluted through a size-exclusion column to remove unbound compounds, and the identity of the bound compounds was determined by mass spectrometry. By varying the concentration of the target protein, it was possible to rank ligands by affinity for the target protein. The activity of a range of compounds (e.g. **10a-10c**) was confirmed after re-synthesis on a 50 μmol scale.

Table 2: Case studies that exploit technology platforms that integrate small molecule synthesis and biological evaluation.

Synthesis		Purification	Biological Evaluation		Target	Ref.
Reaction	**Format**	**Format**	**Assay type**	**Format**		
S_N2, transesterification, and sulfonamide formation	Microfluidic chip (series)	None[a]	Biochemical	Caliper-chip (series)	T cell tyrosine phosphatase	34
Click chemistry[b]	Microfluidic chip (parallel)	Solid phase extraction (parallel)	b		Bovine carbonic anhydrase II	35
Acylation, Buchwald, and amination	Vial-based (parallel)	HPLC-MS (series)	Biochemical	Plate-based (parallel)	Polycomb Protein EED	36
S_N1 and Boc-deprotection	Flow/Microfluidic (series)	LC-MS (series)	Biochemical	Plate-based (parallel)	DPP4	37
Curtius rearrangement	Flow-based[c] (series)	Prep-HPLC (series)	FAC-MS[d,e]	Flow-based (series)	BRD9 bromodomain	38
Condensation, 5-exo cyclisation, amidation, and saponification	Flow-based (series)	In-line scavenging and silica plug[f] (series)	FAC-MS[d]	Flow-based (series)	Human serum albumin	39
N-acylation	Flow-based (series)	Prep-HPLC and LC-MS (series)	Functional	On-chip (series)	BACE1	40
C-O, C-N, C-C and C-S cross-couplings	Plate-based (parallel)	None	Affinity-selection mass spectrometry	Pools of products (series)	CHK1 kinase	41
Sonogashira coupling	Flow/Microfluidic (series)	Silica cartridge and LC-MS (series)	Biochemical	Plate-based (parallel)	ABL1 kinase	42

[a]The eluted product stream was split into two for LC-UV-MS analysis and direct screening. [b]In situ click chemistry catalysed by the target protein was analysed by MS. [c]Reaction precursors were prepared using batch synthesis. [d]Frontal Affinity Chromatography was used to evaluate ligand binding to immobilised protein. [e]The immobilised protein was stable up to 6 months. [f]Solvent removal and manual preparation were carried out prior to screening. Direct injection of product aliquot into the FAC assay was also possible after appropriate dilution. HPLC, high-performance liquid chromatography; UV, ultra violet; MS, mass spectrometry; LC-MS, liquid chromatography-mass spectrometry; FAC-MS, frontal affinity chromatography-mass spectrometry; DPP4, dipeptidyl peptidase-4; BACE1, Beta-secretase 1.

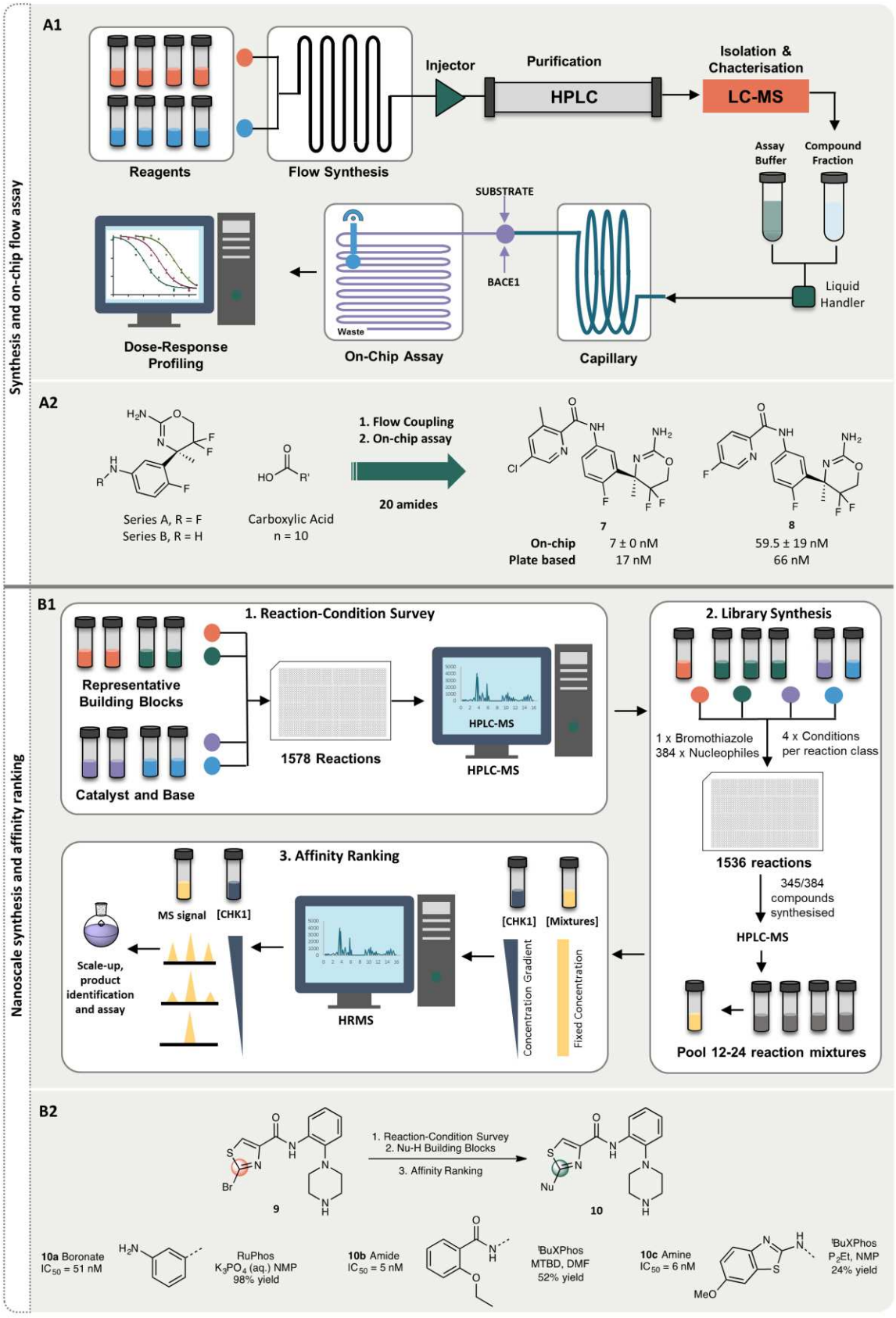

Figure 3 | Platforms that integrate synthesis and biological evaluation to enable bioactive small molecule discovery. (A) Application of an integrated platform in the discovery of BACE1 inhibitors. A1: Schematic illustration of the integrated workflow. A2: Exemplar ligands that were synthesised and tested against BACE1; the on-chip biological evaluation was validated using a plate-based assay. (B) Integration of nanoscale synthesis and affinity ranking in

the discovery of CHK1 inhibitors. B1: Schematic illustration of the integrated workflow. B2: Exemplar CHK1 inhibitors that were discovered.

The efficiency of the experimental exploration of predefined regions within chemical space can be increased using active learning models.[43] Such models are initially built from training data, and are continuously refined in real time in the light of the activity of the new compounds that are synthesised. In a programme focussed on ABI1 kinase inhibitor discovery[42] (Figure 4 and Table 2), active learning was used to drive molecular discovery through integration with microfluidic-based synthesis, purification and biochemical evaluation. A model was initially constructed on the basis of the biological activity of 32 known inhibitors. Using Ponatinib as a template, potential Asp-Phe-Gly (DFG) binding motifs and hinge-binding heterocycles were combined to define a virtual library. Ten aryl iodides (with a potential DFG-binding motif) and 27 alkynes (with a potential hinge-binding motif) were prepared, giving a total of 270 potential Sonogashira coupling products. To establish SAR quickly, and to accelerate the refinement of the model, 29 diverse products were targeted in the first round (Mode A, see Figure 4), of which 22 were successfully prepared and evaluated. The aim of the second round was to optimise potency within the previously identified hotspots (Mode B, see Figure 4): 20 compounds were targeted of which 14 were successfully investigated. In the final round, a combination of the two strategies was used. Overall, in a total of 90 cycles, 64 new compounds (of the 270 possibilities) were synthesised and evaluated with a successful completion rate of 71%. The self-refining active learning model avoided the need for exhaustive SAR studies, and 31 novel inhibitors with IC_{50} < 50 nM (e.g. **11** and **12**) were identified in a total of 135 hr machine time.

Although several platforms have been developed that integrate small molecule synthesis and biological evaluation, throughput is rarely matched at all stages, and some manual operations (e.g. purification, characterisation, quantification and solvent reformatting) are often required. In particular, compound purification is usually performed in series, even when synthesis and biological evaluation are possible in parallel, and can interrupt otherwise integrated workflows. Thus, approaches that enable parallel compound purification, or the preliminary assessment of binding using partially-purified reaction products,[44] may have particular value in integrated discovery workflows.

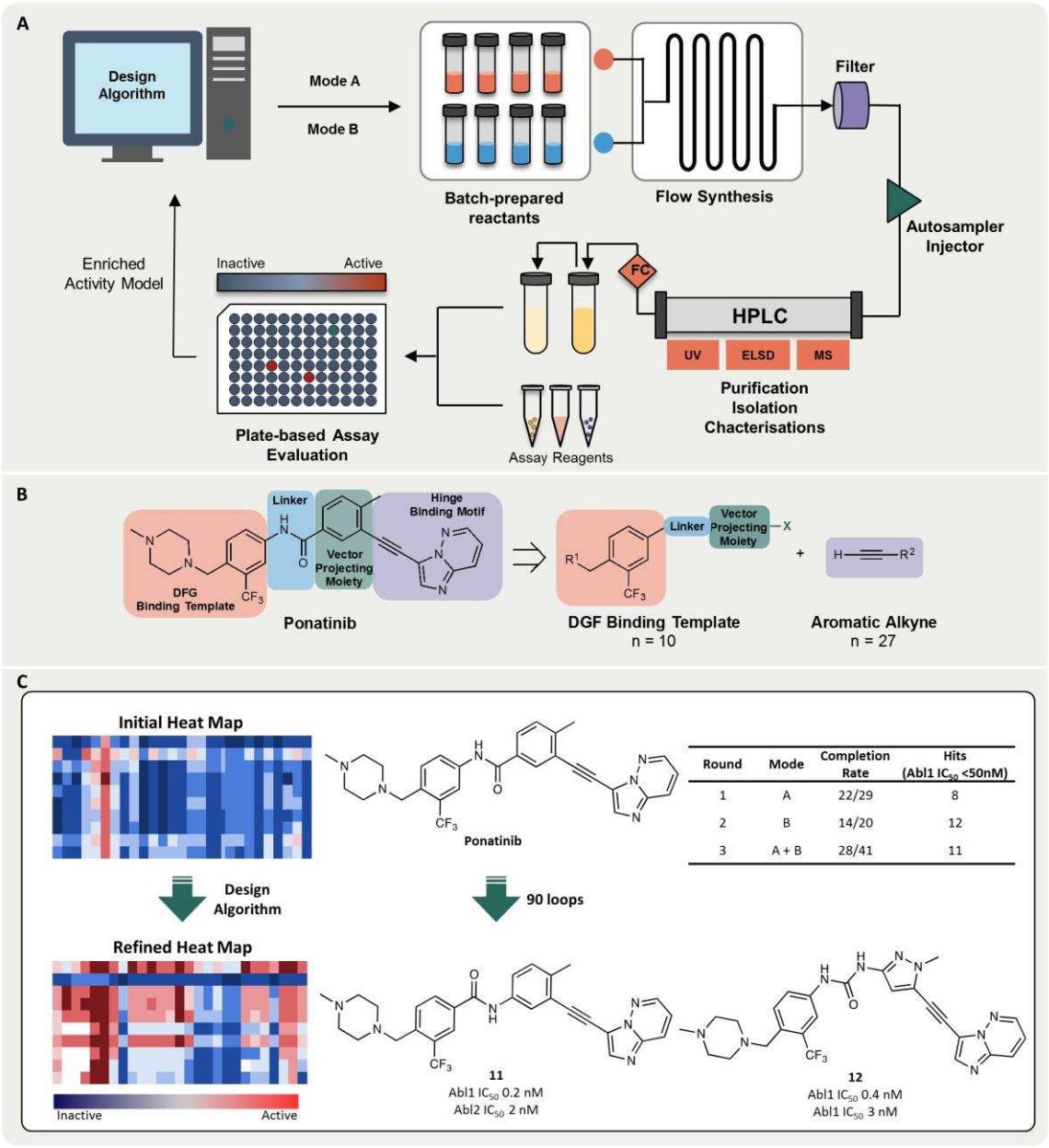

Figure 4 | Integration of active learning, synthesis, purification and evaluation in the discovery of ABI kinase inhibitors. (A) Schematic illustration of the integrated workflow with (i) machine-guided design to identify activity hotspots (Mode A) or optimise activity (Mode B), (ii) flow-based synthesis, (iii) compound purification and reformatting, and (iv) biological evaluation using a plate-based assay. (B) A initial heat map of predicted ABl1 activities was continuously refined through 90 design-synthesis-test loops.

2.3 Function-Directed Molecular Discovery

Current discovery workflows focus on the design of specific molecules that are then deliberately prepared and evaluated. The narrow toolkit of robust reactions that may be accommodated within such workflows can inadvertently limit the diversity of chemical space that may be explored. In stark contrast, the emergence of natural products is structure-blind,[45,46] and is driven by functional benefit to the host organism. Notably, the diverse functions of natural products have inspired the discovery around a third of FDA-approved drugs.[47]

Activity-directed synthesis (ADS) is a function-directed approach that borrows some concepts from the emergence of biosynthetic pathways that yield natural products.[48,49] A conceptually-similar approach – synthetic fermentation – has also been developed, and has been exemplified in the discovery of β-peptide-based inhibitors of hepatitis C virus (HCV) NS3/4A protease.[50] ADS and synthetic fermentation are complementary to dynamic combinatorial chemistry in which the most potent ligands are favoured via synthesis under thermodynamic control in the presence of a target protein.[51]

In ADS, inherently promiscuous reactions – with many possible outcomes – are deliberately harnessed to facilitate the discovery of unexpected bioactive chemotypes. Initially, a reaction array is designed in which the components (substrates, catalysts, solvents) used in reactions are widely varied. After catalyst scavenging, the crude reaction products are evaporated, dissolved in DMSO, and evaluated for biological function. The design of reaction arrays in subsequent rounds is informed by reactions that are known to yield bioactive products. The identity of the responsible bioactive ligands is only revealed after scale-up of the promising reactions, purification and functional validation.

ADS has been exploited in the discovery of agonists of the androgen receptor (AR) (Figure 5). Here, metal-catalysed carbenoid chemistry has been exploited in both intra-[48] and an intermolecular[49] senses. For the intramolecular reactions, twelve substrates were designed that contained a diazo group; the 4-cyano-3-trifluoromethyl phenyl fragment known to target AR; and a potentially reactive side chain. It was shown retrospectively that ADS can drive both optimisation of ligand structure and the corresponding synthetic route. For example, a bioactive β-lactam **14** was produced from **13** in round 1, and its yield was optimised in round 2 through exploitation of a wider range of catalysts and solvents. Finally, introduction of a related substrate **15** in round 3 enabled the identification of a related β-lactam **16**. Over the three rounds, a total of 336 reactions was performed, but only three ligands needed to be purified and structurally elucidated. Crucially, the chemotypes discovered had no annotated AR activity, and the most potent ligands identified had sub-micromolar activity. It was subsequently demonstrated that harnessing intermolecular reactions can expand the

chemical space explored, and facilitate the discovery of further novel and unexpected bioactive chemotypes (e.g. **18** and **19**).[49]

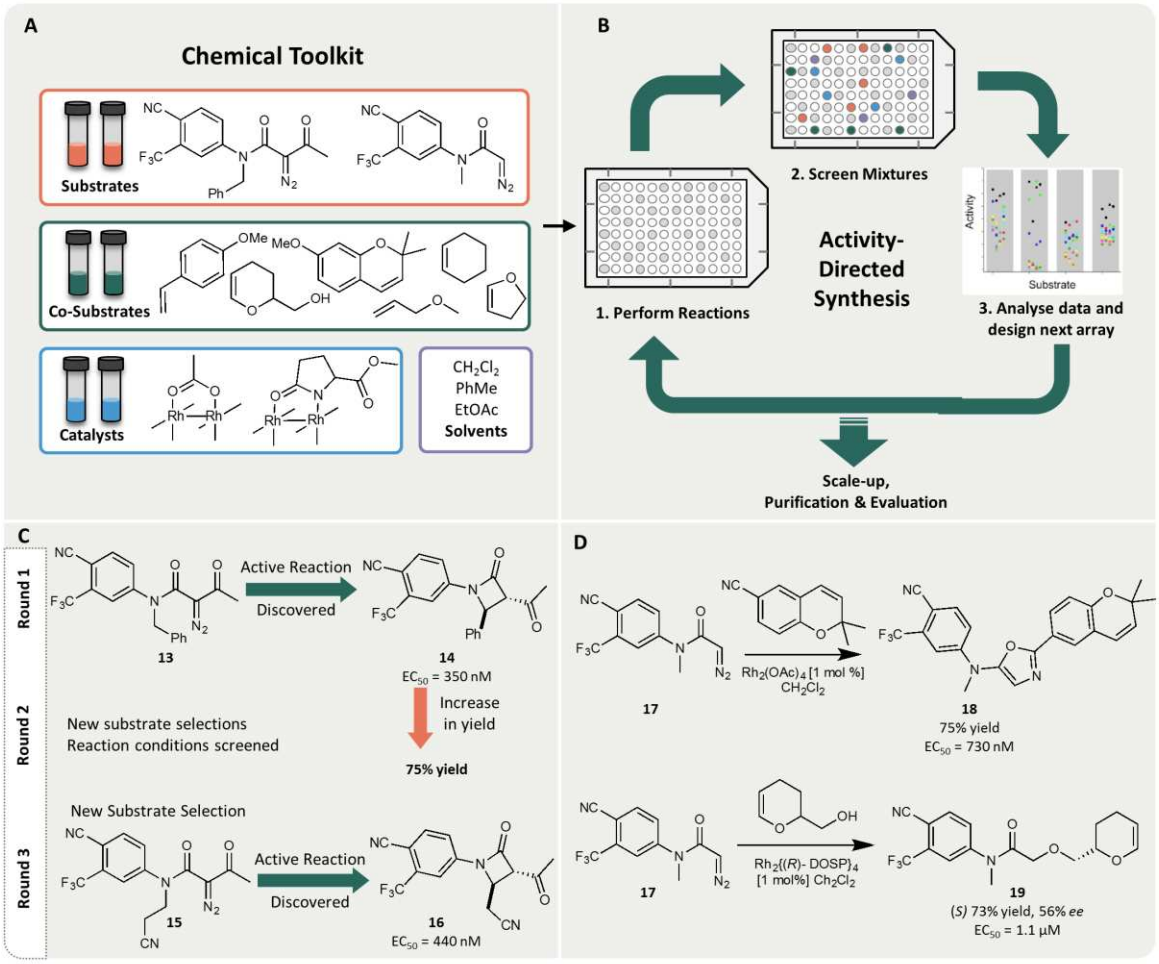

Figure 5 | Activity-directed synthesis of androgen receptor agonists. (A) Reaction array design. (B) Activity-directed synthesis workflow. (C) Evolution of intramolecular reactions that yield bioactive small molecules. (D) Scale-up of prioritised reactions, and product isolation, structural elucidation and biological evaluation; exemplar bioactive **18** and **20** are shown.

Function-directed approaches have the potential to drive the discovery of bioactive small molecules. Such approaches do not require structural information, and can deliberately harness promiscuous reactions to enable diverse regions within chemical space to be explored. All parts of the discovery workflow can be performed in parallel: the design and implementation of reaction arrays, and the scavenging and evaluation of the reaction products. Crucially, resources are then only focused on scaling up and evaluating the products of prioritised reactions that yield active products. Although function-directed approaches have only been exemplified against a limited range of targets, it has been shown than unexpected chemotypes can emerge, in parallel with an associated synthetic route.

3. Conclusions and Outlook

Significant progress has been made towards realising the fully integrated discovery of bioactive small molecules. It has been shown that adjacent stages within design-make-purify-test-analyse cycles can be integrated, and can facilitate the discovery of novel functional molecules. Crucially, the integration of reaction optimisation into discovery workflows has potential to expand significantly the chemical space that may be explored. Specifically, the ability to optimise the synthesis of each target molecule should encourage the adoption of a wider reaction toolkit (for example, C–H functionalisations) within discovery programmes. To maximise the impact on the productivity of discovery, the throughput of the stages within workflows should be both high and matched. To avoid the interruption of otherwise integrated workflows, technologies for parallel purification, and that enable assessment of the activity of partially-purified reaction products, may have particular value. In addition, we note that integration of a wider range of measurements on synthetic products – for example, to capture selectivity and physicochemical properties as well as affinity for the target protein – would have significant value in informing subsequent molecular design.

There are isolated examples of discovery programmes in which reaction optimisation, synthesis and biological evaluation have been fully integrated. Perhaps unsurprisingly, these underpinning technology platforms vary significantly. The discovery of CHK1 kinase inhibitors integrated reaction optimisation (via HPLC-MS) and the assessment of the binding of unpurified reaction products (using affinity-selection mass spectrometry) (Figure 3, Panel B).[41] Although relatively standard coupling chemistry was exploited, the underpinning technology platform may nonetheless enable a wider reaction toolkit to be exploited in discovery. The activity-directed synthesis of androgen receptor agonists exploited metal-catalysed reactions that had many possible outcomes (Figure 5).[49,50] Here, biological evaluation of the crude reaction products enabled both the ligand structure and the corresponding synthetic route to be optimised in parallel.

The realisation of autonomous bioactive molecular discovery would require integration of technologies to drive all stages within iterative cycles of molecular design, synthesis and evaluation. In addition to the experimental platforms described in this Perspective, integration of computational approaches would be required to enable the design of follow-up compounds (and their corresponding syntheses). Computation tools are now available to support the prediction of the outcome of chemical reactions,[52,53] and the design of follow-up molecules with desirable properties (including affinity, selectivity and physicochemical profiles).[2,43,54,55] However, the integration of all of these components remains an unmet challenge that is nonetheless needed to realise fully autonomous molecular discovery!

References

We thank EPSRC (EP/N025652/1) for support.

References

1. Paul, S. M. et al. How to improve R&D productivity: the pharmaceutical industry's grand challenge. Nat. Rev. Drug Discov. **9**, 203-14 (2010).
2. Schneider, G. Automating drug discovery. Nat. Rev. Drug Discov. **17**, 97-113 (2017).
3. Macarron, R. et al. Impact of high-throughput screening in biomedical research. Nat. Rev. Drug Discov. **10** (2011).
4. Ley, S. V., Fitzpatrick, D. E., Ingham, R. J. & Myers, R. M. Organic synthesis: march of the machines. Angew. Chem. Int .Ed. **54**, 3449-64 (2015).
5. Scott, D. E., Bayly, A. R., Abell, C. & Skidmore, J. Small molecules, big targets: drug discovery faces the protein–protein interaction challenge. Nat. Rev. Drug Discov. **15**, 533–550 (2016).
6. Polishchuk, P. G., Madzhidov, T. I. & Varnek A. Estimation of the size of drug-like chemical space based on GDB-17 data. J. Comput. Aided Mol. Des. **27**, 675-9 (2013).
7. Lipkus, A.H. et al. Structural diversity of organic chemistry. A scaffold analysis of the CAS Registry. J. Org. Chem. **73**, 4443-51 (2008).
8. Langdon, S. R., Brown, N. & Blagg, J. Scaffold Diversity of Exemplified Medicinal Chemistry Space. J. Chem. Inf. Model. **51**, 2174-85 (2011).
9. Taylor, R. D., MacCoss, M. & Lawson, A. D. G. Rings in Drugs. J. Med. Chem. **57**, 5845-59 (2014).
10. Lipinski, C.A. Lead- and drug-like compounds: the rule-of-five revolution. Drug. Discov. Today: Technologies **1**, 337-341 (2004).
11. Foley, D.J., Nelson, A. & Marsden, S.P. Evaluating New Chemistry to Drive Molecular Discovery: Fit for Purpose? Angew. Chem. Int. Ed. **55**, 13650-13657 (2016).
12. Walters, W. P., Green, J., Weiss, J.R. & Murcko, M.A. What do medicinal chemists actually make? A 50-year retrospective. J. Med. Chem. **54**, 6405-16 (2011).
13. Schneider, N., Lowe, D. M., Sayle, R. A., Tarselli, M. A. & Landrum, G. A., Big Data from Pharmaceutical Patents: A Computational Analysis of Medicinal Chemists' Bread and Butter. J. Med. Chem. **59**, 4385-4402 (2016).
14. Roughley, S. D. & Jordan, A.M. The medicinal chemist's toolbox: an analysis of reactions used in the pursuit of drug candidates. J. Med. Chem. **54**, 3451-79 (2011).
15. Brown, D. G. & Bostrom, J. Analysis of Past and Present Synthetic Methodologies on Medicinal Chemistry: Where Have All the New Reactions Gone? J. Med. Chem. **59**, 4443-58 (2016)

16. Cooper, T.W., Campbell, I.B. & Macdonald, S.J. 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. **49**, 8082-91 (2010).
17. Blakemore, D. C., Castro, L., Churcher, I., Rees, D. C., Thomas, A. W., Wilson, D. M. & Wood, A. Organic synthesis provides opportunities to transform drug discovery. Nat. Chem. **10**, 383-94 (2018).
18. Goodnow, R. A. Jr., Dumelin, C. E. & Keefe, A. D. DNA-encoded chemistry: enabling the deeper sampling of chemical space. Nat. Rev. Drug Discov. **16**, 131-147 (2016).
19. Li, J., Ballmer, S. G., Gillis, E. P., Fujii, S., Schmidt, M. J. Palazzolo, A. M. E., Lehmann, J. W., Morehouse, G. F. & Burke, M. D. Synthesis of many different types of organic small molecules using one automated process. Science **347**, 1221-6 (2015).
20. Ley, S. V. et al. Multi-step organic synthesis using solid-supported reagents and scavengers: a new paradigm in chemical library generation. J. Chem. Soc., Perkin Trans. 1, 3815-4195 (2000).
21. For example: Carpintero, M., Cifuentes, M., Ferritto, R., Haro, R. & Toledo, M. A. Automated liquid-liquid extraction workstation for library synthesis and its use in the parallel and chromatography-free synthesis of 2-alkyl-3-alkyl-4-(3H)-quinazolinones. J. Comb. Chem. **9**, 818-22 (2007).
22. Ghislieri, D., Gilmore, K. & Seeberger, P. H. Chemical assembly systems: layered control for divergent, continuous, multistep syntheses of active pharmaceutical ingredients. Angew. Chem. Int. Ed. **54**, 678-682 (2015).
23. Nadin, A., Hattotuwagama, C. & Churcher, I. Lead-oriented synthesis: a new opportunity for synthetic chemistry. Angew. Chem. Int. Ed. **51**, 1114-22 (2012).
24. Hwang, Y. J. J. et al. A segmented flow platform for on-demand medicinal chemistry and compound synthesis in oscillating droplets. Chem. Comm. **53**, 6649-6652 (2017).
25. Reizman, B. J., Wang, Y. M., Buchwald, S. L. & Jensen, K. F. Suzuki-Miyaura cross-coupling optimization enabled by automated feedback. React Chem Eng **1**, 658-666 (2016).
26. Cernak, T. et al. Microscale High-Throughput Experimentation as an Enabling Technology in Drug Discovery: Application in the Discovery of (Piperidinyl)pyridinyl-1H-benzimidazole Diacylglycerol Acyltransferase 1 Inhibitors. J. Med. Chem. **60**, 3594-3605 (2017).
27. Buitrago Santanilla, A. et al. Nanomole-scale high-throughput chemistry for the synthesis of complex molecules. Science, **347**, 49-53 (2015).
28. Perera, D. et al. A platform for automated nanomole-scale reaction screening and micromole-scale synthesis in flow. Science **359**, 429-434 (2018).

29. Murray, P. M., Tyler, S. N. G. & Moseley, J. D. Beyond the Numbers: Charting Chemical Reaction Space. Org. Process Res. Dev. **17**, 40-46 (2013).
30. Twilton, J. et al. The merger of transition metal and photocatalysis. Nat. Rev. Chem. **1** (2017).
31. For example: Qin, T. et al. A general alkyl-alkyl cross-coupling enabled by redox-active esters and alkylzinc reagents. Science **352**, 801-805 (2016).
32. Troshin, K. & Hartwig, J. F. Snap deconvolution: An informatics approach to high-throughput discovery of catalytic reactions. Science **357**, 175-181 (2017)
33. Collins, K. D. & Glorius, F. A robustness screen for the rapid assessment of chemical reactions. Nat. Chem. **5**, 597-601 (2013).
34. Hawkes, S. Y. F., Chapela, M. J. V., Montembault, M., Leveraging the Advantages Offered by Microfluidics to Enhance the Drug Discovery Process. QSAR Comb. Sci. **24**, 712-721 (2005).
35. Wang, J. et al, Integrated Microfluidics for Parallel Screening of an In Situ Click Chemistry Library. Angew. Chem. Int. Ed. **45**, 5276 –5281 (2006).
36. Baranczak A. et al. Integrated Platform for Expedited Synthesis−Purification−Testing of Small Molecule Libraries. ACS Med. Chem. Lett. **8**, 461−465 (2017).
37. Czechtizky W. et al. Integrated Synthesis and Testing of Substituted Xanthine Based DPP4 Inhibitors: Application to Drug Discovery. ACS Med. Chem. Lett. 4, 768−772, (2013).
38. Guetzoyan L. et al. Machine-assisted synthesis of modulators of the histone reader BRD9 using flow methods of chemistry and frontal affinity chromatography. Med. Chem. Commun. **5**, 540 (2014).
39. Guetzoyan, L., Nikbin, N., Bexandale, I. R., Ley, S. V. Flow chemistry synthesis of zolpidem, alpidem and other $GABA_A$ agonists and their biological evaluation through the use of in-line frontal affinity chromatography. Chem. Sci. **4**, 764-769 (2013)
40. Werner, M. et al. Seamless Integration of Dose-Response Screening and Flow Chemistry: Efficient Generation of Structure–Activity Relationship Data of b-Secretase (BACE1) Inhibitors. Angew. Chem. Int. Ed. **53**, 1704 –1708 (2014).
41. Gesmundo, N. J., Sauvagnat, B., Curran, P. J., Richards, M. P., Andrews, C. L., Dandliker, P. J. & Cernak, T. Nanoscale synthesis and affinity ranking. Nature **557**, 228-232 (2018).
42. Desai, B. et al. Rapid Discovery of a Novel Series of Abl Kinase Inhibitors by Application of an Integrated Microfluidic Synthesis and Screening Platform. J. Med. Chem. **56**, 3033−3047 (2013).
43. Reker, D., Schneider, G. Active-learning strategies in computer-assisted drug discovery. Drug Discovery Today, **20**, 458-465 (2015).

44. Murray, J. B., Roughley, S. D., Matassova, N. & Brough, P. A. Off-Rate Screening (ORS) By Surface Plasmon Resonance. An Efficient Method to Kinetically Sample Hit to Lead Chemical Space from Unpurified Reaction Products. J. Med. Chem. **57**, 2845-50 (2014).
45. Maplestone, R.A., Stone, M.J. & Williams, D.H. The evolutionary role of secondary metabolites--a review. Gene **115**, 151-7 (1992).
46. Firn, R.D. & Jones, C.G. Natural products - a simple model to explain chemical diversity. Nat. Prod. Rep. **20**, 382 (2003)
47. Newman, D.J. & Cragg, G.M. Natural products as sources of new drugs over the 30 years from 1981 to 2010. J. Nat. Prod, **75**, 311-35 (2012).
48. Karageorgis, G., Warriner, S. & Nelson, A. Efficient discovery of bioactive scaffolds by activity-directed synthesis. Nat. Chem. (2014).
49. Karageorgis, G., Dow, M., Aimon, A., Warriner, S. & Nelson, A. Activity-Directed Synthesis with Intermolecular Reactions: Development of a Fragment into a Range of Androgen Receptor Agonists. Angew. Chem. Int. Ed. **54**, 13538-13544 (2015).
50. Huang, Y. L. & Bode, J. W. Synthetic fermentation of bioactive non-ribosomal peptides without organisms, enzymes or reagents. Nat. Chem. **6**, 877-84 (2014).
51. Mondal, M. & Hirsch, A. K. H., Dynamic combinatorial chemistry: a tool to facilitate the identification of inhibitors for protein targets. Chem. Soc. Rev. **44**, 2455-2488 (2015).
52. Coley, C. W., Barzilay, R., Jaakkola, T. S., Green, W. H. & Jensen, K. F., Prediction of Organic Reaction Outcomes Using Machine Learning. ACS Cent. Sci. **3**, 434-443 (2017).
53. Segler, M. H. S., Preuss, M. & Waller, M. P., Planning chemical syntheses with deep neural networks and symbolic AI. Nature **555**, 604-610 (2018).
54. Yoshida, M. et al., Using Evolutionary Algorithms and Machine Learning to Explore Sequence Space for the Discovery of Antimicrobial Peptides, Chem, **4**, 533-543 (2018).
55. Pickett S. D., Green D. V. S., Hunt D. L., Pardoe D. A., and Hughes I., Automated Lead Optimization of MMP-12 Inhibitors Using a Genetic Algorithm, ACS Med. Chem. Lett., **2**, 28-33 (2011).