

Cite this: *Chem. Sci.*, 2022, 13, 14292

All publication charges for this article have been paid for by the Royal Society of Chemistry

## A vending machine for drug-like molecules – automated synthesis of virtual screening hits†

Angus E. McMillan,<sup>a</sup> Wilson W. X. Wu,<sup>a</sup> Paula L. Nichols,<sup>ab</sup> Benedikt M. Wanner<sup>\*b</sup> and Jeffrey W. Bode<sup>ID \*a</sup>

As a result of high false positive rates in virtual screening campaigns, prospective hits must be synthesised for validation. When done manually, this is a time consuming and laborious process. Large “on-demand” virtual libraries ( $>7 \times 10^{12}$  members), suitable for preparation using capsule-based automated synthesis and commercial building blocks, were evaluated to determine their structural novelty. One sub-library, constructed from iSnAP capsules, aldehydes and amines, contains unique scaffolds with drug-like physicochemical properties. Virtual screening hits from this iSnAP library were prepared in an automated fashion for evaluation against *Aedes aegypti* and *Phytophthora infestans*. In comparison to manual workflows, this approach provided a 10-fold improvement in user efficiency. A streamlined method of relative stereochemical assignment was also devised to augment the rapid synthesis. User efficiency was further improved to 100-fold by downscaling and parallelising capsule-based chemistry on 96-well plates equipped with filter bases. This work demonstrates that automated synthesis consoles can enable the rapid and reliable preparation of attractive virtual screening hits from large virtual libraries.

Received 17th September 2022

Accepted 27th October 2022

DOI: 10.1039/d2sc05182f

rsc.li/chemical-science

### Introduction

As part of drug discovery efforts, including those seeking to exploit novel targets and new IP space, virtual screening provides an efficient method of evaluating a vast chemical space and rationalizing structure activity relationships.<sup>1,2</sup> Contrary to conventional high throughput assays, virtual screening occurs in an entirely digital environment.<sup>3</sup> The near absence of physical constraints allows ultra-high throughput screening and the evaluation of molecules which do not yet physically exist.<sup>4</sup> The enrichment of true positives in samples selected by virtual screening reduces the number of molecules synthetic chemists must prepare but this is limited by a high rate of false positives.<sup>5–7</sup> A recent review of 53 structure based virtual screening campaigns provided a median false positive rate of 83%.<sup>8,9</sup> This necessitates the synthesis of tens to hundreds of putative hit compounds in order to verify one true positive, a task which currently demands a considerable investment in time and resources.<sup>10</sup> Automation of the synthesis steps can increase labour efficiencies, thereby diminishing the barrier to hit validation and enabling the digitalisation of drug discovery.<sup>11</sup>

Automation is particularly well suited to preparing hits from virtual or “on-demand libraries”.<sup>12</sup> Such libraries are derived from a bank of building blocks and a defined set of high-fidelity reactions.<sup>13</sup> They contain molecules which have a high likelihood of successfully being prepared according to a pre-determined route, ensuring synthetic accessibility, and negating retrosynthetic analysis. For example, Enamine Ltd has performed foundational work in this area by providing diverse building blocks and establishing manual on-demand synthesis of molecules from a library of  $2.9 \times 10^{10}$  members.<sup>14,15</sup> Significant steps towards the fully automated synthesis of virtual screening hits have been achieved by Eli Lilly and Strateos with Idea2Data.<sup>16</sup> This infrastructure unites the Proximal Lilly Collection virtual library with their automated synthesis laboratory.<sup>17,18</sup>

Although such synthesis platforms offer the potential to accelerate discovery, they require large capital investments, physical infrastructure and technical staff, rendering them

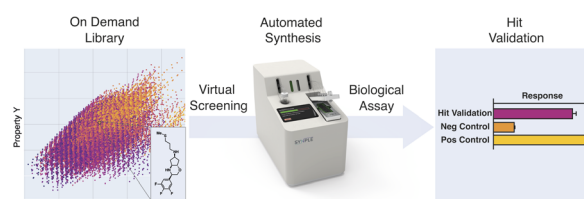


Fig. 1 The concept of using console based automated synthesis to validate virtual screening hits.

<sup>a</sup>Laboratory for Organic Chemistry, Department of Chemistry and Applied Biosciences, ETH Zürich, Zürich 8093, Switzerland. E-mail: bode@org.chem.ethz.ch

<sup>b</sup>Synple Chem AG, Kemptpark 18, Kempththal 8310, Switzerland. E-mail: wanner@synplechem.com

† Electronic supplementary information (ESI) available. CCDC 2169258, 2169260, 2169267 and 2169554. For ESI and crystallographic data in CIF or other electronic format see DOI: <https://doi.org/10.1039/d2sc05182f>



beyond the reach of all but the largest drug-discovery organizations. With the development of smaller, lower-cost, multi-functional automated synthesis consoles, it is now possible that the role of these costly, highly specialized automated synthesis labs may be substituted by more widely available platforms (Fig. 1).<sup>19–21</sup>

As part of our efforts to provide inexpensive, flexible, and predictable synthesis of drug-like compounds, we have engineered compact, capsule-based instruments for fully automated chemical synthesis.<sup>19</sup> By employing fully automated assembly reactions, combined with bifunctional building blocks and commercial reagents, we can access a broad scope of attractive small molecules, with minimal time and manual operation. Limiting the range of assembly reactions does not preclude complexity or novelty, thanks to the vast scope of available

building blocks.<sup>22</sup> Analogous to a vending machine, this system cannot prepare any molecule imaginable, but if a virtual screening hit is identified from the virtual “on-demand” library, then it can be rapidly obtained as a physical sample with minimal effort.

## Results and discussion

We first sought to determine the structure and number of products which could be assembled from commercial building blocks using a linear sequence of three automated synthetic steps. To construct a virtual library of the possible products, we selected commercial building blocks, sourced from “in stock” catalogues, which had been filtered to remove problematic or incompatible functional groups. These materials were then

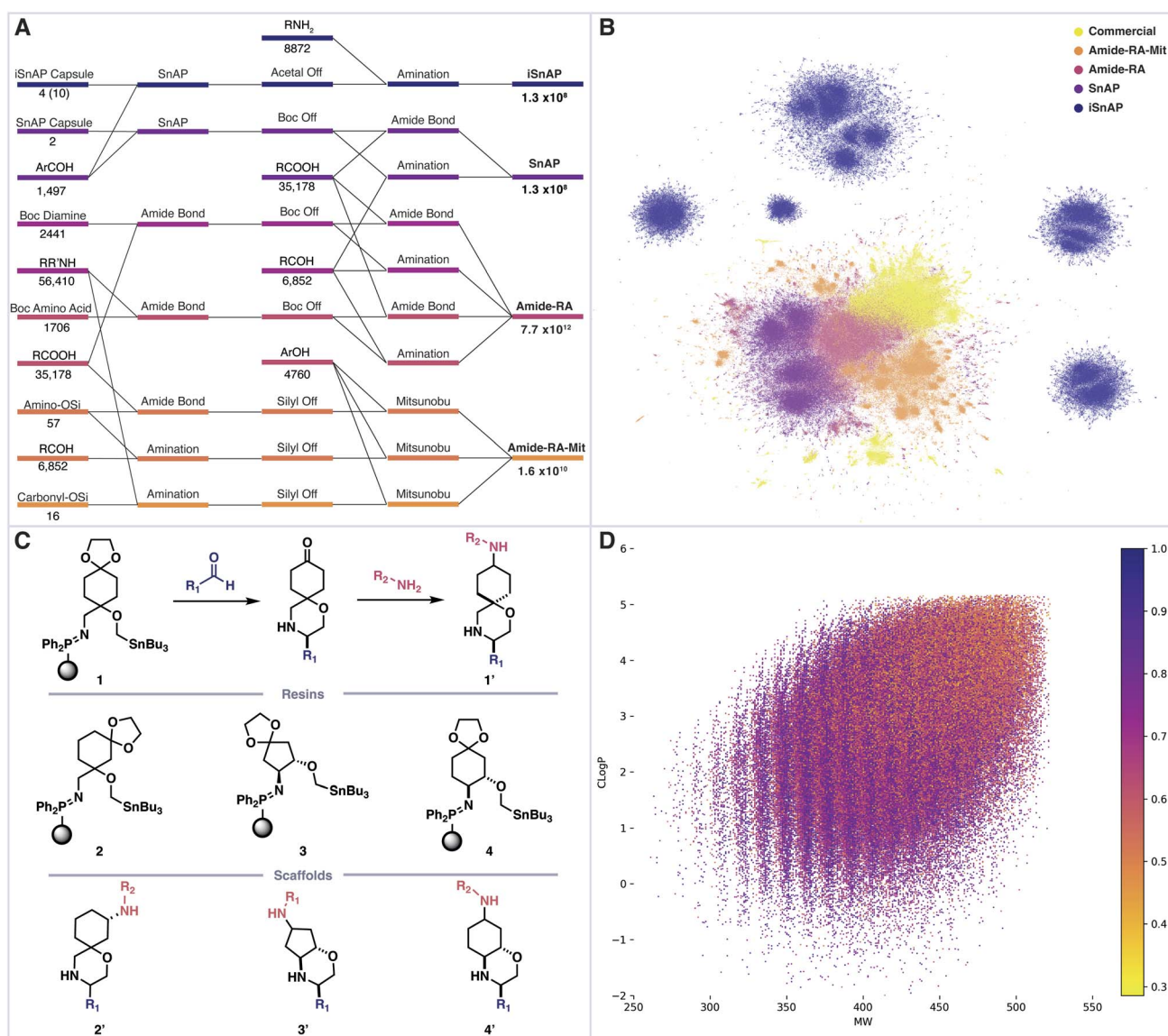


Fig. 2 (A) An enumeration workflow, including number of initial building blocks and final library members. (B) UMAP dimensionality reduction of Morgan fingerprints from 1 x 10<sup>5</sup> randomly selected members from virtual libraries and Zinc now, n\_neighbours = 10, min\_distance = 0.25. (C) iSnAP library is prepared from iSnAP resins, commercial aldehydes and amines. (D) iSnAP library properties cLogP, molecular weight and fraction of sp<sup>3</sup> carbons.



assembled using various combinations of chemical reactions, all of which would be amenable to complete automation (setup, execution, workup, purification) using Synple's reaction capsules and instruments. The selected assembly reactions included reductive amination, amide bond formation, Mitsunobu and Sn amine protocol (SnAP) chemistry, along with appropriate automated deprotections.<sup>23–26</sup> The resulting full library of decorated bifunctional scaffolds exceeds 7 trillion ( $7 \times 10^{12}$ ) members (Fig. 2A).

A sample of  $1 \times 10^5$  molecules was randomly taken from each sub-library, along with ZINC now as a source of commercially available compounds.<sup>27</sup> Morgan fingerprints (radius 6, bits 2048) were generated for each entry and the dimensionality reduction method UMAP was used to produce a 2D plot, in which structural similarity trends could be observed (Fig. 2B).<sup>28,29</sup> Each library displayed varying degrees of structural similarity to the ZINC commercial compounds, with the exception of the iSnAP library which was largely isolated.

This particular library focused on the use of iSnAP resins to build unique spirocyclic or fused morpholines containing a ketal-protected ketone (Fig. 2C). We previously reported a capsule for the preparation of ketones, derived from resin 1, which uses the SCX-2 resin, typically used for automated purification in SnAP capsules, to affect the ketal deprotection.<sup>19</sup> The resulting ketone can be reacted with amines, using a reductive amination capsule, to provide derivatives of scaffold 1'. Resins 1, 3 and 4 gave the final compounds as a mixture of two diastereomers, with the reductive amination being identified as the divergent step; whereas resin 2 provided access to all four diastereomers. The 10 accessible diastereomers of scaffolds 1'–4' were enumerated in full using commercial aldehydes and amines to afford an iSnAP library of  $1.3 \times 10^8$  compounds, from which the randomly selected  $1 \times 10^5$  members were drawn for structural comparison.

Further analysis revealed that the structural novelty of the iSnAP virtual library likely stemmed from the use of unique scaffolds; substructure analysis against Zinc now returned no results for the iSnAP scaffolds represented as Murcko frameworks with alpha substitution.<sup>30</sup> Analysis of the physicochemical properties showed good coverage of drug-like chemical space and application of Lipinski's rule of five filters proceeded with retention of 98.6% of the library (Fig. 2D).<sup>31</sup> Furthermore, the average fraction of  $sp^3$  carbons was found to be 0.70. Such a high value is desirable in prospective drug libraries as saturation has been shown to improve physicochemical properties, such as solubility, and tends to provide less promiscuous drugs by providing more specific vectors for interactions.<sup>32,33</sup>

We provided this iSnAP sub-library to collaboration partners with the interest and infrastructure for virtual screening. In one example, the BASF Open Innovation Platform used the library to identify compounds of interest against 12 target organisms. As part of our collaboration, we prepared these 20 compounds, along with 21 other examples, using fully automated, capsule-based synthesis (Scheme 1).

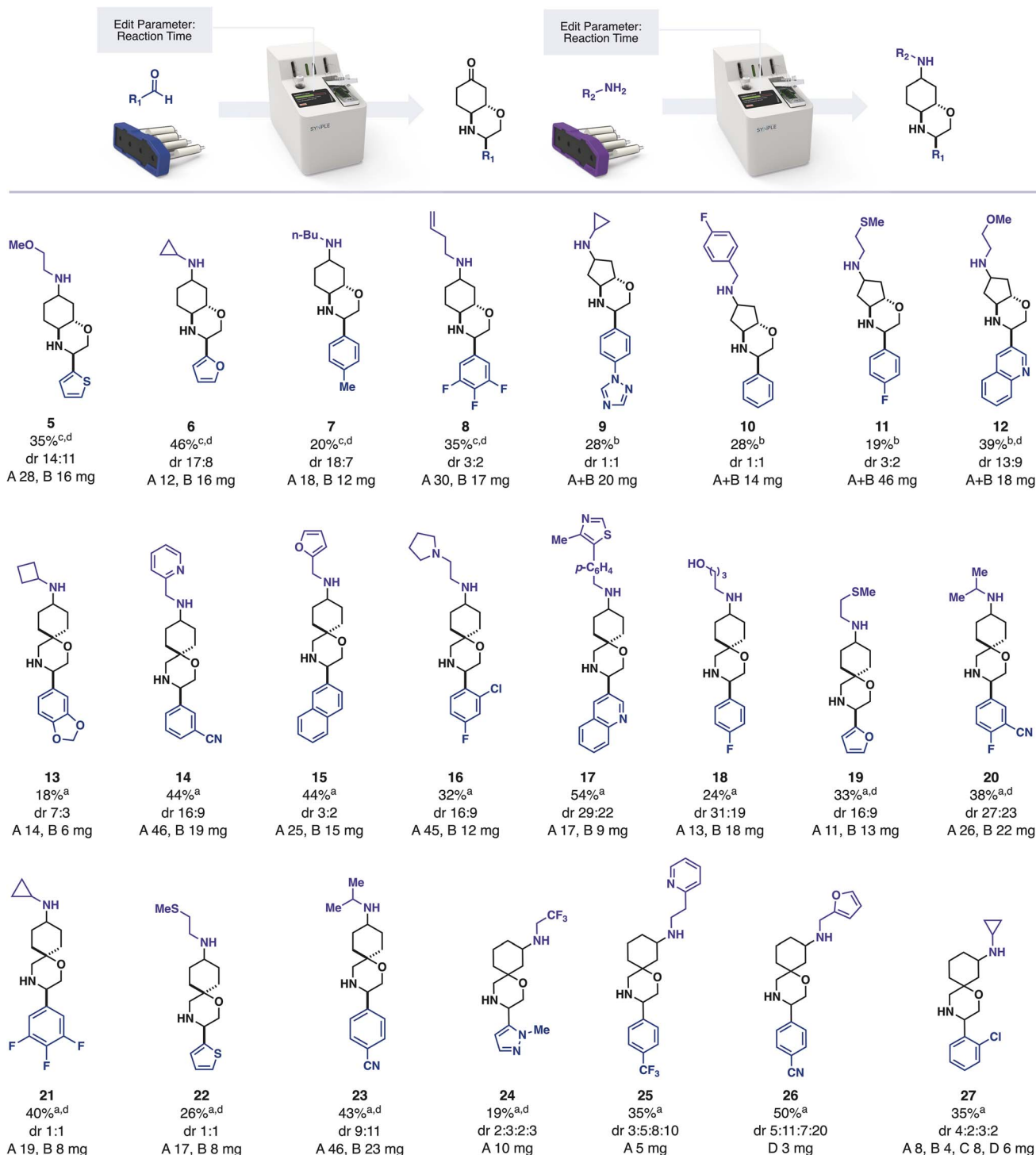
In a typical automated synthetic workflow, the starting aldehyde was added to the reaction vial and the appropriate reaction program was loaded by scanning the iSnAP capsule

using the RFID scanner on the console. Upon pressing "START" the full program of operations required for the reaction and product isolation were automatically initiated, with no further user input required. The system first dissolved the aldehyde in trifluorotoluene and then circulated the solution through the first cartridge of the capsule, which contained one equivalent of the resin-bound iSnAP iminophosphorane at 80 °C. Although the rate of imine formation can be substrate dependent, we used 5 hours of circulation and found that it provided a suitable quantity of output material in all cases. After washing the resin and cooling of the reaction media, the system added one equivalent of  $Cu(OTf)_2$  and lutidine·HOTf by dissolving them out of the second cartridge in HFIP. Upon reaction completion, the automated program removed copper salts by passing the reaction media through a cartridge containing silica. After washing the silica with MeOH, the resulting solution was automatically loaded onto SCX-2 resin. Tributyltin byproducts and other non-basic impurities were washed from the SCX-2 before a solution of aqueous acetone was circulated over the SCX-2 to affect the ketal deprotection while the product was immobilized on the resin.

The products of resins 3 and 4 (fused morpholines) required longer deprotection times compared to resins 1 and 2. Prior to starting the program the reaction time could be edited by selecting "Edit Parameter" on the touch screen interface. Using this feature, it was found that ketones derived from resin 2 were unstable under prolonged deprotection conditions, with no ketone being detected by  $^{13}C$  NMR when the reaction time was increased to 24 hours. This demonstrates a further advantage of automation, in that experimental variables, such as reaction time, can be controlled and tuned without requiring manual actions such as removing a heat source. Upon completion of deprotection, the ketone was eluted from the SCX-2 back into the starting vial using triethylamine in acetonitrile. In the sole manual-intervention step, the solution was concentrated on a rotary evaporator and the selected amine added to the vial. This vial – containing the two reaction partners – was subjected to reductive amination using the appropriate capsule and reaction program on the console. In this reaction, the console circulated the reaction mixture over silica-supported cyanoborohydride in  $CH_2Cl_2$ /HFIP (4 : 1). Any unreacted amine was scavenged by automatically passing the reaction mixture through pre-swelled polymer-supported benzaldehyde and non-basic impurities were removed using SCX-2 catch and release. After elution of the product, the user concentrated the reaction mixture and added the internal standard 1,3,5-trimethoxybenzene to determine the crude yield of the reaction *via* quantitative  $^1H$  NMR. In addition to crude yield, these spectra also informed us that the automated work-up was extremely effective at removing tributyltin residues. Isolation of the final products, including separation of the diastereomers formed, was achieved using standard preparative HPLC methods.

The iSnAP library products could be produced as separable diastereomeric mixtures, further adding to the library diversity. While it is often favourable in early discovery to prepare molecules as racemic mixtures, an assignment of the relatively stereochemistry for diastereoisomers is essential in order to





**Scheme 1** Preparation of library members; crude yield prior to resolution as determined by  $^1\text{H}$  NMR using 1,3,5-trimethoxybenzene as an internal standard; dr determined by HPLC; isolated mass after resolution in mg; <sup>a</sup> deprotection 5 h, reductive amination 3 h <sup>b</sup> deprotection 24 h, reductive amination 12 h <sup>c</sup> deprotection 12 h, reductive amination 12 h <sup>d</sup> pre-selected for biological activity.

establish a link between ligand structure and activity. The gold standard for stereochemical assignment is X-ray crystallography but growing suitable crystals is a time- and material-consuming process. To avoid negating the time-saving benefits of automation, we sought to establish a method of assignment using

data from an analytical technique routinely used for characterization.

In the case of the products derived from resins **3** and **4** (fused morpholines) the assignment of relative stereochemistry could be achieved using NOE NMR experiments (Fig. 3A). Products of



resins **1** and **2** (spirocyclic morpholines) were more difficult to assign, on account of their remote and tertiary stereocentres. The NMR shielding parameters from DFT optimized conformers of compound **15** were Boltzmann weighted to provide predicted  $^{13}\text{C}$  chemical shifts for **15 DiaA** and **15 DiaB**. Although the predictions provided the correct assignment with low corrected mean average errors (AA' 1.39 vs. AB' 1.73, BA' 1.90 vs. BB' 1.66 ppm, B3LYP-D3 6-31G\*\*) the low deviation between experimental results reduced the confidence in this method of assignment.

To assuredly differentiate between the two very similar diastereomeric products, we sought to identify the spectroscopic signal with maximal divergence, and make that signal the target of our predictions. We reasoned that the preference for the scaffold substituents to adopt an axial conformation would shepherd the morpholine methylene into an axial conformation for one of the diastereomers, resulting in a large diastereotopic splitting (Fig. 3A). The lowest energy conformer for both diastereomers were optimized with DFT (B3LYP 6-311G\*\*). The difference in the  $^1\text{H}$  NMR shielding parameters showed good correlation, when compared to the measured diastereotopic splitting values, and consistently differentiated the two diastereomers (Fig. 3B). Furthermore, a browser-based tool with negatable calculation time also provided useable predictions of the diastereotopic splitting, albeit with less accuracy.<sup>34,35</sup> This approach became the default method of assigning the relative stereochemistry of spirocyclic morpholines, augmented in examples with three stereocenters by NOE interactions. The method was validated using X-ray crystallography and can be run in parallel to the automated synthesis so as to not adversely influence hit preparation time.<sup>36,37</sup> Upon compound isolation, only the acquisition of a  $^1\text{H}$  NMR spectrum is needed to make the assignment.

With the assignments in hand, the synthesized hit compounds were screened *in vivo* against the 12 target organisms selected by the BASF Open Innovation Platform (Scheme 1).<sup>38</sup> Compound **10 DiaB** demonstrated average activity against *Aedes aegypti* in the open innovation lead finder assay, where insects were sprayed with 2500 ppm in the entrance screen. Compounds **21 DiaA**, **22 DiaB** and **23 DiaA** also showed average activity in a hit finder assay against *Phytophthora infestans* at 31 ppm. The identification of hit molecules with even modest activity was encouraging and demonstrated that it is indeed possible to use a limited range of robust, automated reaction classes to assemble commercial building blocks into desirable molecules.

We calculated the efficiency, for both manual and automated hit preparation, as user time per molecule prepared.<sup>39</sup> We estimate that manual preparation can be conducted at a rate of 2.75–5.50 user hours per molecule, not including solvent evaporation and lost time from mismatches in operation scheduling and working hours. In the automated process, preparation of the same molecules takes place at a rate of 0.28–0.55 user hours per molecule. This approximates to a 10-fold improvement in efficiency. An important distinction is that, thanks to their compact nature, the automation rate can be scaled linearly by assembling banks of synthesis consoles.

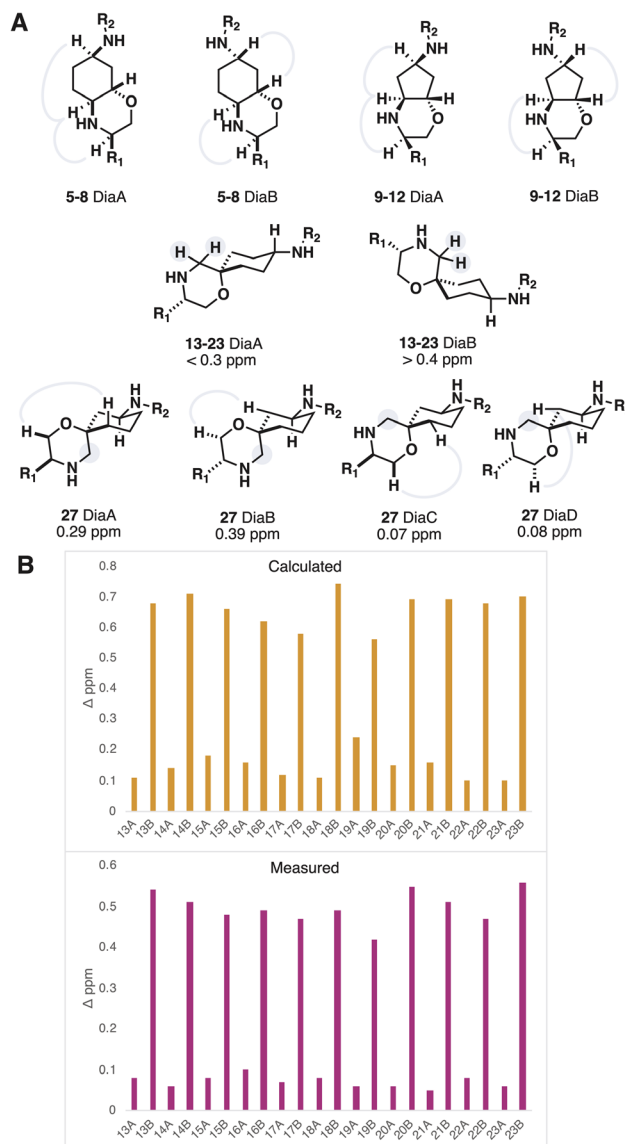
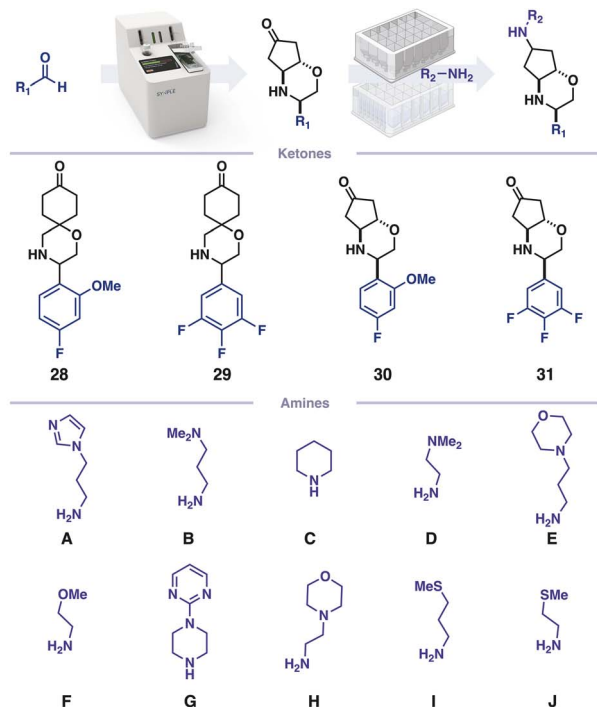


Fig. 3 (A) Relative stereochemistry determination of accessible stereoisomers, NOE interactions and splitting of characteristic diastereotopic protons highlighted. (B) Calculated and measured diastereotopic splitting values for 13–23 using B3LYP 6-311G\*\*.

For many screening campaigns, and in the pursuit of large datasets for machine learning, a further increase in throughput would be beneficial. Capsule based chemistry can be miniaturised to a 4  $\mu\text{mol}$  scale (100-fold decrease) and performed in parallel using 96-well-plates equipped with filter floors in place of a capsule (Fig. 4). To demonstrate this, four ketones were prepared on the console and isolated. Each ketone was combined with 10 different amines in  $\text{CH}_2\text{Cl}_2/\text{HFIP}$  (3 : 1) and transferred to a 96-well filter plate containing silica supported cyanoborohydride. The reactions were agitated for 12 h and the crude material was subsequently obtained by centrifuging the reaction solutions through the filter into a collection plate, which was concentrated prior to analysis by high throughput LCMS.



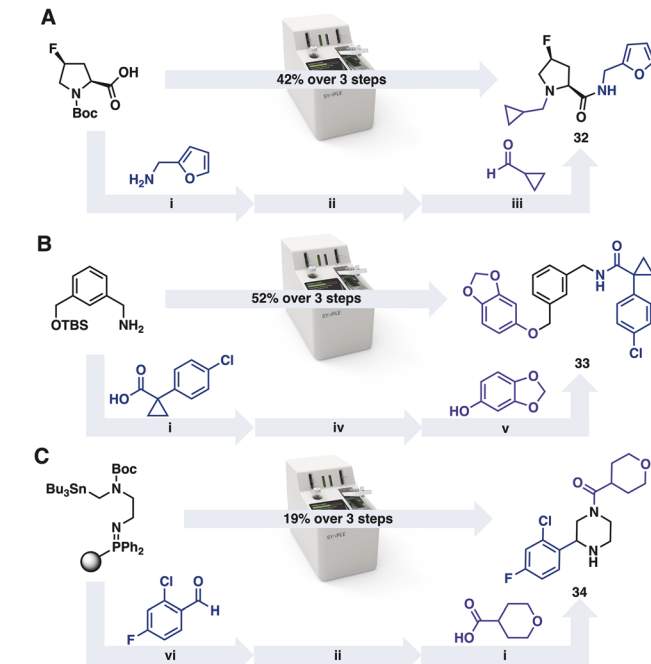


	A	B	C	D	E	F	G	H	I	J
28	63.7	32.6	16.5	74.6	51.5	61.4	60.5	98.0	67.7	37.4
29	93.0	95.0	42.4	87.4	91.6	78.4	86.1	96.1	82.3	70.0
30	88.1	94.1	74.2	99.4	95.2	94.6	97.0	99.4	89.4	95.5
31	93.2	96.3	91.1	97.6	97.5	98.0	93.6	99.8	97.4	88.6

Fig. 4 Rapid diversification using filter plate-based reductive amination; product area percentage as calculated from mass ion extracted chromatograms of ketone, alcohol byproduct and desired product.

Automated extraction of mass ion chromatograms was used for the qualitative detection of remaining starting material and an alcohol byproduct. This analysis informs subsequent screening assays which negative controls should be included, thereby allowing the method to prepare hits at a rate exceeding 0.04–0.07 user hours per molecule, giving a final 100-fold efficacy increase. This rate can be developed further by including more amines in the parallel diversification step. As the plate-based diversification has no resolution of diastereomers, this increase in throughput comes at the cost of confounding relative stereochemical effects. However, the established trends can be conveniently ratified by preparing key molecules using the initial console-based workflow.

In addition to preparing molecules from the iSnAP library, commercial reaction capsules can be applied iteratively to prepare members from the other on-demand libraries (Scheme 2).<sup>40</sup> The default reaction programs were used without optimisation. After completion of each transformation, the console performed an automated purification of the product. The resulting material was concentrated and used directly in the subsequent automated reaction. In the examples tested, ample material for screening was obtained from each of the three step reaction sequences.



Scheme 2 Automated synthesis of members from virtual libraries (A) Amide-RA, (B) Amide-RA-Mit and (C) SnAP libraries; i amide bond forming capsule, ii *tert*-butyloxycarbonyl (Boc) deprotection capsule, iii reductive amination capsule, iv *tert*-butyldimethylsilyl (TBS) ether deprotection capsule, v Mitsunobu capsule, vi SnAP capsule.

## Conclusions

Synthesis consoles can use iterative automated reactions to rapidly assemble molecules from virtual libraries exceeding  $7 \times 10^{12}$  members. Structural comparison indicates that this library covers a similar space to a source of commercial compounds in addition to containing sub-libraries with unique structural elements. One structurally unique sub-library derived from commercial aldehydes, amines and iSnAP capsules was found to contain novel scaffolds with drug-like physicochemical properties. The automated preparation of screening hits from this library was demonstrated and a streamlined method of relative stereochemical assignment was devised to augment the rapid synthesis. Comparison to the manual preparation of the same molecules, the automated workflow provided a 10-fold improvement in user efficiency, which was further increased to >100-fold by using a parallel diversification strategy. Overall, the integration of console based automated synthesis and on-demand virtual libraries provides a widely deployable platform for the rapid validation of virtual screening hits.

## Data availability

Experimental data is provided in the ESL.†

## Author contributions

A. E. M. and W. W. X. W. prepared and analysed the virtual libraries. A. E. M. performed the experiments and diastereomer



assignment. The manuscript was written by A. E. M., P. L. N. and J. W. B.

## Conflicts of interest

B. M. W., P. L. N. and J. W. B. are listed as inventors of a patent (WO Pat., WO2017121724A1, 2016) related to this technology and are co-founders of Synple Chem AG.

## Acknowledgements

Financial support was provided by Innosuisse (innovation project 27406.1 PFLS-LS). We thank Daniel Wirz, Louis Bertschi, the Molecular and Biomolecular Analysis Service (MoBIAS) and the Small Molecule Crystallography Center of ETH Zürich for their analytical support. We are grateful to Marc-Olivier Ebert, Thomas Stadelmann and Sereina Riniker for their useful discussions on the stereochemistry of spirocyclic compounds. Open Innovation Platform Agro screening data supplied by courtesy of BASF SE—used with BASF SE's permission.

## References

- 1 A. Lavecchia and C. Giovanni, Virtual screening strategies in drug discovery: a critical review, *Curr. Med. Chem.*, 2013, **20**, 2839.
- 2 K. Ogata, T. Isomura, H. Yamashita and H. Kubodera, A Quantitative Approach to the Estimation of Chemical Space from a Given Geometry by the Combination of Atomic Species, *QSAR Comb. Sci.*, 2007, **26**, 596.
- 3 A. Varela-Rial, M. Majewski and G. D. Fabritiis, Structure based virtual screening: Fast and slow, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2021, **12**, e1544.
- 4 C. Gorgulla, A. Boeszoermyeni, Z.-F. Wang, P. D. Fischer, P. W. Coote, K. M. Padmanabha Das, Y. S. Malets, D. S. Radchenko, Y. S. Moroz, D. A. Scott, K. Fackeldey, M. Hoffmann, I. Iavniuk, G. Wagner and H. Arthanari, An open-source drug discovery platform enables ultra-large virtual screens, *Nature*, 2020, **580**, 663.
- 5 V.-K. Tran-Nguyen, G. Bret and D. Rognan, True Accuracy of Fast Scoring Functions to Predict High-Throughput Screening Data from Docking Poses: The Simpler the Better, *J. Chem. Inf. Model.*, 2021, **61**, 2788.
- 6 H. Li, K.-H. Sze, G. Lu and P. J. Ballester, Machine-learning scoring functions for structure-based virtual screening, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2021, **11**, e1478.
- 7 J. Jiménez, M. Škalič, G. Martínez-Rosell and G. De Fabritiis, KDEEP: Protein–Ligand Absolute Binding Affinity Prediction via 3D-Convolutional Neural Networks, *J. Chem. Inf. Model.*, 2018, **58**, 287.
- 8 Y. O. Adeshina, E. J. Deeds and J. Karanicolas, Machine learning classification can reduce false positives in structure-based virtual screening, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 18477.
- 9 J. J. Irwin and B. K. Shoichet, Docking Screens for Novel Ligands Conferring New Biology: Miniperspective, *J. Med. Chem.*, 2016, **59**, 4103.
- 10 D. C. Blakemore, L. Castro, I. Churcher, D. C. Rees, A. W. Thomas, D. M. Wilson and A. Wood, Organic synthesis provides opportunities to transform drug discovery, *Nat. Chem.*, 2018, **10**, 383.
- 11 J. Boyd, Robotic Laboratory Automation, *Science*, 2002, **295**, 517.
- 12 W. P. Walters, Virtual Chemical Libraries: Miniperspective, *J. Med. Chem.*, 2019, **62**, 1116.
- 13 S. D. Roughley and A. M. Jordan, The Medicinal Chemist's Toolbox: An Analysis of Reactions Used in the Pursuit of Drug Candidates, *J. Med. Chem.*, 2011, **54**, 3451.
- 14 O. O. Grygorenko, Enamine Ltd: The Science and Business of Organic Chemistry and Beyond, *Eur. J. Org. Chem.*, 2021, 6474.
- 15 O. O. Grygorenko, D. S. Radchenko, I. Dziuba, A. Chuprina, K. E. Gubina and Y. S. Moroz, Generating Multibillion Chemical Space of Readily Accessible Screening Compounds, *iScience*, 2020, **23**, 101681.
- 16 C. A. Nicolaou, C. Humblet, H. Hu, E. M. Martin, F. C. Dorsey, T. M. Castle, K. I. Burton, H. Hu, J. Hendle, M. J. Hickey, J. Duerksen, J. Wang and J. A. Erickson, Idea2Data: Toward a New Paradigm for Drug Discovery, *ACS Med. Chem. Lett.*, 2019, **10**, 278.
- 17 C. A. Nicolaou, I. A. Watson, H. Hu and J. Wang, The Proximal Lilly Collection: Mapping, Exploring and Exploiting Feasible Chemical Space, *J. Chem. Inf. Model.*, 2016, **56**, 1253.
- 18 A. G. Godfrey, T. Masquelin and H. Hemmerle, A remote-controlled adaptive medchem lab: an innovative approach to enable drug discovery in the 21st Century, *Drug Discovery Today*, 2013, **18**, 795.
- 19 T. Jiang, S. Bordi, A. E. McMillan, K.-Y. Chen, F. Saito, P. L. Nichols, B. M. Wanner and J. W. Bode, An integrated console for capsule-based, automated organic synthesis, *Chem. Sci.*, 2021, **12**, 6977.
- 20 S. H. M. Mehr, M. Craven, A. I. Leonov, G. Keenan and L. Cronin, A universal system for digitization and automatic execution of the chemical synthesis literature, *Science*, 2020, **370**, 101.
- 21 J. Li, S. G. Ballmer, E. P. Gillis, S. Fujii, M. J. Schmidt, A. M. E. Palazzolo, J. W. Lehmann, G. F. Morehouse and M. D. Burke, Synthesis of many different types of organic small molecules using one automated process, *Science*, 2015, **347**, 1221.
- 22 A. Tomberg and J. Boström, Can easy chemistry produce complex, diverse, and novel molecules?, *Drug Discovery Today*, 2020, **25**, 2174.
- 23 O. I. Afanasyev, E. Kuchuk, D. L. Usanov and D. Chusov, Reductive Amination in the Synthesis of Pharmaceuticals, *Chem. Rev.*, 2019, **119**, 11857.
- 24 V. R. Pattabiraman and J. W. Bode, Rethinking amide bond synthesis, *Nature*, 2011, **480**, 471.
- 25 O. Mitsunobu and M. Yamada, Preparation of Esters of Carboxylic and Phosphoric Acid via Quaternary Phosphonium Salts, *Bull. Chem. Soc. Jpn.*, 1967, **40**, 2380.



- 26 M. U. Luescher, C.-V. T. Vo and J. W. Bode, SnAP Reagents for the Synthesis of Piperazines and Morpholines, *Org. Lett.*, 2014, **16**, 1236.
- 27 T. Sterling and J. J. Irwin, ZINC 15 – Ligand Discovery for Everyone, *J. Chem. Inf. Model.*, 2015, **55**, 2324.
- 28 H. L. Morgan, The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service, *J. Chem. Doc.*, 1965, **5**, 107.
- 29 L. McInnes, J. Healy and J. Melville, *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*, arXiv, 2020, pp. 1802–03246.
- 30 G. W. Bemis and M. A. Murcko, The Properties of Known Drugs. 1. Molecular Frameworks, *J. Med. Chem.*, 1996, **39**, 2887.
- 31 C. A. Lipinski, Lead- and drug-like compounds: the rule-of-five revolution, *Drug Discovery Today: Technol.*, 2004, **1**, 337.
- 32 F. Lovering, J. Bikker and C. Humblet, Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success, *J. Med. Chem.*, 2009, **52**, 6752.
- 33 F. Lovering, Escape from Flatland 2: complexity and promiscuity, *MedChemComm*, 2013, **4**, 515.
- 34 D. Banfi and L. Patiny, Resurrecting and processing NMR spectra on-line, *Chimia*, 2008, **62**, 280, <https://www.nmrdb.org/>.
- 35 Y. Binev, M. M. B. Marques and J. Aires-de-Sousa, Prediction of  $^1\text{H}$  NMR coupling Constants with Associative Neural Networks Trained for Chemical Shifts, *J. Chem. Inf. Model.*, 2007, **47**, 2089.
- 36 *Cambridge Structural Database*, CCDC 2064034, ARABED, DOI: [10.5517/ccdc.csd.cc278sr2](https://doi.org/10.5517/ccdc.csd.cc278sr2), accessed November 2022.
- 37 *Cambridge Structural Database*, CCDC 2064033, ARABAZ, DOI: [10.5517/ccdc.csd.cc278sq1](https://doi.org/10.5517/ccdc.csd.cc278sq1), accessed November 2022.
- 38 Open Innovation Platform Agro screening data supplied by courtesy of BASF SE – used with BASF SE's permission.
- 39 See ESI† for further details.
- 40 Synple Chem, <https://www.synplechem.com/solutions/cartridges>, accessed October 2022.

