Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Digital Discovery*, 2026, **5**, 153

One step retrosynthesis of drugs from commercially available chemical building blocks and conceivable coupling reactions

Babak Mahjour, [a] Felix Katzenburg, [ac] Emil Lammi [b] and Tim Cernak *[ab]

In this report, the pharmaceuticals listed in DrugBank were structurally mapped to a commercial catalog of chemical feedstocks through reaction agnostic one step retrosynthetic decomposition. Enumerative combinatorics was utilized to retrosynthesize target molecules into commercially available building blocks, wherein only the bond formed and the minimal substructure template of each building block class are considered. In contrast to the status quo in automated retrosynthesis, our algorithm may suggest reactions that do not yet exist but, if they did, could enable the synthesis of drugs in just one reaction step from commercial feedstocks. Cross-referencing synthons to commercial datasets can thus reveal valuable reaction classes for development in addition to streamlining drug production. Decomposed synthons were linked to target molecules by transformations that form one bond after the elimination of each synthon's respective reactive functional handle, as indicated by their building block class. Specific reactivities were analyzed after *post hoc* refinement and clustering of commercial synthons. Maps between boronates, bromides, iodides, amines, acids, chlorides, alcohols, and various C–H motifs to form alkyl–alkyl, alkyl–aryl, and aryl–aryl carbon–carbon, carbon–nitrogen, and carbon–oxygen bonds are reported herein, with specific examples for each provided.

Received 14th July 2025
Accepted 17th November 2025

DOI: 10.1039/d5dd00310e

rsc.li/digitaldiscovery

Introduction

Retrosynthesis is a central task in molecular design, and computer-aided synthesis planning has established itself as a valuable tool to create synthetic chemical recipes efficiently.[1–8] The resurgence of automated retrosynthesis is driven by advances in machine learning and the increasing abundance of machine-readable reaction data. Contemporary retrosynthetic algorithms, often trained on large reaction datasets or relying on reaction templates and rules, can efficiently propose plausible routes to disconnect target molecules into accessible chemical feedstocks. When the goal is, however, to explore unknown chemical space or develop more efficient and novel transformations, biases toward well-established chemistries pose a disadvantage.[9–13] Furthermore, medicinal chemistry practices are largely biased towards a very limited subset of reaction types including amide coupling, Suzuki coupling, Buchwald–Hartwig coupling, and Boc-deprotection.[14,15] Enumerative combinatorics, in contrast, can generate hypothetical transformations that have yet to be discovered between substrate classes.[16,17] In reaction targeting, enumerated transformations are evaluated for value and feasibility through a variety of chemoinformatic means such as chemical space access analysis, retrosynthetic value, and property or topological distribution analysis. We have previously explored amine–acid coupling reactions[16,17] and C–H bonds with this style of enumerative combinatorics,[18] and here extend the analysis to other groups of popular chemical feedstocks including alcohols, halides, amines, acids, aldehydes and boronates.

Our enumerative algorithm was designed to score value and feasibility and is agnostic to the mechanism of hypothetical reactions. We consider a reaction valuable if it can utilize a class of commercial building blocks available in high diversity, and provide access to many drug or druglike structures. Fig. 1A shows the prevalence of building blocks within a commercial dataset collected using SMARTS patterns that have been written to encode functional handles commonly used in medicinal chemistry (*e.g.*, amines, halides, *etc.*). C–H bonds, alcohols and acids are among the most prevalent functional groups, so reactions that can leverage these as functional reaction handles for cross-coupling are of high potential impact. To identify concrete examples of reactivity utilizing abundant starting materials, we employed an algorithmic one-step retrosynthetic strategy that decomposes bonds in target molecules, enumerates functional motifs at the clipped bond positions, and then finds exact matches for each "synthon", now converted into a building block within the MilliporeSigma catalog. Retrosynthetic entries are visualized in bulk *via* chord diagrams, where one building block (synthon a) is arrayed along the

[a]*Department of Medicinal Chemistry, University of Michigan, Ann Arbor, MI, USA. E-mail: tcernak@umich.edu*
[b]*Department of Chemistry, University of Michigan, Ann Arbor, MI, USA*
[c]*Organisch-Chemisches Institut, Universität Münster, Münster, Germany*

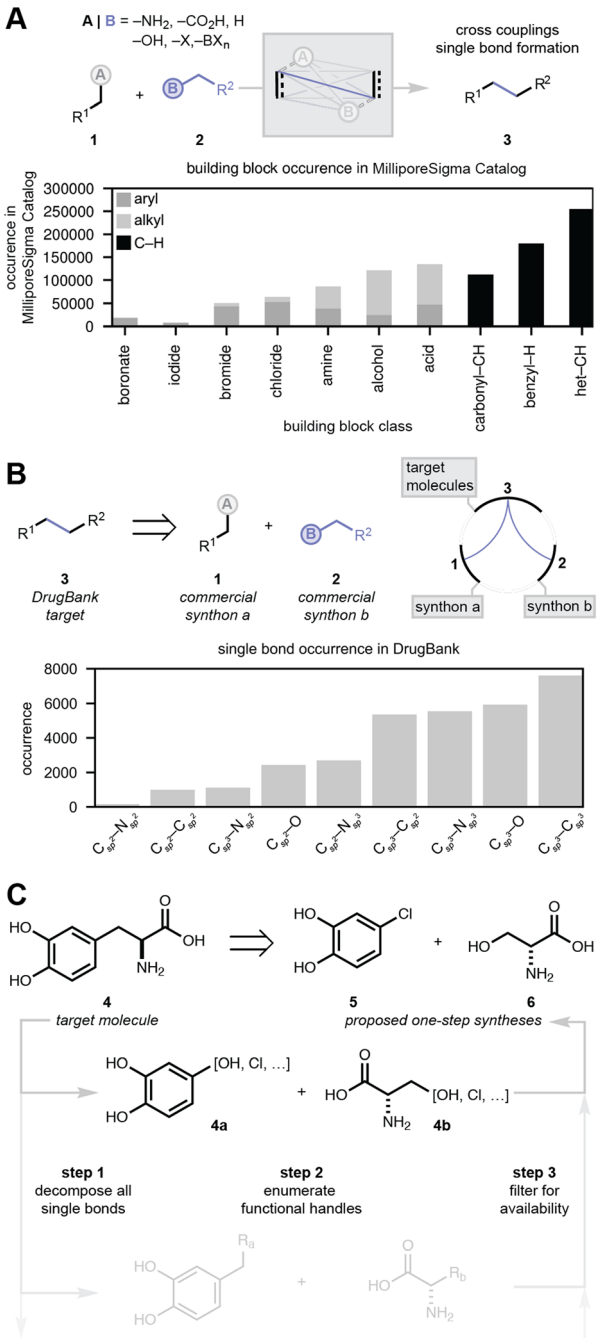

Fig. 1 An enumerative combinatorics analysis of commercial building blocks as one step synthons for reported drug structures. (A) Enumerative combinatorics can identify target bonds between building blocks. The bar chart shows the prevalence of building block classes in the MilliporeSigma catalog (aryl: dark grey, alkyl: light grey, C–H: black). In this study we focus on cross-couplings that form just one bond after the elimination of both building block handles: $^{2|3}A^{\alpha\text{-}A}/^{2|3}A^{\alpha\text{-}B}$ according to ref. 17. (B) Retrosynthesis of the target is represented as a chord diagram where a connecting line means the target can be accessed in one-step by coupling synthon a and synthon b. The bar chart below shows the prevalence of bond types in DrugBank compounds. *X*-axis labels indicate the bond class. The alkyl–alkyl C–C bond is the most common bond type found, followed by the alkyl–alkyl C–N and C–O bonds and then the alkyl–aryl C–C bond. (C) An example of the workflow finding a hit by decomposing L-DOPA (**4**) into commercial starting materials **5** and **6**.

bottom-left arc, the other building block (synthon b) is arrayed along the bottom-right arc, and target molecules are arrayed along the top arc. A chord connecting a synthon and the target molecule indicates that a commercial building block can be used to form the target in one step when merged with a compound found in the other synthon arc. (Fig. 1B).

We focus on the general class of cross-coupling reactions that form a single bond between carbon, nitrogen, or oxygen atoms after the elimination of a building block handle, for instance by coupling **1** and **2** to form structure **3** (Fig. 1B). We also show in Fig. 1B an analysis of the frequency of single bonds that exist within DrugBank. By extending the enumerative combinatorics algorithm to a variety of common building block classes such as amines, acids, alcohols, aldehydes, halides, and boronates, which are popular functional handles in established cross coupling chemistries, we scaffold our feasible reaction space by limiting to coupling modes with some precedent. For example, deaminative and decarboxylative chemistries have been reported, but their tandem reactivity is less established.[16,17,19–23] However, by limiting the specificity of the enumeration to functional handles, we yield a wealth of hypothetical yet realistic reactivities that could maximize the utility of available feedstocks.

Our enumerative combinatorics algorithm uses single or parallel processing to link commercial building blocks to synthetic targets in target databases, in this case DrugBank,[24] *via* hypothetical cross-coupling reactions. An example schematic of the workflow is shown in Fig. 1C. First, each bond in L-DOPA (**4**) is traversed, and each time a single bond is found, the molecule is copied but with that single bond clipped. At the attachment points of the deleted bond, an abstract R group representing common functional handles is placed, as shown in **4a** and **4b**. Then finally, for each combination of enumerated building blocks, a commercial catalog, in this case Aldrich® Market Select from MilliporeSigma[25] is checked for exact matches. For instance, synthon **4a** matches chloride **5** and **4b** matches serine (**6**). The algorithm runs quickly and scales with the number of targets, while remaining invariant to the size of the commercial catalog. In this study, we parallelized the algorithm to decompose all 9082 compounds listed in DrugBank–which were desalted and filtered to those with molecular weights less than 500 g mol^{-1} – into synthons, which were enumerated as building blocks and cross-referenced against the commercial catalog. For both datasets, all duplicates were removed. In the case where two commercial compounds could couple to yield a drug molecule, an entry was recorded. *Post hoc* refinement of the dataset enables more precise analysis of building blocks. For instance, alcohol building block hits can be grouped into primary, secondary, tertiary, or aryl subclasses, and C–H building blocks can be split into benzylic and non-benzylic motifs. The results reveal a wealth of viable reactivities that can form druglike molecules in one step *via* cross-coupling between two commercial substrates, some of which have been previously reported as methodologies. Using our analysis, a total of 2573 of the 9082 drugs (28%) within DrugBank were found to be synthesizable in one step solely using building blocks in the MilliporeSigma catalog. Indeed, all the

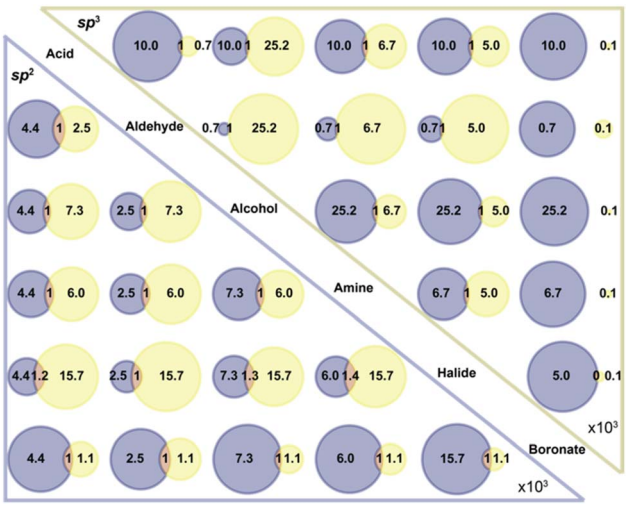

Fig. 2 Comparative analysis of common building block classes after abstraction of functional handles from a commercial catalog. Purple and yellow Venn regions reflect disparate core structure frequencies of occurrence whereas the orange intersection reflects shared core structure occurrence. Bottom left: Venn analysis of shared and disparate core structures of commercially available sp^2 compounds. The values in the purple circles indicate the number of core structures in the building block class above that do not exist in the building block class on the right, and *vice versa* for the values in the yellow circles. The orange overlapping intersection indicates the number of commercially available shared structures between the two classes. Top right: Venn analysis of shared (purple, y-tick labels; yellow, x-tick labels) *versus* disparate core structures (overlapping, orange) of commercially available sp^3 compounds. Occurrence of each functional group in a building block is counted independently.

proposed reactions are single step cross coupling transformations between purchasable compounds, many may not yet be reported with any current methodology but could in principle be developed. While the development of novel reaction methods remains a largely experimental task, but we posit that modern artificial intelligence tools for literature mining and high throughput experimentation[3,26,27] could accelerate the process.

In identifying promising but unrealized reaction methods, a critical metric is the diversity of available building blocks that can undergo these transformations. Indeed, a matched pair analysis of the MilliporeSigma catalog shows that there is in fact little overlap in the availability of building blocks from different families (Fig. 2). It follows that coupling reactions to generate identical chemical linkages will access different product spaces depending on the reactive functional groups employed – for instance, a deaminative–decarboxylative amine–acid coupling from anilines and benzoic acids may generate a compound that couldn't be accessed through Suzuki coupling of commercially available aryl halides and aryl boronates, despite forming the same C–C bond motif.

By grouping the data by building block or bond formed, new and desirable reactivities can be identified. Before the analysis, it was hypothesized that the most valuable couplings would utilize the most abundant building blocks to form the most common types of bonds found in the target dataset. Examples of various retrosynthesis reactions, split by common building block, are shown in Fig. 3.

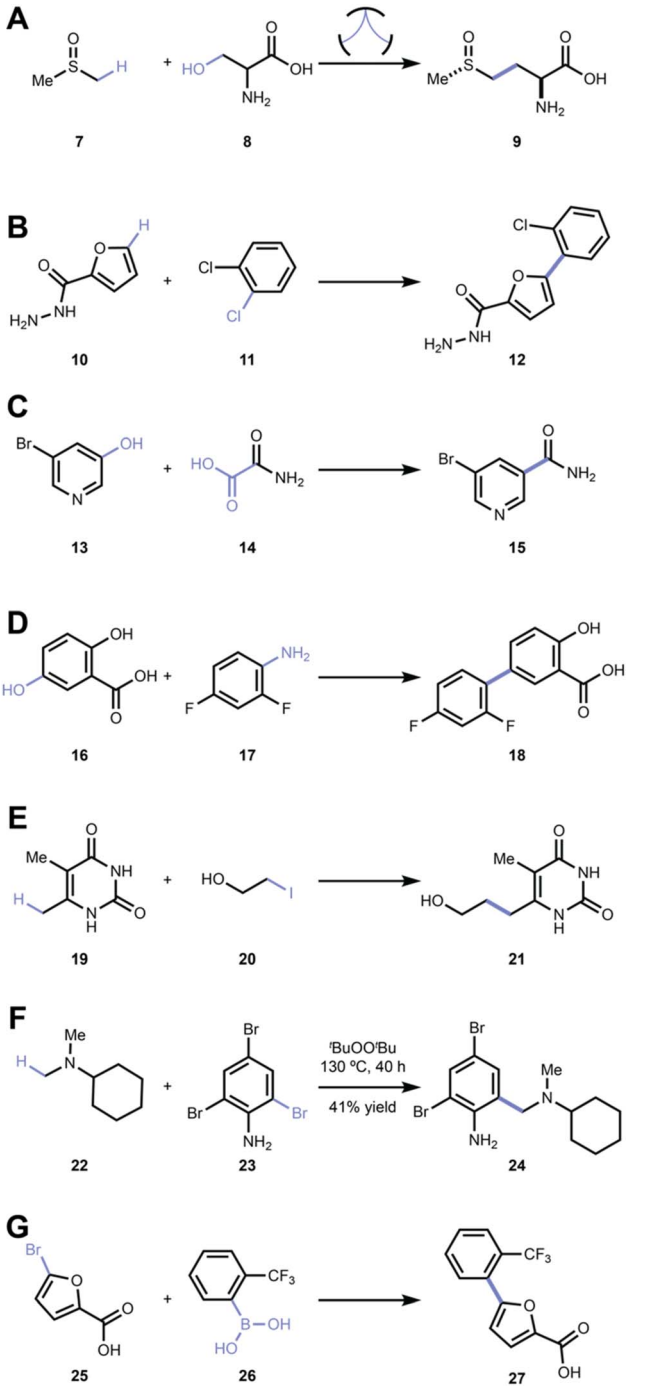

Fig. 3 (A–G) Examples of one step forward syntheses found in the analysis. Each exemplified reaction shows a drug target from DrugBank as the product and two commercially available building blocks from the MilliporeSigma catalog as starting materials. The coupling reaction needed to unite the building blocks to give the product may or may not have been already developed as a reaction method. Each example reaction corresponds to a single grey line in Fig. 4. (f) The synthesis of bromhexine identified in this analysis was previously validated experimentally.[7]

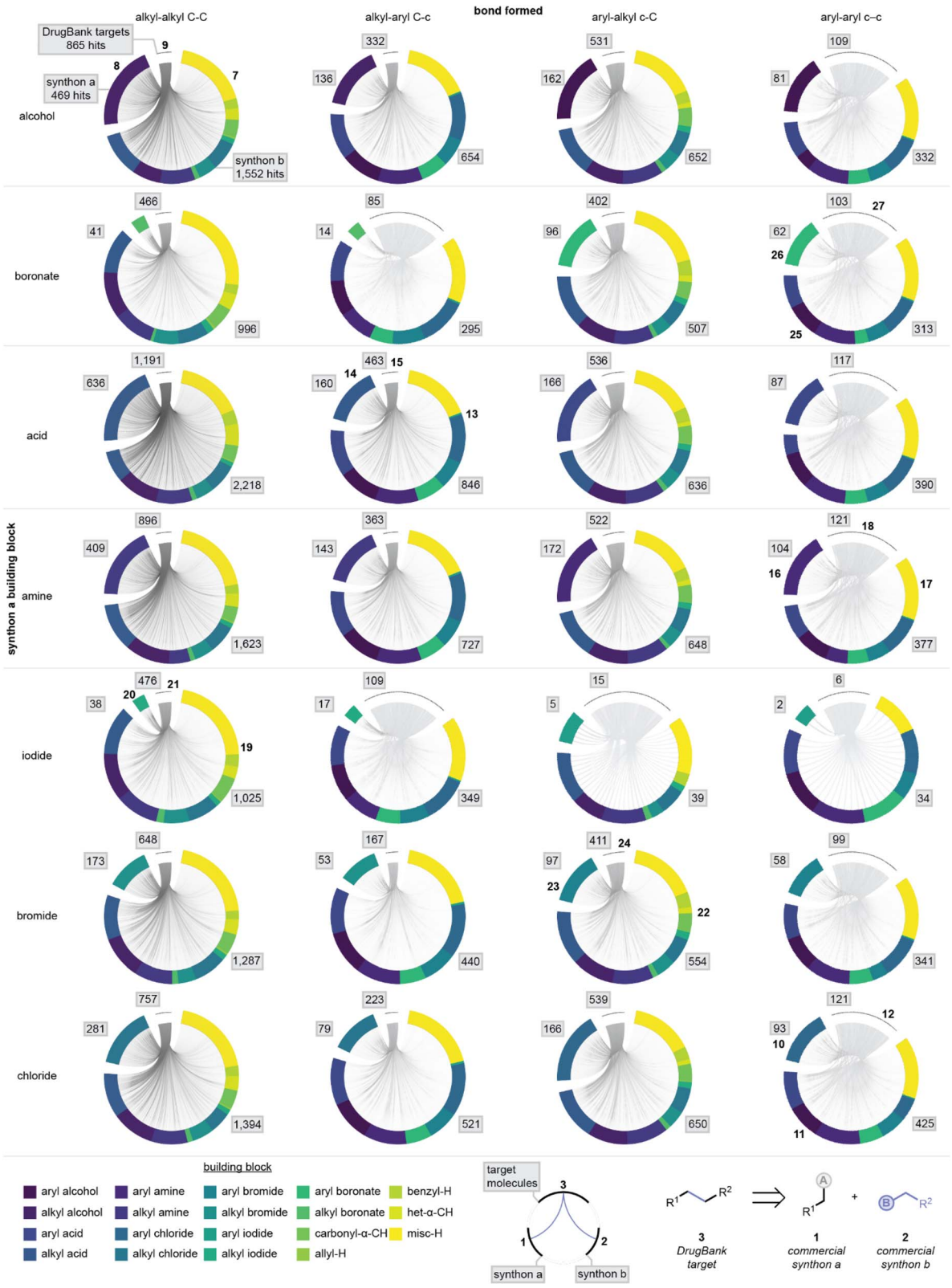

Fig. 4 One-step retrosynthesis maps between DrugBank and purchasable compounds in MilliporeSigma's catalog, trellised by synthon a building block and bond formed. In the first column, both synthons link at an alkyl carbon. In the second column, synthon a, which is the building block group for the row, is alkyl and the remaining synthon b is aryl. In the third column, the building block arc contains aryl synthons and synthon b contains alkyl synthons. In the final column, both synthons are aryl. Compound numbers, in bold, refer to retrosynthetic reactions shown in Fig. 3. An example of a one-step synthesis of bromhexine (**24**), which was experimentally validated,[7] is shown in Fig. 3 and indicated on the bromide aryl–alkyl C–C bond chord plot. Each grey line on the plot corresponds to a single retrosynthesis, such as those shown in Fig. 3. Numbers in grey boxe are the number of targets and synthons, a or b, that yield one-step retrosyntheses.

The analysis is extended to all building block classes and the single bond formation between alkyl and aryl carbon atoms in the trellised Fig. 4. Here, the rows are striated by the building block of synthon b and the columns are split by the bond formed. The length of the rightmost arcs and topmost arcs can be directly compared to gauge the utility of various building block classes in forming a particular bond. Most classes of building blocks were found to have potential use in cross-coupling transformations to form DrugBank compounds. Notably, no aryl iodides were found to form aryl–aryl bonds in compounds found in DrugBank, given our commercial dataset. A complementary functional-group-centric version of Fig. 2–4 that excludes C–H bonds as coupling partners, which we have extensively analyzed previously,[18] is found in the SI. Achieving selectivity in C–H functionalization reactions is a significant challenge and may thwart the ability to realize many proposed retrosynthetic disconnections. We choose to include them in our analysis as there have been many advancements in selective C–H functionalization protocols, particularly in the development of specific catalysts for C–H activation, suggesting that a high value retrosynthesis requiring a selective C–H functionalization could conceivably be achieved through advanced catalyst development.

In the first row of Fig. 4, all dehydroxygenative reactions are considered that form a carbon–carbon bond with another building block to yield a drug in one step. From left to right, the chord diagrams represent the reactions that form an alkyl–alkyl bond, alkyl–aryl bond, and aryl–aryl bond, respectively. 6989 alkyl–alkyl reactions were found to form 865 drugs *via* 469 commercial alcohols (synthon a) and 1552 commercial building blocks (synthon b). An example is shown in Fig. 3A, where DMSO (7) can react with commercial alkyl alcohol **8** to form the drug **9** by a dehydrating C−H functionalization. In the alkyl–aryl chord diagram, 1690 reactions were found to form 332 drugs from 136 commercial alcohols and 654 commercial building blocks. Finally, the aryl–aryl chord diagram reveals 568 reactions that form 109 drugs from 81 commercial alcohols and 332 other commercial building blocks.

In the second row, boronates are considered as building blocks. 2426 alkyl–alkyl reactions were found to form 466 drugs between 41 alkyl boronates and 996 other substrates. As well, 376 alkyl–aryl reactions can form 85 drugs between 14 boronates and 295 other building blocks. Meanwhile, 471 aryl–aryl reactions were found to form 103 drugs using 62 aryl boronates and 313 other aryl compounds, such as the hypothetical Suzuki reaction to form **27** from **25** and **26**.

In the alkyl–alkyl carboxylic acid analysis, 9973 reactions were identified to form 1191 drugs through the coupling of 636 commercial acids and 2218 other building blocks. Likewise in the alkyl–aryl regime, 2321 reactions form 463 drugs between 160 acids and 846 other commercial substrates, as shown by the formation of drug **15** *via* the coupling of phenol **13** and alkyl acid **14**. Finally, 638 aryl–aryl reactions were found to form 117 drugs between 87 aryl acids and 390 aryl building blocks.

Deaminative cross-couplings are identified in the fourth row. 6827 deaminative alkyl–alkyl reactions were found to form 896 drugs when reacting 409 alkyl amines with 1623 other alkyl

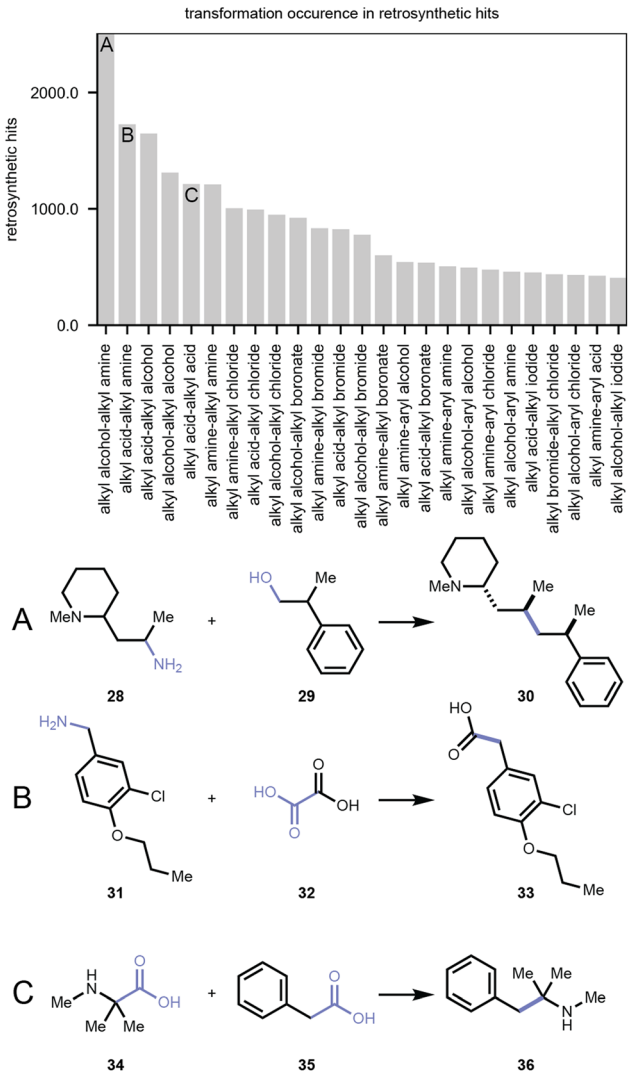

Fig. 5 The top 25 reactions between building block classes identified in the analysis. For the top hit, 2510 retrosynthetic reactions between alkyl alcohols and alkyl amines were found. Three exemplar reactions are shown in A–C.

building blocks. 1827 alkyl–aryl reactions form 363 drugs with 143 amines and 727 building blocks. Lastly, 673 reactions were found to form 121 drugs between 104 aryl amines and 377 aryl building blocks, such as the formation of drug **18** from phenol **16** and aryl amine **17**.

In the deiodinative row, only alkyl–alkyl and alkyl–aryl reactions were found. 2225 alkyl–alkyl reactions can form 476 drugs using 36 iodides and 1025 other substrates, such as the formation of **21** from **19** and **20**. 495 alkyl–aryl reactions were also found, forming 109 drugs between 17 iodides and 349 other building blocks.

Bromides prove to be a versatile building block, with 4397 alkyl–alkyl reactions forming 648 drugs between 173 bromides and 1287 synthon a building blocks. Likewise, 771 alkyl–aryl reactions were identified to form 167 drugs using 53 bromides and 440 other compounds. An example of an aryl–alkyl carbon–carbon bond formed found in this analysis that directly forms

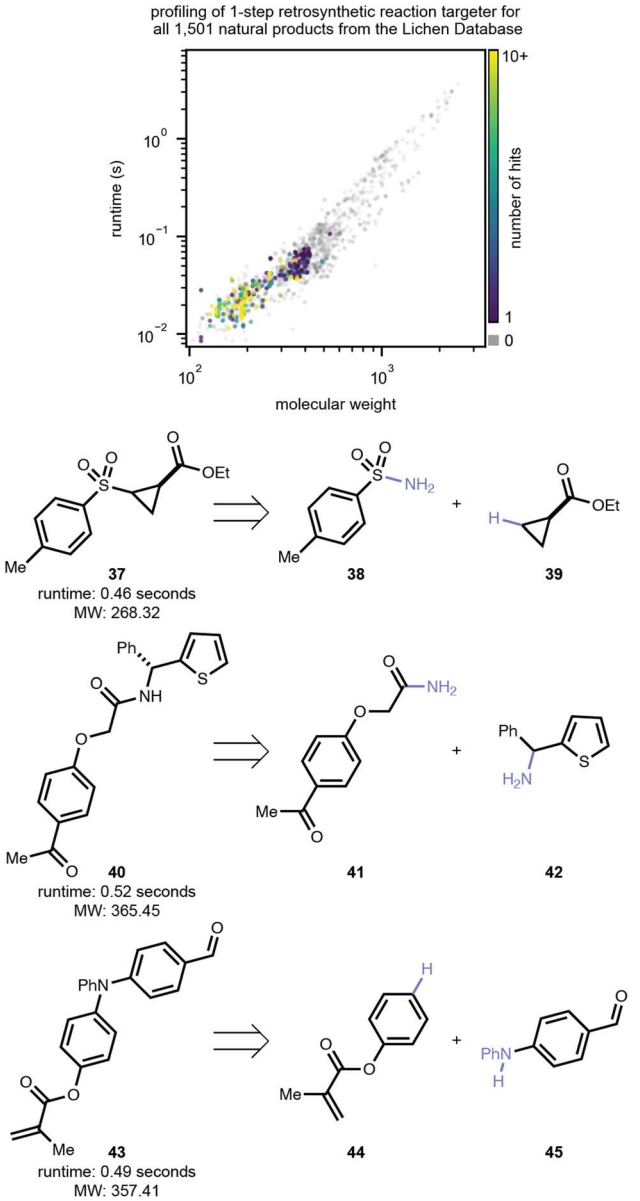

Fig. 6 Natural products from the Coconut Lichen Natural Product database were assessed for one-step retrosynthetic accessibility and to demonstrate the algorithm's runtime and performance. Notably, this analysis exhaustively cross-references a commercial dataset of 399 350 compounds. All hits were found in less than 1 second, and only for the largest molecules over 1000 Da did the runtime exceed ~1–2 seconds, despite the exhaustiveness of the enumeration, which guarantees all solutions to be found. Examples of retrosynthetic hits are shown below. Grey dots are experiments ran with no hits found.

a drug in one step is shown as the previously reported coupling of **22** and **23** to form bromhexine (**24**).[7] Finally, 488 aryl–aryl reactions using 58 bromides and 341 other substrates were identified to form 99 drugs.

In the final row, chloride building blocks are analyzed. In all four instances the chloride-based building blocks can access more target drugs than corresponding iodide and bromide-based building blocks. In the alkyl–alkyl analysis, 5126 dechlorinative reactions were found to form 757 drugs in one step using 281 commercial chlorides and 1394 commercial building blocks. In the alkyl–aryl analysis, 1037 reactions form 223 drugs using 79 chlorides and 521 other building blocks. Finally, the aryl–aryl chord diagram reveals 681 reactions to form 121 drugs using 93 aryl chlorides and 425 other aryl building blocks, such as the reaction to form **12** from furan **10** and chloride **11**.

Retrosynthetic reactions were clustered by building block classes of both synthon a and synthon b to gauge the most valuable reactivities based on transformation occurrence in one-step retrosynthesis of drug targets, and the resulting histogram is shown in Fig. 5. Example retrosynthetic reactions for 3 of the top 25 scoring building block classes include the coupling of alkyl amine **28** with alkyl alcohol **29** to form **30**, the deaminative and decarboxylative coupling of **31** and **32** to form **33**, and the double decarboxylative coupling of **34** and **35** to form phenethylamine **36.** This latter reaction class, **C**, coupling two alkyl carboxylic acids, corresponds to the Kolbe coupling reaction, a classic reaction that has recently seen updating to enable mild electrochemical[28,29] and metalphotoredox-catalyzed[30] transformations for complex molecule diversification. Meanwhile, amine–acid couplings to make C–C bonds, class **B**, have been a focus of our lab.[19–21]

As a final case study to showcase the algorithm's performance in natural product chemical space, we profile the outcome and runtime when running the algorithm against all 1501 natural products collected in the Coconut Lichen Natural Product database.[31] In this case study, the full non-filtered commercial catalog of about 400 000 compounds was used. The result of this analysis is shown in Fig. 6.

When assessing the time to completion for each of the 1501 natural products, no analysis took over ~10 seconds, and the vast majority of compounds and all hits were found in under one second. To identify reactions that synthesize products that cannot be formed with known chemistry, each product with a hit was assessed with two complementary retrosynthesis machine learning models.[32,33] Products that could be formed in a single step using known chemistry were then filtered out. Since we only allow a one-step retrosynthesis, few targets with a molecular weight >500 g mol^{-1} led to viable disconnections into commercially available building blocks, although many targets smaller than 400 g mol^{-1} were successfully disconnected. Three example disconnections are shown below the plot of Fig. 6. While products **37**, **40**, and **43** are not known to be synthesizable with a single reaction, our workflow reveals that they can be formed from commercial compounds should the highlighted transformation be developed. A deaminative coupling to form **37** from sulfonamide **38** and C–H substrate **39** was found in 0.46 seconds. A transamidation to form **40** from **41** and **42** can be approached from deamination of either substrate, as both purchasable compounds have a primary amine at the desired reactive site. Notably the carboxylic acid congener of **41** that one would use with **42** to reach the desired amide bond was not commercially available in the vendor catalog we used. Finally, a C–H amination between **44** and **45** forms **43**, which was identified in 0.43 seconds.

Conclusion

A method to rapidly form single-bond cross coupling maps between target molecules and commercial catalogs is reported and showcased *via* an analysis of compounds in DrugBank or the Coconut Lichen Natural Product database and the commercial catalog provided by MilliporeSigma. Coupling maps are grouped and filtered to reveal various perspectives of reaction targeting. Several reactivities that have been previously reported were showcased as transformations found in the analysis, and examples of prospective chemistries are shared. Mining maps of coupling transformations will continue to reveal valuable reactions to target for synthetic chemistry method development, and additional heuristics to organize the results by estimated cost, scalability, and proposed reaction conditions will continue to refine and strengthen such workflows. While many of the reactivities highlighted here may be currently unfeasible or cannot be achieved with acceptable regioselectivity using modern methods, many of the most popular reaction methods in use today did not exist two decades ago. Future advances in synthetic chemistry and catalysis will continue to push the boundaries of what is possible, and the reaction targeting method disclosed here is one method to guide the hunt for new reactivities and synthesis strategies.

Conflicts of interest

The Cernak Lab has received research funding or in-kind donations from MilliporeSigma, Relay Therapeutics, Janssen Therapeutics, AbbVie, SPT Labtech and Merck & Co., Inc. T. C. is a co-Founder and equity holder of Iambic Therapeutics, Inc.

Data availability

Code and data used and collected in this study is available at https://github.com/cernak-lab/one-step-retro (DOI: https://www.doi.org/10.5281/zenodo.17544767). Details regarding access to and download of the commercial catalog are available upon request from MilliporeSigma (https://www.sigmaaldrich.com).

Supplementary information is available. See DOI: https://doi.org/10.1039/d5dd00310e.

Acknowledgements

This work was supported with start-up funds from the University of Michigan College of Pharmacy and the National Science Foundation (CHE-2236215 to T. C.). B. M. acknowledges a predoctoral fellowship from the ACS MEDI division. F. K. thanks the Deutsche Forschungsgemeinschaft (SPP 2363) and the Deutscher Akademischer Austauschdienst for scholarships.

References

1 E. J. Corey, X.-M. Cheng, *The Logic of Chemical Synthesis*, 1989.
2 Y. Shen, J. E. Borowski, M. A. Hardy, R. Sarpong, A. G. Doyle and T. Cernak, Automation and Computer-Assisted Planning for Chemical Synthesis, *Nat. Rev. Methods Primers*, 2021, **1**(1), 23, DOI: 10.1038/s43586-021-00022-5.
3 E. Shim, J. A. Kammeraad, Z. Xu, A. Tewari, T. Cernak and P. M. Zimmerman, Predicting Reaction Conditions from Limited Data through Active Transfer Learning, *Chem. Sci.*, 2022, **13**(22), 6655–6668, DOI: 10.1039/D1SC06932B.
4 S. Szymkuć, E. P. Gajewska, T. Klucznik, K. Molga, P. Dittwald, M. Startek, M. Bajczyk and B. A. Grzybowski, Computer-Assisted Synthetic Planning: The End of the Beginning, *Angew. Chem., Int. Ed.*, 2016, **55**(20), 5904–5937, DOI: 10.1002/anie.201506101.
5 S. F. Kim, C. Amber, G. L. Bartholomew and R. Sarpong, Skeletal Editing Strategies Driven by Total Synthesis, *Acc. Chem. Res.*, 2025, **58**(11), 1786–1800, DOI: 10.1021/acs.accounts.5c00174.
6 Y. Lin, R. Zhang, D. Wang and T. Cernak, Computer-Aided Key Step Generation in Alkaloid Total Synthesis, *Science*, 2023, **379**(6631), 453–457, DOI: 10.1126/science.ade8459.
7 Y. Lin, Z. Zhang, B. Mahjour, D. Wang, R. Zhang, E. Shim, A. McGrath, Y. Shen, N. Brugger, R. Turnbull, S. Trice, S. Jasty and T. Cernak, Reinforcing the Supply Chain of Umifenovir and Other Antiviral Drugs with Retrosynthetic Software, *Nat. Commun.*, 2021, **12**(1), 7327, DOI: 10.1038/s41467-021-27547-3.
8 W. R. Gutekunst and P. S. Baran, C–H Functionalization Logic in Total Synthesis, *Chem. Soc. Rev.*, 2011, **40**(4), 1976–1991, DOI: 10.1039/C0CS00182A.
9 Y. Jiang, Y. Yu, M. Kong, Y. Mei, L. Yuan, Z. Huang, K. Kuang, Z. Wang, H. Yao, J. Zou, C. W. Coley and Y. Wei, Artificial Intelligence for Retrosynthesis Prediction, *Engineering*, 2023, **25**(6), 32–50, DOI: 10.1016/j.eng.2022.04.021.
10 M. H. Lin, Z. Tu and C. W. Coley, Improving the Performance of Models for One-Step Retrosynthesis through Re-Ranking, *J. Cheminf.*, 2022, **14**(1), 15, DOI: 10.1186/s13321-022-00594-8.
11 T.-L. Phan, K. Weinbauer, M. E. G. Laffitte, Y. Pan, D. Merkle, J. L. Andersen, R. Fagerberg, C. Flamm and P. F. Stadler, SynTemp: Efficient Extraction of Graph-Based Reaction Rules from Large-Scale Reaction Databases, *J. Chem. Inf. Model.*, 2025, **65**(6), 2882–2896, DOI: 10.1021/acs.jcim.4c01795.
12 C. W. Coley, L. Rogers, W. H. Green and K. F. Jensen, Computer-Assisted Retrosynthesis Based on Molecular Similarity, *ACS Cent. Sci.*, 2017, **3**(12), 1237–1245, DOI: 10.1021/acscentsci.7b00355.
13 C. Li and R. A. Shenvi, Total Synthesis of 25 Picrotoxanes by Virtual Library Selection, *Nature*, 2025, **638**(8052), 980–986, DOI: 10.1038/s41586-024-08538-y.
14 N. Gesmundo, K. Dykstra, J. L. Douthwaite, Y. Kao, R. Zhao, B. Mahjour, R. Ferguson, S. Dreher, B. Sauvagnat, J. Saurí and T. Cernak, *Nat. Synth.*, 2023, **2**(11), 1082–1091.
15 J. Boström, D. G. Brown, R. J. Young and G. M. Keserü, Expanding the Medicinal Chemistry Synthetic Toolbox, *Nat. Rev. Drug Discovery*, 2018, **17**(10), 709–727.
16 B. Mahjour, Y. Shen, W. Liu and T. Cernak, A Map of the Amine–Carboxylic Acid Coupling System, *Nature*, 2020, **580**(7801), 71–75, DOI: 10.1038/s41586-020-2142-y.

17 R. Zhang, B. Mahjour, A. Outlaw, A. McGrath, T. Hopper, B. Kelly, W. P. Walters and T. Cernak, *Commun. Chem.*, 2024, **7**(1), 22.
18 B. Mahjour, K. M. Flynn, S. S. Stahl and T. Cernak, Impact of C–H Cross-Coupling Reactions in the One-Step Retrosynthesis of Drug Molecules, *J. Org. Chem.*, 2024, **89**(21), 15387–15392, DOI: 10.1021/acs.joc.4c01567.
19 A. McGrath, H. Huang, J.-F. Brazeau, Z. Zhang, C. O. Audu, N. A. Vellore, L. Zhu, Z. Shi, J. D. Venable, C. F. Gelin and T. Cernak, Modulating the Potency of BRD4 PROTACs at the Systems Level with Amine-Acid Coupling Reactions, *J. Med. Chem.*, 2025, **68**(1), 405–420, DOI: 10.1021/acs.jmedchem.4c02047.
20 J. L. Douthwaite, R. Zhao, E. Shim, B. Mahjour, P. M. Zimmerman and T. Cernak, Formal Cross-Coupling of Amines and Carboxylic Acids to Form Sp3–Sp2 Carbon–Carbon Bonds, *J. Am. Chem. Soc.*, 2023, **145**(20), 10930–10937, DOI: 10.1021/jacs.2c11563.
21 Z. Zhang and T. Cernak, The Formal Cross-Coupling of Amines and Carboxylic Acids to Form Sp^3–Sp^3 Carbon–Carbon Bonds, *Angew. Chem., Int. Ed.*, 2021, **60**(52), 27293–27298, DOI: 10.1002/anie.202112454.
22 Y. Shen, B. Mahjour and T. Cernak, Development of Copper-Catalyzed Deaminative Esterification Using High-Throughput Experimentation, *Commun. Chem.*, 2022, **5**(1), 83, DOI: 10.1038/s42004-022-00698-0.
23 A. McGrath, R. Zhang, K. Shafiq and T. Cernak, Repurposing Amine and Carboxylic Acid Building Blocks with an Automatable Esterification Reaction, *Chem. Commun.*, 2023, **59**(8), 1026–1029, DOI: 10.1039/D2CC05670D.
24 D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox and M. Wilson, DrugBank 5.0: A Major Update to the DrugBank Database for 2018, *Nucleic Acids Res.*, 2018, **46**(D1), D1074–D1082, DOI: 10.1093/nar/gkx1037.
25 Aldrich Market Select, Filtered to Those with Molecular Weight Less than 200 g/Mol, https://www.aldrichmarketselect.com/, accessed 2024-01-17.
26 B. Mahjour, R. Zhang, Y. Shen, A. McGrath, R. Zhao, O. G. Mohamed, Y. Lin, Z. Zhang, J. L. Douthwaite, A. Tripathi and T. Cernak, Rapid Planning and Analysis of High-Throughput Experiment Arrays for Reaction Discovery, *Nat. Commun.*, 2023, **14**(1), 3924, DOI: 10.1038/s41467-023-39531-0.
27 B. Mahjour, Y. Shen and T. Cernak, Ultrahigh-Throughput Experimentation for Information-Rich Chemical Synthesis, *Acc. Chem. Res.*, 2021, **54**(10), 2337–2346, DOI: 10.1021/acs.accounts.1c00119.
28 Y. Hioki, M. Costantini, J. Griffin, K. C. Harper, M. P. Merini, B. Nissl, Y. Kawamata and P. S. Baran, Overcoming the Limitations of Kolbe Coupling with Waveform-Controlled Electrosynthesis, *Science*, 2023, **380**(6640), 81–87, DOI: 10.1126/science.adf4762.
29 B. Zhang, Y. Gao, Y. Hioki, M. S. Oderinde, J. X. Qiao, K. X. Rodriguez, H.-J. Zhang, Y. Kawamata and P. S. Baran, Ni-Electrocatalytic Csp3–Csp3 Doubly Decarboxylative Coupling, *Nature*, 2022, **606**(7913), 313–318, DOI: 10.1038/s41586-022-04691-4.
30 A. V. Tsymbal, L. D. Bizzini and D. W. C. MacMillan, Nickel Catalysis via SH2 Homolytic Substitution: The Double Decarboxylative Cross-Coupling of Aliphatic Acids, *J. Am. Chem. Soc.*, 2022, **144**(46), 21278–21286, DOI: 10.1021/jacs.2c08989.
31 V. Chandrasekhar, K. Rajan, S. R. S. Kanakam, N. Sharma, V. Weißenborn, J. Schaub and C. Steinbeck, COCONUT 2.0: A Comprehensive Overhaul and Curation of the Collection of Open Natural Products Database, *Nucleic Acids Res.*, 2025, **53**(D1), D634–D643, DOI: 10.1093/nar/gkae1063.
32 M. E. Fortunato, C. W. Coley, B. C. Barnes and K. F. Jensen, Data Augmentation and Pretraining for Template-Based Retrosynthetic Prediction in Computer-Aided Synthesis Planning, *J. Chem. Inf. Model.*, 2020, **60**(7), 3398–3407, DOI: 10.1021/acs.jcim.0c00403.
33 I. V. Tetko, P. Karpov, R. Van Deursen and G. Godin, State-of-the-Art Augmented NLP Transformer Models for Direct and Single-Step Retrosynthesis, *Nat. Commun.*, 2020, **11**(1), 5575, DOI: 10.1038/s41467-020-19266-y.