

Impact of C–H Cross-Coupling Reactions in the One-Step Retrosynthesis of Drug Molecules

Babak Mahjour, Kaitlyn M. Flynn, Shannon S. Stahl,* and Tim Cernak*



Cite This: *J. Org. Chem.* 2024, 89, 15387–15392



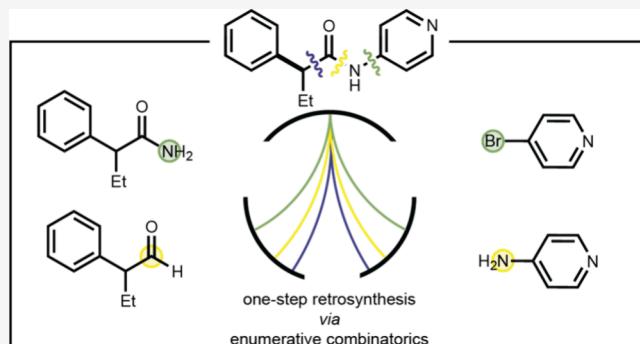
Read Online

ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: Pharmaceutical synthesis requires a diversity of chemical reactions. The discovery of new reactions enable novel retrosynthetic disconnections, potentially expediting access to complex molecules. This Synopsis demonstrates the use of enumerative combinatorics to find impactful underdeveloped reactions for drug synthesis. By mapping pharmaceutical target molecules onto available building blocks using just one retrosynthetic disconnection even if the requisite reaction is not yet known, we highlight the importance of site-selective C–H cross-coupling methods. This cheminformatics-driven retrosynthetic analysis identifies novel reaction methods of value to the synthesis toolbox.



Modern computer-aided synthesis enables efficient identification of retrosynthetic routes to small molecule targets.¹ This approach facilitates the design of synthetic routes to complex molecules, but the algorithms generally rely on established reaction methods.² Enumerative combinatorics is a technique that can explore all possible bond formations for a given pair of building blocks, without requiring that each coupling reaction maps onto a known synthetic method (Figure 1a). This approach has been used recently to demonstrate the array of coupling products that may be accessed from amines and carboxylic acids, extending far beyond conventional amides.^{3,4} In this Synopsis, we show how a similar approach may be adapted for the retrosynthesis of pharmaceuticals, highlighting opportunities for development of reactions that provide more efficient access to drug molecules. The analysis considers one-step retrosynthetic disconnections of DrugBank compounds into commercially available building blocks via formation of a single-bond. Activated C(sp³)–H bonds, including those adjacent to carbonyl, aromatic, and heteroatom groups, are good reactive handles and prominently featured at disconnection sites in the building blocks, highlighting the potential utility of new C(sp³)–H cross-coupling reactions in pharmaceutical synthesis. More broadly, this analysis shows how emerging cheminformatics tools may be used to identify synthetic methods that could have strategic impact in the preparation of drug molecules and analogs thereof.

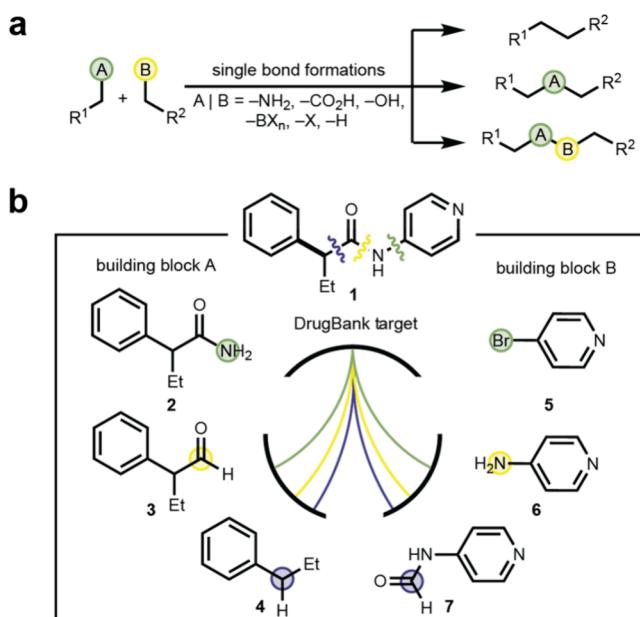


Figure 1. Introduction to enumerative combinatorics for retrosynthesis. (a) Enumerative combinatorics for the formation of a single bond between two coupling partners. (b) Adaptation of enumerative combinatorics for retrosynthetic analysis.

Received: June 21, 2024

Revised: September 21, 2024

Accepted: October 4, 2024

Published: October 14, 2024

RETROSYNTHETIC ANALYSIS OF DRUG MOLECULES USING ENUMERATIVE COMBINATORICS

The concepts highlighted herein take advantage of two available databases: drug compounds from the DrugBank database⁵ and commercially accessible compounds from MilliporeSigma's Aldrich Market Select⁶ catalog. An example retrosynthetic disconnection of a DrugBank compound into commercially available synthons is shown in Figure 1b. Through a single bond formation, drug **1** can be directly accessed from three different pairs of commercially available compounds – **2 + 5**, **3 + 6**, or **4 + 7**. A simple way to represent the array of viable retrosyntheses is through the use of chord diagrams (Figure 1b, center), in which one synthon (building block A) is arrayed along the bottom-left arc, the other synthon (building block B) is arrayed along the bottom-right arc, and the target molecule(s) are arrayed along the top arc. A chord between a synthon and the target molecule indicates that the synthon can be used to form the target in one step when united with a compound found in the other synthon arc. “Above-the-arrow” reaction conditions⁷ associated with individual disconnections may be known, if it maps onto a published protocol, or it may reflect a conceivable or aspirational reaction method that has not yet been reported.

Analysis of the DrugBank compounds reveal the frequency of occurrence of single-bond types and show that sp^3 – sp^3 bonds are more prevalent than sp^2 – sp^2 bonds (Figure 2a). This outcome is noteworthy, given the prevalence of sp^2 – sp^2 coupling methods routinely used in drug discovery and pharmaceutical synthesis.^{8,9} Viewed through a lens of one-step retrosynthesis, this observation highlights the importance of reactions that create sp^3 – sp^3 bonds and the priority of new reactions that functionalize sp^3 -hybridized carbon atoms when targeting drugs and drug-like chemical space.

Here, we demonstrate an algorithm that disconnects DrugBank molecules into two synthons. The DrugBank molecules were first filtered to be limited to those with a molecular weight less than 500 g/mol (after desalting in the RDKit python package). In this process, every carbon–carbon, carbon–nitrogen, and carbon–oxygen single bond was deleted and replaced with generic functional group placeholders on both ends of the original bond (cf. A and B in Figure 1a). Subsequently, building block functional groups (acid, alcohol, boronate, amine, iodine, bromine, chlorine, and hydrogen) were enumerated at the disconnection point of both synthons (A or B). The full combination of both synthon enumerations was compiled, and each pair was cross-referenced against the 114,300 commercial compounds found in MilliporeSigma's Aldrich Market Select⁶ catalog, which had been previously filtered to those synthons with a molecular weight less than 200 g/mol. If both synthons were found to be purchasable, the retrosynthetic reaction was recorded. Once all possible single-step retrosynthetic reactions between drugs and purchasable compounds were recorded, posthoc refinement was performed to group the reactions into synthetically relevant bins. For instance, each transformation containing a synthon with a functional handle other than hydrogen (e.g., deaminative, decarboxylative, etc.) was grouped as an aryl or nonaryl coupling reaction (by checking the α -carbon to be equal to a “c” or “C” in SMARTS notation, respectively). Synthons containing C–H bonds were separated into activated and nonactivated compounds. In the present analysis, “activated” C–H bonds include C(sp^3)–H bonds adjacent to a (hetero)arene (i.e.,

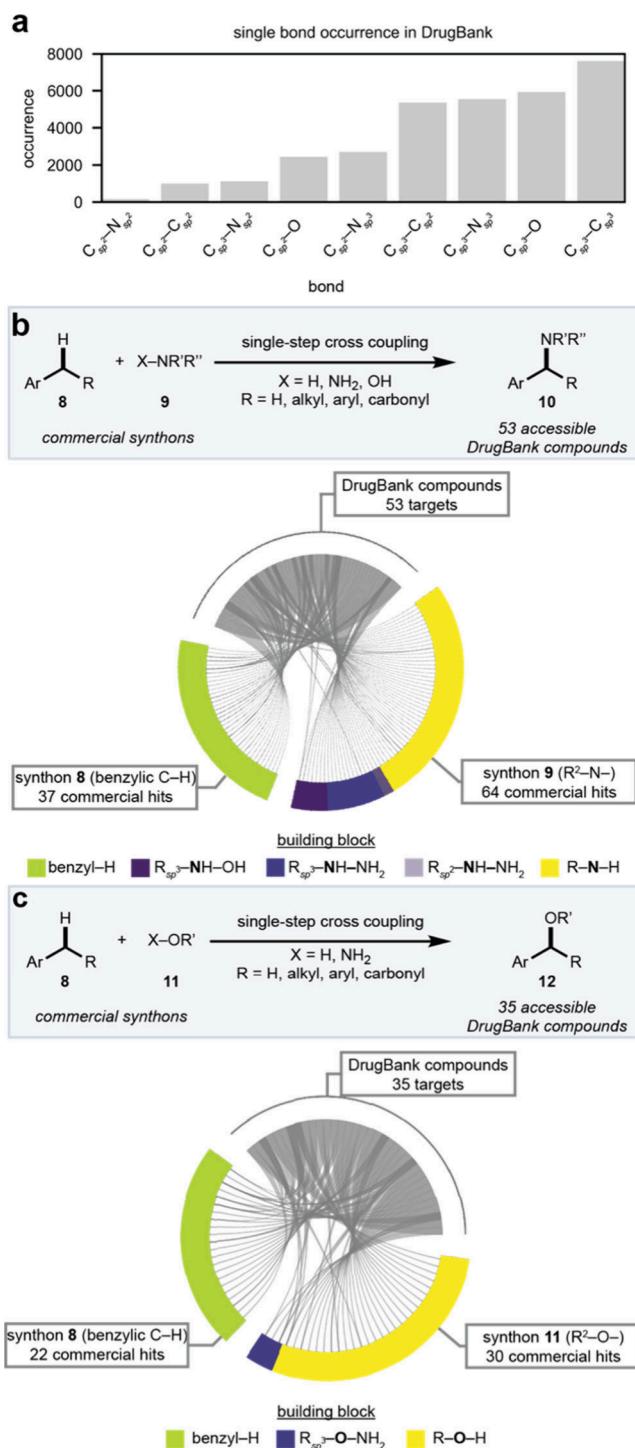


Figure 2. (Hetero)benzylic C–N and C–O retrosynthetic analysis. (a) Number of drugs among 9,082 DrugBank molecules containing an instance of different types of single bonds. (b) All hypothetical single-bond C–H cross-couplings between (hetero)benzylic carbon atoms and nitrogen atoms to form DrugBank compounds. Fifty-three drugs could be synthesized in one step from commercial synthons. (c) All hypothetical single-bond C–H cross-couplings between (hetero)benzylic carbon atoms and oxygen atoms to form DrugBank compounds. Thirty-five drugs could be synthesized in one step from commercial synthons.

benzylic), a carbonyl group, or a heteroatom. The remaining C–H synthons, including aryl and unactivated aliphatic hydrogens, were labeled “miscellaneous”. To produce chord diagrams, all

retrosynthetic transformations were collected, and the synthons containing the functional group of interest were defined as “synthon a” with their corresponding reaction partners described as “synthon b”. Each target drug would then have two collections of chords: one set drawn to the collection of synthon a molecules and the other to the synthon b molecules.

We then used this one-step retrosynthetic analysis of DrugBank compounds to evaluate drug molecules with (hetero)benzylic C–N or C–O bonds to assess the frequency with which they could be disconnected into (hetero)benzylic C–H synthons and N- or O-derived coupling partners. This analysis revealed that 53 compounds could be directly accessed through benzylic C–N bond formation ($8 + 9 \rightarrow 10$, Figure 2b), while 35 could be accessed through benzylic C–O bond formation ($8 + 11 \rightarrow 12$, Figure 2c). Among these 88 disconnections, most of the nitrogen or oxygen motifs would be sourced from the parent amine or alcohol, with a small number of disconnections using deaminative chemistry from *O*-alkyl hydroxylamines or hydrazines, or deoxygenative chemistry from *N*-linked hydroxylamines, to accomplish the coupling reaction.

BUILDING BLOCKS AND EXISTING COUPLING REACTIONS RELEVANT TO ONE-STEP SYNTHESSES OF DRUG MOLECULES

Coupling partners with halide, boronic acid, alcohol, carboxylic acid, and amine functional groups are well-established synthons in modern chemical synthesis.¹⁰ In recent years, activated C–H building blocks have been increasingly used for site-selective bond formation.¹¹ A survey of the different building blocks (Figure 3a) shows that the commercial availability of compounds with activated C(*sp*³)–H bonds exceeds even the most abundant coupling partners, such as alcohols, carboxylic acids and amines. Focusing on the (hetero)benzylic C(*sp*³)–H motif, there is relatively little overlap between these commercially available structures and the corresponding benzyl amines, halides, and alcohols (Figure 3b). A Venn diagram shows that of the 178,144 available benzylic C–H compounds, only 13,796 examples have a corresponding “traditional” cross-coupling synthon analog (i.e., a benzyl amine, halide, or alcohol). This analysis highlights the wealth of valuable chemical space that could be accessed through benzylic or other activated C–H cross-coupling reactions.¹²

There has been a substantial expansion of chemistry utilizing diverse classes of activated C–H bonds. Figure 3c shows some recent examples of benzylic C–H functionalization to form C–N and C–O bonds.^{13,14} For instance, heterobenzylic C–H compound 13 was aminated with 14 to give 15 in 49% yield (as a 7.2:1 mixture of regioisomers) under copper catalysis with *N*-fluorosulfonyl imide (NFSI) as an oxidant.¹⁰ Meanwhile, related conditions leveraging the Biox ligand allowed the smooth formation of benzylic ether 18 from 16 and 17 in 87% yield.¹¹ These examples, along with other transformations from benzylic C–H bonds, could enable access to vast libraries of drug-like compounds.

ASPIRATIONAL COUPLING REACTIONS RELEVANT TO ONE-STEP SYNTHESSES OF DRUG MOLECULES

Extending this analysis of C–H building blocks that could be used in one-step retrosynthesis highlights the importance of C–H cross-coupling reactions that prioritize C–C bond formation (Figure 4). Enolate precursors (α -carbonyl) are featured in 428

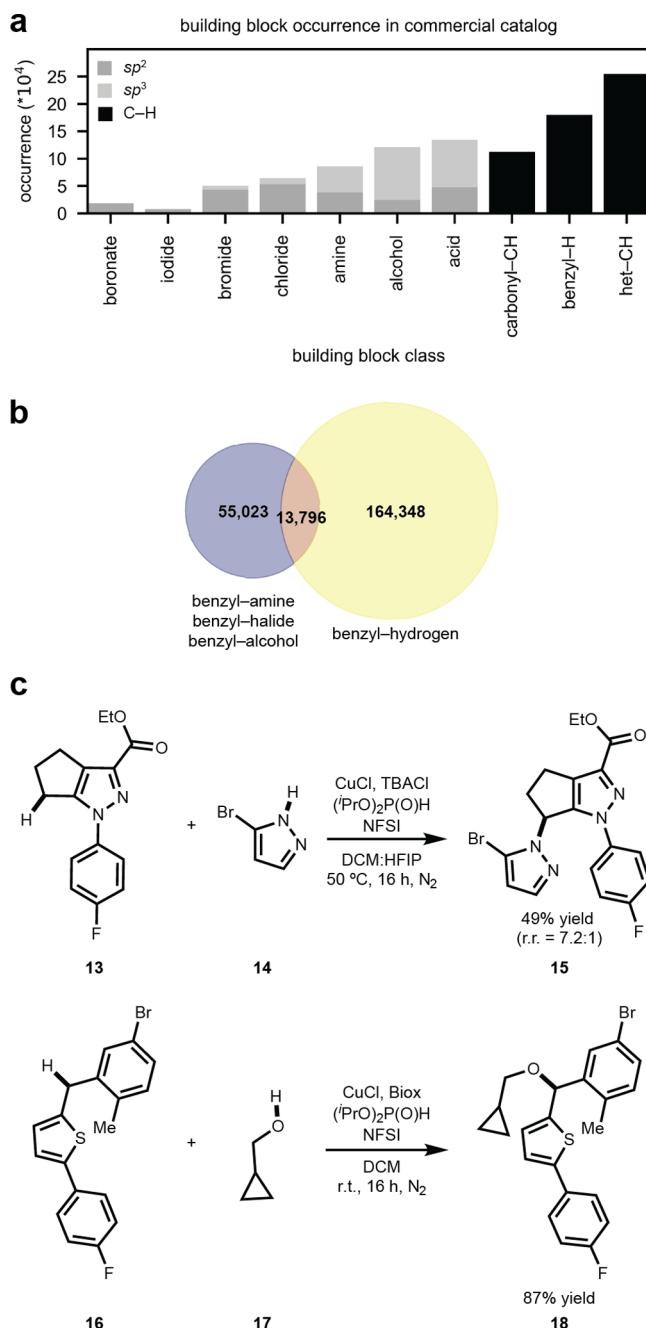


Figure 3. C–H motifs represent valuable building blocks for cross-coupling. (a) Commercial availability of different coupling partners. (b) Comparison of commercially available benzylic building blocks. (c) Examples of benzylic C–H functionalization used to access drug-like compounds.

disconnections of DrugBank compounds: 361 forming C(*sp*³)–C(*sp*³) bonds and 67 forming C(*sp*³)–C(*sp*²) bonds. Many of these reactions should be feasible through established enolate chemistry. Looking to (hetero)benzylic C–H synthons reveals that there are an additional 264 disconnections when extending the study to C–C bond formation. This analysis highlights the significant value that could be gained from development of site-selective C–H benzylic alkylation or arylation with each of the activated C–H coupling partners.¹⁵ C–H bonds adjacent to a heteroatom (N, O, S, etc.) are the most abundant activated C–H synthon identified from this retrosynthetic analysis. A total of 800 DrugBank molecules could be disconnected into

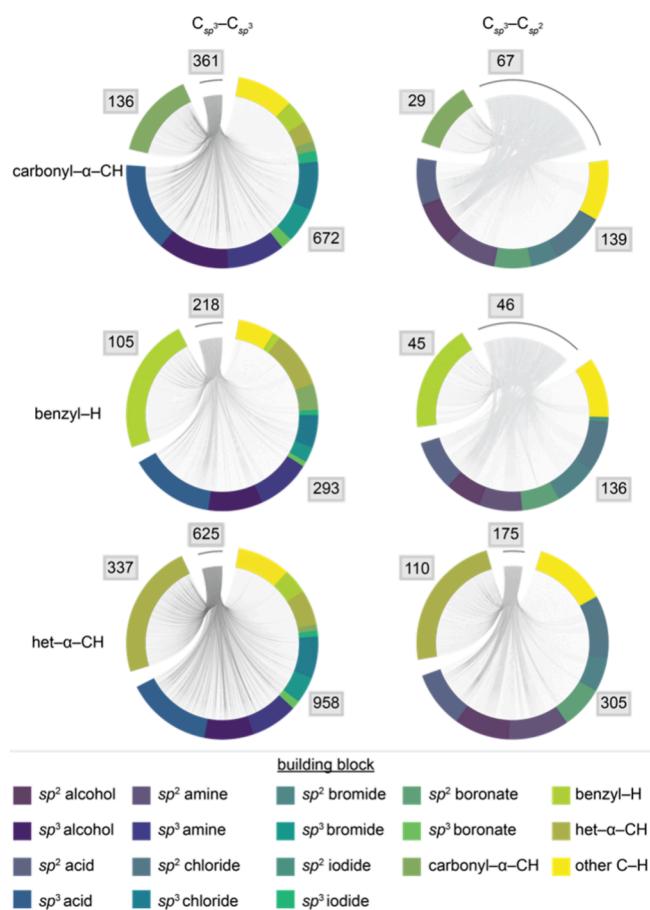


Figure 4. Utilizing activated C–H motifs for C–C bond cross-coupling. Analyzing the DrugBank compounds that can be accessed from commercially available building blocks through a one-step alkylation or arylation of an activated C–H bond.

commercially available compounds with such sites, and the most prominent reaction type features formation of C(sp³)–C(sp³) bonds (625 examples). Indeed, this reaction class has recently emerged,^{16,17} and additional variants that enhance performance on increasingly diverse substrates will be valuable additions to the toolbox.

REPRESENTATIVE APPLICATION OF RETROSYNTHETIC ANALYSIS TO THE SYNTHESIS OF DRUG TARGET AND ANALOGUES THEREOF

The retrosynthetic analysis outlined above shows how activated C–H bonds are useful in drug synthesis. This concept has been recently demonstrated in a preparation of the mucolytic drug bromhexine (**21**), which was disconnected into 2,4,6-tribromoaniline (**19**) and the α-heteroatom C–H building block **20** (Figure 5a). Oxidative conditions have been identified to prepare **21** in a single step using synthons that are considerably less expensive than those used in commercial production of the drug.¹⁸ While these conditions are not ideal for process scale-up, milder conditions can be envisioned.¹⁹ More importantly, this example shows how a cheminformatics-driven retrosynthetic analysis can identify streamlined routes to target drug molecules, in addition to motivating development of impactful new synthetic transformations for pharmaceutical process chemistry.

The same analysis can be used to guide reaction discovery for medicinal chemistry applications. As novel reaction methods are

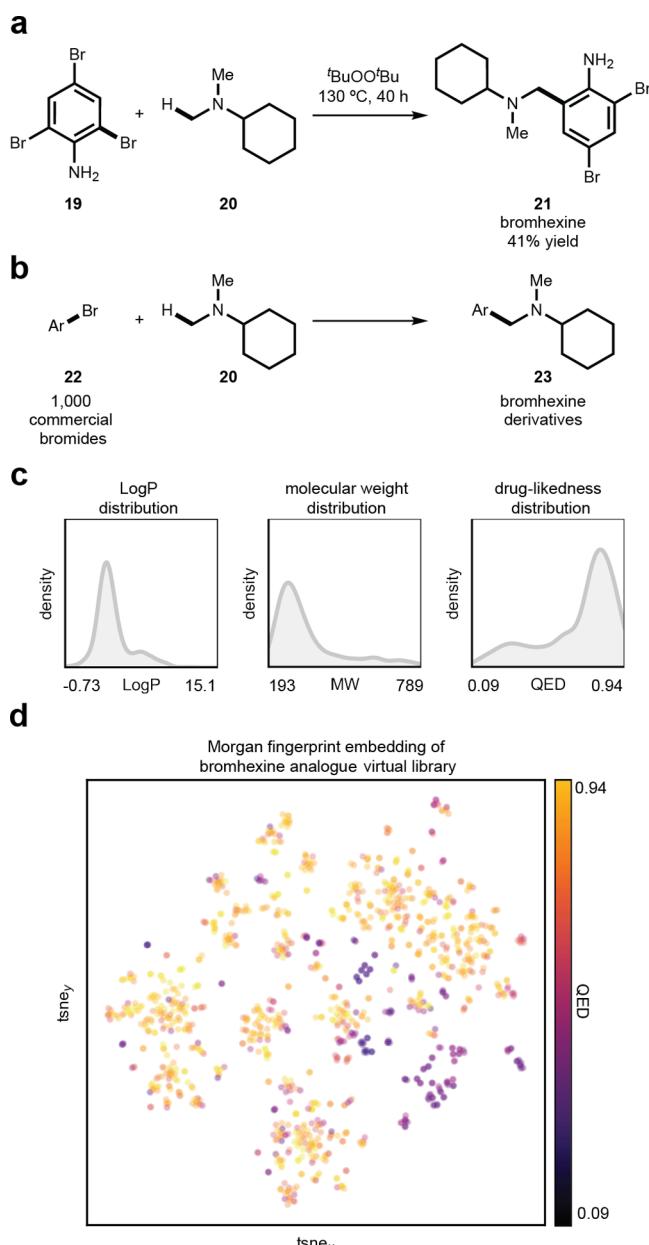


Figure 5. Bromhexine derivative virtual library analysis. (a) Synthesis of bromhexine based on retrosynthetic analysis of C–H arylation reactions. (b) Enumeration of 1,000 aryl bromides to compose a virtual library of bromhexine derivatives. (c) Medicinal chemistry descriptors of a bromhexine derivative library. (d) Chemical space accessed by the bromhexine derivative library, plotted as a tSNE based on 2,048-bit Morgan fingerprints. MW = molecular weight, QED = quantitative estimate of drug-likeness.

developed, pharmacophore-specific libraries of compounds can be accessed using high-throughput synthesis to explore focused domains of chemical space.²⁰ This concept is illustrated here through the design of a virtual library derived from C–H substrate **20** and a set of 1,000 commercially available aryl bromides (**22**) (Figure 5b). The α-amino C–H arylation reaction in Figure 5a provides access to a virtual library of bromhexine derivatives (**23**) that exhibit diverse physicochemical properties. Depending on the objective, specific property distributions can be targeted such as the optimization of lipophilicity, molecular weight, quantitative estimate of drug-likeness (QED), or other physicochemical properties (Figure 5c). The resulting virtual library (**23**) is plotted as a tSNE-based chemical space (Figure 5d) to illustrate the diversity of the generated compounds.

Sc). A map of the compounds obtained via *t*-distributed stochastic neighbor embedding (tSNE, Figure 5d) reveals the diversity of chemical space that can be accessed from such compounds.

CONCLUSION

In this Synopsis, we show how enumerative combinatorics may be applied to the retrosynthesis of drug molecules. This approach contrasts contemporary generative algorithms that prioritize synthetic feasibility, as determined by the availability of published reaction methods.²¹ Instead, it considers all possible disconnections, without regard for pre-existing synthetic methods, with the benefit that it can illuminate strategic opportunities for new reaction development. The results of this analysis reveal the prevalence of activated C(sp³)–H bonds, including those adjacent to carbonyl, aromatic, and heteroatom groups, at the sites of retrosynthetic disconnection. This outcome aligns with the commercial availability of many corresponding synthons, thus enabling synthetic routes to DrugBank compounds to be reimagined through the use or development of new site-selective C–H cross-coupling reactions. The same analysis provides the basis for streamlined access to drug analog libraries that sample novel chemical space.

The information-rich approach to reaction space evaluation presented here is inspired by the mapping and filtering of chemical space that has become routine in drug discovery. By analogy, a reaction mapping approach could highlight aspirational targets for new reaction discovery. This methodology will be impactful for medicinal chemists when reaction methods are identified that utilize abundant building blocks with high chemical diversity. For process chemists, specific reaction methods could be identified to streamline the synthesis of target pharmaceuticals, ideally condensing a multistep synthesis campaign into a single step protocol through the availability or development of a novel reaction method. We are currently experiencing a golden age of reaction method discovery where an experimental protocol for any conceivable chemical transformation seems possible. Such prospects are supported by rapidly developing experimental and computational insights into reaction mechanisms; the broad and ever-expanding availability of reagents and homogeneous, heterogeneous, and biological catalysts for selective chemical reactions; in addition to the growth of useful energy inputs to promote new reactivity, such as thermal, electrochemical, photochemical, and mechanochemical sources. Meanwhile, the coupling of high-throughput experimental methods to machine learning statistical tools continues to accelerate the discovery and optimization of reaction methods. If we imagine a world where any reaction method can be invented, strategies will be needed to identify priorities for specific reactions from the near-infinite reaction cosmos. The analysis presented herein highlights C–H cross-coupling as a privileged opportunity, while the enumerative combinatoric methodology provides the basis for further discovery.

ASSOCIATED CONTENT

Data Availability Statement

The data underlying this study are available in the published article and openly available at https://github.com/cernak-lab/reaction_targeting.

AUTHOR INFORMATION

Corresponding Authors

Shannon S. Stahl – Department of Chemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706, United States;  orcid.org/0000-0002-9000-7665; Email: stahl@chem.wisc.edu

Tim Cernak – Department of Medicinal Chemistry, University of Michigan, Ann Arbor, Michigan 48109, United States;  orcid.org/0000-0001-5407-0643; Email: tcernak@med.umich.edu

Authors

Babak Mahjour – Department of Medicinal Chemistry, University of Michigan, Ann Arbor, Michigan 48109, United States

Kaitlyn M. Flynn – Department of Chemistry, University of Wisconsin—Madison, Madison, Wisconsin 53706, United States

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.joc.4c01567>

Notes

The authors declare no competing financial interest.

Biographies



Babak Mahjour completed his Ph.D. in medicinal chemistry with Prof. Tim Cernak at the University of Michigan in 2023. He joined the Coley Research group at MIT soon after to continue studying chemical synthesis and lab automation, with an eye toward AI-driven translational research.



Kaitlyn M. Flynn studied at University of Wisconsin—Madison and received her M.S. in chemistry in 2023 under the advisement of Prof. Shannon S. Stahl studying photochemical C–H chlorination reaction development. After completing her studies, she joined GSK as a Senior

Scientist in Drug Substance Development where she works on process development for small molecule assets.



Shannon S. Stahl is currently the Steenbock Professor of Chemical Sciences at the University of Wisconsin—Madison, where he began his independent career in 1999. He obtained his B.S. in chemistry at the University of Illinois Urbana-Champaign and his Ph.D from Caltech in 1997, under the supervision of Prof. John Bercaw. He was a National Science Foundation postdoctoral fellow with Prof. Stephen Lippard at the Massachusetts Institute of Technology from 1997 to 1999. He and his research group specialize in redox catalysis, with an emphasis on the catalytic chemistry of molecular oxygen, selective oxidation reactions, and C—H functionalization and cross-coupling. These topics are complemented by efforts focused on electrocatalytic reactions for chemical synthesis and energy conversion.



Tim Cernak is a chemist with an interest in pharmaceuticals, synthesis, informatics, and environmental issues. He currently is an Associate Professor of Medicinal Chemistry at the University of Michigan, with additional appointments in the Department of Chemistry, the Program in Chemical Biology, the Michigan Institute of Data Science, and the Center for Global Health Equity. Prior to working at the University of Michigan, Dr. Cernak worked as a drug hunting chemist at Merck & Co., Inc. He has served on the scientific advisory board of the National Science Foundation Center for Selective C—H Functionalization, the University of Dundee Drug Discovery Unit, and Scorpion Therapeutics. He is a co-founder of Iambic Therapeutics. Tim is passionate about the preservation of biodiversity.

■ REFERENCES

- (1) Shen, Y.; Borowski, J. E.; Hardy, M. A.; Sarpong, R.; Doyle, A. G.; Cernak, T. Automation and computer-assisted planning for chemical synthesis. *Nat. Rev. Methods Primers* **2021**, *1* (1), 23.
- (2) Szymkuć, S.; Gajewska, E. P.; Klucznik, T.; Molga, K.; Dittwald, P.; Startek, M.; Bajczyk, M.; Grzybowski, B. A. Computer-Assisted Synthetic Planning: The End of the Beginning. *Angew. Chem., Int. Ed.* **2016**, *55* (20), 5904–5937.
- (3) Mahjour, B.; Shen, Y.; Liu, W.; Cernak, T. A map of the amine—carboxylic acid coupling system. *Nature* **2020**, *580* (7801), 71–75.
- (4) Zhang, R.; Mahjour, B.; Outlaw, A.; McGrath, A.; Hopper, T.; Kelley, B.; Walters, W. P.; Cernak, T. Exploring the combinatorial explosion of amine—acid reaction space via graph editing. *Commun. Chem.* **2024**, *7* (1), 22.
- (5) Wishart, D. S.; Feunang, Y. D.; Guo, A. C.; Lo, E. J.; Marcu, A.; Grant, J. R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; Assempour, N.; Iynkkaran, I.; Liu, Y.; Maciejewski, A.; Gale, N.; Wilson, A.; Chin, L.; Cummings, R.; Le, D.; Pon, A.; Knox, C.; Wilson, M. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **2018**, *46* (D1), D1074–d1082.
- (6) MilliporeSigma. <https://www.aldrichmarketselect.com/> (accessed 2024-01-17).
- (7) Davies, I. W. The digitization of organic synthesis. *Nature* **2019**, *570* (7760), 175–181.
- (8) Brown, D. G.; Boström, J. Analysis of Past and Present Synthetic Methodologies on Medicinal Chemistry: Where Have All the New Reactions Gone? *J. Med. Chem.* **2016**, *59* (10), 4443–4458.
- (9) Lovering, F.; Bikker, J.; Humblet, C. Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success. *J. Med. Chem.* **2009**, *52* (21), 6752–6756.
- (10) Campeau, L.-C.; Hazari, N. Cross-Coupling and Related Reactions: Connecting Past Success to the Development of New Reactions for the Future. *Organometallics* **2019**, *38* (1), 3–35.
- (11) Golden, D. L.; Suh, S.-E.; Stahl, S. S. Radical C(sp³)—H functionalization and cross-coupling reactions. *Nat. Rev. Chemistry* **2022**, *6* (6), 405–427.
- (12) Chen, S.-J.; Krska, S. W.; Stahl, S. S. Copper-Catalyzed Benzyl C—H Cross-Coupling Enabled by Redox Buffers: Expanding Synthetic Access to Three-Dimensional Chemical Space. *Acc. Chem. Res.* **2023**, *56* (24), 3604–3615.
- (13) Chen, S.-J.; Golden, D. L.; Krska, S. W.; Stahl, S. S. Copper-Catalyzed Cross-Coupling of Benzyl C—H Bonds and Azoles with Controlled N-Site Selectivity. *J. Am. Chem. Soc.* **2021**, *143* (36), 14438–14444.
- (14) Hu, H.; Chen, S.-J.; Mandal, M.; Pratik, S. M.; Buss, J. A.; Krska, S. W.; Cramer, C. J.; Stahl, S. S. Copper-catalysed benzyl C—H coupling with alcohols via radical relay enabled by redox buffering. *Nat. Catalysis* **2020**, *3* (4), 358–367.
- (15) Zhang, W.; Wu, L.; Chen, P.; Liu, G. Enantioselective Arylation of Benzyl C—H Bonds by Copper-Catalyzed Radical Relay. *Angew. Chem., Int. Ed.* **2019**, *58* (19), 6425–6429.
- (16) Ma, P.; Liu, Y.; Chen, L.; Zhao, X.; Yang, B.; Zhang, J. Photocatalyst- and additive-free decarboxylative alkylation of N-aryl tetrahydroisoquinolines induced by visible light. *Org. Chem. Front.* **2021**, *8* (11), 2473–2479.
- (17) Bergonzini, G.; Schindler, C. S.; Wallentin, C.-J.; Jacobsen, E. N.; Stephenson, C. R. J. Photoredox activation and anion binding catalysis in the dual catalytic enantioselective synthesis of β -amino esters. *Chem. Sci.* **2014**, *5* (1), 112–116.
- (18) Lin, Y.; Zhang, Z.; Mahjour, B.; Wang, D.; Zhang, R.; Shim, E.; McGrath, A.; Shen, Y.; Brugger, N.; Turnbull, R.; Trice, S.; Jasty, S.; Cernak, T. Reinforcing the supply chain of umifenovir and other antiviral drugs with retrosynthetic software. *Nat. Commun.* **2021**, *12* (1), 7327.
- (19) Ahneman, D. T.; Doyle, A. G. C—H functionalization of amines with aryl halides by nickel-photoredox catalysis. *Chem. Sci.* **2016**, *7* (12), 7002–7006.
- (20) Gasmundo, N. J.; Sauvagnat, B.; Curran, P. J.; Richards, M. P.; Andrews, C. L.; Dandliker, P. J.; Cernak, T. Nanoscale synthesis and affinity ranking. *Nature* **2018**, *557* (7704), 228–232.
- (21) Bellmann, L.; Penner, P.; Gastreich, M.; Rarey, M. Comparison of Combinatorial Fragment Spaces and Its Application to Ultralarge Make-on-Demand Compound Catalogs. *J. Chem. Inf. Model.* **2022**, *62* (3), 553–566.