Check for updates

RESEARCH ARTICLE

ORGANIC CHEMISTRY

# Computer-aided key step generation in alkaloid total synthesis

Yingfu Lin[1], Rui Zhang[2], Di Wang[1], Tim Cernak[1,2]*

**Efficient chemical synthesis is critical to satisfying future demands for medicines, materials, and agrochemicals. Retrosynthetic analysis of modestly complex molecules has been automated over the course of decades, but the combinatorial explosion of route possibilities has challenged computer hardware and software until only recently. Here, we explore a computational strategy that merges computer-aided synthesis planning with molecular graph editing to minimize the number of synthetic steps required to produce alkaloids. Our study culminated in an enantioselective three-step synthesis of (−)-stemoamide by leveraging high-impact key steps, which could be identified in computer-generated retrosynthesis plans using graph edit distances.**

Efficient total synthesis converts commercially available starting materials into complex target molecules in as few steps as possible. In practice, this is often achieved by optimizing key steps to simultaneously form many of the requisite target bonds, thereby rapidly generating structural complexity. Common examples include cycloadditions, cascades (*1*), or multicomponent coupling reactions (*2*). Although the notion of a key step is well known to practitioners of total synthesis, it has not been formalized in computer-aided synthesis planning (CASP). Modern CASP strategies aim to minimize protecting-group manipulations and maximize convergency, but the focus of automated retrosynthesis has been on encoding reaction rules for maximum reliability in experimental realization of predicted routes, which favors robust, well-precedented reactions. Meanwhile, state-of-the-art human synthetic strategies maximize step and atom economy by targeting innovative but riskier key steps and minimizing low-impact concession steps such as protecting group manipulations (*3*), unnecessary redox operations (*4*), and functional group interconversions (*5*). Here we explore the synergy of automatically generating many plausible key steps using CASP, with manual selection for high-impact steps, which are encoded as the reduction in graph edit distance, to minimize overall step count in total synthesis.

Experimental demonstration of routes designed with modern CASP has been achieved in pharmaceutical synthesis (*6*–*9*) but has seen considerably less use in the complex setting of alkaloid total synthesis. Notable exceptions include a recent 30-step synthesis of weisaconitine D (*10*) and a 15-step synthesis of (+)-tacamonidine (*11*), representing the state of the art. We chose stemoamide (**1**, Fig. 1), isolated from *Stemonaceae* plants (*12*), as an ideal target molecule because its four stereocenters and fused-ring structure would challenge modern CASP (*11*), while its precedence in thirty-two historic and contemporary syntheses (*13*–*21*) provides a strong benchmark for comparison.

[1]Department of Medicinal Chemistry, University of Michigan, Ann Arbor, MI 48109, USA. [2]Department of Chemistry, University of Michigan, Ann Arbor, MI 48109, USA.
*Corresponding author. Email: tcernak@med.umich.edu

## Computer-aided retrosynthetic plans

The first retrosynthetic plan removes the α-methyl group and breaks the C–N bond of the azepine ring to give cyclized alkene **2** (Fig. 1), which was itself produced by a key organocatalyzed Mannich-allylation-lactonization sequence recommended by CASP. Advanced intermediate **2** was further simplified to starting materials **3**, **4** and two equivalents of aldehyde **5**. A second route to **1** resulted from the evolution of our CASP and graph editing strategy. In this route, a Schmidt-Aubé rearrangement recommended by CASP is a key simplifying element. The resultant cyclobutanone intermediate **6** was further broken down into **7**, **8**, **9**, and **10**, leveraging a key Michael addition and alkylation sequence, which was identified as a key step in a separate CASP search of **1**. The computationally discovered routes were edited and experimentally realized as described below.

## Revelation of the Mannich reaction as a key step

Modern CASP platforms favor reactions with well-established conditions. As a result, predicted routes generally have lower risk but can be biased toward low-impact transformations. A system for recognizing synthetic ideality (*22*, *23*) could synergize with the power of modern CASP platforms even though additional experimental optimization may be required to realize riskier high-impact key steps.

As a first attempt to minimize step count from CASP-generated routes, (−)-**1** was subjected to automated retrosynthesis in the software SYNTHIA with a scoring function that promoted chemoselectivity, regioselectivity, and stereoselectivity and demoted the use of protecting groups. An organocatalyzed Mannich reaction appeared as a proposed transformation in every predicted route (Fig. 2A), an unexpected outcome because this well-known reaction had not featured in any of the 32 prior syntheses of **1**. Nonetheless, even the shortest calculated route (fig. S1) was seven steps long, which was competitive with, but not better than, the shortest human-derived enantioselective route (*14*). We thus chose to logically edit calculated routes for brevity by maximizing high-impact transformations and minimizing low-impact transformations. This required a new system to measure step impact.

Obvious inefficiencies such as protecting-group manipulations could be minimized by the CASP software in the calculated routes, but it was nonobvious whether other specific steps among the hundreds of calculated retrosynthetic routes analyzed were impactful or not. In principle, the impact of a given synthetic

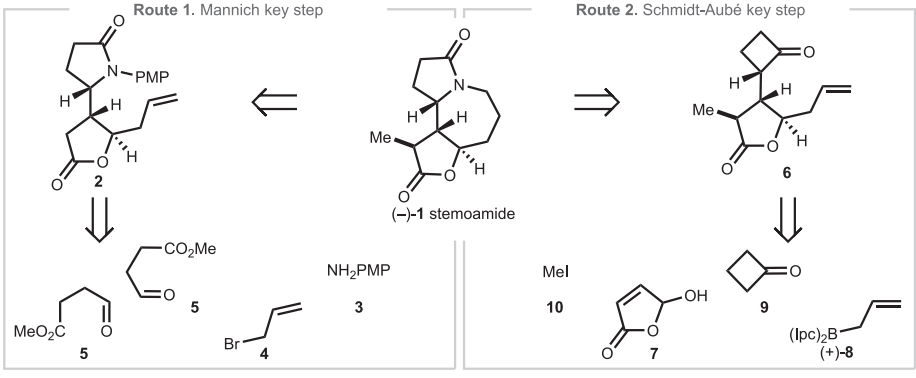


**Fig. 1. Two retrosynthetic plans of 1 with computer-aided key step generation.** PMP, *p*-methoxyphenyl; Ipc, isopinocampheyl.

Downloaded from https://www.science.org at University Van Amsterdam on April 13, 2025

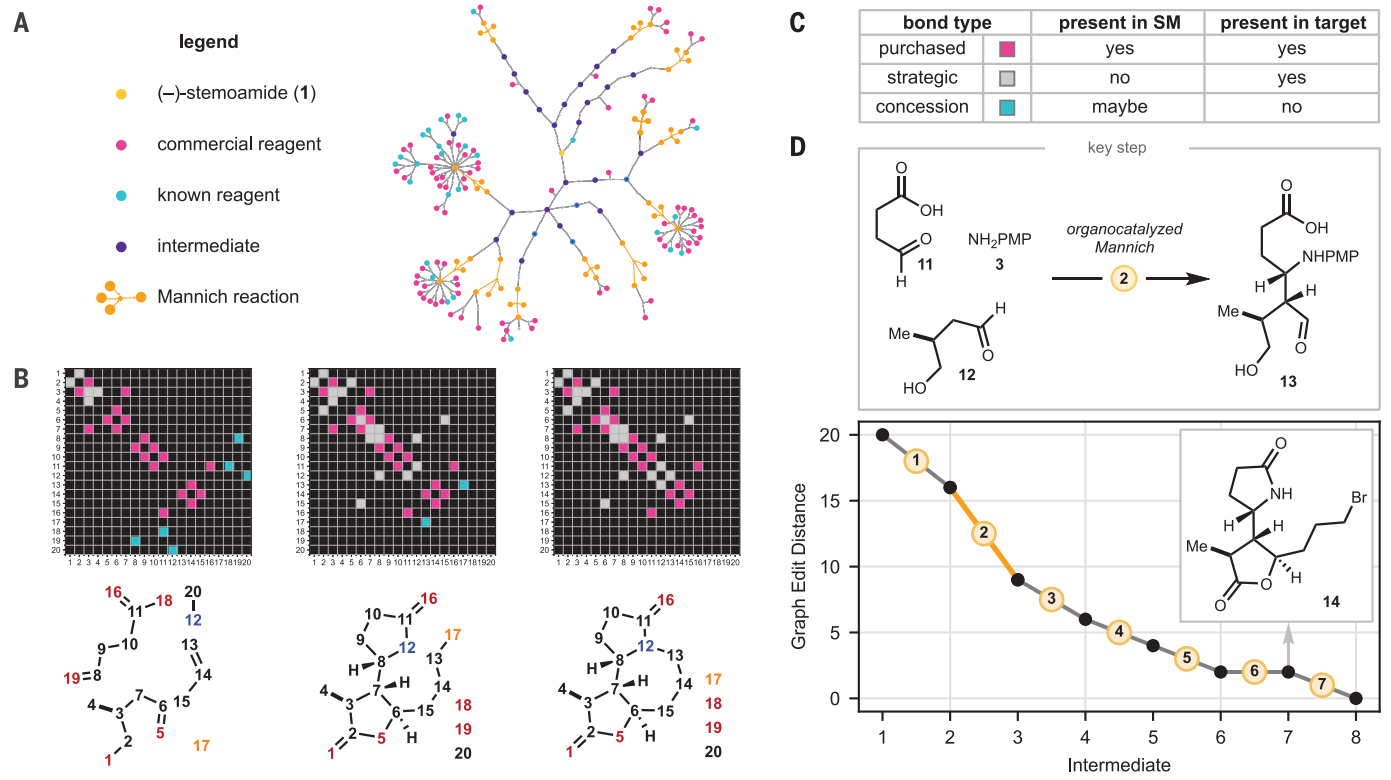


**Fig. 2. Identifying key steps by graph edit analysis.** (**A**) Network analysis of 50 SYNTHIA-predicted routes to (–)-**1**. In this search result, the Mannich reaction featured as a consistent disconnection, as highlighted by a cluster of four orange dots in each route. (**B**) Adjacency matrices of **3**, **11**, **12**, **8**, and HBr (left); **14** and disconnected concession atoms (middle); and **1** with disconnected concession atoms (right). Matrix width and length are determined by the total number of heavy atoms in the overall synthesis. For concession groups, only the attached atom is considered in the graph. (**C**) Classification of bond type used in graph edit analysis. (**D**) Graph edit distance plots of a retrosynthetic route to **1**, produced by CASP, highlighting Mannich key step.

step in a multistep sequence can be measured by the reduction in graph edit distance (*24*) between molecular graphs (*25*) of intermediates and the target molecule. Efficient multistep synthesis converts the bonds and atoms of commercially available starting materials into those of the target with key steps assembling a large portion of the target bonds and stereocenters simultaneously. Thus, high-impact reaction steps should have a steep slope in a molecular graph edit distance plot between given intermediates en route to the final target.

We encoded molecular graphs of each intermediate, including starting materials and the final target, as individual adjacency matrices (Fig. 2B and fig. S2) in which the number of rows and columns is equal to the total number of heavy atoms and groups used in the entire synthetic route. In this way, all bonds of the final target and their progression from starting materials, as well as any concession groups used in the synthesis, are mapped exactly in each individual matrix and in relation to the final target's matrix (Fig. 2C). A simple comparison of the matrix for **1** (Fig. 2B, right) reveals that it shares more entries in common (99%) with the calculated penultimate intermediate **14** (Fig. 2B, middle) than it does with the matrix for starting materials **3**, **11**, **12**, **8** and HBr (Fig. 2B, left) (93%). Accordingly, key steps maximize the reduction in graph edit distance from a given intermediate to the target, which is equivalent to maximizing formation of target bonds while minimizing reaction manipulations on concession groups. This graph formalization shares similarity with established concepts of synthetic ideality but is machine readable, carries exact atom and bond mapping, and requires no labeling of reaction type to separate low-impact reactions such as redox manipulations from impactful cycloadditions or cascade reactions. Because graphs capture the exact location of every bond, graph edit distance was superior to other metrics, such as Tanimoto distance based on Morgan fingerprints (fig. S3), at highlighting the impact of key transformations. A survey of published total syntheses by graph edit distance (fig. S4) shows that diverse key steps are readily visualized. A full graph analysis of the shortest calculated route to **1** (fig. S5) reveals the impact of the Mannich coupling (Fig. 2D), which appears as the steepest declining step (yellow) in the graph edit distance plot.

### Six-step total synthesis of stemoamide

Although the computed routes to **1** focused our attention on the Mannich disconnection (Fig. 2A) as an impactful key step, we recognized opportunities for improvement. For instance, C2 and C11 are both in the carboxyl oxidation state in **1**, so by considering redox economy (*4*), the oxidation state of **11** and **12** could be harmonized to excise two calculated steps. This realization unveiled a hidden symmetry element within **1**, where two equivalents of commercially available aldehyde **5** (Fig. 3A) unite in a self-Mannich reaction (*26*, *27*). This would require installation of the chiral α-methyl group at a later stage, and fortuitously a diastereoselective methylation had already been reported as a viable final step in several syntheses of **1** (*13*, *14*).

In experimental practice (Fig. 3A), stirring **3** with a fourfold excess of aldehyde **5** and 20 mol % L-proline in *N*,*N*-dimethylformamide at –15°C, then adding a mixture of allyl bromide **4**, zinc, and bismuth chloride directly to the reaction vessel and warming to room temperature, cleanly delivered lactone **16** as the major product via intermediacy of Mannich-adduct **15**. Rather than isolate **16**, excess zinc and insoluble materials were removed from the reaction mixture by filtration, and the crude filtrate was treated with trifluoroacetic acid (TFA) to provide lactam **17**, which upon purification by silica gel chromatography was

Downloaded from https://www.science.org at University Van Amsterdam on April 13, 2025

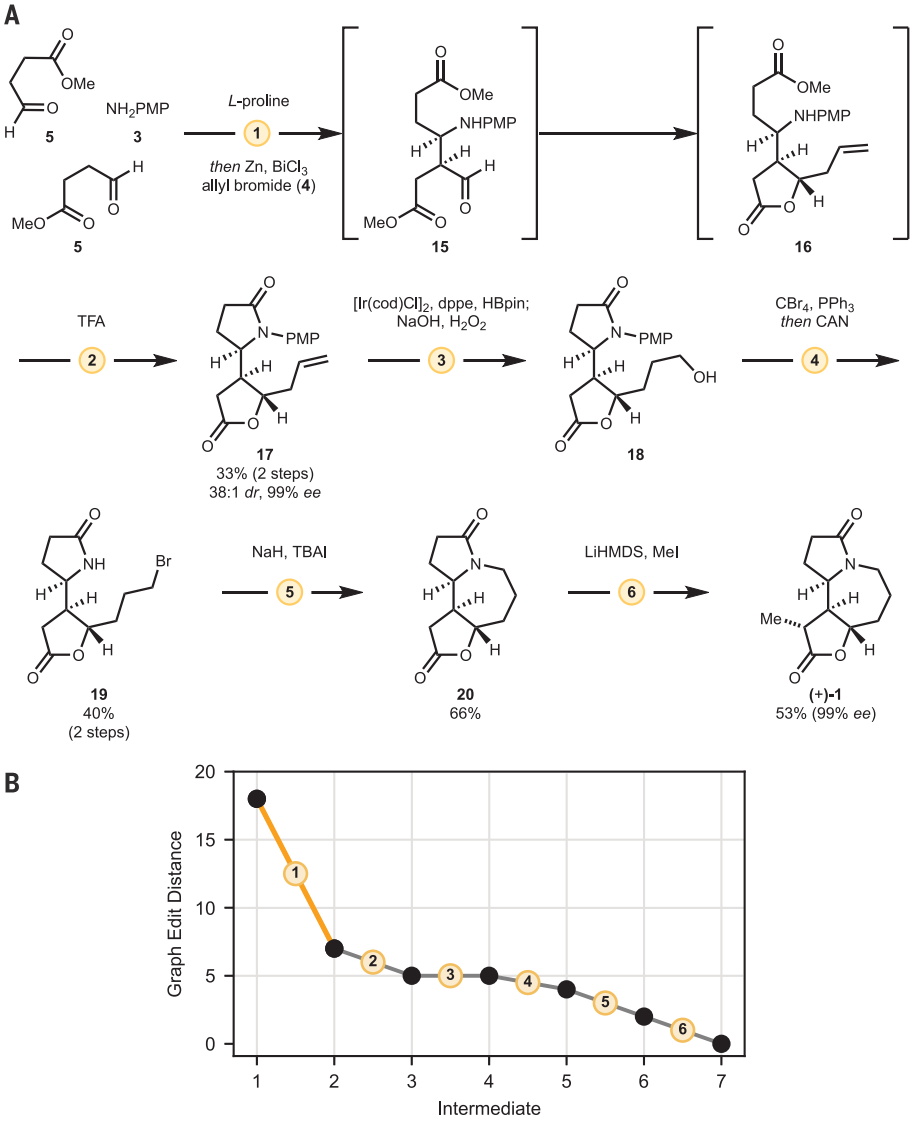


**Fig. 3. Total synthesis of 1 relying on an asymmetric organocatalyzed Mannich reaction.** (**A**) Total synthesis of (+)-**1** in six steps. dppe, 1,2-bis(diphenylphosphino)ethane; CAN, ceric ammonium nitrate; TBA, tetra-*n*-butylammonium; HMDS, hexamethyldisilazide. (**B**) Graph edit distance analysis of the route.

obtained in 38:1 diastereomeric ratio (dr), 99% enantiomeric excess (ee), and isolated in 33% overall yield from **3**. Thus, five bonds, two rings, and three stereocenters were rapidly formed in high selectivity via a two-step sequence requiring just one chromatographic purification. Next, it was necessary to perform the hydrobromination of **17**. Experimentally, the CASP suggestion to use hydrobromic acid (*28*) led to an intractable mixture. The best experimental protocol that we identified involved conversion of alkene **17** to primary alcohol **18** followed by bromination and in situ removal of the *p*-methoxyphenyl group, giving **19**. Two final steps, based on known reactions (*13*, *14*), were required to form the azepane ring (**20**) and install the C10 methyl group via diastereoselective enolate alkylation, giving (+)-**1**. Our synthesis was performed in six steps and required only four chromatographic purifications, marking the shortest enantioselective total synthesis of **1**. For convenience, we chose to make (+)-**1** by using L-proline, although the software accurately recognized that D-proline would lead to (−)-**1**. The experimental route that arose through our modifications of the calculated route was subjected to a graph edit analysis (Fig. 3B). Here the high impact of the Mannich-allylation key step is readily apparent, as is the low impact of the ensuing functional-group interconversions required to complete the synthesis.

### Three-step total synthesis of stemoamide

To best our own result in terms of step count, we generated hundreds of additional calculated routes to **1**, as well as related late-stage intermediates such as **20**, searching for key steps that could be repurposed. In our first synthesis (Fig. 3), we aimed to stay as true to the CASP-predicted route as possible, but made some modifications as described. In our second route, we focused on merging interesting key steps from multiple calculated routes. Graph edit distance analysis was used to direct our attention to key steps of CASP-calculated routes and facilitate the combination of multiple key steps. One computed strategy of interest involved a peculiar cyclobutanone intermediate (an analog of **6**, Fig. 4A), which was primed for Schmidt-Aubé rearrangement to access **1**. Rather than apply the proposed four-step sequence (fig. S13) to access an analog of key intermediate **6**, we realized that an analogous diastereoselective Michael addition–alkylation sequence could be feasible to access **6** by adding cyclobutanone as a nucleophile into Michael acceptor **21** and quenching the intermediate enolate with methyl iodide. A similar Michael addition–alkylation sequence also emerged as a key step in a calculated route to **1** when the Mannich reaction was specifically excluded from the CASP search (fig. S6). We elected to perform the Michael addition via the Enders (*S*)-1-amino-2-methoxymethylpyrrolidine (SAMP) hydrazone of cyclobutanone (**22**) (*29*), as the SAMP auxiliary method was frequently proposed in a variety of computed routes. Thus, our edited retrosynthetic strategy maximized high-impact steps and minimized low-impact steps to arrive at a concise proposal for the synthesis of **1**.

Experimentally (Fig. 4A), we produced **21** from commercially available **7** via Brown allylation in 89% ee and 58% yield. The optimized protocol for the conjugate addition comprised deprotonation of **22** with *n*-butyllithium, addition of **21** to a cold solution of the anion, and trapping of the intermediate enolate with methyl iodide. Quenching of the reaction mixture with aqueous hydrochloric acid yielded ketone **6** in 88% yield as a 4:1 mixture of diastereomers at C9a. Subsequent iodosobenzoic acid (IBA) catalyzed *anti*-Markovnikov hydroazidation with trimethylsilyl azide (*30*, *31*) gave **23**, and Lewis acid–induced intramolecular Schmidt–Aubé rearrangement (*32*–*34*) of presumed intermediate **24** led to (−)-**1**. The longest linear sequence was three steps, giving the target in 22% overall yield, thereby halving our prior step count.

Step impact can be readily observed in our six-step synthesis of **1,** where the first organocatalyzed Mannich-allylation step markedly increases the graph similarity of **3**, **5**, and **4** to **1**, installing 45% of the bonds required to produce **1** (Fig. 3B). This high-impact step is then followed by a series of low-impact steps, such as protecting-group manipulations and functional-group interconversions, which are

Downloaded from https://www.science.org at University Van Amsterdam on April 13, 2025

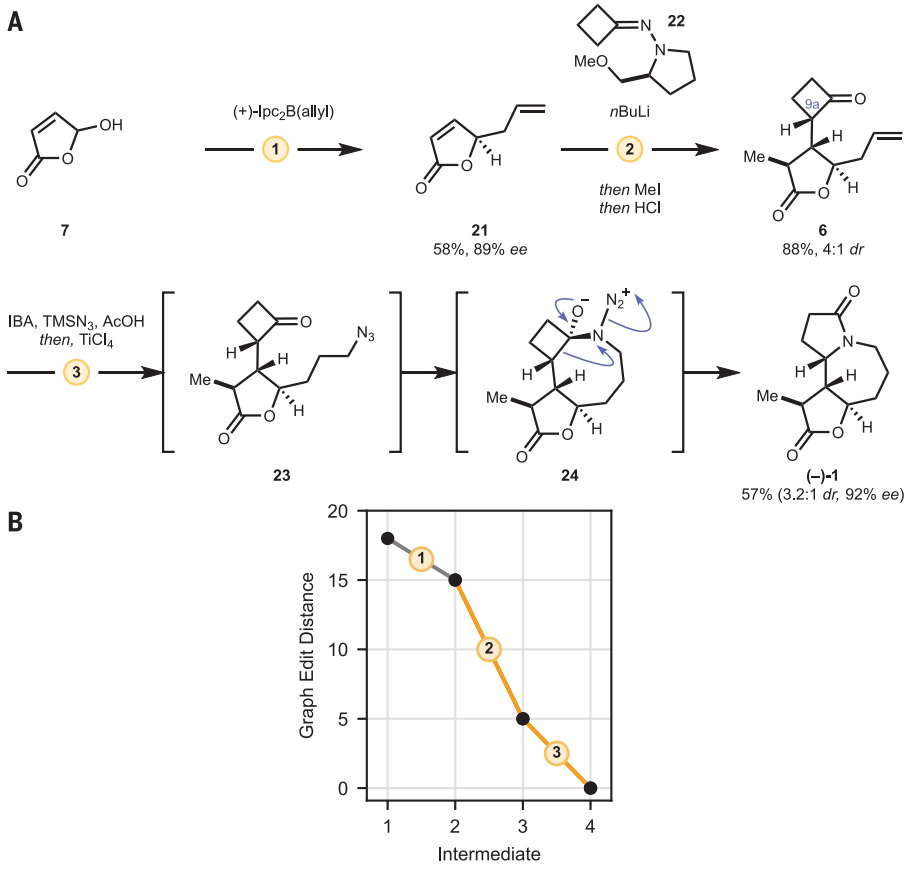


**Fig. 4. Total synthesis of 1 featuring a Schmidt-Aubé rearrangement.** (**A**) Total synthesis of (−)-**1** in three steps. (+)-Ipc$_2$B(allyl), (+)-*B*-allyldiisopinocampheylborane; IBA, iodosobenzoic acid; TMS, trimethylsilyl. (**B**) Graph edit distance analysis of the route.

easily recognized by their shallow slope in Fig. 3B. Meanwhile, our 3-step route is much more efficient with the sequence of steps contributing 17, 55, and 28%, respectively, to the graph similarity of intermediates to **1** (Fig. 4B). The key steps that we used were selected from an analysis of more than a thousand calculated retrosynthetic routes.

In addition to highlighting key steps, the graph edit distance technique can be used to highlight shortcuts in the route, which may require the invention of new reactions. This is easily achieved by grouping neighboring transformations with a modest slope in the graph edit plot into a single shortcut step. For instance, steps 3, 4 and 5 in our Mannich route (Fig. 3) could in principle be grouped into an overall *anti*-Markovnikov hydroamidation (fig. S7). It was possible to quench the TFA-promoted lactamization (step 2) with ceric ammonium nitrate (CAN) to produce an analog of **17** with the PMP group removed in 33% overall yield from **3**. This analog could be converted to **20** in a single step following invention of an *anti*-Markovnikov hydroamidation reaction, ultimately leading to a four-step synthesis of **1**. This strategy can be generally applied to suggest specific new reactions, with their respective atom mappings, that provide a shortcut in any synthetic route.

Modern CASP has demonstrated the ability to produce practicable routes to modestly complex targets, but to date route proposals do not challenge the brevity of modern human-derived routes. Our method shows that it is possible to unite multiple high-impact steps from diverse CASP route proposals, as shown for (−)-**1** in Fig. 4, to arrive at concise synthetic routes. We focused on step count as a sole optimization metric in the current study, but important real-world metrics such as reagent cost, building block availability, or predicted yield could be easily incorporated as a weighted distance metric. As advances in automated retrosynthesis make complex molecules more accessible, it is likely that precision pharmaceuticals will become increasingly available.

## REFERENCES AND NOTES


1. K. C. Nicolaou, D. J. Edmonds, P. G. Bulger, *Angew. Chem. Int. Ed.* **45**, 7134–7186 (2006).
2. B. B. Touré, D. G. Hall, *Chem. Rev.* **109**, 4439–4486 (2009).
3. T. Newhouse, P. S. Baran, R. W. Hoffmann, *Chem. Soc. Rev.* **38**, 3010–3021 (2009).
4. N. Z. Burns, P. S. Baran, R. W. Hoffmann, *Angew. Chem. Int. Ed.* **48**, 2854–2867 (2009).
5. S. W. M. Crossley, R. A. Shenvi, *Chem. Rev.* **115**, 9465–9531 (2015).
6. T. Klucznik *et al.*, *Chem* **4**, 522–532 (2018).
7. Y. Lin *et al.*, *Nat. Commun.* **12**, 7327 (2021).
8. A. Wołos *et al.*, *Nature* **604**, 668–676 (2022).
9. C. W. Coley *et al.*, *Science* **365**, eaax1566 (2019).
10. C. J. Marth *et al.*, *Nature* **528**, 493–498 (2015).
11. B. Mikulak-Klucznik *et al.*, *Nature* **588**, 83–88 (2020).
12. L. Wang *et al.*, *Phytochem. Rev.* **21**, 835–862 (2022).
13. G. A. Brito, R. V. Pirovani, *Org. Prep. Proced. Int.* **50**, 245–259 (2018).
14. M. Yoritate *et al.*, *J. Am. Chem. Soc.* **139**, 18386–18391 (2017).
15. X. Yin, K. Ma, Y. Dong, M. Dai, *Org. Lett.* **22**, 5001–5004 (2020).
16. J. H. Siitonen, D. Csókás, I. Pápai, P. M. Pihko, *Synlett* **31**, 1581–1586 (2020).
17. Z. Guo *et al.*, *Angew. Chem. Int. Ed.* **60**, 14545–14553 (2021).
18. F. Cao *et al.*, *Org. Lett.* **23**, 6222–6226 (2021).
19. T. Shi *et al.*, *Org. Chem. Front.* **9**, 771–774 (2022).
20. X. Wang *et al.*, *Org. Chem. Front.* **9**, 3818–3822 (2022).
21. G. Bernardi Rosso, B. Matos Paz, R. Aloise Pilli, *Eur. J. Org. Chem.* **2022**, e202200585 (2022).
22. D. S. Peters *et al.*, *Acc. Chem. Res.* **54**, 605–617 (2021).
23. J. B. Hendrickson, *J. Am. Chem. Soc.* **97**, 5784–5800 (1975).
24. A. Sanfeliu, K. Fu, *IEEE Trans. Syst. Man Cybern.* **SMC-13**, 353–362 (1983).
25. L. David, A. Thakkar, R. Mercado, O. Engkvist, *J. Cheminform.* **12**, 56 (2020).
26. W. Notz *et al.*, *J. Org. Chem.* **68**, 9624–9634 (2003).
27. Y. Hayashi *et al.*, *Angew. Chem. Int. Ed.* **42**, 3677–3680 (2003).
28. P. J. Kropp *et al.*, *J. Am. Chem. Soc.* **112**, 7433–7434 (1990).
29. D. Hazelard, A. Fadel, *Tetrahedron Asymmetry* **16**, 2067–2070 (2005).
30. H. Li, S.-J. Shen, C.-L. Zhu, H. Xu, *J. Am. Chem. Soc.* **141**, 9415–9421 (2019).
31. X. Li, P. Chen, G. Liu, *Sci. China Chem.* **62**, 1537–1541 (2019).
32. J. Aube, G. L. Milligan, *J. Am. Chem. Soc.* **113**, 8965–8966 (1991).
33. G. L. Milligan, C. J. Mossman, J. Aube, *J. Am. Chem. Soc.* **117**, 10449–10459 (1995).
34. K. J. Frankowski, R. Liu, G. L. Milligan, K. D. Moeller, J. Aubé, *Angew. Chem. Int. Ed.* **54**, 10555–10558 (2015).
35. Y. Lin, R. Zhang, D. Wang, T. Cernak, Code and Data for Computer-Aided Key Step Generation in Alkaloid Total Synthesis, version 1.1, Zenodo (2022); https://doi.org/10.5281/zenodo.7449332.


## ACKNOWLEDGMENTS


The authors thank N. Brugger, S. Jasty, A. Tathe, R. Turnbull, and L. Rickershauser (MilliporeSigma) for helpful discussions and assistance with graphics. We also thank S. Trice (previously of MilliporeSigma) for early discussions and support of the work. **Funding:** This work was funded by MilliporeSigma, start-up funds from the University of Michigan College of Pharmacy, and the National Science Foundation (CHE-2236215) **Author contributions:** Y.L. performed synthetic chemistry experiments, Y.L. and T.C. performed SYNTHIA searches. R.Z. and D.W. developed the graph edit distance technique. All authors reviewed and interpreted the data and wrote the manuscript. T.C. supervised the work. **Competing interests:** The Cernak Lab has received research funding or in-kind donations from MilliporeSigma, Relay Therapeutics, Janssen Therapeutics, SPT Labtech, and Merck & Co., Inc. T.C. holds equity in Scorpion Therapeutics and is a cofounder of and equity holder in Entos, Inc. The Regents of the University of Michigan have filed a provisional patent on some aspects of this work. **Data and materials availability:** Code for graph editing techniques is available at Zenodo (*35*). SYNTHIA is available at https://www.synthiaonline.com.


**License information:** Copyright © 2023 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse


## SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.ade8459
Materials and Methods
Figs. S1 to S71
Table S1
NMR Spectra
References (*36*–*59*)


Submitted 14 September 2022; accepted 4 January 2023
10.1126/science.ade8459


Downloaded from https://www.science.org at University Van Amsterdam on April 13, 2025