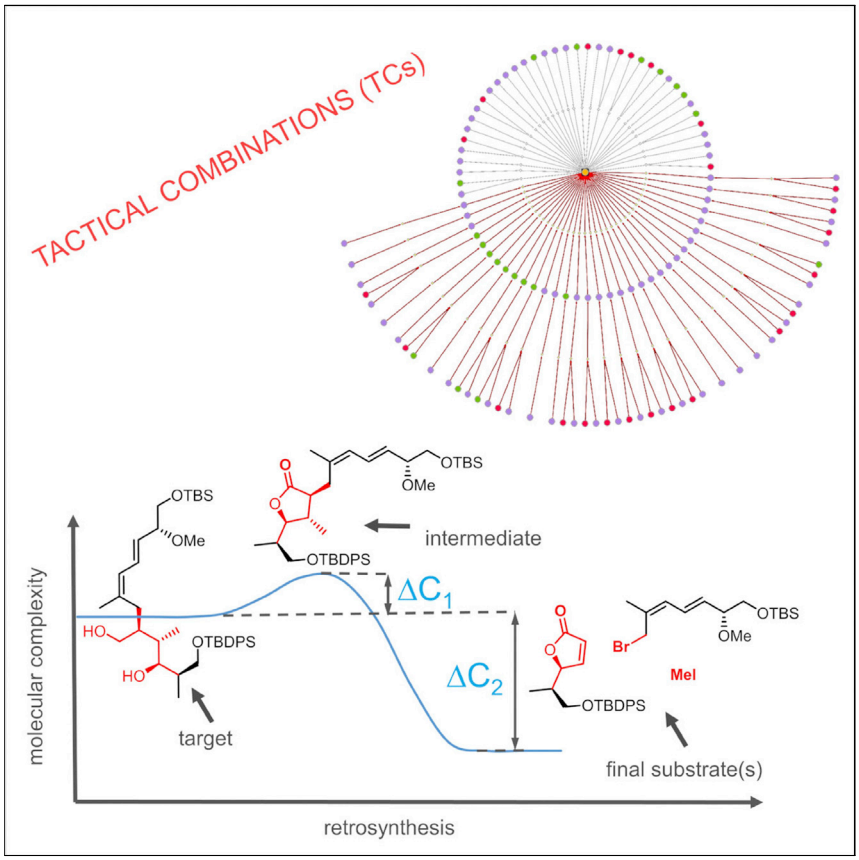
Chem

CellPress

Article

Algorithmic Discovery of Tactical Combinations for Advanced Organic Syntheses



Tactical combinations (TCs) are two-step reaction sequences that first, in a retrosynthetic direction, complexify the target structure but, by doing so, enable significant structural simplification in subsequent steps. TCs are often hard to identify, and only a few hundred have been cataloged to date. This work shows that computers can now discover TCs having no literature precedent systematically and in large numbers, in effect unlocking new and elegant synthetic approaches.

Ewa P. Gajewska, Sara Szymkuć, Piotr Dittwald, Michał Startek, Oskar Popik, Jacek Mlynarski, Bartosz A. Grzybowski

jacek.mlynarski@gmail.com (J.M.)
nanogrzybowski@gmail.com (B.A.G.)

HIGHLIGHTS

Only ~500 synthetically powerful "tactical combinations" have been cataloged to date

Computer discovers millions of such elegant, useful, and viable reaction sequences

Examples illustrate how the newly discovered combinations streamline synthetic planning

Gajewska et al., Chem 6, 1–14
January 9, 2020
https://doi.org/10.1016/j.chempr.2019.11.016

Please cite this article in press as: Gajewska et al., Algorithmic Discovery of Tactical Combinations for Advanced Organic Syntheses, Chem (2019), https://doi.org/10.1016/j.chempr.2019.11.016

Chem

CellPress

Article

Algorithmic Discovery of Tactical Combinations for Advanced Organic Syntheses

Ewa P. Gajewska,[1,5] Sara Szymkuć,[1,5] Piotr Dittwald,[1] Michał Startek,[2] Oskar Popik,[1] Jacek Mlynarski,[1,*] and Bartosz A. Grzybowski[1,3,4,6,*]

SUMMARY

Whereas most organic molecules can be synthesized from progressively simpler substrates, syntheses of complex organic targets often involve counterintuitive sequence of steps that first complexify the structure but, by doing so, open up possibilities for pronounced structural simplification in subsequent, downstream steps. Such complexifying/simplifying reaction sequences, called tactical combinations (TCs), can be quite powerful and elegant but also inherently hard to spot—indeed, only some 500 TCs have so far been cataloged, and even fewer are routinely used in synthetic practice. This paper describes computer-driven discovery of large numbers of viable TCs (over 46,000 combinations of reaction classes and ~4.85 million combinations of reaction variants), the vast majority of which have no prior literature precedent. Examples—including a concise wet lab synthesis of a small natural product—are provided to illustrate how the use of these newly discovered TCs can streamline the design of syntheses leading to important drugs and/or natural products.

INTRODUCTION

When planning syntheses of complex organic molecules, it is often not sufficient to gradually simplify the structure—instead, it may be beneficial to go through an intermediate that does not, per se, produce any immediate gain (or even intermittently increases structural complexity) but sets the scene for a "downstream" disconnection offering a significant structural simplification (Figures 1 and 2A). A few decades ago, Corey and Cheng[1] christened such sequences as "tactical combinations" (TCs)—since then, some of them have become a part of mainstream retrosynthetic thinking[2–6] (Figures 1A and 1B), some are less obvious and require a trained eye to spot[7,8] (Figures 1C–1E), and yet some others, used as key steps in syntheses of complex natural products, are truly remarkable, making one wonder how the authors were inspired to identify such an elegant combination[9,10] (Figures 1F and 1G). Indeed, the notion that TCs are "inspired" rather than "discovered" is reinforced by the fact that the largest collection cataloged in Ott[11] provides only ca. 500 combinations of suitable reaction types (see Section S3). On the other hand, TCs are composed of known reaction types, and one might reasonably hypothesize that more than just 500 are to be found among the myriad of possible two-step reaction combinations. Here, we validate this hypothesis via big-data analyses that allow us to enumerate and discover TCs in a systematic manner. Inspecting close to a billion two-step putative reaction sequences, we identify and rank over 46,000 combinations of reaction classes (and ca. 4.85 million combinations of reaction variants) that meet the definition of a TC. Remarkably, the vast majority of these TCs

The Bigger Picture

Although computers have recently made remarkable progress in autonomous synthetic planning, their ability to strategize over multiple steps, essential for the synthesis of complex targets, is still limited. One form of such "strategizing" is the so-called tactical combinations (TCs)—which are sequences of steps that first complexify the target but, by doing so, enable elegant and simplifying downstream disconnections. TCs are often counterintuitive even to human experts and, over several decades, only about 500 of them have been cataloged. Here, we show that computers can systematically discover much larger numbers (millions) of previously unreported yet valid TCs that unlock new and elegant synthetic approaches. These results and accompanying experimental demonstrations indicate that computers can now assist chemists not only by processing and adapting existing synthetic approaches (e.g., via artificial intelligence tools learning from prior art) but also by uncovering new ones.

(>95%) have not yet been reported in the literature, opening new avenues for elegant synthetic design by humans (with the help of a searchable strategist repository of TCs available to the academic community at http://strategist.grzybowskigroup.pl) or by computers, as illustrated by machine-designed syntheses in Figure 6. These opportunities are further illustrated by an experimental execution of a concise synthesis of both enantiomers of a small natural product in which the use of a newly discovered TC halves the number of steps compared to previously described routes. In a broader context, the current study is an example of how large-scale computer analyses can not only learn to mimic existing synthetic approaches (as in the now popular machine-learning methods[12–14]) but also intentionally create new organic-synthetic knowledge—an ability that, to date, has been an exclusive prerogative of organic-synthetic experts.

RESULTS AND DISCUSSION

Our current results build on over a decade of work on the Chematica platform for retrosynthetic design.[17–21] At the heart of this program is the knowledge base of ca. 75,000 reaction transforms/rules, each coded by expert chemists based on the underlying reaction mechanism (i.e., not just a particular reaction precedent) and carefully delineating substituent scope as well as contextual information about potential reactivity conflicts, protection requirements, selectivity issues, etc. (for examples, see Szymkuć et al.,[17] Supplemental Information to Klucznik et al.,[18] and Section S1). The quality of these reaction rules—essential for identifying meaningful TCs—has been discussed in Molga et al,[20] and corroborated by successful experimental execution of numerous Chematica-planned, multistep syntheses of high-value, medicinally relevant targets[18] and, most recently, non-trivial natural products.[22]

Using these rules for individual reactions, our objective here is to find and rank those two-step sequences in which the first reaction does not simplify, or even complexifies the structure (be it in terms of the number of heavy atoms, rings, or stereocenters), but the second one simplifies it considerably (Figure 2A). We note that the degree of structural complexification or simplification cannot usually be deduced based solely on the reaction rule that spans only the reaction center and its chemical environment (see examples in Figure 2B versus Figure 2C). Consequently, we have applied our reaction transforms to large numbers of structurally diverse scaffolds—initially, to 50,000 molecules chosen randomly from published journal articles and patents (stored in the Network of Organic Chemistry [NOC][19,21]), and later when the number of new TCs discovered in this collection of relatively simple molecules began to plateau—to an additional collection of 135,000 natural products and their synthetic precursors,[23] 1,800 approved drugs,[16] and finally to a collection of molecules we have assembled over the years to match all our reaction transforms (i.e., serve as test-cases during Chematica's development). Overall, we considered close to 220,000 molecules. For each of them, we applied Chematica's transforms matching the reaction cores as well as stereo- and regioselectivity. To probe as much of the synthetic space as possible, we treated the molecules either as targets from which we expanded two retrosynthetic generations (Figures 2D and 2E) and/or as intermediates in two-step sequences (one step in the retro and one in the forward direction, the latter using Chematica's inverted transforms). These operations were repeated for all targets, and in the end we generated in the order of a billion viable two-step reaction sequences; these calculations took several months of continuous operations on a 64-core workstation.

From this set of plausible two-step sequences, the TCs were selected and ranked according to the multiple criteria illustrated in Figure 3A. First, we kept only those for

[1]Institute of Organic Chemistry, Polish Academy of Sciences, ul. Kasprzaka 44/52, 01-224 Warsaw, Poland

[2]Faculty of Mathematics, Informatics, and Mechanics, University of Warsaw, 02-097 Warsaw, Poland

[3]Department of Chemistry, Ulsan National Institute of Science and Technology, 50 UNIST-gil, Ulsan 44919, Republic of Korea

[4]IBS Center for Soft and Living Matter, Ulsan National Institute of Science and Technology, 50 UNIST-gil, Ulsan 44919, Republic of Korea

[5]These authors contributed equally

[6]Lead Contact

*Correspondence: jacek.mlynarski@gmail.com (J.M.), nanogrzybowski@gmail.com (B.A.G.)

https://doi.org/10.1016/j.chempr.2019.11.016

Please cite this article in press as: Gajewska et al., Algorithmic Discovery of Tactical Combinations for Advanced Organic Syntheses, Chem (2019), https://doi.org/10.1016/j.chempr.2019.11.016

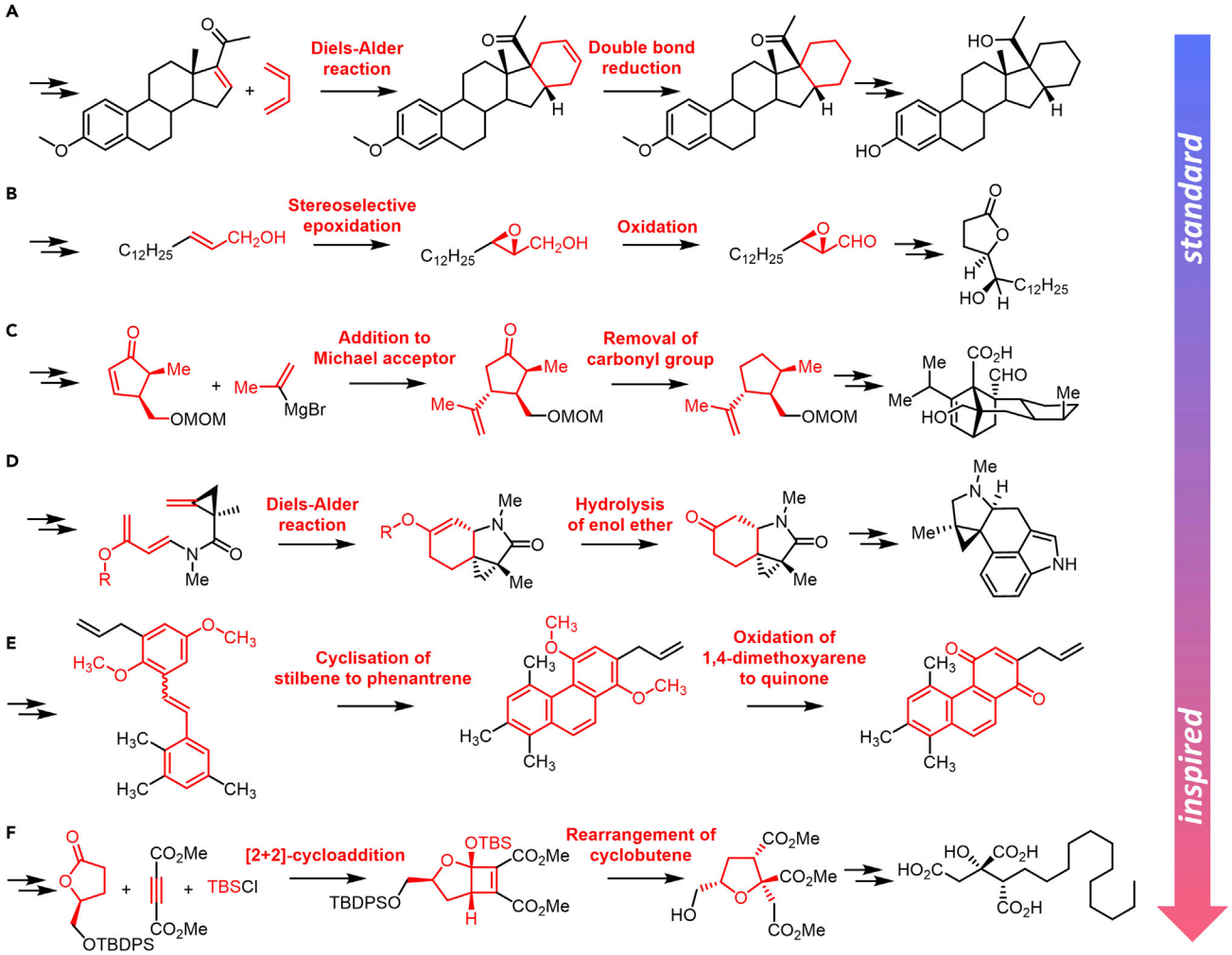


Figure 1. Examples of Known Tactical Combinations Ranging from "Standard" to "Inspired," All Also Rediscovered Algorithmically

The core of each TC is colored in red; the step closer to the target is not, by itself, structure simplifying but is essential for the key disconnection downstream.

(A) Thinking in the retrosynthetic direction (i.e., from the target to the substrates), one easily recognizes that the Diels-Alder reaction requires a double bond to be first introduced. This very popular TC is used here to synthesize a dihydroxy steroid, a new estrogen receptor α (ERα) antagonist, proposed by Kuznetsov et al.[4]

(B) One of the key intermediates in the synthesis of (4R,5R)-muricatacin[6] is an epoxy-aldehyde—seeing it, one is prompted to consider oxidation of an epoxy alcohol (no complexity decrease) anticipating a structure-simplifying asymmetric epoxidation reaction in the preceding step (large complexity gain, setting up two stereocenters and building a ring).

(C) In a perhaps less obvious TC, having a carbonyl on the cyclopentane intermediate is essential for stereoselective addition to the Michael acceptor—this sequence was used in the synthesis of sordaricin.[7]

(D) Retrosynthetic conversion of a ketone into an enol ether enables a Diels-Alder reaction. This TC was used by McCabe and Wipf[8] in the total synthesis of (−)-cycloclavine, allowing for stereoselective construction of a substituted 6-membered ring.

(E and F) Two TCs we consider quite "inspired": (E) the retrosynthetic "move" from quinone to 1,4-dimethoxyarene might, by itself, seem quite unproductive but is essential for the preceding cyclisation of stilbene into phenantrene. This TC was used[9] in the synthesis of a natural product derivative, 3-deoxyplectranthon A. (F) The acid-mediated rearrangement of cyclobutene diester sets the scene for the [2+2]-cycloaddition, constructing two stereocenters. This TC was essential for the synthesis of a citrate natural product[10]—(−)-CJ-13,982. For the statistics of known TCs most often used in the synthetic literature, see Section S3.

which some atoms are shared between the reaction cores of individual transformations. If two reactions occur at two distant loci of a molecule, they are in principle independent operations (give and take possible adjustments in the protecting groups) rather than a "synchronized" sequence that requires a particular order of steps, one

Please cite this article in press as: Gajewska et al., Algorithmic Discovery of Tactical Combinations for Advanced Organic Syntheses, Chem (2019), https://doi.org/10.1016/j.chempr.2019.11.016

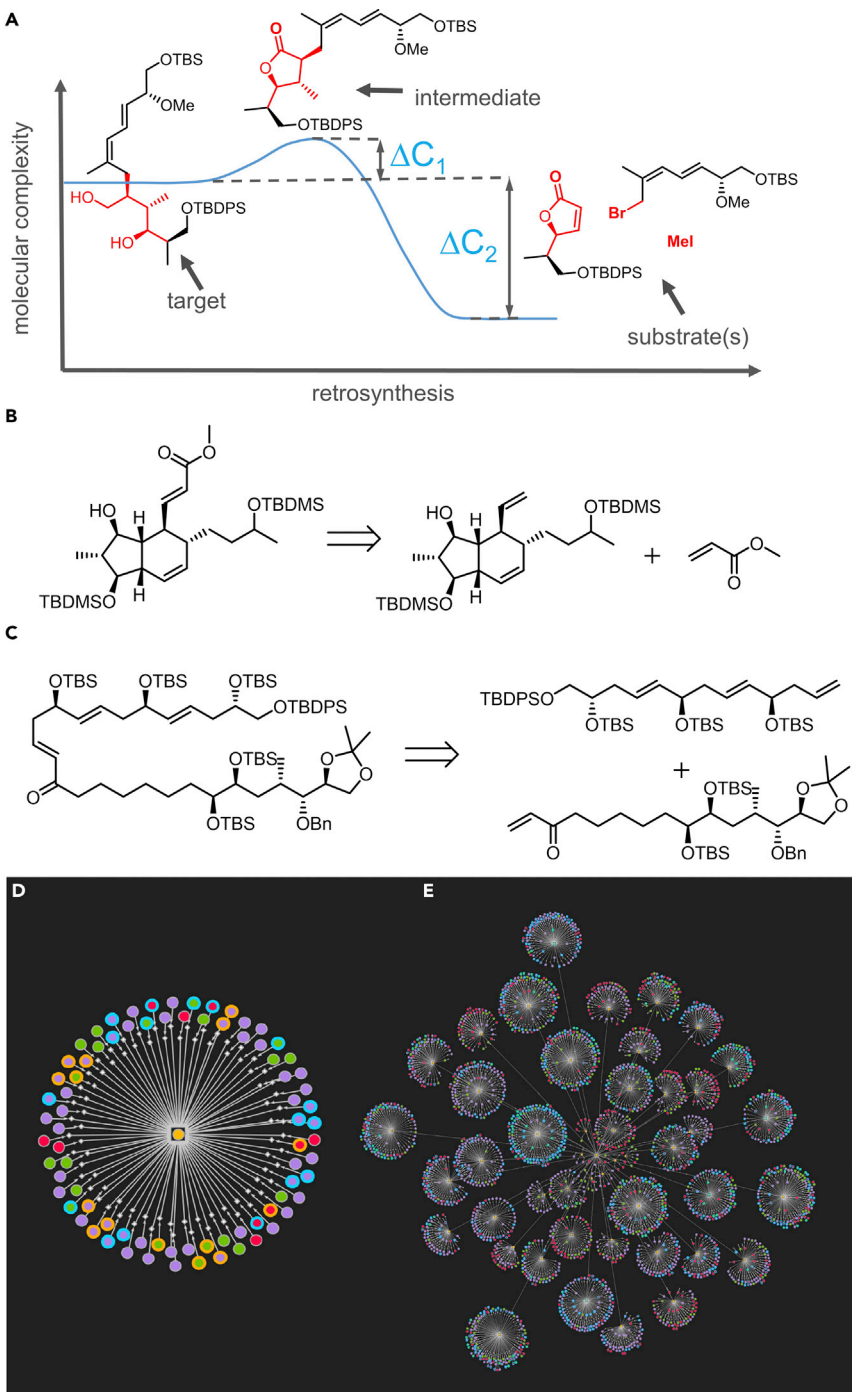


Figure 2. Quantification of and Searches for Tactical Combinations

(A) A plot illustrating how molecular complexity changes over a typical TC. The first retrosynthetic step does not simplify the structure or can even complexify it—here, it introduces a 5-membered ring in the intermediate. For this step, the complexity change in the retrosynthetic direction in non-negative, $\Delta C_1 = C_{intermediate} - C_{target} \geq 0$. The second retrosynthetic step offers significant structural simplification—here, alkenylation-alkylation sequence. For this step, the complexity change compared to the target (and, of course, the intermediate) is negative, $\Delta C_2 = C_{substrate} - C_{target} < 0$. We observe that these definitions are more stringent than for some of the TCs Corey cataloged[3,11]; therein, both steps could simplify the structure. However, such monotonic-complexity-decrease combinations of steps are, arguably, easily identified while

Please cite this article in press as: Gajewska et al., Algorithmic Discovery of Tactical Combinations for Advanced Organic Syntheses, Chem (2019), https://doi.org/10.1016/j.chempr.2019.11.016

Figure 2. *Continued*

planning chemical pathways one-step-at-a-time, and there is no need for combining them into sequences.

(B and C) When considering the measures of complexity and complexity change, it should be remembered that the same transformation can produce different complexity changes for different targets. A metathesis reaction used in the synthesis of hexacyclinic acid[15] in (B) offers only moderate complexity reduction but a significant one when used to make an intermediate of amphidinol[16] shown in (C).

(D) Chematica's screenshot in (D) shows a typical set of synthetic possibilities within one step of the central target molecule.

(E) When searching for TCs, the daughter nodes are further expanded (here, only a fraction of the nodes are expanded) and all two-step sequences are examined according to the criteria described in the main text and in Figure 3. Color coding of the nodes: yellow, target molecule; violet, unknown molecule; green, molecule known in literature; red, a commercially available chemical; orange halo, reactivity conflict detected; and blue, protection needed. Small, diamond-shaped nodes over reaction arrows signify reaction operations (for the theoretical background of these two-node-type, bipartite graphs of chemical reactions, see Szymkuć et al.,[17] Klucznik et al.,[18] Fialkowski et al.,[19] Molga et al.,[20] and Kowalik, et al.[21]).

enabling the other. Second, we inspected whether there exists a single reaction transform that can replace the two-step sequence—if so, this sequence is not an efficient TC. Third, we quantified the complexity increase/decrease between target, intermediate, and the substrate(s) and, with reference to Figure 2A, required that $\Delta C_1 = C_{intermediate} - C_{target} \geq 0$ but $\Delta C_2 = C_{substrate(s)} - C_{target} < 0$. As the measures of complexity, we considered the number of rings in a molecule, #R; the numbers of stereocenters, #S; and the lengths of molecules' SMILES (corrected for equivalent-complexity symbols such as @ and @@), which is a more faithful metric of structural complexity than just molecular mass (which can be dominated by few heavy atoms). We note that our conditions for ΔCs eliminated all trivial and non-productive protection/deprotection steps or sequences that produced the same substrates via two reactions gradually decreasing complexity (see Figure 3A). Fourth, we removed those sequences for which the core of the first transformation contained a functionality inherently incompatible (i.e., cross-reactive) with the second transformation (Figure S2). We emphasize that although there can certainly be conflicting groups present outside of the reaction core, such groups do not disqualify a general TC template as such but only its application to a particular target. Such out-of-the-core reactivity conflicts are recognized and removed from consideration during planning of specific syntheses. Altogether, these filtering procedures left ca. 46,000 TC candidates in terms of different reaction classes (and 250,000 unique pairs of reaction names).

We further extended this set by capitalizing on two interrelated facts: (1) that all ~75,000 of our expert-coded reaction rules are named and assigned to specific reaction types/classes (e.g., "1,4-conjugated addition," "Catellani reaction," etc.) and (2) there can be several reaction variants in each class that differ in substituent scope, reaction conditions, etc. Our searches might have identified a particular combination of variants (matching only a specific target) but missed other plausible combinations of variants belonging to the same reaction classes and matching other targets we have not inspected. We therefore considered such variants and counted them as TCs provided they matched other criteria outlined above (see example in Figure 3B). In this way, our set of TCs was extended to ~4.85 million combinations of reaction variants.

Finally, we ranked our TCs according to the degree of structural simplification they offer. In such rankings, we focused on the TCs' inherent simplifying power and made it independent of the target molecule (cf. Figures 2B and 2C). As discussed in detail

Please cite this article in press as: Gajewska et al., Algorithmic Discovery of Tactical Combinations for Advanced Organic Syntheses, Chem (2019), https://doi.org/10.1016/j.chempr.2019.11.016

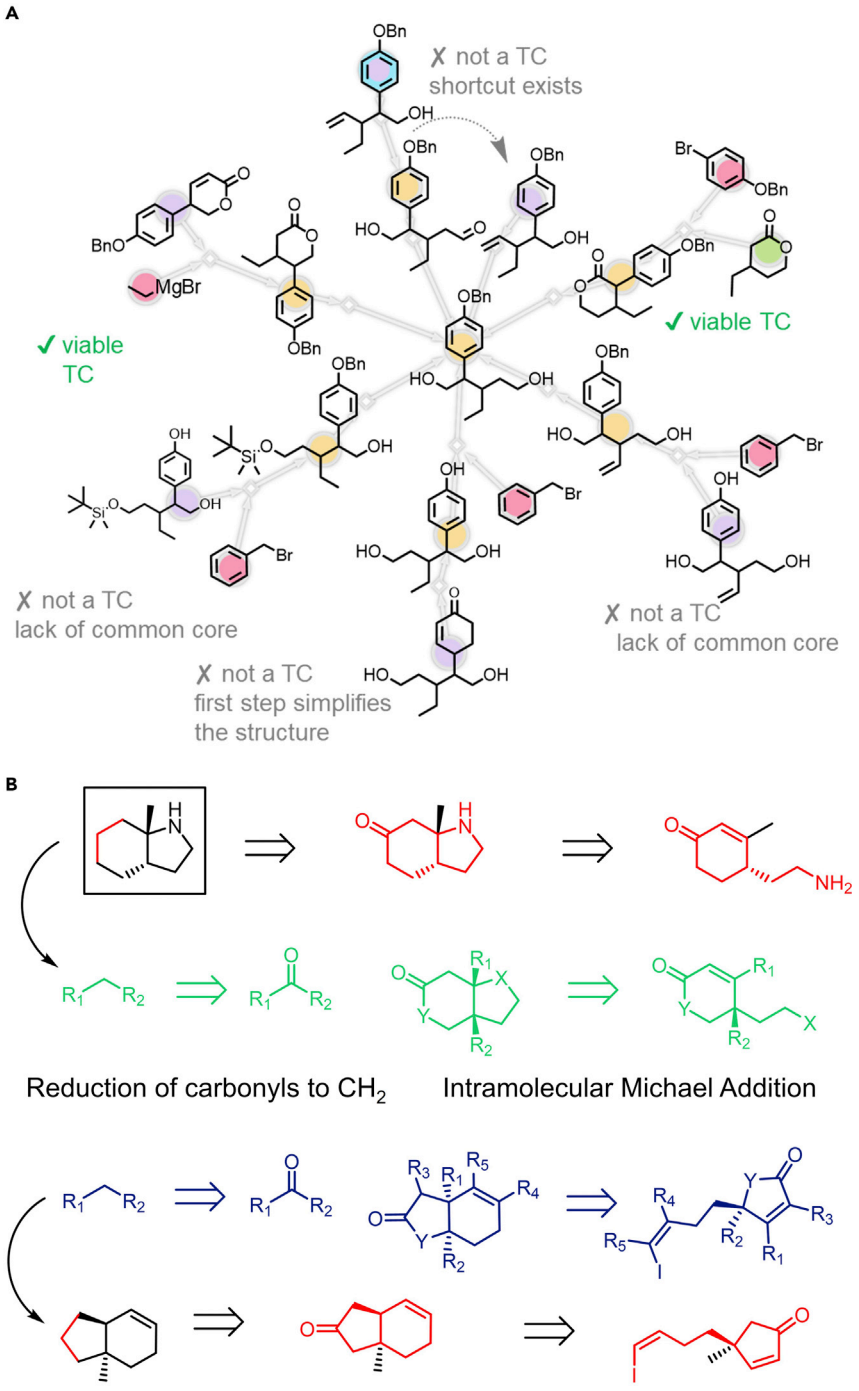


Figure 3. Selection and Extension of Tactical Combinations

(A) Scheme illustrating several key criteria used to distinguish "true" tactical combinations from "ordinary" two-step sequences (for discussion, see main text).

(B) Extrapolation from a TC found for a particular target to another combination of reaction variants from the same reaction classes. Here, a TC relying on the removal of a carbonyl oxygen and intramolecular Michael addition was identified for the specific target (octahydroindole derivative) shown in the upper-left frame. The combination of generalized reaction transforms is shown below (in green), along with class names. An analogous combination of transforms is also plausible,

Please cite this article in press as: Gajewska et al., Algorithmic Discovery of Tactical Combinations for Advanced Organic Syntheses, Chem (2019), https://doi.org/10.1016/j.chempr.2019.11.016

Figure 3. *Continued*
differing in the variant of the Michael addition (navy colored). For this variant, the ring size, stereochemistry requirements, and nucleophile (a vinyl organolithium reagent generated *in situ* from vinyl iodide, Piers et al.[24]) are different. The bottom panel illustrates how this extended strategy is applied to a target that was not in the original dataset.

in Section S2, we used three different measures and created three partly overlapping sets—one quantifying the simplification according to the number of rings created in the second retrosynthetic step, one according to the number of stereocenters created, and one according to the "centrality" of division of the target molecule into smaller fragments.

The TCs identified by the above analyses comprise the already-known TCs (minority, 3% in terms of reaction variants and 5% in terms of reaction types) as well as those not previously reported in the literature (majority, 97% and 95%, respectively). While the known TCs are, arguably, less interesting for the current work, their statistical analyses – based on literature precedents – reveal some important trends summarized in Figure 4 (see also Section S4 for additional correlations). With reference to this figure we observe that:

(1) The TCs are not biased to a specific collection of targets to which we matched reaction rules. In particular, Figure 4A indicates that TCs popular in the NOC[19,21] (i.e., matching many target molecules from the NOC) are also popular with respect to other, non-overlapping collections of targets (e.g., Zinc database).
(2) The popularity of a TC does not correlate with its score (Figure 4B)—in other words, if a given TC is used widely, it does not necessarily mean it offers a very significant structural simplification. What this finding also implies is that some rare combinations—including those that have not yet been recognized by the community—are potentially synthetically quite powerful (cf. below).
(3) Not surprisingly, more popular TCs allow, in general, production of more diverse targets (Figure 4C) and are also comprised of individual reactions that are themselves popular in synthetic practice (Figure 4D).

Taken together, these trends indicate that the existing repertoire of TCs is dominated by "popular" combinations of popular reactions—in fact, rather a limited number of familiar combinations listed in Section S3 and all rediscovered by our algorithm. Moreover, point (2) hints that this selection might just reflect expedience and/or habit rather than objective "simplifying power" of the TCs. In contrast, the millions of new TCs we identified are unbiased by prior art and many of them offer a very high degree of simplification (e.g., among the ring-creating TCs alone, there are 450 combinations with scores higher than any known example from the NOC; see Figure S11). Most importantly, these newly discovered TCs offer elegant and counterintuitive ways of making diverse scaffolds, as illustrated by examples in Figure 5.

While such examples indicate that the new TCs can considerably expand the scope and elegance of synthetic approaches, the question remains how to make practical use of this newly acquired knowledge—of course, remembering the millions of reaction combinations and applying them to one's target of interest "from memory" is not feasible. Accordingly, we created a Strategist webapp (available at http://strategist.grzybowskigroup.pl), whereby the user can query our TC collection by the names, keywords, or substructures characterizing the individual reactions or desired targets. We envision this easy-to-use portal (cf. Section S5 for a short user

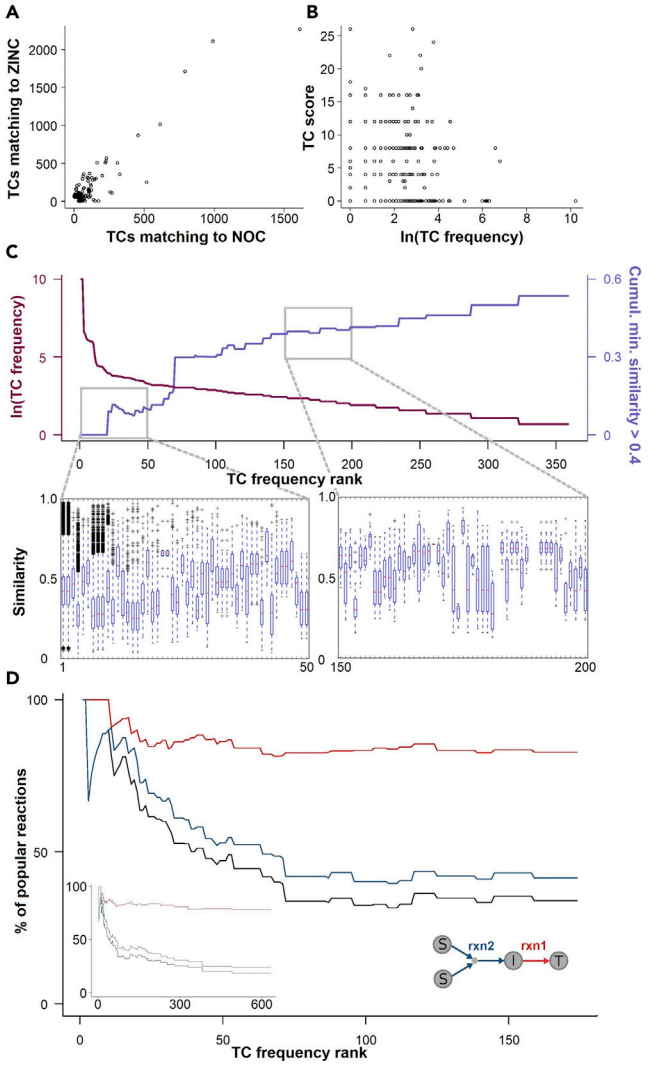


Figure 4. Statistical Trends within the Sets of Known Ring-Forming Tactical Combinations

(A) 25,000 molecules were chosen at random from the Network of Organic Chemistry (NOC)[19,21] and 25,000 different molecules from the zinc database (both sets composed of molecules of SMILES length <200 and containing at least one ring). The axes quantify to how many target molecules from a given collection (x axis, NOC and y axis, zinc) a given TC can be matched. For instance, a TC composed of the reduction of C=C double bond and Diels-Alder reactions can be applied to ~1,600 targets from the NOC and ~2,200 targets from zinc (i.e., molecules having a cyclohexane ring potentially prone to the TC) The 0.92 correlation (p value < 0.0001) in the graph provides evidence that the applicability of TCs does not depend on the set of targets.

(B) On the other hand, the frequency with which TCs are used (as reported in the literature) does not correlate with the score quantifying their simplifying power. Some rarely used TCs can offer very significant structural simplification. The frequency data on the x axis are based on the frequencies of known TCs in the NOC (~600 TC matches).

(C) The more popular strategies are applicable to more diverse targets. The x axis plots the rank of a TC (#1 = most popular), the left y axis and the pink line quantify the number of times a given TC appears in the NOC, and the right y axis and blue line quantify the structural diversity of the targets this TC is applicable to. This diversity measure, traced by the blue line, is the cumulative number of TCs for which the minimal similarity between each TC's targets is above 40%—as seen, for the most popular TCs, there are always pairs of very dissimilar targets whereas for less popular TCs, the targets are becoming more similar to each other. This trend is emphasized by the insets below showing raw target similarity distributions for TC's ranked 1–50 and 150–200 (boxplots show upper and lower quartiles as the boxes' borders, median as the red line, and outliers marked above/below

Please cite this article in press as: Gajewska et al., Algorithmic Discovery of Tactical Combinations for Advanced Organic Syntheses, Chem (2019), https://doi.org/10.1016/j.chempr.2019.11.016

Figure 4. *Continued*
whiskers). The similarity between any two targets is quantified by the Dice similarity coefficient based on Morgan fingerprints.
(D) Popular TCs are, in general, composed of synthetically popular reactions. Here, we consider an individual reaction to be popular if it has at least 300 literature precedents in the NOC. In the two most popular TCs (ranks #1 and #2 on the x axis), the reactions closer to the target (rxn1) are popular and so they receive a score of 100% (red line); same scores are for the second reactions (rxn2; blue line) and the TCs (i.e., both reactions composing a TC are popular, black line). In the third-most popular TC, only the first reaction is popular and so the cumulative scores are now 100% for the first reaction, 66.67% for the second, and 66.67% for TCs up to rank #3. The rest of the cumulative distribution is then constructed in an analogous manner up to a desired TC rank (~#170 in the main plot and ~#600 in the bottom-left inset). As seen, the less popular TCs are comprised of progressively less and less popular and more specialized individual reactions (especially the structure-simplifying reactions further from the target; rxn2, blue line).

manual) will serve as a useful idea generator for chemists wishing to test whether a particular intermediate they encounter during retrosynthetic planning might (or might not) fit a template of one or more TCs.

In parallel, and to further validate the correctness and usefulness of our TC collection in synthetic design, we incorporated it into the retrosynthetic Chematica platform. Previously, the program—akin to a competent but not yet expert chemist—was able to design efficient routes relying on gradual structural simplification of the intermediates it created.[18] At the same time, it was not really able to find truly "inspired" complexifying/simplifying sequences. Examples in Figures 6A–6C—autonomously designed by the computer and including pathways that replicate experimentally performed syntheses—demonstrate that upon incorporation of the TC collection, it now developed such expert ability.

For example, while designing the synthesis of levomilnacipran, an antidepressant drug marketed by Forest Laboratories, the machine made use of a TC—marked by red reaction arrows in Figure 6A—which first (in the retrosynthetic direction) forms a bicyclic system that is then simplified into epichlorohydrin and phenylacetonitrile. Remarkably, the pathway designed autonomously by Chematica is virtually identical to the patented route[34] and differs only in one of the starting materials (epichlorohydrin versus epihydrinamine). The synthesis of (+)-γ-lycorane, the degradation product of lycorine, an amaryllidaceae alkaloid shown in Figure 6B hinges on two TCs—early on, epoxide opening followed by oxidation of the alcohol and later, enantioselective reduction of cycloketone followed by an Appel reaction. Reassuringly, the route found by the machine also closely mirrors the literature approach of Liu et al.[35]

Moving away from routes closely resembling those already reported in the literature (and thus directly confirming the correctness of our TCs), Figure 6C shows Chematica's synthetic plan for a terpene-derived natural product, aphanamal, in which the sequence alkenylation-alkylation of the enone (employing *trans*-directing effect of the isopropyl group present in the 5-membered ring) followed by the reduction of the carbonyl group sets up two key stereocenters, including one quaternary center. This TC is known in literature[36] but, so far, has not been used for this or similar targets.

Most importantly, the TCs can help improve and shorten previously described routes. As a case in point, we considered a platelet aggregation inhibitor called imperanene.[33,37–42] To date, the syntheses allowing access to both the natural, (*S*) and unnatural (*R*) isomers entailed 8 to 11 steps[39,33] (see bottom portion of Figure 6D and Section S7). With the use of a previously unknown TC—that is, a previously

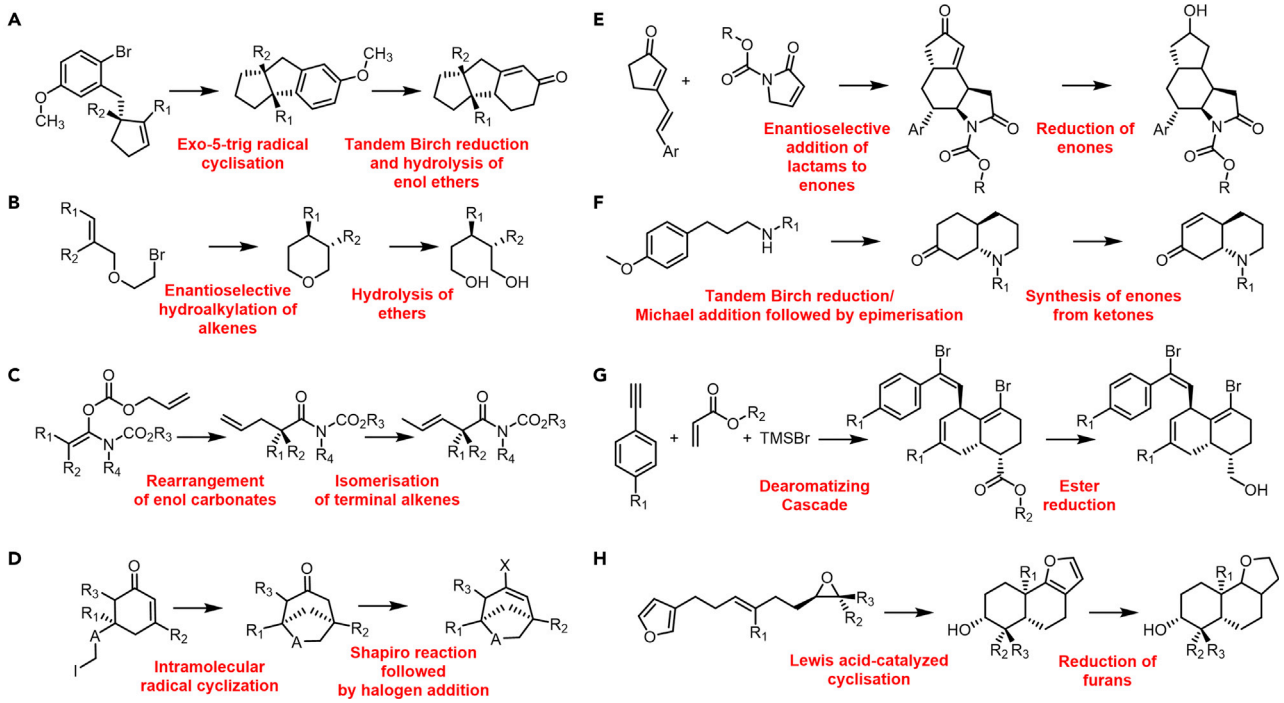


Figure 5. Examples of Tactical Combinations Unknown in the Literature

(A) A sequence of exo-5-trig radical cyclisation followed by tandem Birch reduction and hydrolysis of an enol ether constructs a *cis*-tricyclic ring system, found in the triquinane sesterterpenes.[25,26]

(B) Formation of the tetrahydropyran ring in the first retrosynthetic step sets the scene for the intramolecular, enantioselective hydroalkylation reaction, which installs two stereocenters. The structural motif produced by this sequence of reactions can be used in the synthesis of 3,4-disubstituted piperidine scaffolds[27] often present in biologically active compounds such as famoxetine.

(C) Enantioselective rearrangement of an enol carbonate constructs an acyclic quaternary carbon stereocenter. Subsequent isomerization of a terminal to an internal alkene forms a structural motif that can be further utilized in the synthesis of alkaloids.[28]

(D) A combination of intramolecular radical cyclisation and Shapiro reaction followed by halogen addition constructs a bicyclo<3.2.1>oct-2-ene skeleton from a simple cyclohexanone derivative. "A" stands for a carbon, oxygen, silicon, or nitrogen atom.

(E) Enantioselective addition of a lactam to an enone forms a fused tricyclic core with four stereocenters. Reduction of an enone to an unsaturated alcohol finalizes the TC.

(F) Tandem Birch reduction/Michael addition followed by epimerization and subsequent oxidation of a ketone to an enone enables the synthesis of a *trans* bicycle.

(G) Tactical combination of a dearomatizing cascade (involving two alkynes) and subsequent reduction of the ester group stereoselectively constructs a bicyclic core.

(H) A strategy based on cationic cyclisation of an epoxyalkene and reduction of the furan. This combination of reactions constructs a trialicyclic motif with diterpenoid moiety present in numerous natural products such as lanosterol[29] or myriceric acid.[30] The first reaction of the strategy—cationic cyclisation of epoxyalkene—has previously been used in the synthesis of natural products of this type (steviol, crotogoudin,[31] or aphidicolin[32]) but, to our best knowledge, has never been followed by the furan ring reduction in the second step of the strategy shown. We note that the tetrohydrofuran moiety formed in this TC via reduction of the furan enables functionalizations impossible to perform with the aromatic ring. For additional examples, see Figure S11. To query for other TCs, the reader is encouraged to use the webapp at http://strategist.grzybowskigroup.pl.

unreported complexifying/simplifying sequence of known reactions—comprising enantioselective addition of an epoxide to an aldehyde and reduction of benzyl alcohol, Chematica identified a significantly shorter (three steps plus two protections) approach that, for the sake of its elegance, we committed to experimental validation. The sequence detailed in the middle part of Figure 6D worked as predicted and with conditions as suggested by the machine. Specifically, the first step relied on Krische's methodology (using an Ir catalyst, based on (*R*)-SEGPHOS and 4-cyano-3-nitrobenzoate[43]), enabling enantioselective coupling of a vinyl epoxide (**1**) and protected vanillin (**2**) to give a diol (66% yield, 9:1 *anti*:*syn dr*) whose primary alcohol was tritylated in a 79% yield. Subsequently, Lewis-acid-catalyzed removal of the benzylic hydroxyl with concomitant deprotection of both primary and phenolic hydroxyl

Chem

CellPress

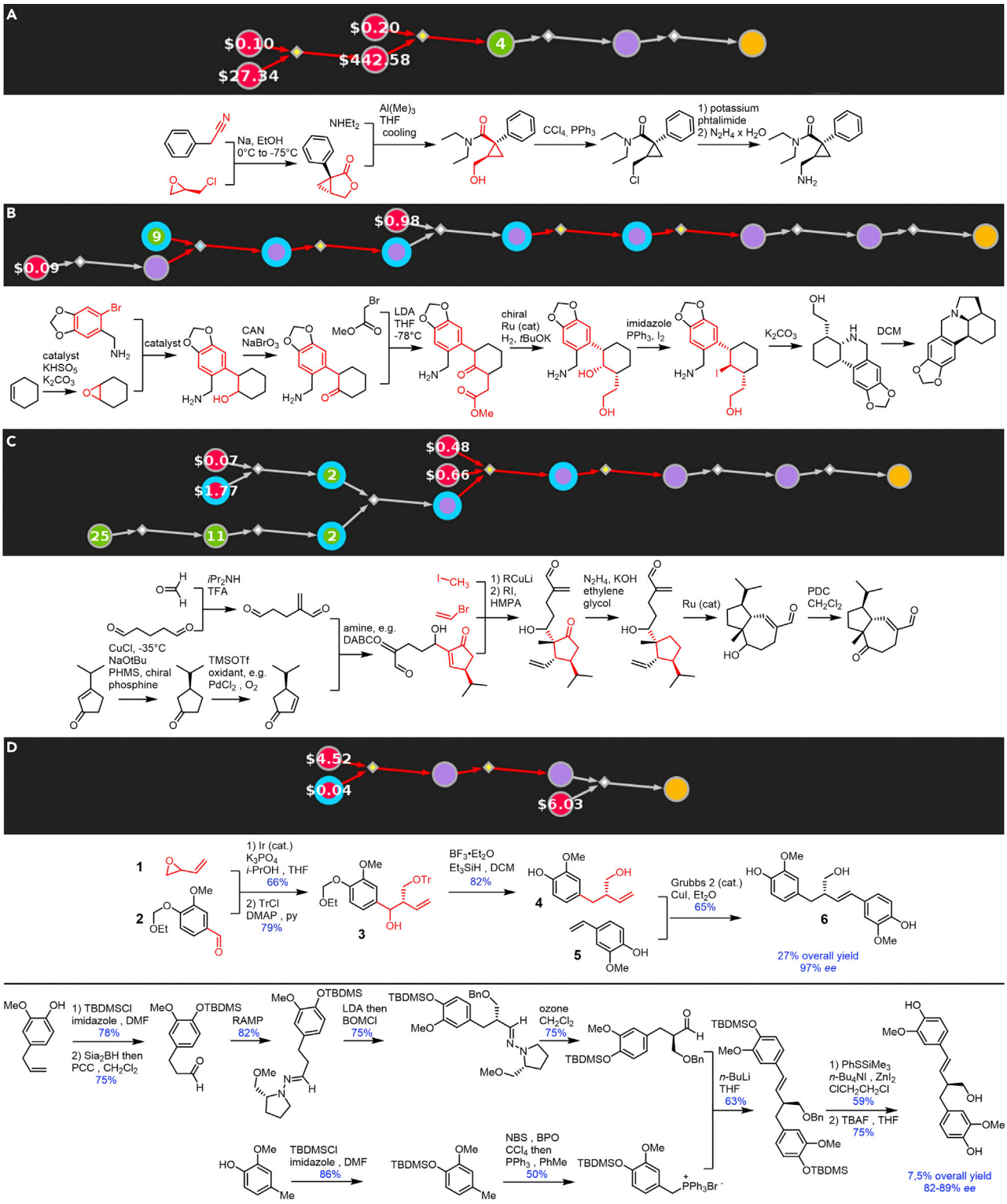


Figure 6. Syntheses of Medicinally Relevant Molecules and Natural Products Designed Autonomously by the Chematica Program with the Use of the TC Collection

Graphs on black background are screenshots from Chematica. Red arrows denote steps constituting a TC. Colors of the nodes are: yellow, target molecule; violet, unknown molecule; green, molecule known in literature; red, a commercially available chemical; and blue halo, protection needed. Numbers on red nodes are prices per gram from the Sigma-Aldrich catalog. Numbers on green nodes denote in how many ways this known molecule

Figure 6. *Continued*
has been previously made (the so-called "synthetic popularity"). All searches were set to terminate only upon finding relatively inexpensive and/or synthetically popular starting materials. The searches took a few minutes on a 64-core workstation. For all four targets, steps forming a TC are indicated by red reaction arrows in Chematica's screenshots and red-colored TC cores in the corresponding intermediates. The targets are: (A) levomilnacipran, an antidepressant drug; (B) (+)-γ-lycorane, the degradation product of lycorine (an Amaryllidaceae alkaloid); (C) aphanamal, a terpene-derived natural product; and (D) imperanene, a platelet aggregation inhibitor. In (D), the top portion is Chematica's miniature, whereas the middle portion is the corresponding, experimentally executed plan (with conditions and yields given next to the reaction arrows and with the TC core colored red). The bottom portion is the shortest literature-reported route, from Shattuck et al.,[33] allowing for the preparation of both (*R*) and (*S*) enantiomers of imperanene. For a summary of other syntheses of imperanene, see Section S7.

groups from compound (**3**) produced alkene (**4**) in an 82% yield. The final step was olefin metathesis, using a second-generation Grubbs catalyst and copper(I) iodide,[44] of (**4**) with a commercially available styrene derivative (**5**). Metathesis worked in a 65% yield and gave the (*S*)-imperanene target with a 97% *ee* and overall pathway yield of 27%. We note that by using an enantiomeric iridium catalyst, the route is easily adapted to unnatural (*R*)-imperanene, which we also obtained (98% *ee*, 28% overall yield; for all synthetic details, see Section S8).

Conclusions

Such synthetic examples lead us to conclude that the use of the TCs we discovered can, indeed, become a valuable addition to the toolkit of organic-synthetic experts. We recognize that at least some TCs we described could also be identified by human experts; however, computers can automate this discovery process and perform it on a scale that an individual brain cannot handle. In this context, we note that our approach could be extended to the discovery of longer sequences (e.g., up/up/down or up/no-change/down in terms of complexity), but it must be remembered that searches for such sequences would require expansion of more synthetic generations, *N*, and would thus scale as M^N (where $M \sim O(100)$ is the average number of synthetic possibilities at each step[17,18]). From a practical point of view, we believe that resources such as Chematica or Strategist can foster implementation of new and creative synthetic approaches—in particular, if a steadily simplifying sequence leading to one's compound of interest is problematic, one can now query, within seconds, Strategist's TCs' compendium to rapidly find a complexifying/simplifying reaction sequence with which to overcome the limiting synthetic "roadblock."

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.chempr.2019.11.016.

ACKNOWLEDGMENTS

The authors thank the U.S. DARPA for generous support under the "Make-It" award, 69461-CH-DRP #W911NF1610384. B.A.G. gratefully acknowledges personal support from the Institute for Basic Science Korea, project code IBS-R020-D1. J.M. and O.P. thank the Foundation for Polish Science for financial support under award TEAM/2017-4/38.

AUTHOR CONTRIBUTIONS

E.P.G. and S.S. coded a large portion of Chematica's reaction rules, designed the criteria to filter and rank the strategies, inspected the results, and selected the synthetic examples used in the paper. P.D. and M.S. developed computer codes for strategy searches, selection, and filtering. P.D. developed the webapp portal. O.P. synthesized both enantiomers of imperanene. J.M. supervised the synthesis. B.A.G. conceived the project and supervised the research. All authors participated in the writing of the manuscript.

Please cite this article in press as: Gajewska et al., Algorithmic Discovery of Tactical Combinations for Advanced Organic Syntheses, Chem (2019), https://doi.org/10.1016/j.chempr.2019.11.016

DECLARATION OF INTERESTS

The authors declare no competing interests. While Chematica was originally developed and owned by B.A.G.'s Grzybowski Scientific Inventions, LLC, neither he nor the co-authors hold any stock in this company, which is now property of Merck KGaA, Darmstadt, Germany. The authors continue to collaborate with Merck KGaA within the DARPA "Make-It" award. All queries about access options to Chematica (now rebranded as Synthia), including academic collaborations, should be directed to Dr. Sarah Trice at sarah.trice@sial.com.

Received: August 21, 2019
Revised: October 2, 2019
Accepted: November 18, 2019
Published: December 19, 2019

REFERENCES

1. Corey, E.J., and Cheng, X.-M. (1989). The Logic of Chemical Synthesis (Wiley).

2. Corey, E.J. (1988). Robert Robinson lecture. Retrosynthetic thinking—essentials and examples. Chem. Soc. Rev. *17*, 111–133.

3. Long, A.K., and Kappos, J.C. (1994). Computer-assisted synthetic analysis. Performance of tactical combinations of transforms. J. Chem. Inf. Model. *34*, 915–921.

4. Kuznetsov, Y.V., Levina, I.S., Scherbakov, A.M., Andreeva, O.E., Fedyushkina, I.V., Dmitrenok, A.S., Shashkov, A.S., and Zavarzin, I.V. (2018). New estrogen receptor antagonists. 3,20-Dihydroxy-19-norpregna-1,3,5(10)-trienes: synthesis, molecular modeling, and biological evaluation. Eur. J. Med. Chem. *143*, 670–682.

5. Nicolaou, K.C., Snyder, S.A., Montagnon, T., and Vassilikogiannakis, G. (2002). The Diels-Alder reaction in total synthesis. Angew. Chem. Int. Ed. *41*, 1668–1698.

6. van Aar, M.P.M., Thijs, L., and Zwanenburg, B. (1995). Synthesis of (4R,5R)-muricatacin and its (4R,5S)-analog by sequential use of the photo-induced rearrangement of epoxy diazomethyl ketones. Tetrahedron *51*, 11223–11234.

7. Mander, L.N., and Thomson, R.J. (2003). Total synthesis of Sordaricin. Org. Lett. *5*, 1321–1324.

8. McCabe, S.R., and Wipf, P. (2017). Eight-step enantioselective total synthesis of (−)-cycloclavine. Angew. Chem. Int. Ed. *56*, 324–327.

9. Kaliakoudas, D., Eugster, C.H., and Rüedi, P. (1990). Synthese von Plectranthonen, diterpenoiden phenanthren-1,4-chinonen. Helv. Chim. Acta *73*, 48–62.

10. Atkin, L., Chen, Z., Robertson, A., Sturgess, D., White, J.M., and Rizzacasa, M.A. (2018). Synthesis of alkyl citrates (−)-CJ-13,981, (−)-CJ-13,982, and (−)-L-731,120 via a cyclobutene diester. Org. Lett. *20*, 4255–4258.

11. Ott, M. (2004). LHASA (Centre for Molecular and Biomolecular Informatics). http://cheminf.cmbi.ru.nl/cheminf/lhasa/.

12. Coley, C.W., Green, W.H., and Jensen, K.F. (2018). Machine learning in computer-aided synthesis planning. Acc. Chem. Res. *51*, 1281–1289.

13. Segler, M.H.S., Preuss, M., and Waller, M.P. (2018). Planning chemical syntheses with deep neural networks and symbolic AI. Nature *555*, 604–610.

14. Liu, B., Ramsundar, B., Kawthekar, P., Shi, J., Gomes, J., Luu Nguyen, Q.L., Ho, S., Sloane, J., Wender, P., and Pande, V. (2017). Retrosynthetic reaction prediction using neural sequence-to-sequence models. ACS Cent. Sci. *3*, 1103–1113.

15. Stellfeld, T., Bhatt, U., and Kalesse, M. (2004). Synthesis of the A, B,C-ring system of hexacyclinic acid. Org. Lett. *6*, 3889–3892.

16. 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., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 2017 Nov 8. https://doi.org/10.1093/nar/gkx1037. https://www.drugbank.ca/drugs.

17. Szymkuć, S., Gajewska, E.P., Klucznik, T., Molga, K., Dittwald, P., Startek, M., Bajczyk, M., and Grzybowski, B.A. (2016). Computer-assisted synthetic planning: the end of the beginning. Angew. Chem. Int. Ed. *55*, 5904–5937.

18. Klucznik, T., Mikulak-Klucznik, B., McCormack, M.P., Lima, H., Szymkuć, S., Bhowmick, M., Molga, K., Zhou, Y., Rickershauser, L., Gajewska, E.P., et al. (2018). Efficient syntheses of diverse, medicinally relevant targets planned by computer and executed in the laboratory. Chem *4*, 522–532.

19. Fialkowski, M., Bishop, K.J.M., Chubukov, V.A., Campbell, C.J., and Grzybowski, B.A. (2005). Architecture and evolution of organic chemistry. Angew. Chem. Int. Ed. *44*, 7263–7269.

20. Molga, K., Gajewska, E.P., Szymkuć, S., and Grzybowski, B.A. (2019). The logic of translating chemical knowledge into machine-processable forms: a modern playground for physical-organic chemistry. React. Chem. Eng. *4*, 1506–1521.

21. Kowalik, M., Gothard, C.M., Drews, A.M., Gothard, N.A., Weckiewicz, A., Fuller, P.E., Grzybowski, B.A., and Bishop, K.J.M. (2012). Parallel optimization of synthetic pathways within the network of organic chemistry. Angew. Chem. Int. Ed. *51*, 7928–7932.

22. Grzybowski, B.A. (2018). Computational design and experimental validation of synthetic routes created by Chematica. Abstr. Pap. Am. Chem. Soc. *6*.

23. Irwin, J.J., Sterling, T., Mysinger, M.M., Bolstad, E.S., and Coleman R, G. (2012). ZINC: a free tool to discover chemistry for biology. J. Chem. Inf. Model. *52*, 1757–1768.

24. Piers, E., Harrison, C.L., and Zetina-Rocha, C. (2001). Intramolecular conjugate addition of alkenyl and aryl functions to enones initiated by lithium−iodine exchange. Org. Lett. *3*, 3245–3247.

25. Wright, J., Drtina, G.J., Roberts, R.A., and Paquette, L.A. (1988). A convergent synthesis of triquinane sesterterpenes. Enantioselective synthesis of (-)-retigeranic acid A. J. Am. Chem. Soc. *110*, 5806–5817.

26. Paquette, L.A., Wright, J., Drtina, G.J., and Roberts, R.A. (1987). Enantiospecific total synthesis of natural (-)-retigeranic acid A and two (-)-retigeranic acid B candidates. J. Org. Chem. *52*, 2960–2962.

27. Igarashi, J., Ishiwata, H., and Kobayashi, Y. (2004). Concise synthesis of *trans*- and *cis*-3,4-disubstituted piperidines based on regio- and stereoselective allylation of cyclopentenyl esters. Tetrahedron Lett. *45*, 8065–8068.

28. Nidhiry, J.E., and Prasad, K.R. (2013). Enantiospecific total synthesis of indole alkaloids (+)-eburnamonine, (−)-aspidospermidine and (−)-quebrachamine. Tetrahedron *69*, 5525–5536.

29. Tian, Y., Xu, X., Zhang, L., and Qu, J. (2016). Tetraphenylphosphonium tetrafluoroborate/1,1,1,3,3,3-hexafluoroisopropanol (Ph_4PBF_4/HFIP) effecting epoxide-initiated cation–olefin polycyclizations. Org. Lett. *18*, 268–271.

30. Lu, J., Aguilar, A., Zou, B., Bao, W., Koldas, S., Shi, A., Desper, J., Wangemann, P., Xie, X.S., and Hua, D.H. (2015). Chemical synthesis of tetracyclic terpenes and evaluation of antagonistic activity on endothelin-A receptors

and voltage-gated calcium channels. Bioorg. Med. Chem. *23*, 5985–5998.

31. Song, L., Zhu, G., Liu, Y., Liu, B., and Qin, S. (2015). Total synthesis of atisane-type diterpenoids: application of Diels-Alder cycloadditions of podocarpane-type unmasked *ortho*-benzoquinones. J. Am. Chem. Soc. *137*, 13706–13714.

32. Tanis, S.P., Chuang, Y.-H., and Head, D.B. (1985). A formal total synthesis of (±)-aphidicolin. Tetrahedron Lett. *26*, 6147–6150.

33. Shattuck, J.C., Shreve, C.M., and Solomon, S.E. (2001). Enantioselective synthesis of imperanene, a platelet aggregation inhibitor. Org. Lett. *3*, 3021–3023.

34. Nicolas, M., Hellier, P., Diard, C., and Subra, L. (2013). Method for synthesis of (1S, 2R)-milnacipran. US Patent US8604241 B2, filed January 29, 2010, and published December 10, 2013.

35. Liu, C., Xie, J.H., Li, Y.L., Chen, J.Q., and Zhou, Q.L. (2013). Asymmetric hydrogenation of α,α′-disubstituted cycloketones through dynamic kinetic resolution: an efficient construction of chiral diols with three contiguous stereocenters. Angew. Chem. Int. Ed. *52*, 593–596.

36. Barbe, G., and Charette, A.B. (2008). Total synthesis of (+)-Lepadin B: stereoselective synthesis of nonracemic polysubstituted hydroquinolines using an RC-ROM process. J. Am. Chem. Soc. *130*, 13873–13875.

37. Eklund, P.C., Riska, A.I., and Sjöholm, R.E. (2002). Synthesis of R-(−)-imperanene from the natural lignan hydroxymatairesinol. J. Org. Chem. *67*, 7544–7546.

38. Davies, H.M.L., and Jin, Q. (2003). Intermolecular C–H activation at benzylic positions: synthesis of (+)-imperanene and (−)-α-conidendrin. Tetrahedron Asymmetry *14*, 941–949.

39. Carr, J.A., and Bisht, K.S. (2004). Enantioselective synthesis of imperanene via enzymatic asymmetrization of an intermediary 1,3-diol. Org. Lett. *6*, 3297–3300.

40. Doyle, M.P., Hu, W., and Valenzuela, M.V. (2002). Total synthesis of (S)-(+)-imperanene. Effective use of regio- and enantioselective intramolecular carbon−hydrogen insertion reactions catalyzed by chiral dirhodium(II) carboxamidates. J. Org. Chem. *67*, 2954–2959.

41. Takashima, Y., and Kobayashi, Y. (2009). Synthesis of (S)-imperanene by using allylic substitution. J. Org. Chem. *74*, 5920–5926.

42. Egi, M., Sugiyama, K., Saneto, M., Hanada, R., Kato, K., and Akai, S. (2013). A mesoporous-silica-immobilized oxovanadium cocatalyst for the lipase-catalyzed dynamic kinetic resolution of racemic alcohols. Angew. Chem. Int. Ed. *52*, 3654–3658.

43. Feng, J., Garza, V.J., and Krische, M.J. (2014). Redox-triggered C–C coupling of alcohols and vinyl epoxides: diastereo- and enantioselective formation of all-carbon quaternary centers via tert-(hydroxy)-prenylation. J. Am. Chem. Soc. *136*, 8911–8914.

44. Voigtritter, K., Ghorai, S., and Lipshutz, B.H. (2011). Rate enhanced olefin cross-metathesis reactions: the copper iodide effect. J. Org. Chem. *76*, 4697–4702.