nature communications

Article

https://doi.org/10.1038/s41467-024-54611-5

Systematic, computational discovery of multicomponent and one-pot reactions

Received: 15 November 2024

Accepted: 18 November 2024

Published online: 27 November 2024

Check for updates

Rafał Roszak[1,6], Louis Gadina[2,3,6], Agnieszka Wołos[1,6], Ahmad Makkawi[2], Barbara Mikulak-Klucznik[1], Yasemin Bilgi[2,3], Karol Molga[1,2], Patrycja Gołębiowska[2], Oskar Popik[2], Tomasz Klucznik[1], Sara Szymkuć[1], Martyna Moskal[1], Sebastian Baś[2,4], Rafał Frydrych[2,3], Jacek Mlynarski[2], Olena Vakuliuk[2], Daniel T. Gryko[2] ✉ & Bartosz A. Grzybowski[2,3,5] ✉

Discovery of new types of reactions is essential to organic chemistry because it expands the scope of accessible molecular scaffolds and can enable more economical syntheses of existing structures. In this context, the so-called multicomponent reactions, MCRs, are of particular interest because they can build complex scaffolds from multiple starting materials in just one step, without purification of intermediates. However, for over a century of active research, MCRs have been discovered rather than designed, and their number remains limited to only several hundred. This work demonstrates that computers taught the essential knowledge of reaction mechanisms and rules of physical-organic chemistry can design – completely autonomously and in large numbers – mechanistically distinct MCRs. Moreover, when supplemented by models to approximate kinetic rates, the algorithm can predict reaction yields and identify reactions that have potential for organocatalysis. These predictions are validated by experiments spanning different modes of reactivity and diverse product scaffolds.

Computational discovery of new reaction classes is one of the holy grails of chemoinformatics, with first efforts by Ivar Ugi[1–4] dating back to 1970s. In this context, reactions that build complex scaffolds from multiple simple components in one step (i.e., multicomponent reactions, MCRs[5–11]; Fig. 1a) and/or proceed sequentially in one pot[12–14] are of particular interest as they minimize separation and purification operations, and increase the overall step- and atom-economy[15] as well as "greenness"[16,17] of synthesis. However, the number of known MCR classes remains limited to several hundred (Fig. 1b, c), perhaps because the most popular reactivity patterns (e.g., isocyanide, β-dicarbonyl, or imine-based MCRs) and their straightforward combinations[18] and extensions[19–21] have been studied in nearly exhaustive detail. Rational discovery of MCRs remains difficult because it entails understanding and analysis of intricate networks of mechanistic steps spanning multiple substrates, intermediates, and side reactions that can hijack the desired multicomponent sequence. Here, we show that computers equipped with broad knowledge of mechanistic transforms, rules of physical-organic chemistry, and approximations of kinetic rates can perform such network analyses rapidly and in a high-throughput manner, and can guide systematic discovery, ranking, and yield estimation of mechanistically distinct types of MCRs, one-pot sequences and even organocatalytic reactions, several of which we validate by experiment. These results evidence that synthesis-planning algorithms are no longer limited to skillful manipulation of the existing knowledge-base of full reactions[22–28] but can assist in its creative expansion.

Every chemical reaction is a sequence of elementary steps or, at a less precise but very popular representation, of arrow-pushing

[1]Allchemy Inc., Highland, IN, USA. [2]Institute of Organic Chemistry, Polish Academy of Sciences, Warsaw, Poland. [3]Center for Algorithmic and Robotized Synthesis (CARS), Institute for Basic Science (IBS), Ulsan 44919, Republic of Korea. [4]Jagiellonian University, Krakow, Poland. [5]Department of Chemistry, Ulsan Institute of Science and Technology, UNIST, Ulsan 44919, Republic of Korea. [6]These authors contributed equally: Rafał Roszak, Louis Gadina, Agnieszka Wołos. ✉e-mail: daniel.gryko@icho.edu.pl; nanogrzybowski@gmail.com

https://doi.org/10.1038/s41467-024-54611-5

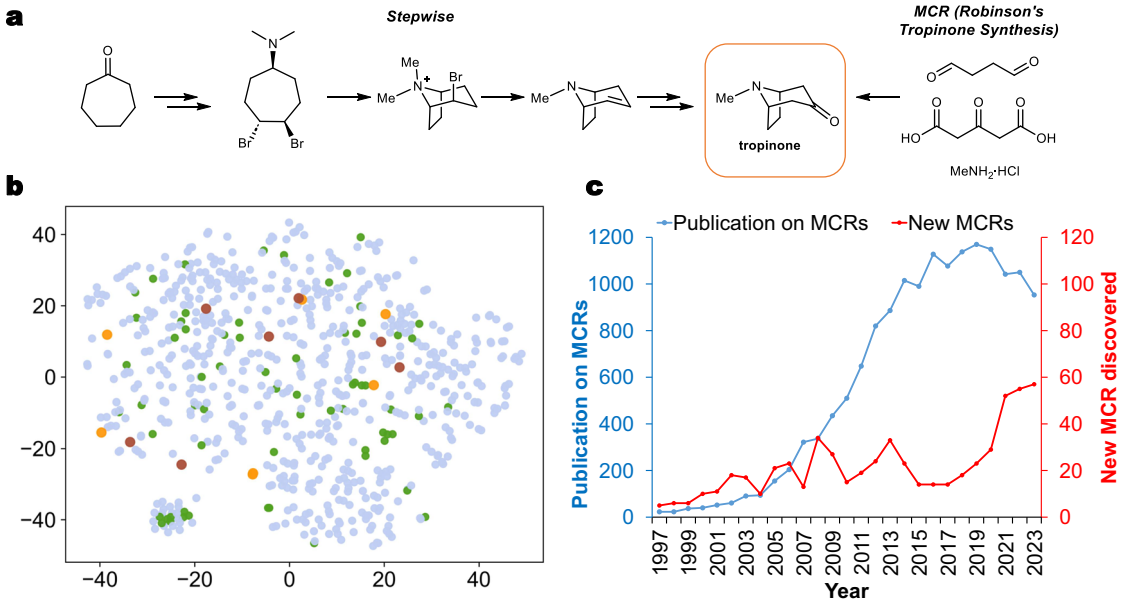


Fig. 1 | Significance and current discovery rate of multicomponent reactions. **a** A classic example illustrating the elegance and efficiency of Robinson's one-step, MCR synthesis of tropinone vs. prior, fifteen-step synthesis[10]. In the latter, only two key steps are shown. **b** t-SNE projection "map" illustrating diversity of 631 known MCR classes/types (smaller blue markers) and 66 one-pot classes (green) vs. the MCRs and one-pots (larger red and orange markers, respectively) discovered in this work and validated by experiment. The known MCR and one-pot classes were curated by our group over the years based on several extensive literature reviews – all this data (631 + 66) is available for download, along with the links to the first publication reporting a given reaction type, as either a .csv or Excel file from https://doi.org/10.5281/zenodo.10817102. In the map, each marker corresponds to t-SNE projection of reaction fingerprints, defined as a difference between fingerprint of the product and fingerprint of the substrates[83] (i.e., difference across a full reaction, not its mechanistic steps). The interactive t-SNE map is deposited at https://mcrmap.allchemy.net. **c** Blue line and left axis quantify the numbers of papers on MCRs published in a given year (based on "multicomponent reaction" query of the Web of Knowledge database, August 2024). Red line and right axis plot are based on the set of 631 MCR types from **b**. For each year, the number of newly discovered MCR types (i.e., published for the first time in this year) is plotted. The number of publications on MCRs peaked around 2019 and has slightly decreased since. On the other hand, the discovery rate of new MCRs seems to have followed cyclical variations. It should be noted, however, that since the nadir in 2015–2017, it is now increasing perceptibly, perhaps signaling renewed interest in multicomponent reactions.

steps[29], which has been used in computational chemistry for decades[30–35] (though in most cases to analyze only certain types of chemistries and with limited accuracy, see Supplementary Section S6 in ref. 36 and Supplementary Section S3 here). As we have recently shown for complex carbocationic rearrangements[36], this level of description is appealing because, compared to quantum methods, it reduces the number of degrees of freedom one needs to consider, while still retaining enough accuracy to rationalize the mechanisms of the vast majority of organic chemical transformations, including the previously unpreported reactions[37,38]. In this work, we use a large and diverse collection of arrow-pushing operators to generate networks of mechanistic steps starting from sets of multiple substrates potentially exhibiting different modes of reactivity. We then aim to identify the mechanistic pathways and conditions that would select only some of these modes and would proceed, in one pot, cleanly into products significantly more complex than the starting materials. Uniquely and mindful of various cross-reactivities possible in multicomponent reaction mixtures, we consider possible by-products, products of side reactions, and further reactions of these species as well as their potential interference with the main mechanistic pathway. We scrutinize these processes for kinetics to ensure that side-processes do not hijack the desired sequence, lowering or even nullifying its yield, which we also aim to approximate. Within this general approach, the problem of designing MCRs or one-pot sequences becomes one of selecting the substrates, expanding the mechanistic networks forward and sideways from these substrates, and performing kinetic analysis to trace conflict-free mechanistic routes (Fig. 2).

Results

Choice of substrates

While the algorithm accepts any user-specified molecules as input, guessing the substrates resulting in productive MCRs may be challenging. Instead, we rely on a high-throughput, computational analyses of substrate combinations from a house-curated collection of ca. 2400 simple, diverse and commercially available small molecules featuring one or two groups reactive in various types of transformations (Fig. 2a and, for details, Methods and Supplementary Section S4).

Mechanistic transforms

To propagate the mechanistic networks, a collection of ~8000 commonly accepted mechanistic transforms was encoded at the aforementioned arrow-pushing level in the SMARTS notation as described before[39,40]. This collection includes a broad range of chemistries although it is certainly not yet without omissions (see Methods). Transforms account for by-products (Fig. 2b) and are categorized according to typical reaction conditions, temperature range and water tolerance, as well as typical speeds (very slow, slow, fast, very fast, and uncertain if conflicting literature data have been reported, VS-S-F-VF-U). Since the focus of the algorithm is to generate scaffolds not yet described in the literature, the algorithm does not consider stereochemistry. For more details on rule coding, see Methods and Supplementary Section S5.

Forward expansion of mechanistic networks

For a given set of substrates (henceforth, synthetic generation G_0), the algorithm applies the mechanistic transforms to create the

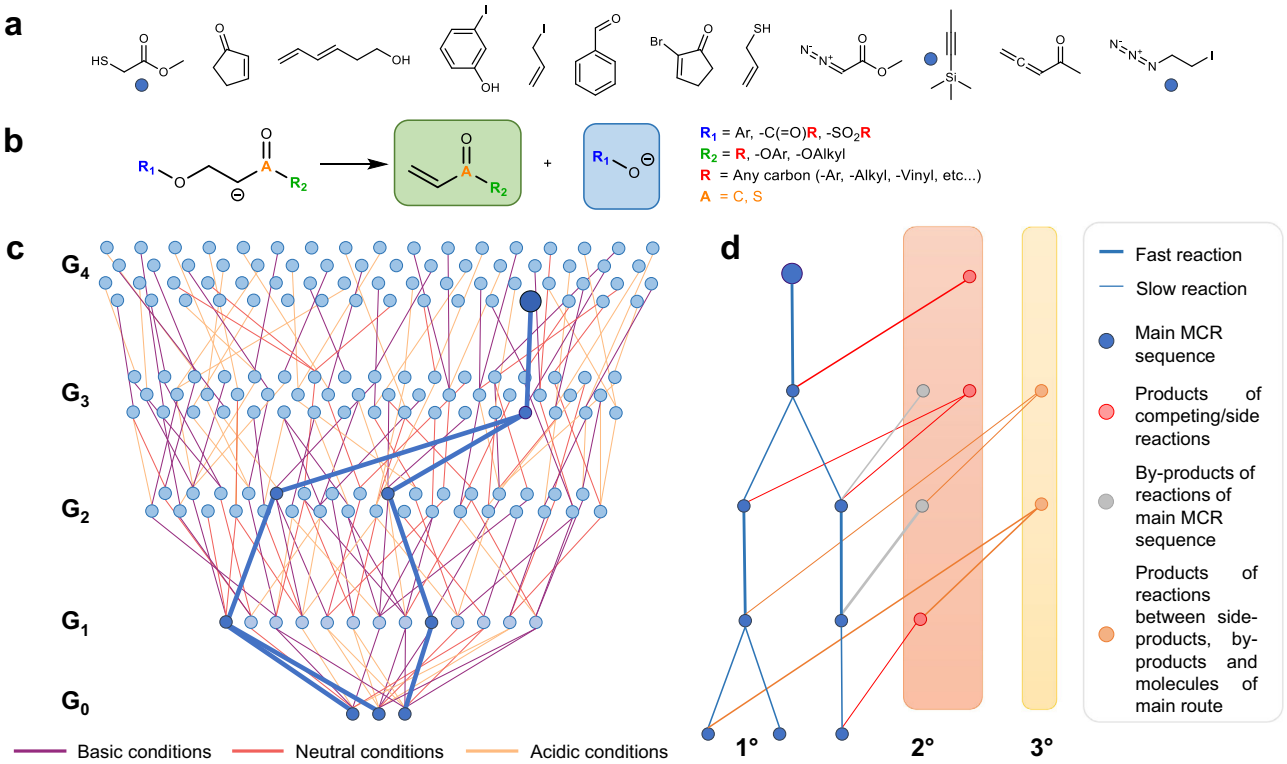


Fig. 2 | Key elements of the MECH algorithm to discover MCRs. a Examples of simple starting materials from the collection of ca. 2400 (see main text and Methods). **b** Abbreviated example of one of ~8000 mechanistic transforms, here E1cB elimination. Different positions can accommodate various substituents, some of which are listed to the right of the reaction scheme. Note that the transform is coded to account for the by-product(s), here a phenoxide, carboxylate or mesylate. Classification of reaction conditions, rates, etc., are also parts of the transform's record but, for clarity, are not shown here. For tutorial of rule coding see Supplementary Section 5 and examples deposited at https://zenodo.org/records/13381201. **c** Application of mechanistic transforms to a given set of starting materials iteratively expands the synthetic generations, G_n, of a network of possible intermediates and immediate by-products. In the schematic miniature drawn, the three circle markers in the bottom row (G_0) may be the three molecules from panel **a**, and the network is expanded to G_4. Different colors of connections between the nodes are intended to denote different types of conditions – to emphasize that this "forward" network expansion probes all conditions' combinations. Within the network thus constructed, the conditions may be matching (i.e., mutually compatible), corresponding to a MCR candidate at the Level 1 of analysis (sequence of steps highlighted in dark blue). **d** Such a sequence is expended sideways, to perform analyses at Levels 2 and higher. Level 2 – branching-out of the main path to include by-products (gray) and products of competing/side reactions possible under the same class of reaction conditions (red; for condition types, see Methods); Level 3 – further branching to account for the reactions between side- and by-products. At Level 3 (and higher, not shown here for clarity but see Fig. 3b), undesired reactions of side-/by-products with each other and with the members of the main pathway are also considered and marked in orange. Faster reactions are represented schematically by thicker connections and it is essential that, at any junction, the side reactions are not faster that the main-path ones.

first-generation, G_1, of products and by-products, which are then iteratively reacted[23,25,36] to give generations G_2, G_3 (up to some user-specified generation n), resulting in rapidly expanding[26] networks of mechanistic steps (Figs. 2c and 3a). At this stage, all classes of reaction conditions are allowed to survey the "synthesizable space" broadly but intermediates containing highly strained scaffolds not known as reaction intermediates (e.g., cyclobutenylene but not benzyne) are eliminated. Molecules can also be checked for the pKa of all C-H bonds[41] to ensure that reactions with electrophiles, such as C-H alkylations, proceed at the most acidic positions. Also, to prevent oligo/polymerization and limit network's size, each substrate is allowed to contribute atoms to any molecule in the network at most twice (see User Manual).

Selection of mutually-compatible MCR/one-pot sequences

Pathways leading to every neutral molecule within the network thus created are traced by Dijkstra-type algorithm; if multiple routes are detected, they are retrieved and ranked according to length. For any of these mechanistic sequences to be suitable candidates for MCR or one-pot reactions, the conditions specified for individual mechanistic steps must be matching. This is the Level 1 of analysis (Figs. 2c and 3a) and the sequences:

(i) Cannot combine steps requiring oxidative and reductive conditions, and cannot use water-sensitive steps after water-requiring ones;

(ii) Should use solvents of the same class, although protic solvents are allowed to be added to aprotic ones (but not vice versa);

(iii) Cannot change multiple times between non-overlapping high and low temperature ranges (which would be experimentally impractical);

(iv) Should allow only for monotonic changes in acidity (e.g., basic-acidic-basic changes are not allowed). Additionally, steps proceeding in strongly basic conditions (with, e.g. LDA) are not allowed if earlier steps required acidic conditions.

Sideways network expansion around main MCR/one-pot routes

If Level 1 analysis identifies a candidate, condition-matching sequence, the aforementioned sideways analysis of potential side reactions is performed (Figs. 2d and 3b). At Level 2, the kinetics of side reactions are examined. Initially, this is done in a rudimentary manner, according to the aforementioned "very slow-slow-fast-very fast-uncertain" categorization of reaction steps (cf. examples in Methods). In particular, warnings are assigned if, for a given reaction of the main path, a side-step possible under the same or similar conditions is faster. Such

https://doi.org/10.1038/s41467-024-54611-5

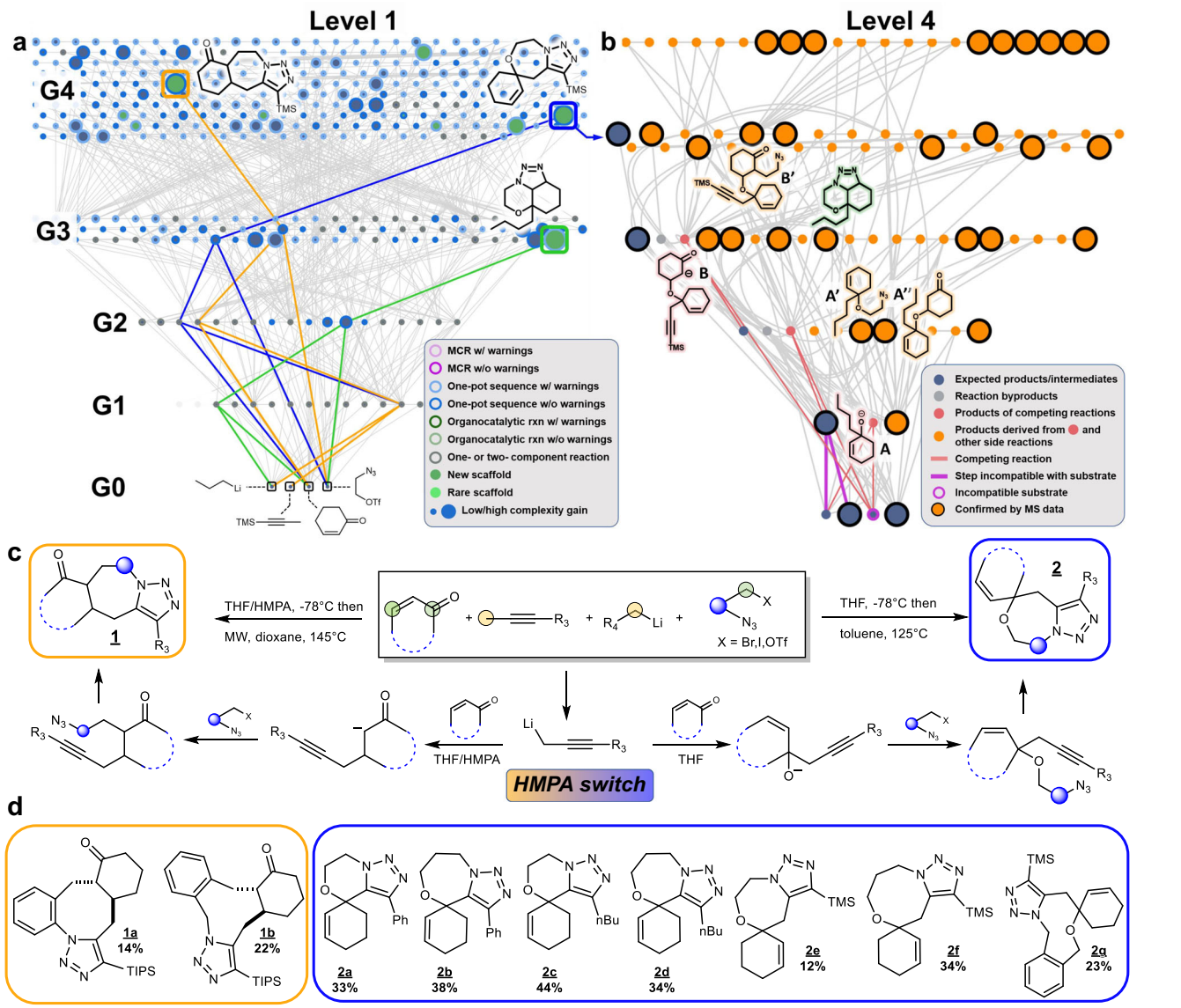


Fig. 3 | Example of algorithmically-discovered one-pot sequences and the corresponding mechanistic network expanded to Level 4. a Screenshot of Level 1 network propagated from cyclohexenone, trimethylsilylpropyne, *n*-butyllithium and azidotriflate substrates to $n = 4$ generations, G_4. The network encompasses all mutually-compatible sequences possible under different types of conditions. Node sizes are proportional to complexity increase per mechanistic step, $\Delta C/n$ (cf. Methods). Colors of the halos define MCR/one-pot sequences with or without warnings. Nodes whose interiors are colored green correspond to scaffolds not described in the literature. Within this network, two sequences (traced in blue and orange) up to G_4 are predicted to be one-pot without warnings and leading to unknown scaffolds 1 and 2 offering marked increase in $\Delta C/n$ (largest green nodes). A path to another complex scaffold in G_3 is also marked (in green). This product is predicted to form from the 1,2-adduct of nBuLi/cyclohexanone/azide cyclizing onto the double bond, and was detected by ESI-MS in the reaction mixture (structure highlighted in *green* in the L4 network in panel **b**). **b** Screenshot of the network branched-out from the blue pathway in **a** and analyzed at Level 4 (for networks analyzed at Levels 2 and 3, see Supplementary Fig. S158; interactive network expandable to Level 4 is deposited at https://mcrchampionship.allchemy.net). This Level 4 network encompasses various by- and side-products (gray and red nodes, respectively) and their further reactions (products marked in orange) between themselves and with the "parent" pathway. Larger orange nodes are likely structural assignments of peaks observed in the ESI-MS of the crude-reaction mixture. Interestingly, although the peaks corresponding to some predicted byproducts (e.g., A, B; structures drawn here with *pink* highlights) were not manifest in the ESI-MS spectra, their formation is corroborated by further products (A', A", B'; structures drawn with orange highlights) that can only be derived from these undetected species. For more structural assignments, see Supplementary Fig. S158. Also, the key cross-reactivity mandating sequential addition of reagents rather than MCR (i.e., reaction of alkyllithium with enone during metalation of alkyne) is highlighted by brighter pink connections at the bottom of the network. **c** General scheme and intermediates of the blue and orange one-pot pathways (leading to scaffolds 1 and 2, as in **a**) along with reaction conditions. In the substrates, the available nucleophilic and electrophilic sites are marked yellow and green, respectively, while the dark blue circle and the dotted arcs denote linkers between the azide and (pseudo)halides and a cyclic or acyclic fragment of the enone, respectively. The regioselectivity of addition (1,2- vs 1,4-) of propargyllithium reagent is controlled by the addition of HMPA as co-solvent. **d** Specific derivatives 1a, 1b and 2a–2g synthesized according to the general protocol along with the corresponding isolated yields. Note that the yields are low, as indeed predicted by the algorithm (see main text). Compounds 1a and 1b were isolated as single diastereoisomers. THF tetrahydrofuran, HMPA hexamethylphosphoramide, MW microwave, OTf triflate, TMS trimethylsilyl, TIPS triisopropylsilyl.

cases are flagged but not permanently removed from the mechanistic network since it is sometimes possible to generate thermodynamic products via a slower reaction (e.g., slow 1,4-addition of cyanide to methyl-vinyl ketone vs. fast 1,2-addition). Additional warnings are assigned if any of the by-products shows cross-reactivity with the main pathway or the reaction mixture becomes too complex (e.g., if three or more metals from catalysts or reagents are present and there is a possibility for unforeseen complexation of active species or deactivation of catalysts by ligand exchange). The by- and side-products from Levels 1 and 2 are allowed to react further, to give species at higher Levels, for which similar cross-reactivity analyses are performed. Importantly, the algorithm also analyzes whether reactivity conflicts between forming intermediates and yet unreacted substrates exist. If all substrates contributing atoms to the final product can be present in the reaction vessel from the beginning, the sequence is categorized as a plausible MCR (with possible condition changes obeying (i)-(iv) above); if, however, some intermediates are found to be cross-reactive with the substrates, then the algorithm suggests a one-pot option with sequential addition of the problematic substrate. In the current work, we focus on MCRs and one-pot sequences that entail no unresolved conflicts or warnings within Level 4 networks (see realistic examples in Fig. 3b and Supplementary Figs. S158–S162).

Prioritization and post-design evaluation

Because even for small substrate sets, the networks thus constructed may span large numbers of plausible MCR/one-pot products (Fig. 3a), additional analyses are performed to identify those that offer maximal complexification of the scaffold, those producing previously unknown scaffolds, those that are similar to approved drugs, and more (see Methods). The algorithm can also read in the positions of experimentally recorded mass-spectrometric signals and map them onto the Level 2-4 networks, which often facilitates analysis of experimental reaction mixtures (cf. Fig. 3b, Supplementary Fig. S158, and Supplementary Section S1).

Estimation of yields

Finally, once a desired MCR/one-pot candidate is selected, the algorithm performs a more in-depth kinetic analysis aimed at the estimation of reaction's yield. Since experimental kinetic rate constants for the vast majority of mechanistic steps are not available, we developed a physical-organic model grounded in free-energy linear relationships and approximating the rate constants of mechanistic steps using Mayr's nucleophilicity indices (see refs. 42,43 and Methods).

Experimental validations

From amongst the multitude of putative MCRs the algorithm has thus far identified, we focused on those that offer mechanistic uniqueness (i.e., substantial difference vs. known MCRs) and high substrate-to-product complexity increase, start from simple (commercially available or easy-to-make) substrates, and produce scaffolds of potential usefulness. Another factor was the conciseness of these protocols vs. traditional retrosynthetic planning that is based on full reactions rather than mechanistic steps and cannot capitalize on the use of reactive intermediates. Accordingly, for all one-pot/MCR products, we also ran the state-of-the-art retrosynthetic program (Chematica/Synthia[22,24]) which either planned multistep routes (on average 4 and up to 11 steps; all deposited at https://zenodo.org/records/10817102) or did not suggest any syntheses at all. All sequences are named "Mach" to highlight their machine-driven discovery (and to allude to its speed).

One-pot, non-MCR sequences

We begin with an example that is, admittedly, simple but serves to illustrate various modalities of the algorithm. Starting from enone, alkyllithium, azidohalide and alkyne, the mechanistic network propagated to G_4 (Fig. 3a) contains conditions-matched sequences leading to 391 products with MW < 500. Two compounds in G_4 correspond to previously unreported, tricyclic scaffolds **1** and **2**, both characterized by large per-step complexity increase from the substrates, and with **2** featuring a spiro system akin to that in some drugs and bioactive agents[44–47]. The mechanistic sequences to these products diverge at the initial step. The Mach1 route (blue) proceeds via the 1,2 addition, generation of the alkoxide intermediate, *O*-alkylation, and click reaction closing two rings. The Mach2, orange route starts with 1,4 Michael-type addition creating a carbanion at the alpha carbon, followed by *C*-alkylation and click reaction. The algorithm predicts that these sequences (1) may be performed only as one-pot rather than MCR (with enone added only after the complete consumption of the alkyllithium substrate); (2) the initial steps in both routes can be carried out using propargyllithium, with HMPA acting as a switch (Fig. 3c) to promote the 1,4 addition[48]; and (3) that both sequences will result in poor yields, ca. 20–40%. All these predictions turned out correct, with the isolated yields of derivatives **1a**,**1b** and **2a**–**2g** shown in Fig. 3d ranging from 12 to 44%. Of note, the computer-predicted competing reactivity modes were also congruent with ESI-MS analyses – in Fig. 3b and Supplementary Fig. S158, larger orange nodes denote side-/by-products with masses matching the spectra.

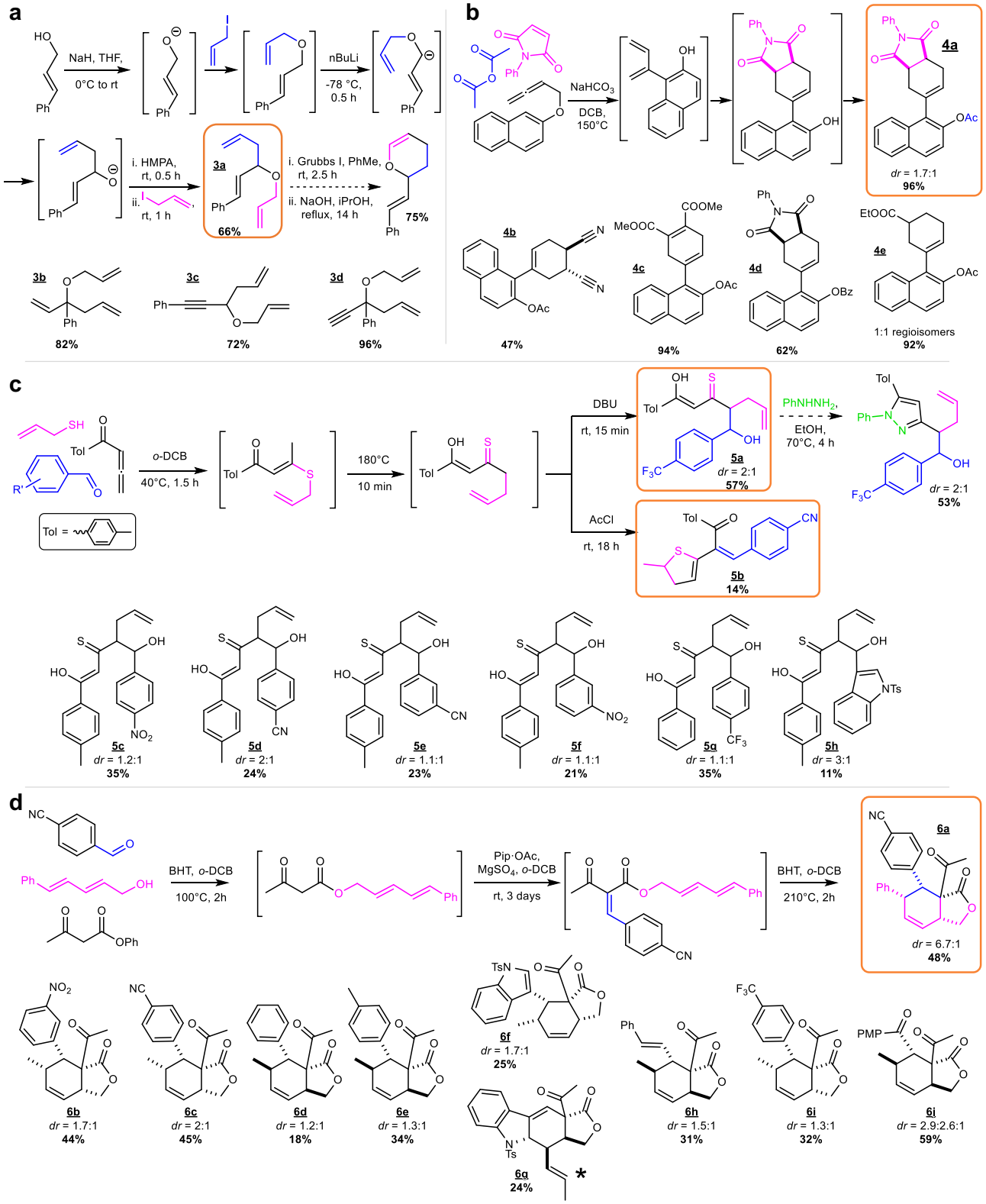
Another prediction for a one-pot, Mach3 sequence relying on a 2,3-Wittig rearrangement and leading to branched diallylic ethers **3a**–**3d**, is illustrated in Fig. 4a and Supplementary Fig. S159. This sequence was committed to experiment because, after metathesis (not compatible with one-pot conditions and carried out separately, dashed reaction arrow in Fig. 4a), it affords access to cyclic enol ether scaffolds that are used in various medicinal syntheses[49–52]. This sequence was predicted to proceed in good, ~68% yield vs. 66–96% yields we obtained.

MCR sequences

Turning to MCRs rather than one-pot sequences, Fig. 4b and Supplementary Fig. S160 illustrate a Mach4 sequence, in which an allene, a maleimide derivative, and a carboxylic acid anhydride engage in a sequence of Claisen rearrangement, aromatization, Diels-Alder cycloaddition, deprotonation and acylation to yield a 1-(1-cyclohexenyl)naphthalene, atropisomeric scaffold **4a** familiar from various types of drugs[53,54]. Scaffolds of this type are typically prepared via various multistep protocols[55–59]. The MCR approach shortens these procedures while commencing from substrates of similar complexity and does not require transition metal catalysts or pre-functionalized aryl systems. The experimental yields for **4a** and its analogs **4b**–**4e** were generally quite satisfactory and in most cases >90% (for the originally predicted sequence, the algorithm predicted 99% vs. 96% in experiment).

Another pair of MCRs using allene as one of the substrates is illustrated in Fig. 4c and begins with a nucleophilic addition of an allyl thiol to the allene and isomerization followed by thio-Claisen rearrangement. Network analysis detailed in Supplementary Figs. S161 and S162 indicates that the sequence can then diverge. In Mach5 MCR, addition of excess base results in straightforward condensation with an aromatic aldehyde occurring at the less acidic methylene group of the thioketone and leading to **5a** in 57% yield (vs. predicted 43%). This product or its analogs **5c**–**5h** can further react (outside of the MCR, dashed reaction arrow) with phenyl hydrazine[60] to give substituted pyrazoles which are popular motifs of many drugs. By contrast, in Mach6, addition of acetyl chloride triggers a relatively rare[61] sequence of acetylation of alcohol, acidic elimination of acetic acid catalyzed by the in-situ generated HCl to give the Knoevenagel-type adduct, thioketo-enol tautomerization followed by spontaneous cyclization. The 2,3-dihydrothiophene products **5b** are obtained in significantly lower yields (~10% and up to 14% for the cyano derivative vs. 12% predicted yield, though these experimental values are affected

https://doi.org/10.1038/s41467-024-54611-5



by partial decomposition of the product during purification), and their applications are less conspicuous[62].

The sequence underlying Mach7 MCR shown in Fig. 4d – leading to a scaffold akin to oblongolide natural products considered as potential algicide, herbicide[63] and antiviral[64] agents – is perhaps familiar to a retrosynthetically-trained eye. Indeed, the succession of transesterification of sorbic alcohol, Knoevenagel condensation and Diels-Alder reaction has also been found by Chematica/Synthia. However, the MECH algorithm correctly predicted that it could be folded-up into a one-step MCR leading to **6a**–**6j**. The yields of racemic mixtures were up to 59% (compared to 55% predicted by the algorithm and 13–38% for multistep syntheses of similar scaffolds reported in refs. 65,66) and with the procedure readily scalable to gram scale (Supplementary Section S6.7). Also, one less obvious outcome predicted by the algorithm is that for the indole-3-carbaldehyde substrate, the Knoevenagel adduct can engage in a reverse-demand Diels-Alder

Fig. 4 | Computer-discovered one-pot sequences and MCRs. For details of mechanistic networks, see Supplementary Figs. S159–S162. **a** Scheme of a one-pot sequence for the synthesis of branched allyl ethers. The sequence is detected as one-pot rather than the MCR because excessive allyl iodide would react with *n*-butyllithium, hampering deprotonation and subsequent Wittig rearrangement (cf. Supplementary Fig. S159 marking this conflict). Non-isolated intermediates are shown in brackets and the isolated product **3a** is framed in orange. This product has been separately cyclized via ring-closing metathesis to afford cyclic enol ether. Additional derivatives **3b**–**3d** were prepared from allyl iodide and other commercially available β,γ-unsaturated alcohols. **b** Scheme of a MCR producing unsaturated β-naphthol esters. Key non-isolated intermediates are shown in brackets and the isolated product **4a** is framed in orange. Additional derivatives **4b**–**4e** were prepared using different commercially available dienophiles and acylating agents. **c** Scheme of a MCR producing unsaturated hydroxylated monothio-β-diketones (existing in the thioenol tautomeric form) under basic conditions (top) or 2,3-dihydrothiophenes under acidic conditions (bottom) applied during the last step. Non-isolated intermediates are shown in brackets and the isolated products (**5a** originally predicted for the top MCR and highest yielding **5b** for the bottom one are framed in orange. The monothio-β-diketone product has been separately reacted (dashed arrow) with phenylhydrazine (green) to afford a substituted pyrazole. Additional products **5c**–**5h** were prepared by the top MCR. **d** Scheme of the MCR producing unsaturated bicyclic lactones. Key non-isolated intermediates are shown in brackets and the isolated product **6a** is framed in orange. Additional derivatives **6b**–**6j** were prepared using different commercially available aldehydes and dienes. BHT butylhydroxytoluene, DCB dichlorobenzene, THF tetrahydrofuran, DBU 1,8-diazabicyclo(5.4.0)undec-7-ene, HMPA hexamethylphosphoramide, Pip·OAc piperidinium acetate, dr diastereomeric ratio. Note: **4a** was observed as a 1.7:1 mixture of diastereoisomers with two distinct ^{1}H NMR signals (separated by 0.5 ppm) for **Me**-OAc protons. These signals can be attributed to known through-space shielding by **Ph**-N in one of the diastereoisomers. However, no distinct signals allowing for determination of dr's were observed for structurally similar (OBz vs. OAc) **4d**. The product of reverse-demand Diels-Alder cyclization **6g** is marked with a star and was isolated as a single diastereoisomer. Percentage values in all panels are isolated yields.

cycloaddition to give a relatively complex, tetracyclic scaffold **6g** isolated in 24% yield.

Substrate-reusing and organocatalytic sequences

The next two examples are interesting for the unique ways in which they use or reuse some of the substrates. In the Mach8 sequence shown in Fig. 5a, b, the phenol substrate is first used to form an activated ester that then reacts with 2-allylcyclohexanone to give a spiro β-lactone which, upon addition of $MgBr_2$, undergoes a ring-expanding rearrangement into a substituted hexahydro-2(3H)-benzofuranone **7a** in 31% yield (vs. predicted 48%). Such motifs are found in various natural products and bioactive compounds[67] and the particular scaffold, upon metathesis and reduction, could create a ring system present in lancifonins. However, when iodo-substituted phenols and cyclohexanone (instead of 2-allylcyclohexanone) are used as substrates and the network is propagated to higher generations, iodophenol is regenerated as a by-product of the spirocyclization step and then – upon product's decarboxylation – is reused in situ as a substrate in Heck reaction, to complete Mach9 MCR yielding **7b** in up to 35% yield (vs. predicted 35%).

In turn, Fig. 5c–e illustrate a Mach10 reaction that was predicted and then confirmed as organocatalytic. With the initial set of substrates (α-bromo-α,β-unsaturated ester, methyl thioglycolate and sodium azide), the algorithm suggested an MCR that could lead to a dihydrothiophenecarboxylate scaffold **8a** similar to some GABA receptor inactivators[68]. However, the program also indicated that that the C-H pKa of the α-azidoester be higher than that of the α-thioester – that is, the deprotonation (either by azide anion[69] or sodium methoxide) at the former locus should be preferred and could lead to rapid elimination (green arrow in Fig. 5c, blue arc connection in the L2 network in Fig. 5d) rather than cyclization. This elimination sets a feedback loop regenerating the thiol (colored pink in Fig. 5d), which effectively acts as an organocatalyst sustaining azide substitution at vinylic α-position. This was, indeed, verified in experiment with the original reaction to **8b** proceeding under mild conditions in 67% yield (vs. algorithm-predicted 47%), and with the further scope of **8c**–**8f** illustrated in Fig. 5e. For alkyl ketones, 10 mol% of the thiol is optimal, while for β-aryl ketones, 35 mol% thiol load is necessary due to the trapping of the thiol catalyst in the S_N2 reaction with the alkyl bromide (obtained after 1,4-addition of thiol to Michael acceptor).

Discussion

The above experimental examples cover only a tiny fraction of substrate combinations that can give rise to MCRs or one-pot sequences. To broaden and speed up the discovery process, we have automated the choices of substrate triplets and quartets (from the aforementioned set of ca. 2400 reactive molecules) as well as subsequent network expansion and analysis. With tens of thousands of substrate combinations now probed and with further searches ongoing, the list of the currently 50 top-ranked (by complexity increase per step metric, see Methods) MCR candidates is maintained at https://mcrchampionship.allchemy.net. Users of Allchemy's MECH can perform searches with their own substrates of choice, and can opt to "compete" and post their results therein (if the scores place them within top-50), in the world' first "championship" for computerized reaction design.

It is our hope that, in the fullness of time, this resource will enable discovery of MCRs in quantities that may have significant impact on the practice of synthetic chemistry. This said, algorithms like ours do not replace all of chemists' insights and the need for conditions' optimization (e.g., in terms of screening for optimal temperatures, solvents, etc.). There is also plenty of room for further improvements (see Supplementary Section S2), for an example of incorrect MCR prediction) and extensions of the algorithm, e.g., to incorporate radical-based mechanisms and additional catalytic transformations, or to adapt the workflow to the retrosynthetic direction (to suggest imaginative disconnections of specific scaffolds).

Methods

Mechanistic rules

As outlined in the main text, the mechanistic transforms are encoded in the SMARTS notation and account for by-products (a tutorial on coding the rules is provided in Supplementary Section 5). The templates are generalized – that is, they do not encompass just a single reaction precedent (as in the recently published repository of mechanistic steps for popular radicalic reactions[70]) but each specifies the scope of admissible substituents at various positions of the SMARTS template as well as a list of incompatible groups. These explicitly defined incompatibilities help limit the sizes of the networks and remove from analysis at least the obviously problematic steps, in which two or more motifs would react on commensurate time scales, inevitably leading to undesired complex reaction mixtures and ruining a "clean" MCR.

Furthermore, rules are accompanied by information about reaction conditions that is essential to later wire-up individual mechanistic steps into mutually compatible sequences. In this context, each transform is categorized according to general conditions (basic, neural, acidic), solvent class (protic/aprotic and polar/non-polar), temperature range (very low = <−20 °C, low = −20 to 20 °C, r.t., high = 40 to 150 °C, and very high = >150 °C); and water tolerance (yes, no, water is required). One transform can have more than one categorization (e.g., Diels-Alder cycloaddition can be carried out either under neutral conditions at high temperature or at very low, low or room temperatures using a Lewis acid catalyst) – in such cases,

https://doi.org/10.1038/s41467-024-54611-5

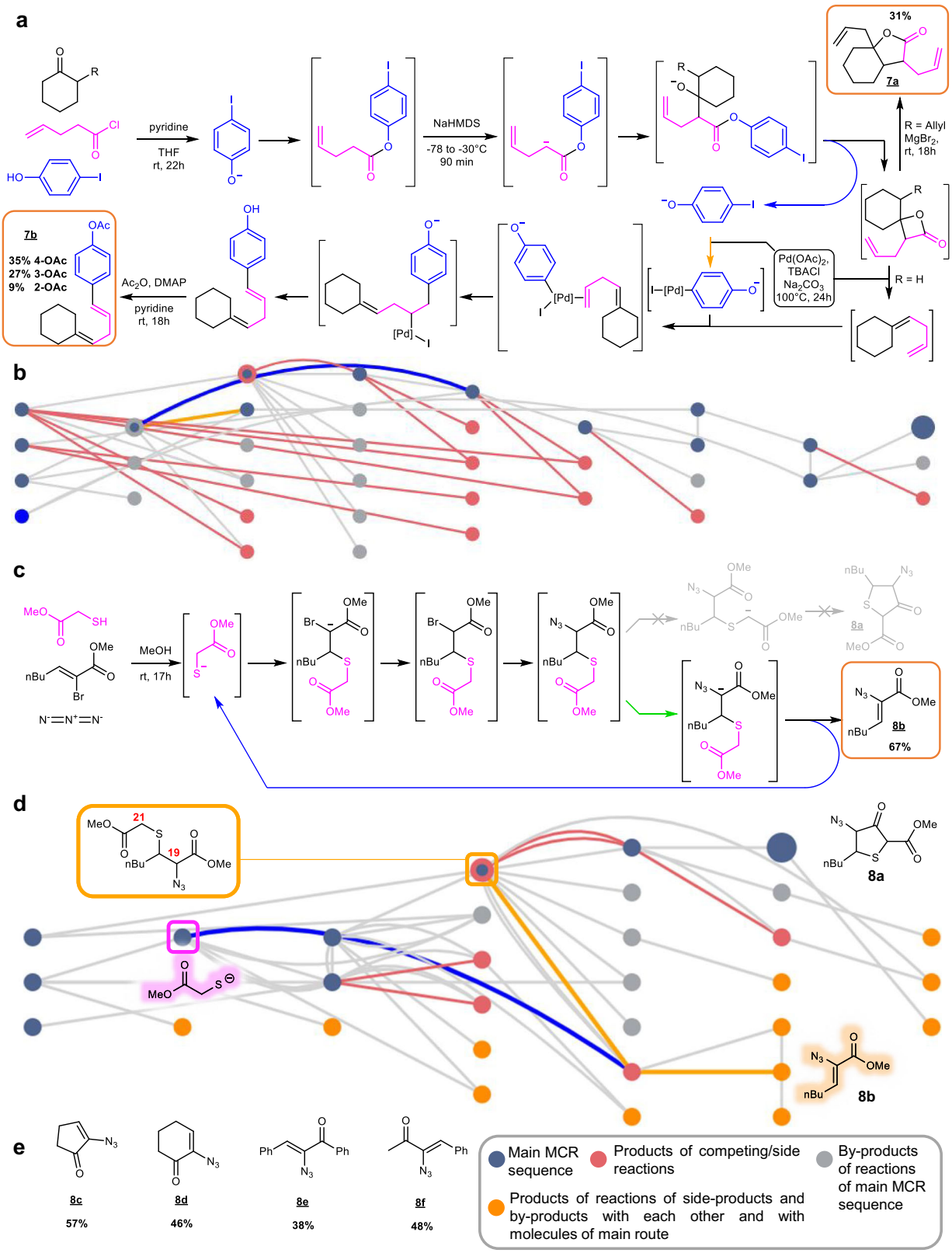


Fig. 5 | Computer-discovered substrate-reusing MCRs and an organocatalytic reaction. a Scheme of a MCR for the synthesis of arylated skipped dienes. Non-isolated intermediates are shown in brackets and the isolated products are framed in orange. The obtained dienes were separately acetylated for the purpose of purification. The bicyclic lactone **7a** (upper right) was obtained from substituted cyclohexanone (R = allyl) and phenol substrates when $MgBr{\cdot}Et_2O$ was used instead of the Pd-catalyst. **b** The Level 2 graph view of the path leading to the arylated diene from **a**. Reuse of iodophenol byproduct in Heck-coupling (with oxidative addition step marked orange) is marked with the blue arc. **c** Scheme of organocatalytic thiol-catalyzed sp^2-azidation. Non-isolated intermediates are shown in brackets and the isolated product **8b** is framed in orange. **d** The Level 3 graph view of the path from **c**. Reuse of thiol, acting as an organocatalyst, is marked with the blue arc. **e** Additional vinyl azides **8c–f** prepared by the MCR from **c** using different α-bromoenones. Abbreviations: DMAP, 4-dimethylaminopyridine; TBACl, tetrabutylammonium chloride.

multiple conditions are provided and, when considering sequences of compatible steps, are treated as logical alternatives. Each transform also contains specific suggestions for reagents commonly used in reactions involving this mechanistic step (e.g., diethylaluminium chloride in Claisen rearrangement, *n*-butyllithium in [2,3]-Wittig rearrangement, etc.).

Regarding the initial and rough categorization of kinetics, each transform is assigned a typical speed category (very slow, slow, fast, very fast, uncertain). A "very slow" step (conversion time above ca. 24 hrs) is, for example, addition of amines to trisubstituted Michael acceptors. Steps categorized as "slow" (few to ca. 24 h) are, e.g., reaction of a deprotonated nitro compound with a ketone, addition of an alcohol to a protonated nitrile, or 1,3-dipolar cycloaddition of imine and nitrile oxide. Examples of "fast" steps (minutes to few hrs) include deprotonation of alcohols, alcoholysis of anhydrides, or addition of organocuprates to activated alkenes. "Very fast" steps (seconds to minutes) are, for example, decomposition of oxaphosphetanes to alkenes and phosphine oxide, elimination of a chloride anion from an adduct of amine and acyl chloride, tautomerizations leading to aromatic compounds (e.g., 2,4-cyclohexadienone to phenol). "Uncertain" steps are those for which literature provides conflicting reaction data (i.e., wide range of reaction times and/or rates strongly influenced by substrate structures or small changes in reaction conditions) or those for which literature is insufficient to determine the reaction rate of an individual mechanistic step. One example from this category is addition of an imine to phenolic compounds, for which reaction rate strongly depends on the activity/nucleophilicity of phenolic component but even more on reaction conditions, resulting in time spans from 5 minutes to 9 hours for reactions involving the same substrates (see ref. 71 – 9 h[72]; – 7.5 h[73]; – 3 h[74]; – 5 min). Another example is S_N2 reaction of a secondary bromide with cyanide anion, for which the reaction rate is strongly influenced by the character and size of substituents on the halide component and the type of solvent used, with polar aprotic solvents facilitating the reaction and polar protic solvents impeding it. For instance, reaction of 2-bromo-2-(2-methylphenyl)-1-(morpholin-4-yl)ethanone with sodium cyanide in methanol takes 24 h[75], while reaction of a similar molecule, methyl 2-(1-bromo-2-methoxy-2-oxoethyl) benzoate, with potassium cyanide in DMF takes only 1 h[76]).

The rules covered in the current version of the MECH module span a broad range of acid-base catalyzed steps (including Lewis acids), substitutions, eliminations, additions, rearrangements, pericyclic reactions as well as basic transformations catalyzed by transition metals (e.g., mechanistic steps of Suzuki, Buchwald-Hartwig, Heck, and Pauson-Khand reactions). Basic carbocationic chemistry is included but not exhaustively (a separate HopCat module dedicated to such rearrangements is available in our recent publication[36]). Also, radical mechanistic steps are not (yet) included since their proper application requires generalization (cf. short discussion in Supplementary Section S3) and likely additional heuristics based on thermodynamic and molecular-mechanical considerations (akin to those we described in the HopCat paper[36]). Some rare types of steps involving π-complexes had to be simplified in notation since they are not properly handled by RDKit (they are encoded as 3-membered rings rather than interaction between metal and multiple bonds, e.g., during Heck reaction).

Additional details of network expansion

During expansion of mechanistic networks, the program generally uses the individual steps, e.g., imine formation is divided into 1) ketone protonation, 2) imine addition to the protonated ketone, 3) proton transfer from nitrogen to oxygen, 4) formation of an iminium cation via elimination of water, 5) deprotonation of the iminium cation (Supplementary Fig. S163a). However, because the networks expand very rapidly with the number of steps ("synthetic generations"), such step-by-step expansions may be inefficient in exploring longer mechanistic sequences – for instance, the five-step imine formation is only part of, say, the Ugi multicomponent reaction. To reduce computational cost, we have also encoded some shortcut steps that, for popular transformation types, concatenate individual mechanistic steps (those occurring in a rapid sequence and/or those leading to unstable intermediates; see example in Supplementary Fig. S163b). When executed as one "super-step", the shortcuts keep all the information about by-products of all individual steps. The network expansions then use both the step-by-step and shortcut strategies. Of note, if a given substrate can engage in a very-fast, VF, mechanistic step (e.g., tautomerization, elimination leading to an aromatic product, etc.), only this rapid step is performed under given reaction conditions. Other competing mechanistic steps can be applied to this substrate only if they proceed under different class of conditions.

Further details of route prioritization and post-design evaluation

The MECH module offers multiple options to filter, analyze, and prioritize the one-pot/MCR pathways within the mechanistic networks. As described in detail in Supplementary Section S1, the user can filter off those products that are formed via mechanistic steps having non-overlapping "cores" (reactions occurring on disjoint parts of the molecule will likely yield "linear" structures and will not complexify the starting scaffold), or those that do not involve any rearrangements or pericyclic reactions.

To easier identify and prioritize sequences that offer the highest degree of complexification, nodes in the network can be sized in proportion to the increase of structural complexity per step, $\Delta C/n$, where ΔC is calculated along an atom-mapped path as (a·#Rearrangements + a·#RingsFormed + #BondsCreated + #BondsDisconnected), where $a = 5$ is used here to strongly favor formation of cyclic scaffolds and sequences containing rearrangements. Furthermore, the nodes can be colored as molecules known/unknown in the literature or, more generally, according to whether the scaffold is without precedent in prior literature. The algorithm to determine scaffold uniqueness first defines a scaffold "base" as a set of connected rings, whereby a ring is considered connected if it fulfills either of the two criteria: a) it shares at least one atom with any of the other rings in the base, b) is connected with a double bond to any of the other rings. The final scaffold is obtained from this base by inclusion of atoms connected to the base with double bond (i.e., oxygen from carbonyl group or exomethylene double bond). Note that this definition inherits both elements and bond orders from the parent molecule such that, for instance, cyclohexane, cyclohexene, cyclohexanone and cyclohexanethione are all considered as different scaffolds. Finally, a scaffold is considered without prior precedent if it is not present in the list of 95,191 scaffolds extracted from the Zinc collection[77]. The nodes within the networks can also be colored by similarity to approved drugs, reaction type, hazardous compounds, and more (see User Manual in Supplementary Section S1). Last but not least, the user can input a list of mass-spectrometric signals recorded in experiment and the likely M + 1 and M + 23 nodes will be marked on Level 1–4 trees (Fig. 2b and Supplementary Fig. S127).

Estimation of yields

To estimate the yields of MCR/one-pot candidates, we developed a physical-organic model grounded in free-energy linear relationships. In this model, to be detailed in a separate publication[78], the rate constants of mechanistic steps are approximated by using Mayr's nucleophilicity, N, and electrophilicity, E, indices[42,43] as $\log k_{20deg} \propto (N + E)$, which are further fine-tuned by corrections capturing relative reactivities, stoichiometries and amounts of various species in the mechanistic networks, $\ln k_i = \ln k_i^{Mayr} + \sum \text{corrections}(r_i)$. The weights of the individual corrections were trained on the mechanistic networks of 20 diverse MCRs reported before (chosen to represent both low- and high-yielding ones), and the model was then used to predict the yields of the

mechanistically distinct MCRs described in the current publication. For the training set of the known MCRs, the Pearson correlation coefficient (ρ^2) between the experimental and modeled yields was 0.80 with mean absolute error of 10.5. For the test set of reactions used in this study, $\rho^2 = 0.86$ and MAE = 7.3. These metrics compare quite favorably with generally unsatisfactory correlations observed for various machine learning models trained on full, substrate-to-product reactions without any mechanistic knowledge[79–82].

Pre-curated collection of substrates available through Allchemy's user interface

Although arbitrary substrates can be input in Allchemy's MECH module, we have also curated a list of ~2400 simple and commercially available substrates that, in our experience, improve the chances of finding MCR reactions. To begin with, the Zinc collection[77] was pruned to retain only molecules with, at most, 15 heavy atoms. After removing stereochemistry, ~410,000 unique entries were left. Molecules containing either poorly reactive fragments (94 patterns, e.g., heterocycles, polycyclic systems, ethers) or several unfunctionalized carbon atoms were removed, as they only introduced unnecessary structural complexity without desired reactivity. The remaining molecules were queried for the presence of one or two reactive groups defined by experienced synthetic chemists (164 patterns of FGs listed in Supplementary Tables S2, S3) – there were 36,294 such molecules of which 16,631 had one reactive FG and 19,663 had two reactive FGs. In the latter, we only kept molecules in which the FGs were separated by, at most, three atoms – in this way, when these molecules reacted, they were more likely to form smaller rings rather than macrocycles. For some FG combinations, there were many hits (e.g., the algorithm identified 97 commercially available isocyanates and 94 compounds possessing both aryl bromide and secondary amine FGs). In such cases, the compound with the lowest molecular mass was retained.

Data availability

The list of reactions and literature sources of known MCRs and one-pots is deposited as .csv and Excel files at Zenodo under accession code https://zenodo.org/records/10817102. All 3108 uniqe reactions from all networks are deposited (along with condition classification, rate categorization and optimized rate parameters) at Zenodo under accession code https://zenodo.org/records/13381381. Multistep synthesis plans produced by Chematica/Synthia for targets made here via MCRs and one-pots are deposited at Zenodo under accession code https://zenodo.org/records/10817102 (note: no syntheses were found for Mach6 and for one of the two variants of Mach2). The X-ray crystallographic coordinates for structures reported in this study have been deposited at the Cambridge Crystallographic Data Centre (CCDC), under deposition numbers 2402793. These data can be obtained free of charge from The Cambridge Crystallographic Data Centre via www.ccdc.cam.ac.uk/data_request/cif. User manuals are available in Supplementary Section S1. Interactive networks for all examples described in the text and results of MCR Championships are available for analysis at https://mcrchampionship.allchemy.net under restricted access. The interactive t-SNE map of known MCRs and one-pots is available under restricted access at https://mcrmap.allchemy.net. All searches we described or any other searches one may wish to execute, can be performed under restricted access at https://mech.allchemy.net. Access to all restricted services can be obtained by academic users by sending a request to admin@allchemy.net from an academic address. The restrictions are dictated by server capacity so the access can be provided to twenty concurrent academic users on a rolling basis and two-week slots.

Code availability

Codes for network expansion and MCR analysis are deposited at https://zenodo.org/records/13381201. Codes for the estimation of kinetic rates and calculation of yields are deposited at https://zenodo.org/records/13381381. The same repository (https://zenodo.org/records/13381381) houses codes for the optimization of kinetic parameters as well as 30 digitized mechanistic networks on which the rate-prediction model was trained and tested (with details of the model development described in ref. 78). Interactive Allchemy MECH web-app is freely available at https://mech.allchemy.net (given server capacity, to twenty concurrent academic users on a rolling basis and two-week slots).

References

1. Ugi, I. et al. New applications of computers in chemistry. *Angew. Chem. Int. Ed.* **18**, 111–123 (1979).
2. Bauer, J. & Ugi, I. Chemical-reactions and structures without precedent generated by computer-program. *J. Chem. Res.-S* **11**, 298-298 (1982).
3. Bauer, J., Herges, R., Fontain, E. & Ugi, I. IGOR and computer assisted innovation in chemistry. *Chimia* **39**, 43–53 (1985).
4. Ugi, I. K. et al. Computer assistance in the design of syntheses and a new generation of computer programs for the solution of chemical problems by molecular logic. *Pure. Appl. Chem.* **60**, 1573–1586 (1988).
5. Dömling, A. & Ugi, I. Multicomponent reactions with isocyanides. *Angew. Chem. Int. Ed.* **39**, 3168–3210 (2000).
6. Dömling, A., Wang, W. & Wang, K. Chemistry and biology of multicomponent reactions. *Chem. Rev.* **112**, 3083–3135 (2012).
7. Ganem, B. Strategies for innovation in multicomponent reaction design. *Acc. Chem. Res.* **42**, 463–472 (2009).
8. D'Souza, D. M. & Müller, T. J. J. Multi-component syntheses of heterocycles by transition-metal catalysis. *Chem. Soc. Rev.* **36**, 1095–1108 (2007).
9. Phelps, J. M. et al. Multicomponent synthesis of α-branched amines via a zinc-mediated carbonyl alkylative amination reaction. *J. Am. Chem. Soc.* **146**, 9045–9062 (2024).
10. Medley, J. W. & Movassaghi, M. Robinson's landmark synthesis of tropinone. *Chem. Comm.* **49**, 10775–10777 (2013).
11. Brandner, L. & Müller, T. J. J. Multicomponent synthesis of chromophores – The one-pot approach to functional π-systems. *Front. Chem.* **11**, 1124209 (2023).
12. Zhao, W. Y. & Chen, F. E. One-pot synthesis and its practical application in pharmaceutical industry. *Curr. Org. Synth.* **9**, 873–897 (2012).
13. Broadwater, S. J., Roth, S. L., Price, K. E., Kobašlija, M. & McQuade, D. T. One-pot multi-step synthesis: a challenge spawning innovation. *Org. Biomol. Chem.* **3**, 2899–2906 (2005).
14. Wheeldon, I. et al. Substrate channeling as an approach to cascade reactions. *Nat. Chem.* **8**, 299–309 (2016).
15. Hayashi, Y. Time and pot economy in total synthesis. *Acc. Chem. Res.* **54**, 1385–1398 (2021).
16. Paul, B., Maji, M., Chakrabartia, K. & Kundu, S. Tandem transformations and multicomponent reactions utilizing alcohols following dehydrogenation strategy. *Org. Biomol. Chem.* **18**, 2193–2214 (2020).
17. Cioc, R. C., Ruijter, E. & Orru, R. V. A. Multicomponent reactions: advanced tools for sustainable organic synthesis. *Green Chem* **16**, 2958–2975 (2014).
18. Ruijter, E., Scheffelaar, R. & Orru, R. V. A. Multicomponent reaction design in the quest for molecular complexity and diversity. *Angew. Chem. Int. Ed.* **50**, 6234–6246 (2011).
19. Hayashi, H. et al. In silico reaction screening with difluorocarbene for *N*-difluoroalkylative dearomatization of pyridines. *Nat. Synth.* **1**, 804–814 (2022).
20. Novikov, M. S., Khlebnikov, A. F., Krebs, A. & Kostikov, R. R. Unprecedented 1,3-dipolar cycloaddition of azomethine ylides derived from difluorocarbene and imines to carbonyl compounds –

Synthesis of oxazolidine derivatives. *Eur. J. Org. Chem.* **1**, 133–137 (1998).
21. Dong, S., Fu, X. & Xu, X. [3+2]-Cycloaddition of catalytically generated pyridinium ylide: A general access to indolizine derivatives. *Asian J. Org. Chem.* **9**, 1133–1143 (2020).
22. Mikulak-Klucznik, B. et al. Computational planning of the synthesis of complex natural products. *Nature* **588**, 83–88 (2020).
23. Wołos, A. et al. Computer-designed repurposing of chemical wastes into drugs. *Nature* **604**, 668–676 (2022).
24. Lin, Y., Zhang, R., Wang, D. & Cernak, T. Computer-aided key step generation in alkaloid total synthesis. *Science* **379**, 453–457 (2023).
25. Wołos, A. et al. Synthetic connectivity, emergence, and autocatalysis in the network of prebiotic chemistry. *Science* **369**, eaaw1955 (2020).
26. Gajewska, E. P. et al. Algorithmic discovery of tactical combinations for advanced organic syntheses. *Chem* **6**, 280–293 (2020).
27. Molga, K. et al. A computer algorithm to discover iterative sequences of organic reactions. *Nat. Synth.* **1**, 49–58 (2022).
28. Gothard, C. M. et al. Rewiring chemistry: algorithmic discovery and experimental validation of one-pot reactions in the network of organic chemistry. *Angew. Chem. Int. Ed.* **51**, 7922–7927 (2012).
29. Grossman, R. B. *The Art of Writing Reasonable Organic Reaction Mechanisms* (Springer International Publishing, 2019).
30. Gund, T. M., Schleyer, P. R., Gund, P. H. & Wipke, W. T. Computer assisted graph theoretical analysis of complex mechanistic problems in polycyclic hydrocarbons. The mechanism of diamantane formation from various pentacyclotetradecanes. *J. Am. Chem. Soc.* **97**, 743–751 (1975).
31. Marsili, M. *Computer chemistry* (CRC Press, 1990).
32. Chen, J. H. & Baldi, P. No electron left behind: a rule-based expert system to predict chemical reactions and reaction mechanisms. *J. Chem. Inf. Model.* **49**, 2034–2043 (2009).
33. Kayala, M. A. & Baldi, P. ReactionPredictor: Prediction of complex chemical reactions at the mechanistic level using machine learning. *J. Chem. Inf. Model.* **51**, 2526–2540 (2012).
34. Jorgensen, W. L. et al. CAMEO: a program for the logical prediction of the products of organic reactions. *Pure Appl. Chem* **62**, 1921–1932 (1990).
35. Satoh, H. & Funatsu, K. SOPHIA, a knowledge base-guided reaction prediction system-utilization of a knowledge base derived from a reaction database. *J. Chem. Inf. Comput. Sci.* **35**, 34–44 (1995).
36. Klucznik, T. et al. Computational prediction of complex cationic rearrangement outcomes. *Nature* **625**, 508–515 (2024).
37. Roque, J. B., Kuroda, Y., Göttemann, L. T. & Sarpong, R. Deconstructive diversification of cyclic amines. *Nature* **564**, 244–248 (2018).
38. Kennedy, S. H., Dherange, B. D., Berger, K. J. & Levin, M. D. Skeletal editing through direct nitrogen deletion of secondary amines. *Nature* **593**, 223–227 (2021).
39. Taitz, Y., Weininger, D. & Delany, J. J. *Daylight Theory: SMARTS - A language for describing molecular patterns*, https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html (1997).
40. Molga, K., Gajewska, E. P., Szymkuć, S. & Grzybowski, B. A. The logic of translating chemical knowledge into machine-processable forms: a modern playground for physical-organic chemistry. *React. Chem. Eng.* **4**, 1506–1521 (2019).
41. Roszak, R., Beker, W., Molga, K. & Grzybowski, B. A. Rapid and accurate prediction of pKa values of C–H Acids using graph convolutional neural networks. *J. Am. Chem. Soc.* **141**, 17142–17149 (2019).
42. Mayr, H. & Patz, M. Scales of nucleophilicity and electrophilicity: A system for ordering polar organic and organometallic reactions. *Angew. Chem. Int. Ed.* **33**, 938–957 (1994).
43. *Mayr's Database Of Reactivity Parameters - Start page*. (2023) Available at: https://www.cup.lmu.de/oc/mayr/reaktionsdatenbank/ (Accessed: 6th December 2023).
44. Chavan, S. R. et al. Iminosugars spiro-linked with morpholine-fused 1,2,3-triazole: Synthesis, conformational analysis, glycosidase inhibitory activity, antifungal assay, and docking studies. *ACS Omega* **2**, 7203–7218 (2017).
45. Tanaka, N. et al. Isolation and structures of attenols A and B. Novel bicyclic triols from the Chinese bivalve *Pinna attenuata*. *Chem Lett* **28**, 1025–1026 (1999).
46. Chen, D. et al. Discovery, structural insight, and bioactivities of BY27 as a selective inhibitor of the second bromodomains of BET proteins. *Eur. J. Med. Chem.* **182**, 111633 (2019).
47. Teiji, K. et al. Multi-cyclic cinnamide derivatives. Patent US 2007219181A1 (2007).
48. Sikorski, W. H. & Reich, H. J. The Regioselectivity of addition of organolithium reagents to enones and enals: The role of HMPA. *J. Am. Chem. Soc.* **123**, 6527–6535 (2001).
49. Ireland, R. E., Armstrong, I., Lebreton, J. D. J., Meissner, R. S. & Rizzacasa, M. A. Convergent synthesis of polyether ionophore antibiotics: synthesis of the spiroketal and tricyclic glycal segments of monensin. *J. Am. Chem. Soc.* **115**, 7152–7165 (1993).
50. Danishefsky, S. J., DeNinno, S. & Lartey, P. A concise and stereoselective route to the predominant stereochemical pattern of the tetrahydropyranoid antibiotics: an application to indanomycin. *J. Am. Chem. Soc.* **109**, 2082–2089 (1987).
51. Parker, K. A. & Georges, A. T. Reductive aromatization of quinols: synthesis of the C-arylglycoside nucleus of the papulacandins and chaetiacandin. *Org. Lett.* **2**, 497–499 (2000).
52. Gurjar, M. K., Krishna, L. M., Reddy, B. S. & Chorghade, M. S. A versatile approach to anti-asthmatic compound CMI-977 and its six-membered analogue. *Synthesis* **2000**, 557–560 (2000).
53. Banwell, M. G. et al. Small molecule glycosaminoglycan mimetics. Patent WO 2006135973A1 (2006).
54. Mattson, R. J. & Catt, J. D. Piperazinyl-cyclohexanes and cyclohexenes. Patent US 6153611A (2000).
55. Chongquing, P., Zhu, Z., Zhang, M. & Gu, Z. Palladium-catalyzed enantioselective synthesis of 2-aryl cyclohex-2-enone atropisomers: platform molecules for the divergent synthesis of axially chiral biaryl compounds. *Angew. Chem. Int. Ed.* **56**, 4777–4781 (2017).
56. Mahecha-Mahecha, C. et al. Sequential Suzuki–Miyaura coupling/Lewis acid-catalyzed cyclization: an entry to functionalized cycloalkane-fused naphthalenes. *Org. Lett.* **22**, 6267–6271 (2020).
57. Xu, W. & Yoshikai, N. Cobalt-catalyzed directed C–H alkenylation of pivalophenone N–H imine with alkenyl phosphates. *Beilstein J. Org. Chem.* **14**, 709–715 (2018).
58. Kang, D., Kim, J., Oh, S. & Lee, P. H. Synthesis of naphthalenes via platinum-catalyzed hydroarylation of aryl enynes. *Org. Lett.* **14**, 5636–5639 (2012).
59. Zhang, X., Sarkar, S. & Larock, R. C. Synthesis of naphthalenes and 2-naphthols by the electrophilic cyclization of alkynes. *J. Org. Chem.* **71**, 236–243 (2006).
60. Kumar, S. V. et al. Cyclocondensation of arylhydrazines with 1, 3-bis (het) arylmonothio-1, 3-diketones and 1, 3-bis (het) aryl-3-(methylthio)-2-propenones: Synthesis of 1-aryl-3, 5-bis (het) arylpyrazoles with complementary regioselectivity. *J. Org. Chem.* **78**, 4960–4973 (2013).
61. Mohan, C., Singh, P. & Mahajan, M. P. Facile synthesis and regioselective thio-Claisen rearrangements of 5-prop-2-ynyl/enyl-sulfanyl pyrimidinones: transformation to thienopyrimidinones. *Tetrahedron* **61**, 10774–10780 (2005).
62. Splivallo, R. & Ebeler, S. E. Sulfur volatiles of microbial origin are key contributors to human-sensed truffle aroma. *Appl. Microbiol. Biotechnol.* **99**, 2583–2592 (2015).

63. Dai, J. et al. New oblongolides isolated from the endophytic fungus Phomopsis sp. from Melilotus dentata from the shores of the Baltic Sea. *Eur. J. Org. Chem.* **2005**, 4009–4016 (2005).
64. Bunyapaiboonsri, T., Yoiprommarat, S., Srikitikulchai, P., Srichomthong, K. & Lumyong, S. Oblongolides from the Endophytic Fungus Phomopsis sp. BCC 9789. *J. Nat. Prod.* **73**, 55–59 (2010).
65. Shing, T. K. M. & Yang, J. A short synthesis of natural (-)-oblongolide via an intramolecular or a transannular Diels-Alder reaction. *J. Org. Chem.* **60**, 5785–5789 (1995).
66. Magedov, I. V. et al. Reengineered epipodophyllotoxin. *Chem. Commun.* **48**, 10416–10418 (2012).
67. Hur, J., Jang, J. & Sim, J. A Review of the pharmacological activities and recent synthetic advances of γ-butyrolactones. *Int. J. Mol. Sci.* **22**, 2769 (2021).
68. Le, H. V. et al. Design and mechanism of tetrahydrothiophene-based γ-aminobutyric acid aminotransferase inactivators. *J. Am. Chem. Soc.* **137**, 4525–4533 (2015).
69. Kalashnikov, A. I., Sysolyatin, S. V., Sakovich, G. V., Sonina, E. G. & Shchurova, I. A. Facile method for the synthesis of oseltamivir phosphate. *Russ. Chem. Bull.* **62**, 163–170 (2013).
70. Tavakoli, M., Chiu, Y. T. T., Baldi, P., Carlton, A. M. & Van Vranken, D. RMechDB: A public database of elementary radical reaction steps. *J. Chem. Inf. Model.* **63**, 1114–1123 (2023).
71. Tavakoli, H. R., Moosavi, S. M. & Bazgir, A. $ZrOCl_2{\cdot}8H_2O$ as an efficient catalyst for the synthesis of dibenzo [*b*,*i*]xanthene-tetraones and fluorescent hydroxyl naphthalene-1,4-diones. *Res. Chem. Intermed.* **41**, 3041–3046 (2015).
72. Liu, D., Zhou, S. & Gao, J. Room-temperature synthesis of hydroxylnaphthalene-1,4-dione derivative catalyzed by phenylphosphinic acid. *Synth. Commun.* **44**, 1286–1290 (2014).
73. Shaabani, S., Naimi-Jama, M. R. & Maleki, A. Synthesis of 2-hydroxy-1,4-naphthoquinone derivatives via a three-component reaction catalyzed by nanoporous MCM-41. *Dyes Pigm* **122**, 46–49 (2015).
74. Shaterian, H. R. & Mohammadnia, M. Effective preparation of 2-amino-3-cyano-4-aryl-5,10-dioxo-5,10-dihydro-4H-benzo[g]chromene and hydroxyl naphthalene-1,4-dione derivatives under ambient and solvent-free conditions. *J. Mol. Liq.* **177**, 353–360 (2013).
75. Tayama, E., Sato, R., Takedachi, K., Iwamoto, H. & Hasegawa, E. A formal method for the de-N,N-dialkylation of Sommelet–Hauser rearrangement products. *Tetrahedron* **68**, 4710–4718 (2012).
76. Kim, S. H., Lee, H. S., Kim, K. H. & Kim, J. N. An expedient synthesis of poly-substituted 1-arylisoquinolines from δ-ketonitriles via indium-mediated Barbier reaction protocol. *Tetrahedron Lett* **50**, 6476–6479 (2009).
77. Irwin, J. J. et al. A Free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **45**, 177–182 (2005).
78. Szymkuć, S., Wołos, A., Roszak, R. & Grzybowski, B. A. Estimation of multicomponent reactions' yields from networks of mechanistic steps. *Nat. Commun*. (2024) In press.
79. Saebi, M. et al. On the use of real-world datasets for reaction yield prediction. *Chem. Sci.* **14**, 4997–5005 (2023).
80. Liu, Z., Moroz, Y. S. & Isayev, O. The challenge of balancing model sensitivity and robustness in predicting yields: a benchmarking study of amide coupling reactions. *Chem. Sci.* **14**, 10835–10846 (2023).
81. Beker, W. et al. Machine learning may sometimes simply capture literature popularity trends: A case study of heterocyclic Suzuki-Miyaura coupling. *J. Am. Chem. Soc.* **144**, 4819–4827 (2022).
82. Skoraczyński, G. et al. Predicting the outcomes of organic reactions via machine learning: are current descriptors sufficient? *Sci. Rep.* **7**, 3582 (2017).
83. Schneider, N., Lowe, D. M., Sayle, R. A. & Landrum, G. A. Development of a novel fingerprint for chemical reactions and its application to large-scale reaction classification and similarity. *J. Chem. Inf. Model.* **55**, 39–53 (2015).

Acknowledgements

Development of the MECH module within the Allchemy platform (by R.R., A.W., K.M, T.K., B.M.-K., S.S., M.M.) was supported by internal funds of Allchemy, Inc. A.M., L.G., Y.B., P.G., O.P. gratefully acknowledge funding from the National Science Centre, Poland (Award 2018/30/A/ST5/00529). S.B. was supported by a grant from the Priority Research Area Anthropocene under the Strategic Programme Excellence Initiative at the Jagiellonian University. J. M. gratefully acknowledge funding from the Foundation for Polish Science (award TEAM/2017-4/38) – these three awards supported part of experimental validations described in this paper. O.V. and D.T.G. gratefully acknowledges support from the Polish National Science Center, Poland (grants OPUS 2020/37/B/ST4/00017) and the Foundation for Polish Science (TEAM POIR.04.04.00-00-3CF4/16-00). The experimental part of this project also received funding from the European Research Council (ERC) under the European Union's or Horizon Europe research and innovation programme (Grant agreement No. 101097337, ARCHIMEDES to D.T.G.). During paper's revision, L.G., Y.B., and R.F. were also generously supported by the Institute for Basic Science, Korea (Project Code IBS-R020-D1). Analysis of pathways and writing of the paper by B.A.G. was also supported by the Institute for Basic Science, Korea (Project Code IBS-R020-D1).

Author contributions

R.R., A.W., K.M, S.S., M.M. and B.A.G. designed and developed Allchemy platform and performed analyses and calculations described in the paper. A.M. performed syntheses described in Fig. 3, Y.B. and P.G. performed syntheses described in Fig. 4a, B.M.-K. and T.K. performed syntheses described in Fig. 4b, Y.B. performed syntheses described in Fig. 4c, L.G. performed syntheses described in Fig. 4d, O.P. performed syntheses described in Fig. 4a with supervision from J.M., P.G. performed syntheses described in Fig. 4c, S.B. performed syntheses described in Supplementary Section S2 with supervision from J.M. and help from R.F. O.V. and D.T.G. helped with the evaluation of the kinetic networks. B.A.G. conceived and supervised research and wrote the paper with help from other authors.

Competing interests

The authors declare the following competing interests: R.R., A.W., K.M, B.M.-K., T.K., S.S., M.M. and B.A.G. are consultants and/or stakeholders of Allchemy, Inc. Allchemy software and its MECH module is property of Allchemy Inc., USA. All queries about access options to Allchemy, including academic collaborations, should be sent to saraszymkuc@allchemy.net. L.G., A.M., Y.B., P.G., O.P., S.B., R.F., J.M., O.V. and D.T.G. declare no competing interest.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-54611-5.

Correspondence and requests for materials should be addressed to Daniel T. Gryko or Bartosz A. Grzybowski.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.