Ideation and Evaluation of Novel Multicomponent Reactions via Mechanistic Network Analysis and Automation

Babak A. Mahjour[1], Juncheng Lu[2], Jenna Fromer[1], Nicholas Casetti[1], Connor W. Coley[1,3]

[1] Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139
[2] Department of Chemistry, Carnegie Mellon University, Pittsburgh, PA 15213
[3] Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139

Abstract

Novel reactivity is paramount to accessing valuable chemical space. Chemists use mechanistic intuition in conjunction with modern reaction screening techniques to discover, invent, or optimise chemical reactions. We have codified this logic in an automated cheminformatic workflow as one approach to systematic reaction invention. Hundreds of expert-encoded elementary reaction templates were used to construct a highly connected mechanistic network. This network can be used to enumerate reaction pathways for a set of given input substrates and reagents, serving as a qualitative "virtual flask". Our method's predictive capability is first exemplified through the regeneration of mechanistic pathways to the main and potential side products of seven known multicomponent reactions. Then, we showcase its innovative capability in a multicomponent reaction invention pipeline that rapidly screens three component sets of starting materials for scenarios where two components form an intermediate that is captured by a third reactant. Two novel three component transformations proposed by the model were experimentally validated using robotically dosed parallel reaction plates employing a broad range of reaction conditions. We discuss the potential utility of these novel transformations and interrogate the kinetics of both reaction systems with a robot-operated assay.

Main

Novel reactivity can enable alternative access to valuable chemical space and allow for the development of cheaper, faster, or more sustainable syntheses of chemical products or libraries.[1,2] The identification of conditions that enable a specific, selective transformation between molecular structures is a core practice of organic synthesis. While expert chemists routinely optimise reaction conditions to yield known products, developing an entirely new system that enables an unprecedented transformation between substrates remains a significant challenge. We view this intentional development as the essence of 'invention'. On the other hand, identifying the 'initial hit' – the enabling conditions that yield a novel transformation – is often seen as a 'discovery', as reactivity is often complex, non-linear, and sensitive to many variables. Chemists bridge the gap between 'discovery' and 'invention' through experimentation, formulating and testing many hypotheses to maximise the productivity or selectivity toward the outcome of interest. Serendipity has also played a major role in the discovery of transformations and reaction conditions. Modern high throughput experimentation techniques[3,4] and continued improvement of chemistry automation[5,6] have even led to concepts of 'accelerated serendipity',[7-9] where information gain is maximised using minimal resources. Nonetheless, the design space of reaction conditions is virtually infinite, even when limited to the subset of easily automatable chemistry. Thus, we view computer-aided reaction design as an opportunity to guide experimentation away from purely serendipitous discovery and towards the intentional invention of unprecedented chemical transformations.

The computer-aided systematic design of novel reactivity has been researched since the late 1960s,[10-12] but such models have yet to see mainstream use.[13,14] Recently, the increased availability of reaction datasets with large numbers of entries has enabled contemporary attempts at data-driven reactivity models.

https://doi.org/10.26434/chemrxiv-2024-qfjh9 **ORCID:** https://orcid.org/0000-0002-8225-6514 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

These models have been successful for many in-distribution tasks such as yield prediction, regioselectivity prediction, retrosynthesis, and even condition recommendation.[15,16] Certain data-driven methods have been pursued to expand the substrate scope of known transformations,[17,18] but most fail to extrapolate chemical reasoning to both novel and reasonable transformations.[19] An interesting data-driven approach proposed by Segler and Waller uses link prediction within a knowledge graph of millions of reactions to generate reactions between co-reactive substrates, but with varying levels of plausibility judged only on the basis of structural similarity.[20]

Another approach to systematic reaction invention has been through automated experimentation and analysis. Multidimensional screening in conjunction with clever UPLC-MS pooling techniques and spectra processing algorithms can accelerate the identification of novel catalytic reactions.[21,22] Similarly, methods have been proposed to automate the exploration of reaction space, targeting areas of high model uncertainty to maximise the likelihood of enabling a novel transformation.[23-25] While such methods have resulted in unexpected chemistries in controlled systems, they have relied on limiting the reactant pool to a modestly-sized set of reagents predisposed to be highly reactive. Likewise, data-driven, automated exploration of electrochemistry has proven fruitful in identifying reactive pairs when selecting an ideal model substrate.[26]

Reaction discovery can alternatively be treated as the identification of unprecedented mechanistic steps (i.e., elementary reactions) or sequences thereof. For instance, an experimental approach reported by Glorius and coworkers in 2016 targeted a single, specific elementary step by using high throughput screening to test reagents for quenching potential in photocatalysis.[27] Computationally, chemical networks can be generated via quantum chemical exploration of potential energy surfaces and be used to identify potentially novel mechanistic steps at significant cost.[28-31] Alternatively, this mechanism-based strategy of reaction discovery can be formalised and expanded by developing a manually curated set of expert-encoded elementary reactions that serves as a ruleset in algorithms that generate reaction pathways. While reaction network datasets can be tedious to develop, this approach has seen recent success in extrapolation to novel chemistry. In early 2024, a template-based mechanistic model was used by Grzybowski and coworkers to predict the outcome of cationic rearrangements.[32] Contemporaneously to this work, the same group extended their closed-source platform to propose multicomponent and one-pot reactivities.[33] Our goals are similar, but we additionally investigate of the formal multicomponent nature[34] of our newly discovered reactions and pursue more systematic definitions of novelty and utility to select the most interesting pathways for validation.

Herein, we report a generalised method for systematic reaction invention that merges automated experimentation with template-based mechanistic modelling. In our approach, we strike a balance between feasibility and novelty by targeting hypothetical reactions that represent novel combinations of known elementary steps. To do so, we developed an extended version of the SMARTS language to define mechanistic transformations (Extended Data Fig. 1) and created a reaction corpus consisting of hundreds of elementary steps derived from popular reaction types and known named reactions. Each reaction in the corpus, encoded with mechanistic SMARTS, is described by a set of core atoms and abstractable R-groups to flexibly capture the reaction's substrate scope. A network of these generalised templates can be constructed by linking reaction templates that share common substructures (Fig. 1a). Provided a set of input molecules, the network can be used to enumerate viable reaction pathways to intermediates and products. This cheminformatic framework, or "virtual flask", when integrated with the computer aided synthesis planning program ASKCOS,[35] effectively serves as a "digital twin" for experimental organic chemistry, albeit in a qualitative manner that attempts to anticipate only species and not product ratios or rates. For a given set of starting materials, the virtual flask iteratively constructs a "state network", representing the evolution of compounds through the mechanistic network (Fig. 1b). A series of filters and calculated features inform the novelty and feasibility of potential multicomponent reactions for a given set of reactants (Fig. 1c); many combinations of reactants can be evaluated in this manner to prioritise the most novel and feasible combinations for experimental testing.

We first showcase the capabilities of the virtual flask by recreating known chemistries for seven multicomponent reactions. We then demonstrate its use in a cheminformatic pipeline that proposes ideas for novel multicomponent transformations. Ultimately, two novel reactivities proposed by the virtual flask were realised using an automated high throughput assay to rapidly assess reaction conditions. The first reaction

https://doi.org/10.26434/chemrxiv-2024-qfjh9 **ORCID:** https://orcid.org/0000-0002-8225-6514 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

enables rapid access to disubstituted 2-carbamoyl benzoates through the kinetic manipulation of anhydride hydrolysis. The second reaction enables extended access to a chiral quaternary centre which is otherwise inaccessible through typical Mannich reactivity by exploiting an unexpected activation of DMSO with $POCl_3$. Finally, we discuss several kinetic experiments that were performed to study the multicomponent nature of the uncovered chemistries. The modularity and open-source nature of this framework enable it to serve as a general model to assist in predicting, explaining, and discovering chemical reactivity with accuracy and precision that will increase as our computational filters become more robust.

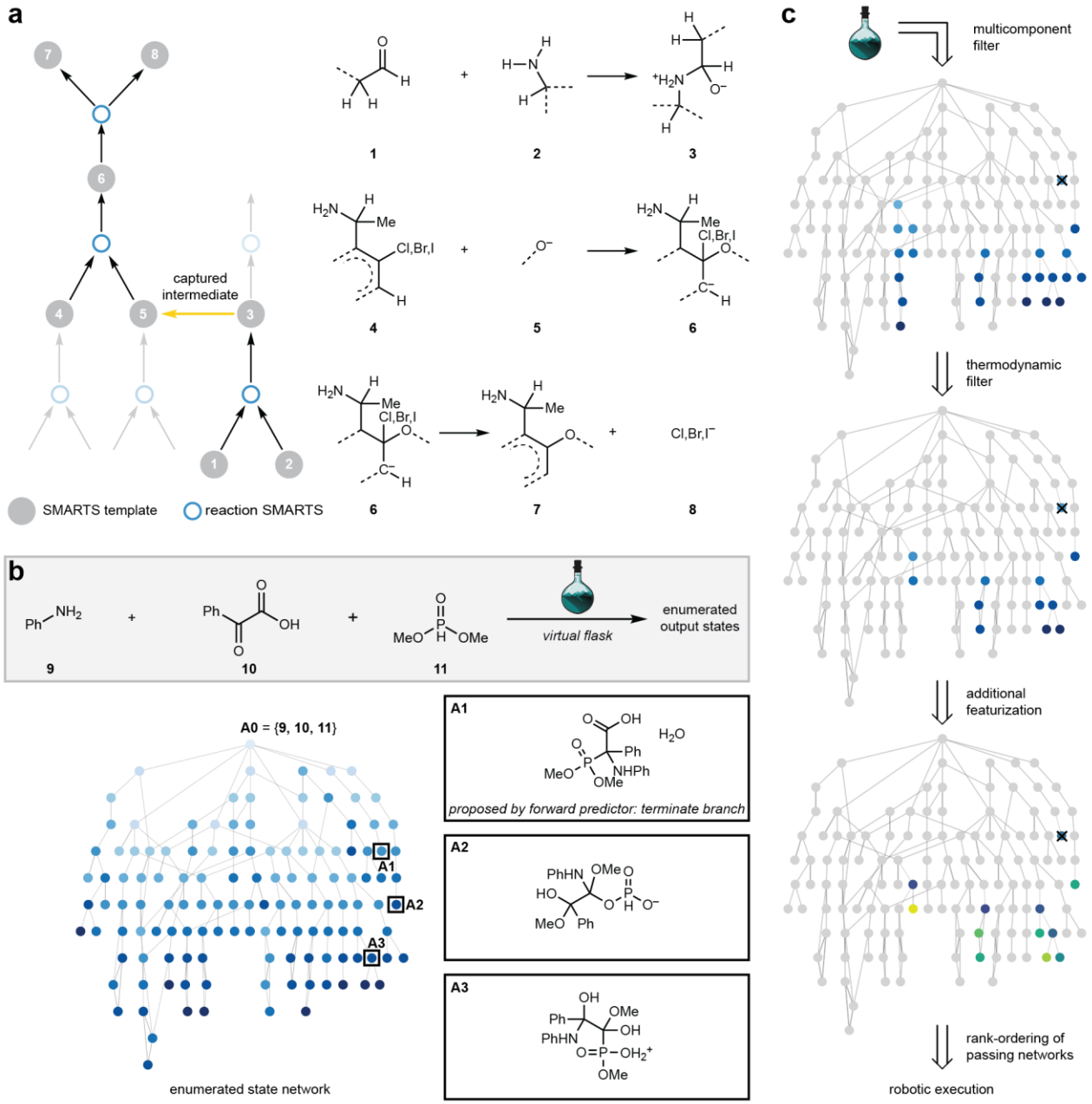


Fig. 1 | The virtual flask screens reactants for novel and feasible multicomponent reactions. a, The encoding scheme used to create a mechanistic network. Elementary reactions, when encoded as SMARTS, can be used to form a network of overlapping intermediates. SMARTS templates are the reactant or product patterns that form a reaction SMARTS when joined by ">>". **b,** Input sets of reactants are fed into the virtual flask, producing a state network that enumerates reaction pathways between the initial reaction state and

https://doi.org/10.26434/chemrxiv-2024-qfjh9 **ORCID:** https://orcid.org/0000-0002-8225-6514 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

intermediate/product states. If a state contains a product proposed by our machine learning-based forward predictor models, the branch is terminated, and the transformation is considered insufficiently novel. **c,** State networks are assessed through a series of post-processing filters and feature calculations. Once the state network is fully generated, each node is assessed for multicomponent character – the existence of a product containing atoms from all three input substrates. States passing this filter are then assessed for thermodynamic feasibility using ground state energy calculations and structural heuristics related to ring strain and charge. Finally, a variety of calculations are computed on nodes passing all previous filters to allow for the organisation and rank ordering of network hits (i.e., to facilitate the prioritisation of which three component reactions to run in the chemistry lab). Features of interest include overall network size and maximum difference of ground state energies between any two intermediates.

Results

The mechanistic network within the virtual flask is constructed based on substructures shared between different reaction templates. For instance, the first elementary reaction of Fig. 1a that reacts aldehyde **1** with amine **2** to form ion **3** is a commonly seen mechanistic step in many iminium-based reactions. The next two transformations shown are the final two steps of the S_NAr reaction, whereby templates **4** and **5** couple to form intermediate **6**, which then decomposes into products **7** and **8**. The mechanistic network identifies compounds which can be products of one elementary step and reactants of another (in this case, starting material **5** could be represented by intermediate **3**). An example output is shown in Fig. 1b, where amine **9**, 2-oxoacid **10**, and phosphite **11** are fed into the virtual flask. The network begins with the initial state **A0**, representing the input set of reactants, and expands to new states after each propagation of the mechanistic network, as indicated by linked connecting state nodes of a lighter colour. The three states **A1**, **A2**, and **A3** are shown and marked on the network, representing possible states in which the input reactants could exist in. Notably, each state maintains a heavy atom mass balance with the initial state and tracks reactive and non-reactive atoms and bonds throughout the generation of the network (Extended Data Fig. 2).

The coverage of the elementary reaction templates in our virtual flask is assessed by recreating known chemistries. Input substrates leading to seven experimentally verified multicomponent reactions were sampled from the methodology literature. For each reaction's set of inputs, the mechanistic network was propagated up to five times to generate the resultant state networks shown on the right of Fig. 2.

In the first example, the Mannich reaction was recreated by feeding amine **12**, aldehyde **13**, and ketone **14** into the virtual flask. This resulted in a state network containing reaction state **15**, consisting of the expected Mannich product and two equivalents of byproduct water (Fig. 2a).[36]

In Fig. 2b, the Greico coupling is recreated with an initial state consisting of aniline **9**, aldehyde **13**, and diene **16**. This example created a simple state network that contained the product state **17**.[37] In Fig. 2c, the Kabachnik-Fields reaction replaces diene **16** with phosphite **11** to generate the expected product state **18**.[38] Interestingly, the resultant state network visually resembles the one produced given the Greico coupling reagents due to shared reactivities.

In Fig. 2d, the Passerini reaction is recreated with acid **19**, aldehyde **20**, and isocyanide **21** to form product state **22**.[39] This reaction contains a unique rearrangement step, possibly reflecting the sparsity of the resultant state network. In Fig. 2e, we show the aryl Petasis reaction that combines aldehyde **23** with amine **24** and boronate **25** to produce state **26** with the expected product and boric acid byproduct.[40] Next the Strecker synthesis was recreated from aldehyde **27**, isocyanate **28**, and amine **29**. This formed state **30**, which was the captured compound in the associated reported chemistry (Fig. 2f).[41] Finally, Fig. 2g shows the recreation of the aldehyde-alkyne-amine reaction, forming state **34** from inputs **31**, **32**, and **33**.[42] By confirming that the virtual flask successfully recapitulated these archetypal multicomponent reactions, we validated both our definition of reaction SMARTS as well as our network traversal algorithm.

https://doi.org/10.26434/chemrxiv-2024-qfjh9 **ORCID:** https://orcid.org/0000-0002-8225-6514 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

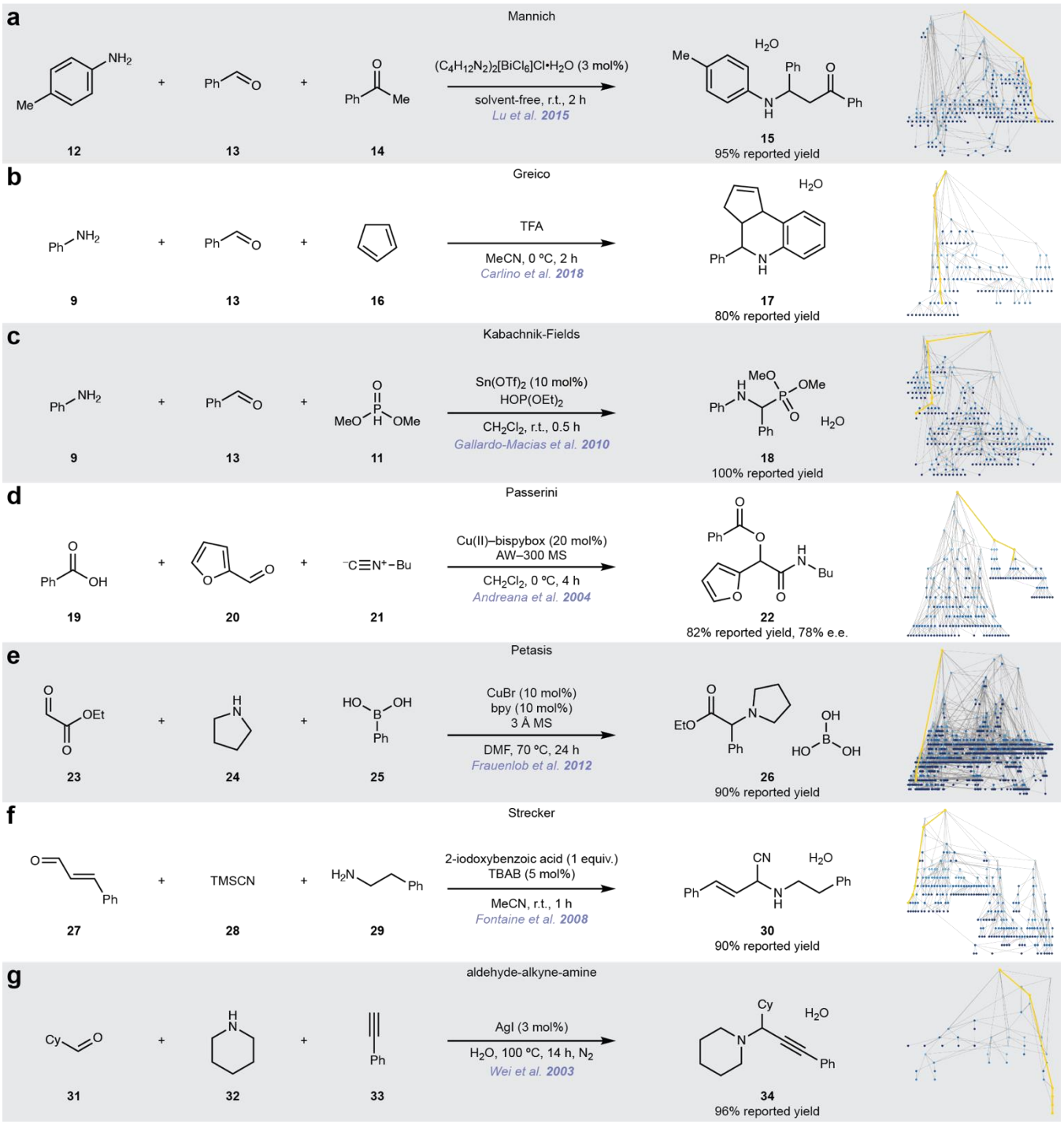


Fig. 2 | The virtual flask correctly generates the products of known reactions as well as potential byproducts. a-g, Virtual flask inputs and corresponding product states. A mechanistic pathway from substrates to product is highlighted in gold. Reaction conditions shown above and below the arrow were not considered during mechanistic reaction propagation.

Next, we prospectively demonstrated the pipeline's use in the discovery of novel multicomponent reactions. triplets of substrates were enumerated from a random set of 39 compounds from our in-house inventory, primarily containing amines, acids, and various common nucleophiles (Fig. S9). The substrate

https://doi.org/10.26434/chemrxiv-2024-qfjh9 **ORCID:** https://orcid.org/0000-0002-8225-6514 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

library was not predisposed towards reactive species. Each set of three compounds was fed into the virtual flask to generate a state network with a maximum depth of five mechanistic propagations. 2,603 state networks were generated, filtered, and evaluated for novelty and feasibility (see Methods and Supplementary Information §5). An interactive two-dimensional histogram binning passing hits by overall network size (as an indication of potential mechanistic complexity) and maximum difference in ground state energies found in proposed pathways (as an indication of overall thermodynamic feasibility) was developed to allow for manual analysis and down selection (Fig. 3a-b). We focused on hits with small network sizes and low calculated max ΔG, inspecting mechanistic pathways and cross-referencing reaction databases online to ensure the reactions were truly novel and not false positives. An example of a passing mechanistic sequence as shown in the analysis dashboard can be found in Extended Data Fig. 3.

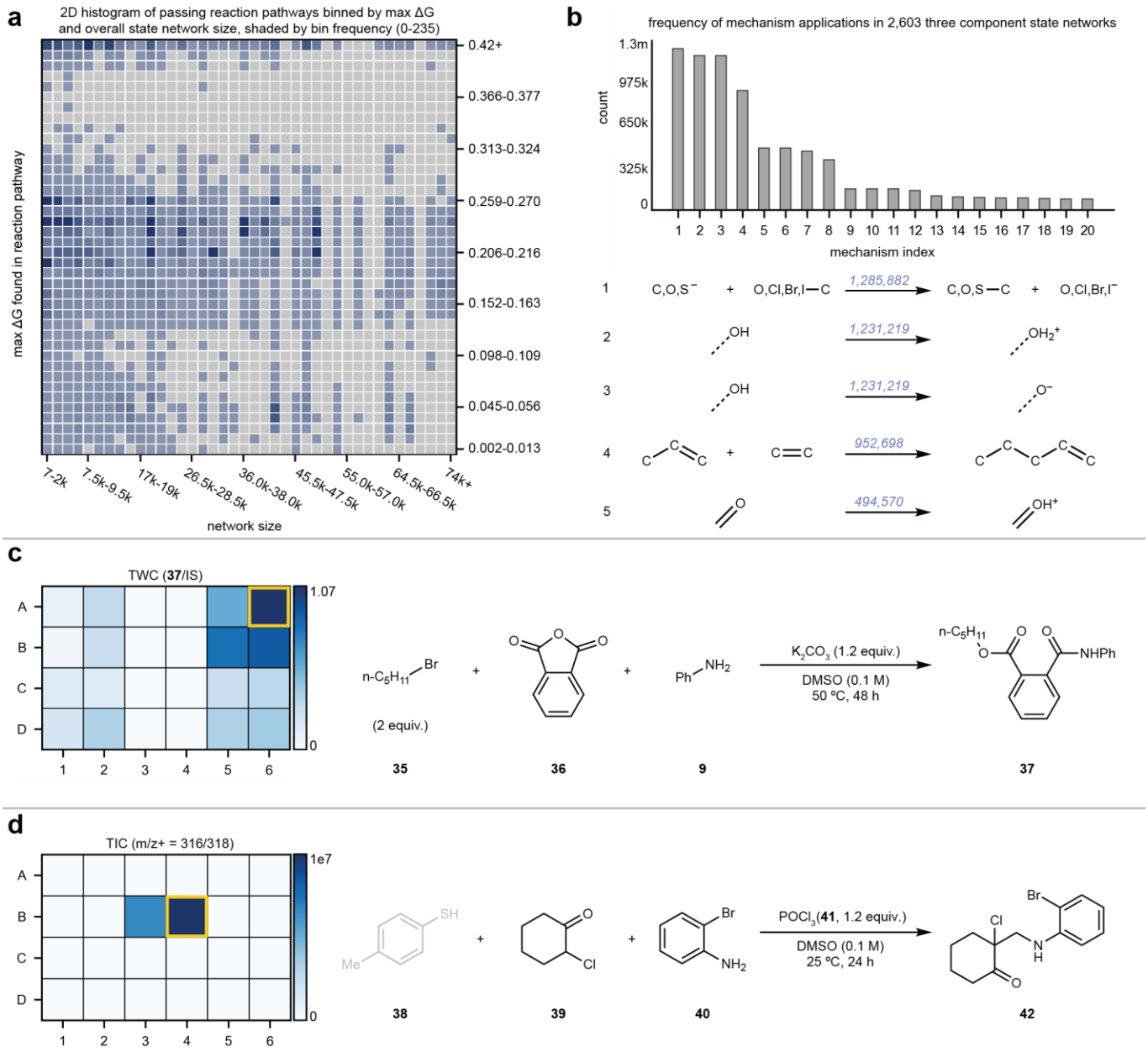


Fig. 3 | An automated assay rapidly identifies enabling conditions. a, An interactive 2-dimensional heatmap showing the distribution of passing hits identified by the model. Reaction pathways are binned into

https://doi.org/10.26434/chemrxiv-2024-qfjh9 **ORCID:** https://orcid.org/0000-0002-8225-6514 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

buckets based on calculated scores. In this case, the x-axis bins pathways by their state network's overall size, reflecting potential for side reactivities, and the y-axis bins by the maximum difference in ground state energies between intermediates in reaction pathways, serving as a metric for feasibility. Greyed out bins have no hits, otherwise, the bin's opacity correlates with the number of hits it contains. During manual downselection, hypotheses with low scores for both were given higher priority. **b,** Given our inventory, certain mechanistic steps were frequently utilised in networks with passing hits (max: 235). The top 20 mechanisms found in the dataset are shown in the bar chart, and the top five are visualised below along with the number of times the mechanism was applied to a state across all passing networks. **c,** A ring-opening reaction captured via an S_N2 step. The high throughput condition screen reveals that the multicomponent transformation is strongly controlled by the reaction conditions. Tin chloride, used in columns 3 and 4, deactivates the pathway to **37** entirely, while the addition of potassium carbonate led to its highest yield. **d,** An unexpected Swern oxidation-like activation of DMSO with $POCl_3$ in conjunction with a Mannich addition led to the formation of a quaternary centre. Thiol **38** has been greyed out as further experimentation revealed it did not participate in the formation of product **42**. A legend defining the 24-condition heatmap screens in **c** and **d** can be found in Extended Data Fig. 5. TWC, total wavelength chromatograph; IS, internal standard; TIC, total ion count.

We showcase two reactions that were identified and selected for automated execution after our analysis of the hits proposed by the model (Fig. 3c-d). As each step of the proposed mechanistic sequences generated by the virtual flask (Extended Data Fig. 4) were derived from templates inspired by known reactions, reasonable reaction conditions were generated for each experiment with a machine learning-based reaction context recommender (Supplementary Information §7-8).[43] While the mechanism proposed by the virtual flask that forms both products may not be entirely accurate, they are plausible enough to use as a basis for reaction condition screening. Conditions amenable to parallel screening using a single 24-well high throughput experiment were selected (Extended Data Fig. 5), and reaction mixtures were analysed with LC-MS as well as 1D and 2D NMR (see Methods and Supplementary Information §11-12).

The first verified reaction was a ring opening sequence shown in Fig. 3a, where phthalic anhydride **36** is attacked by aniline **9** leading to hydrolysis and formation of a carboxylate which undergoes S_N2 with bromopentane **35** to generate **37.** While chemically simple, and further testing with forward synthesis machine learning models showed that they were able to predict **37** given **9**, **35**, and **36** as inputs, this reaction does not appear to have been reported previously (Figs. S10-S13 in Supplementary Information). We suspect that this style of reactivity can be extended to more efficiently introduce complexity to natural product analogues such as artesunate,[44] a powerful antimalarial.[45]

Additionally, we identify a Mannich aminomethylation sequence where bromoaniline **40** is methylated by an unexpected but mechanistically tractable chlorosulfonium ion generated from $POCl_3$ (**41**)-activated DMSO,[46,47] which is subsequently quenched through the Mannich addition of **39** to form product **42**. Originally, the combination of **38**, **39**, and **40** yielded an unexpected mass hit in the presence of both solvent and $POCl_3$ that we then observed, through an ablation study, did not depend on thiol **38.** However, that observed mass hit, later confirmed to be product **42**, was successfully predicted by the virtual flask when DMSO and $POCl_3$ were included alongside the starting materials. To assess whether our reaction usefully expands the scope of Mannich-like chemistry, we attempted to form the same product using **39** and **40** with typical Mannich conditions (in DMSO at room temperature for 48 h,[48] as well at 90 ºC for 24 h) and formaldehyde instead of $POCl_3$/DMSO. No product was observed. Furthermore, very few examples of direct aminomethylations onto alpha substituted cyclohexanones have ever been reported (see Supplementary Information §10). Of these few examples, only a single reaction was found to utilize an aryl amine.[49]

We anticipate this method to serve as a parallel method to forming quaternary centres to the catalytic asymmetric Mannich system developed by Toste and coworkers[50] in addition to serving as a potential method to insert methylenes to form C-C-N bonds from carbonyl and amine building blocks, in a similar vein to Liu and coworkers' recent report on aminative Suzuki-Miyaura couplings.[51] We expect further development of this reactivity to enable novel access to chemical space around valuable small molecule therapeutics that may

https://doi.org/10.26434/chemrxiv-2024-qfjh9 **ORCID:** https://orcid.org/0000-0002-8225-6514 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

have been difficult to readily access with standard Mannich chemistries, such as tramadol and its derivatives.[52] We are continuing to evaluate the scope of the reactivity and developing chiral ligands to maximise enantioselectivity. Assay details for both reactions are shown in Extended Data Fig. 5.

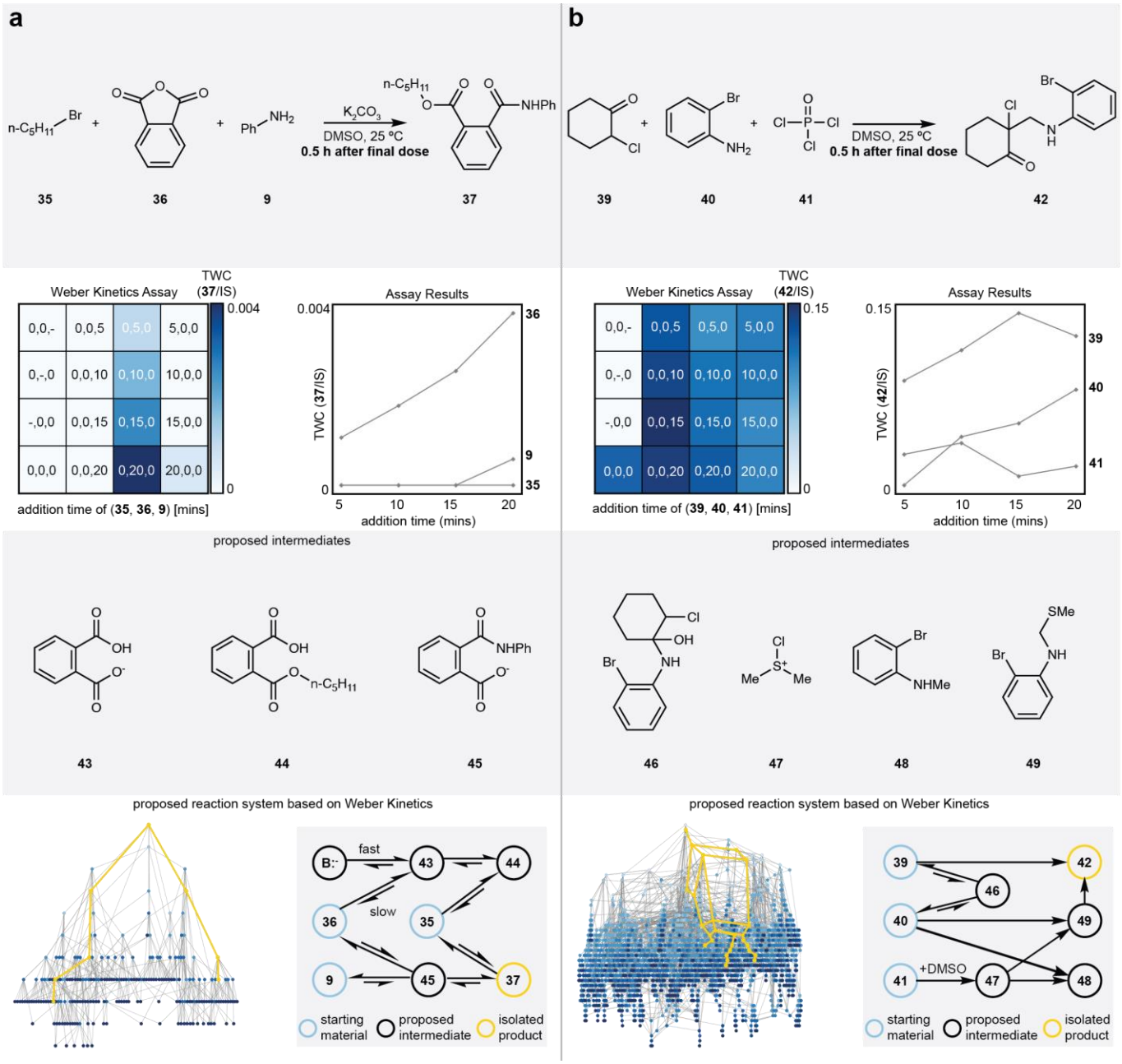


Fig. 4 | A robotic kinetics assay assesses the formal multicomponent nature of reactivity. Heatmaps show normalised product abundance under different addition orders and timing of three (or two) components. In the first column, all two component reactions are tested in addition to the three-component reaction with all components being dosed simultaneously. In the following three columns, two components are added initially while the addition of the third component is delayed in 5-minute intervals, testing all possible permutations of the kinetics assay. We showcase proposed and identified intermediates and hypothesize a reaction system based off the kinetic assay results and the species believed to exist throughout the time course of each reaction. **a,** The kinetics assay performed on the first hit reported in this study. Each reaction is coloured by the UV integration ratio between the product and the internal standard. Formation of the product **43** is fastest

https://doi.org/10.26434/chemrxiv-2024-qfjh9 **ORCID:** https://orcid.org/0000-0002-8225-6514 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

when **35** and **9** are dosed before **36**. **b,** The kinetics assay was performed on the reaction between **39**, **40**, and $POCl_3$ (**41**). The formation of the product is fastest when **39** and **40** are dosed before **41** is added. TWC, total wavelength chromatograph; IS, internal standard.

It can be difficult to determine whether a reaction is a formal multicomponent coupling or a series of two component reactions in one pot.[34] To better understand the reactivity of the transformations identified in this study, an assay detailed in ref. 34 (herein referred to as a "Weber kinetic assay") was adapted to our robotics platform to assess the kinetics of the coupling. In addition to testing each two-component permutation and the reaction where all three components are dosed in tandem, we robotically executed time series experiments wherein the third component was added in 5-minute intervals after the other two components were initially dosed. Reaction mixtures were analysed via LC/MS (see Methods and Supplementary Information §11f-11j). The sensitivity of the observed product quantity with respect to various dosing schedules reveals information about reaction network dynamics. In Fig. 4a, the assay results indicate that formation of the product **37** is fastest when aniline **9** and bromopentane **35** are dosed before the addition of anhydride **36**. We reason this result as a phenomenon emerging from competing reaction pathways. In this experiment, potassium carbonate was added to each well before any substrate addition. We suspect that due to the initial basic conditions of the reaction vessel, anhydride is quickly converted into **43** and **44** if allowed to pre-react before all components are dosed; side reactivities consume (or occupy) the starting materials before the formation of the product is realised within the time scale of the experiment. As this side reactivity is reversible, we anticipate that the product would form eventually, given its thermodynamic stability. We performed a complimentary experiment where potassium carbonate was added in rapid succession after the final compound of each reaction was dosed. (Supplementary Information §11i) No product in any reaction over the time scale of the experiment was observed, perhaps indicating the role of a basic pre-environment in the activation of the intended reaction pathway.

Similarly, in Fig. 4b, a kinetic profile is revealed with a noticeably faster formation of product **42** when **39** and **40** are dosed before the addition of $POCl_3$ (**41**). Again, this is reasoned through hypothetical side reactivities, in addition to the existence of side products identified in the reaction crude. Most of the reactivity in this system is rooted in the formation of the chlorosulfonium ion **47** generated from DMSO after activation via $POCl_3$. We note that this activation is not a competing pathway due to the preparation of stock solutions. The formation of intermediate **46** from **39** and **40** is transient and does not consume the initial starting materials, but other substrate combinations with **47** led to a variety of side products including the aminomethylated **48** and intermediate **49** (both identified by UPLC-MS, see Supplementary Information §11e). However, in the series where **46** (and by extension, **47**) is added last, the formation of key intermediate **49** is maximised, leading to a more rapid formation of the final product. This analysis, while not an extensive mechanistic interrogation, provides clues as to how this reaction can be developed further and showcases the utility of robotics in the execution of kinetic snapshot assays probing the order and timing of addition.

Discussion

Using a mechanistic network derived from encoded elementary steps, a virtual flask was developed that can generate reaction state networks given a set of input molecular structures. This methodology was used in a cheminformatic pipeline to screen sets of reactants for combinations likely to produce novel multicomponent reactivity. In conjunction with robotics, hypothesised reactions were able to be rapidly tested in the lab, leading to the discovery of two novel multicomponent transformations given a limited initial inventory set. The second reaction, a previously unreported aminomethylation/Mannich sequence using $POCl_3$, is the subject of ongoing investigations in our lab.

We anticipate that further development of this platform will only increase the model's robustness in reaction exploration and will serve as a valuable addition to the computer-aided synthesis toolkit. We are striving to develop more accurate mechanistic networks to better engineer reaction conditions (including the design of new catalysts) to shift kinetic favourability towards the most desirable reaction pathways. Improvements in the comprehensiveness of the mechanistic template corpus can be made adaptively based

https://doi.org/10.26434/chemrxiv-2024-qfjh9 **ORCID:** https://orcid.org/0000-0002-8225-6514 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

on new experimental data and will enable more accurate anticipation of competing reactions. Integrating retrosynthetic analysis and “reaction targeting” logic into our model architecture will better guide the virtual flask towards valuable regions of chemical space. Finally, to mitigate the analytical bottleneck of robotic experimentation, we are investigating the robustness of automated structure elucidation by tandem mass spectrometry in this context of reaction screening and discovery.

https://doi.org/10.26434/chemrxiv-2024-qfjh9 **ORCID:** https://orcid.org/0000-0002-8225-6514 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Methods

Computational Workflow

Our algorithmic strategy involves identifying opportunities to capture reactive intermediates formed by two components with a third component; this type of multicomponent chemistry is commonly associated with iminium formation and seen in familiar reactions such as the Ugi reaction, the Mannich reaction, and the Petasis reaction. We consider a reaction to be *multicomponent* if the sequential addition of reactants is unable to produce the same product distribution compared to the reaction in which the substrates are added simultaneously.[34] We consider a multicomponent reaction to be *novel* if its products cannot be predicted by separately trained machine learning models for major product prediction.[53] As a novelty filter, any state that contains such products is deemed terminal and is removed. In other words, the chemistry generating this set of reaction outcomes is considered not novel, and the state is no longer propagated as the products are considered stable. The second filter assesses thermodynamic feasibility, eliminating nodes that contain more than three ionic species or one species with a formal charge greater than two absolute charge units. Post-processing calculations further filter the propagated network. First, any remaining nodes that do not contain a product species with at least one atom from each of the input substrates are removed. Non-multicomponent reactions are removed by ensuring that all reactive atoms have participated in a unified transformation to remove hits where reactivities occur at separate, unrelated locations. 3D geometry optimisation is performed on compounds in remaining nodes to assess the stability of hypothetical product species; states that contain species with highly strained conformations or non-stable motifs are removed. As an additional filter, the ground state energy of each intermediate in the proposed mechanistic pathway is calculated with GFN2-xTB,[54] and any route with an absolute difference in ground state energy between two intermediates greater than 10 kcal/mol is removed. Finally, a series of properties calculations are performed on all remaining non-filtered states to rank-order reactions of interest, assisting manual analysis and downselection of hits before experimental validation. In this work, we primarily utilised metrics such as the size of the overall network, calculated ground state energies between intermediates contained in passing mechanistic routes, the complexity of the product based on the generated ring system, and “molecular uniqueness” calculated as a product’s minimum Tanimoto distance to any molecule in DrugBank as features of interest.

Discovery Assay Workflow

Reaction conditions, excluding solvents, were sampled from a context recommender machine learning model[43] given reactions that formed components of the overall mechanistic sequence being proposed by the virtual flask (Supplementary Information §6-7). DMSO was selected as a universal solvent for initial screening due to its automation-friendly nature. Stock solutions recipes were calculated and followed to prepare an automated high throughput assay. An OpenTrons Flex liquid handling robot was used to transfer aliquots of stock solutions to a 24-well reactor block. Reactions were run with stirring at room temperature or a temperature proposed by the context recommender. After 24h or 48h (decided through studying literature using similar reagents, depending on the chemistry), the assay was worked up with water and diluted to 10 mM before being injected into our UPLC-MS system for analysis of the crude mixtures. Passing hits were scaled up in a fume hood, and a semipreparative column was used to isolate the products for full spectroscopic analysis. See Supplementary Information §11a-11e.

Weber Kinetics Assay[34] Workflow

We used an assay developed by Lutz Weber and colleagues as a previously established expert method to interrogate the multicomponent nature of the novel reactivities uncovered in our study.[34] We adapted their experiment to allow our robotic platform to automate the order of addition and time series addition permutations the assay consists of. Stock solutions were calculated and prepared as before, but the robotic workflow was modified to screen the dosing and permutation of reagent additions in a 4x4 reaction grid. For

https://doi.org/10.26434/chemrxiv-2024-qfjh9 **ORCID:** https://orcid.org/0000-0002-8225-6514 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

each reaction sequence permutation of a three-component reaction, 4 dosing times of the third component were tested, in addition to all two component reaction combinations as well as the original three component reaction as a control (where all three components are added in tandem). As before, reactions were worked up, diluted, and equimolar internal standard was added before injection into our UPLC-MS. See Supplementary Information §11h-11m.

Data Availability

All data, including experimental details, spectral data and raw .fid files generated or analysed during this study are included in Supplementary Information. All code and applications used in the study are available at https://github.com/coleygroup/virtual_flask.

Code Availability

All code and applications produced during this study can be found at https://github.com/coleygroup/virtual_flask under an MIT license.

References

1 Mahjour, B., Shen, Y., Liu, W. & Cernak, T. A map of the amine–carboxylic acid coupling system. *Nature* **580**, 71-75 (2020). https://doi.org:10.1038/s41586-020-2142-y

2 Zhang, R., Mahjour, B., Outlaw, A., McGrath, A., Hopper, T., Kelley, B., Walters, W. P. & Cernak, T. Exploring the combinatorial explosion of amine–acid reaction space via graph editing. *Communications Chemistry* **7**, 22 (2024). https://doi.org:10.1038/s42004-024-01101-w

3 Shevlin, M. Practical High-Throughput Experimentation for Chemists. *ACS Medicinal Chemistry Letters* **8**, 601-607 (2017). https://doi.org:10.1021/acsmedchemlett.7b00165

4 Mahjour, B., Zhang, R., Shen, Y., McGrath, A., Zhao, R., Mohamed, O. G., Lin, Y., Zhang, Z., Douthwaite, J. L., Tripathi, A. & Cernak, T. Rapid planning and analysis of high-throughput experiment arrays for reaction discovery. *Nature Communications* **14**, 3924 (2023). https://doi.org:10.1038/s41467-023-39531-0

5 Abolhasani, M. & Kumacheva, E. The rise of self-driving labs in chemical and materials sciences. *Nature Synthesis* **2**, 483-492 (2023). https://doi.org:10.1038/s44160-022-00231-0

6 Shi, Y., Prieto, P. L., Zepel, T., Grunert, S. & Hein, J. E. Automated Experimentation Powers Data Science in Chemistry. *Accounts of Chemical Research* **54**, 546-555 (2021). https://doi.org:10.1021/acs.accounts.0c00736

7 McNally, A., Prier, C. K. & MacMillan, D. W. C. Discovery of an α-Amino C–H Arylation Reaction Using the Strategy of Accelerated Serendipity. *Science* **334**, 1114-1117 (2011). https://doi.org:doi:10.1126/science.1213920

8 Prieto Kullmer, C. N., Kautzky, J. A., Krska, S. W., Nowak, T., Dreher, S. D. & MacMillan, D. W. C. Accelerating reaction generality and mechanistic insight through additive mapping. *Science* **376**, 532-539 (2022). https://doi.org:doi:10.1126/science.abn1885

9 Kanan, M. W., Rozenman, M. M., Sakurai, K., Snyder, T. M. & Liu, D. R. Reaction discovery enabled by DNA-templated synthesis and in vitro selection. *Nature* **431**, 545-549 (2004). https://doi.org:10.1038/nature02920

10 Balaban, A. Chemical graphs. 3. Reactions with cyclic 6-membered transition states. *Rev. Roum. Chim* **12**, 875-902 (1967).

11 Herges, R. Reaction planning: Computer-aided reaction design. *Tetrahedron Computer Methodology* **1**, 15-25 (1988). https://doi.org:https://doi.org/10.1016/0898-5529(88)90005-X

12 Herges, R. & Hoock, C. Reaction Planning: Computer-Aided Discovery of a Novel Elimination Reaction. *Science* **255**, 711-713 (1992). https://doi.org:doi:10.1126/science.255.5045.711

13 Jorner, K., Tomberg, A., Bauer, C., Sköld, C. & Norrby, P.-O. Organic reactivity from mechanism to machine learning. *Nature Reviews Chemistry* **5**, 240-255 (2021). https://doi.org:10.1038/s41570-021-00260-x

https://doi.org/10.26434/chemrxiv-2024-qfjh9 **ORCID:** https://orcid.org/0000-0002-8225-6514 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

14 Houk, K. N. & Liu, F. Holy Grails for Computational Organic Chemistry and Biochemistry. *Accounts of Chemical Research* **50**, 539-543 (2017). https://doi.org:10.1021/acs.accounts.6b00532
15 Shen, Y., Borowski, J. E., Hardy, M. A., Sarpong, R., Doyle, A. G. & Cernak, T. Automation and computer-assisted planning for chemical synthesis. *Nature Reviews Methods Primers* **1**, 23 (2021). https://doi.org:10.1038/s43586-021-00022-5
16 Tu, Z., Stuyver, T. & Coley, C. W. Predictive chemistry: machine learning for reaction deployment, reaction development, and reaction discovery. *Chemical Science* **14**, 226-244 (2023). https://doi.org:10.1039/D2SC05089G
17 Wang, X., Yao, C., Zhang, Y., Yu, J., Qiao, H., Zhang, C., Wu, Y., Bai, R. & Duan, H. From theory to experiment: transformer-based generation enables rapid discovery of novel reactions. *Journal of Cheminformatics* **14**, 60 (2022). https://doi.org:10.1186/s13321-022-00638-z
18 Zhang, Z.-J., Li, S.-W., Oliveira, J. C. A., Li, Y., Chen, X., Zhang, S.-Q., Xu, L.-C., Rogge, T., Hong, X. & Ackermann, L. Data-driven design of new chiral carboxylic acid for construction of indoles with C-central and C–N axial chirality via cobalt catalysis. *Nature Communications* **14**, 3149 (2023). https://doi.org:10.1038/s41467-023-38872-0
19 Bort, W., Baskin, I. I., Gimadiev, T., Mukanov, A., Nugmanov, R., Sidorov, P., Marcou, G., Horvath, D., Klimchuk, O., Madzhidov, T. & Varnek, A. Discovery of novel chemical reactions by deep generative recurrent neural network. *Scientific Reports* **11**, 3178 (2021). https://doi.org:10.1038/s41598-021-81889-y
20 Segler, M. H. S. & Waller, M. P. Modelling Chemical Reasoning to Predict and Invent Reactions. *Chemistry – A European Journal* **23**, 6118-6128 (2017). https://doi.org:https://doi.org/10.1002/chem.201604556
21 Robbins, D. W. & Hartwig, J. F. A Simple, Multidimensional Approach to High-Throughput Discovery of Catalytic Reactions. *Science* **333**, 1423-1427 (2011). https://doi.org:doi:10.1126/science.1207922
22 Troshin, K. & Hartwig, J. F. Snap deconvolution: An informatics approach to high-throughput discovery of catalytic reactions. *Science* **357**, 175-181 (2017). https://doi.org:doi:10.1126/science.aan1568
23 Dragone, V., Sans, V., Henson, A. B., Granda, J. M. & Cronin, L. An autonomous organic reaction search engine for chemical reactivity. *Nature Communications* **8**, 15733 (2017). https://doi.org:10.1038/ncomms15733
24 Gromski, P. S., Henson, A. B., Granda, J. M. & Cronin, L. How to explore chemical space using algorithms and automation. *Nature Reviews Chemistry* **3**, 119-128 (2019). https://doi.org:10.1038/s41570-018-0066-y
25 M. Mehr, S. H., Caramelli, D. & Cronin, L. Digitizing chemical discovery with a Bayesian explorer for interpreting reactivity data. *Proceedings of the National Academy of Sciences* **120**, e2220045120 (2023). https://doi.org:doi:10.1073/pnas.2220045120
26 Zahrt, A. F., Mo, Y., Nandiwale, K. Y., Shprints, R., Heid, E. & Jensen, K. F. Machine-Learning-Guided Discovery of Electrochemical Reactions. *Journal of the American Chemical Society* **144**, 22599-22610 (2022). https://doi.org:10.1021/jacs.2c08997
27 Hopkinson, M. N., Gómez-Suárez, A., Teders, M., Sahoo, B. & Glorius, F. Accelerated Discovery in Photocatalysis using a Mechanism-Based Screening Method. *Angewandte Chemie International Edition* **55**, 4361-4366 (2016). https://doi.org:https://doi.org/10.1002/anie.201600995
28 Barter, D., Clark Spotte-Smith, E. W., Redkar, N. S., Khanwale, A., Dwaraknath, S., Persson, K. A. & Blau, S. M. Predictive stochastic analysis of massive filter-based electrochemical reaction networks. *Digital Discovery* **2**, 123-137 (2023). https://doi.org:10.1039/D2DD00117A
29 Zhao, Q. & Savoie, B. M. Simultaneously improving reaction coverage and computational cost in automated reaction prediction tasks. *Nature Computational Science* **1**, 479-490 (2021). https://doi.org:10.1038/s43588-021-00101-3
30 Dewyer, A. L., Argüelles, A. J. & Zimmerman, P. M. Methods for exploring reaction space in molecular systems. *WIREs Computational Molecular Science* **8**, e1354 (2018). https://doi.org:https://doi.org/10.1002/wcms.1354

https://doi.org/10.26434/chemrxiv-2024-qfjh9 **ORCID:** https://orcid.org/0000-0002-8225-6514 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

31 Wang, L.-P., Titov, A., McGibbon, R., Liu, F., Pande, V. S. & Martínez, T. J. Discovering chemistry with an ab initio nanoreactor. *Nature Chemistry* **6**, 1044-1048 (2014). https://doi.org:10.1038/nchem.2099

32 Klucznik, T., Syntrivanis, L.-D., Baś, S., Mikulak-Klucznik, B., Moskal, M., Szymkuć, S., Mlynarski, J., Gadina, L., Beker, W., Burke, M. D., Tiefenbacher, K. & Grzybowski, B. A. Computational prediction of complex cationic rearrangement outcomes. *Nature* **625**, 508-515 (2024). https://doi.org:10.1038/s41586-023-06854-3

33 Roszak, R., Gadina, L., Wołos, A., Makkawi, A., Mikulak-Klucznik, B., Bilgi, Y., Molga, K., Gołębiowska, P., Popik, O. & Klucznik, T. Systematic, computational discovery of multicomponent reactions and one-pot sequences. (2024).

34 Weber, L., Illgen, K. & Almstetter, M. Discovery of new multi component reactions with combinatorial methods. *Synlett* **1999**, 366-374 (1999).

35 Coley, C. W., Thomas, D. A., Lummiss, J. A. M., Jaworski, J. N., Breen, C. P., Schultz, V., Hart, T., Fishman, J. S., Rogers, L., Gao, H., Hicklin, R. W., Plehiers, P. P., Byington, J., Piotti, J. S., Green, W. H., Hart, A. J., Jamison, T. F. & Jensen, K. F. A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* **365**, eaax1566 (2019). https://doi.org:doi:10.1126/science.aax1566

36 Lu, H., Wu, R., Cheng, H., Nie, S., Tang, Y., Gao, Y. & Luo, Z. An Efficient, Mild, Solvent-Free, One-Pot Three-Component Mannich Reaction Catalyzed by (C4H12N2)2[BiCl6]Cl·H2O. *Synthesis* **47**, 1447-1454 (2015). https://doi.org:10.1055/s-0034-1380278

37 Carlino, L., Christodoulou, M. S., Restelli, V., Caporuscio, F., Foschi, F., Semrau, M. S., Costanzi, E., Tinivella, A., Pinzi, L., Lo Presti, L., Battistutta, R., Storici, P., Broggini, M., Passarella, D. & Rastelli, G. Structure–Activity Relationships of Hexahydrocyclopenta[c]quinoline Derivatives as Allosteric Inhibitors of CDK2 and EGFR. *ChemMedChem* **13**, 2627-2634 (2018). https://doi.org:https://doi.org/10.1002/cmdc.201800687

38 Gallardo-Macias, R. & Nakayama, K. Tin(II) Compounds as Catalysts for the Kabachnik-Fields Reaction under Solvent-Free Conditions: Facile Synthesis of α-Aminophosphonates. *Synthesis* **2010**, 57-62 (2010). https://doi.org:10.1055/s-0029-1217091

39 Andreana, P. R., Liu, C. C. & Schreiber, S. L. Stereochemical Control of the Passerini Reaction. *Organic Letters* **6**, 4231-4233 (2004). https://doi.org:10.1021/ol0482893

40 Frauenlob, R., García, C., Bradshaw, G. A., Burke, H. M. & Bergin, E. A Copper-Catalyzed Petasis Reaction for the Synthesis of Tertiary Amines and Amino Esters. *The Journal of Organic Chemistry* **77**, 4445-4449 (2012). https://doi.org:10.1021/jo3003503

41 Fontaine, P., Chiaroni, A., Masson, G. & Zhu, J. One-Pot Three-Component Synthesis of α-Iminonitriles by IBX/TBAB-Mediated Oxidative Strecker Reaction. *Organic Letters* **10**, 1509-1512 (2008). https://doi.org:10.1021/ol800199b

42 Wei, C., Li, Z. & Li, C.-J. The First Silver-Catalyzed Three-Component Coupling of Aldehyde, Alkyne, and Amine. *Organic Letters* **5**, 4473-4475 (2003). https://doi.org:10.1021/ol035781y

43 Gao, H., Struble, T. J., Coley, C. W., Wang, Y., Green, W. H. & Jensen, K. F. Using Machine Learning To Predict Suitable Conditions for Organic Reactions. *ACS Central Science* **4**, 1465-1476 (2018). https://doi.org:10.1021/acscentsci.8b00357

44 Presser, A., Feichtinger, A. & Buzzi, S. A simplified and scalable synthesis of artesunate. *Monatshefte für Chemie - Chemical Monthly* **148**, 63-68 (2017). https://doi.org:10.1007/s00706-016-1865-9

45 Rosenthal, P. J. Artesunate for the Treatment of Severe Falciparum Malaria. *New England Journal of Medicine* **358**, 1829-1836 (2008). https://doi.org:doi:10.1056/NEJMct0709050

46 Patil, S. M., Kulkarni, S., Mascarenhas, M., Sharma, R., Roopan, S. M. & Roychowdhury, A. DMSO–POCl3: a reagent for methylthiolation of imidazo[1,2-a]pyridines and other imidazo-fused heterocycles. *Tetrahedron* **69**, 8255-8262 (2013). https://doi.org:https://doi.org/10.1016/j.tet.2013.07.017

https://doi.org/10.26434/chemrxiv-2024-qfjh9 **ORCID:** https://orcid.org/0000-0002-8225-6514 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

47 Jadhav, S. D. & Singh, A. Oxidative Annulations Involving DMSO and Formamide: K2S2O8 Mediated Syntheses of Quinolines and Pyrimidines. *Organic Letters* **19**, 5673-5676 (2017). https://doi.org:10.1021/acs.orglett.7b02838
48 Ibrahem, I., Casas, J. & Córdova, A. Direct Catalytic Enantioselective α-Aminomethylation of Ketones. *Angewandte Chemie International Edition* **43**, 6528-6531 (2004). https://doi.org:https://doi.org/10.1002/anie.200460678
49 Neochoritis, C. G., Eleftheriadis, N., Tsiantou, A., Stephanidou-Stephanatou, J. & Tsoleridis, C. A. One-Pot DBU-Promoted Synthesis of Hydroacridinones and Spirohexahydropyrimidines. *Synlett* **24**, 2768-2772 (2013). https://doi.org:10.1055/s-0033-1339922
50 Yang, X., Phipps, R. J. & Toste, F. D. Asymmetric Fluorination of α-Branched Cyclohexanones Enabled by a Combination of Chiral Anion Phase-Transfer Catalysis and Enamine Catalysis using Protected Amino Acids. *Journal of the American Chemical Society* **136**, 5225-5228 (2014). https://doi.org:10.1021/ja500882x
51 Onnuch, P., Ramagonolla, K. & Liu, R. Y. Aminative Suzuki–Miyaura coupling. *Science* **383**, 1019-1024 (2024). https://doi.org:doi:10.1126/science.adl5359
52 Shen, Q., Qian, Y.-y., Xu, X.-j., Li, W., Liu, J.-g. & Fu, W. Design, synthesis and biological evaluation of N-phenylalkyl-substituted tramadol derivatives as novel μ opioid receptor ligands. *Acta Pharmacologica Sinica* **36**, 887-894 (2015). https://doi.org:10.1038/aps.2014.171
53 Tetko, I. V., Karpov, P., Van Deursen, R. & Godin, G. State-of-the-art augmented NLP transformer models for direct and single-step retrosynthesis. *Nature Communications* **11**, 5575 (2020). https://doi.org:10.1038/s41467-020-19266-y
54 Bannwarth, C., Ehlert, S. & Grimme, S. GFN2-xTB—An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *Journal of Chemical Theory and Computation* **15**, 1652-1671 (2019). https://doi.org:10.1021/acs.jctc.8b01176

Acknowledgements

This work was supported by the National Science Foundation under Grant No. CHE-2144153 and the MIT-Novo Nordisk Artificial Intelligence Postdoctoral Fellows Program. The authors thank Jordan Liles and Anji Been for helpful discussions around the two prospective case studies, Bridget Becker and Sarah Willis of MIT DCIF for support in NMR operation and analysis, Joules Provenzano and Runzhong Wang for help with structure elucidation, Guangqi Wu for automation assistance, and Joonyoung Joung for help seeding the mechanistic template collection.

Author Information

Authors and Affiliations

Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139
Babak A. Mahjour, Jenna Fromer, Nicholas Casetti, Connor W. Coley

Department of Chemistry, Carnegie Mellon University, Pittsburgh, PA 15213
Juncheng Lu

Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139
Connor W. Coley

https://doi.org/10.26434/chemrxiv-2024-qfjh9 **ORCID:** https://orcid.org/0000-0002-8225-6514 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Contributions

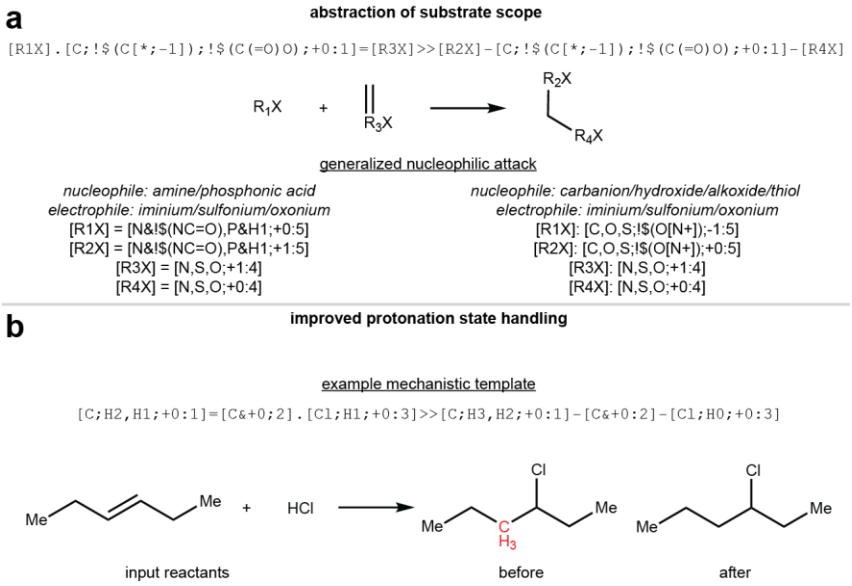
B.M ideated and developed the platform. B.M. and J.L. performed chemistry experiments. J.F. performed cheminformatic analyses. N.C. assisted in thermodynamic calculations. C.W.C. proposed and supervised the study. All authors analysed the data and aided in writing the manuscript.

Ethics Declaration

The authors declare no competing interests.

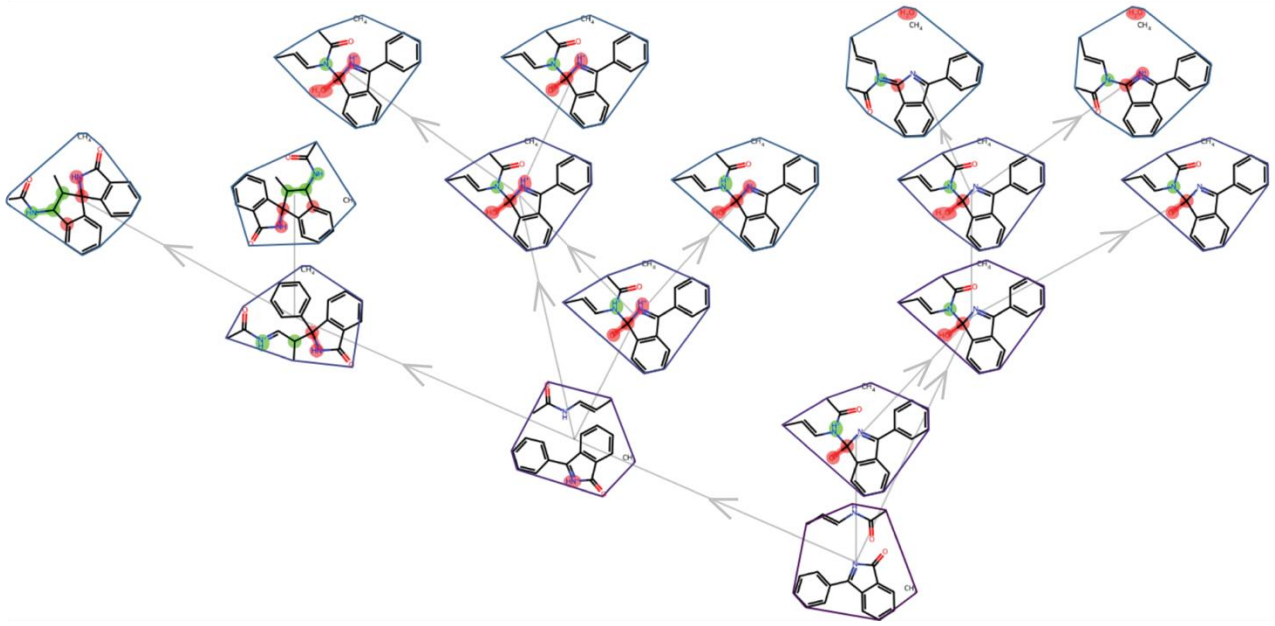
https://doi.org/10.26434/chemrxiv-2024-qfjh9 **ORCID:** https://orcid.org/0000-0002-8225-6514 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0

Extended Data and Figures



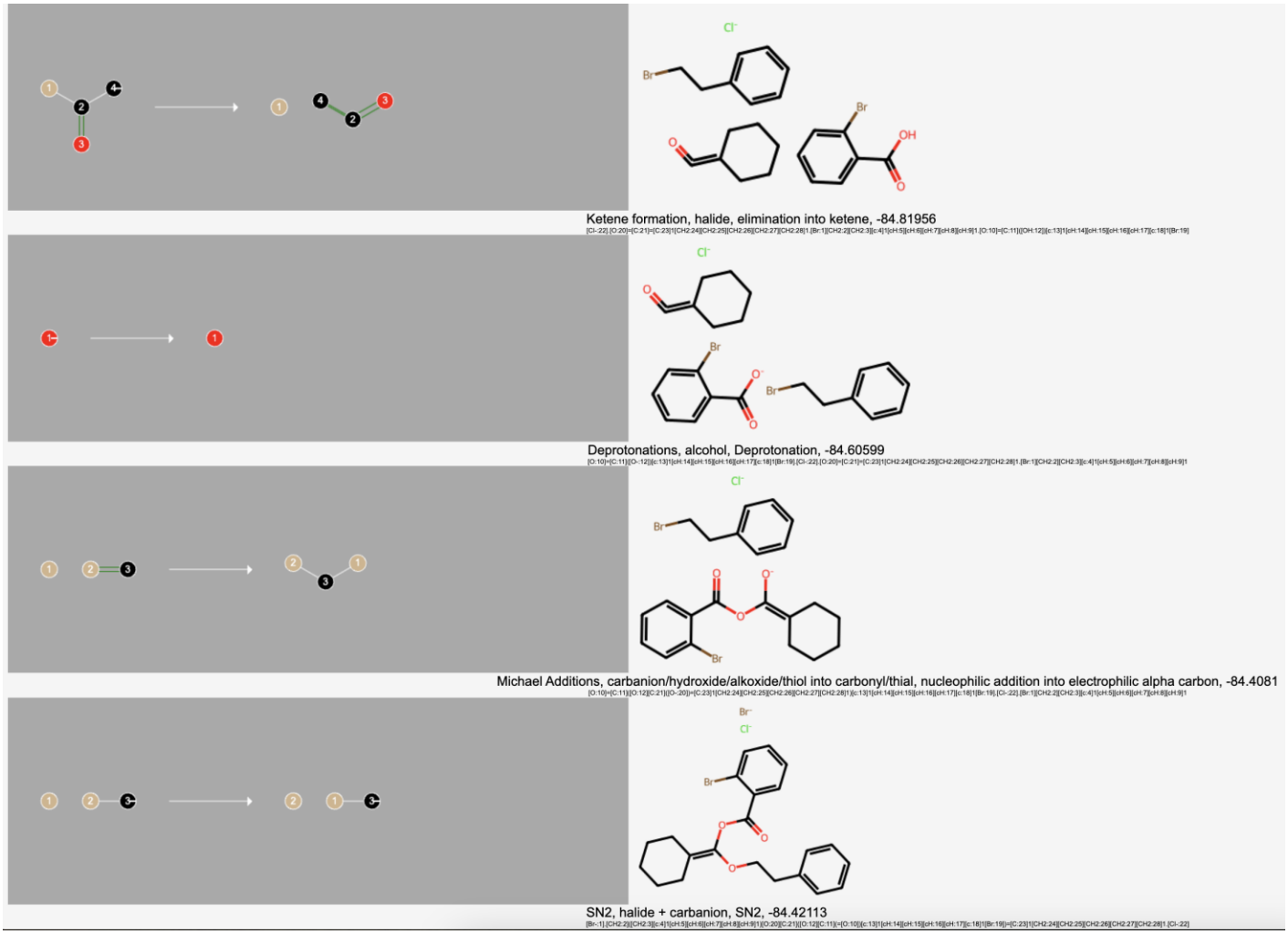
Extended Data Fig. 1 | Enhancements of SMARTS Language to facilitate mechanistic encodings. a, To minimise the size and improve the legibility of our SMARTS corpus, abstractable “R” tokens were introduced to the SMARTS language. While SMARTS are already incredibly expansive, this encoding style clearly delineates the core mechanistic structure and sequence of a reaction from the reactivity’s substrate scope. In a downstream step, all scope-mechanism combinations are created by replacing the R tokens with chemical structures representing the structural environment known to be amenable with a certain mechanistic step. **b,** Often, chemical transformations require specific substitution patterns at reactive centres. To avoid the need to repeat the same mechanistic encoding for each protonation state amenable to a transformation, while maintaining the ability to limit matching reactants, reaction SMARTS template application was modified to allow for multiple hydrogens to be specified in the product. Thus, protons are correctly conserved when the protonation state is not able to be implicitly defined.

https://doi.org/10.26434/chemrxiv-2024-qfjh9 **ORCID:** https://orcid.org/0000-0002-8225-6514 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0



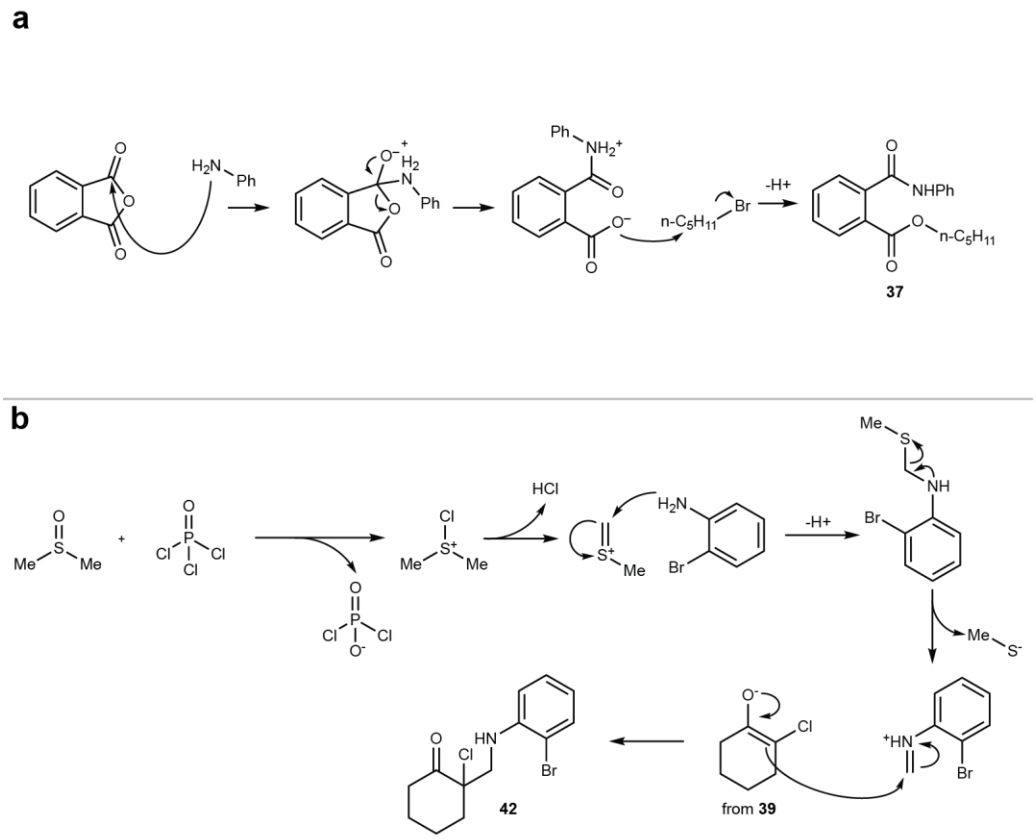
Extended Data Fig. 2 | Exemplary Enumerated State Network. An interactive force directed layout of a state network. Each state is bound by its images' convex hull, which is coloured by a hue correlating to the propagative step in which the state first appeared. Reactive atoms are highlighted with a colour corresponding to the substrate in which the atom originated (in this case, the indolone, red, the amide, green, and methane non-participating).

https://doi.org/10.26434/chemrxiv-2024-qfjh9 **ORCID:** https://orcid.org/0000-0002-8225-6514 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0



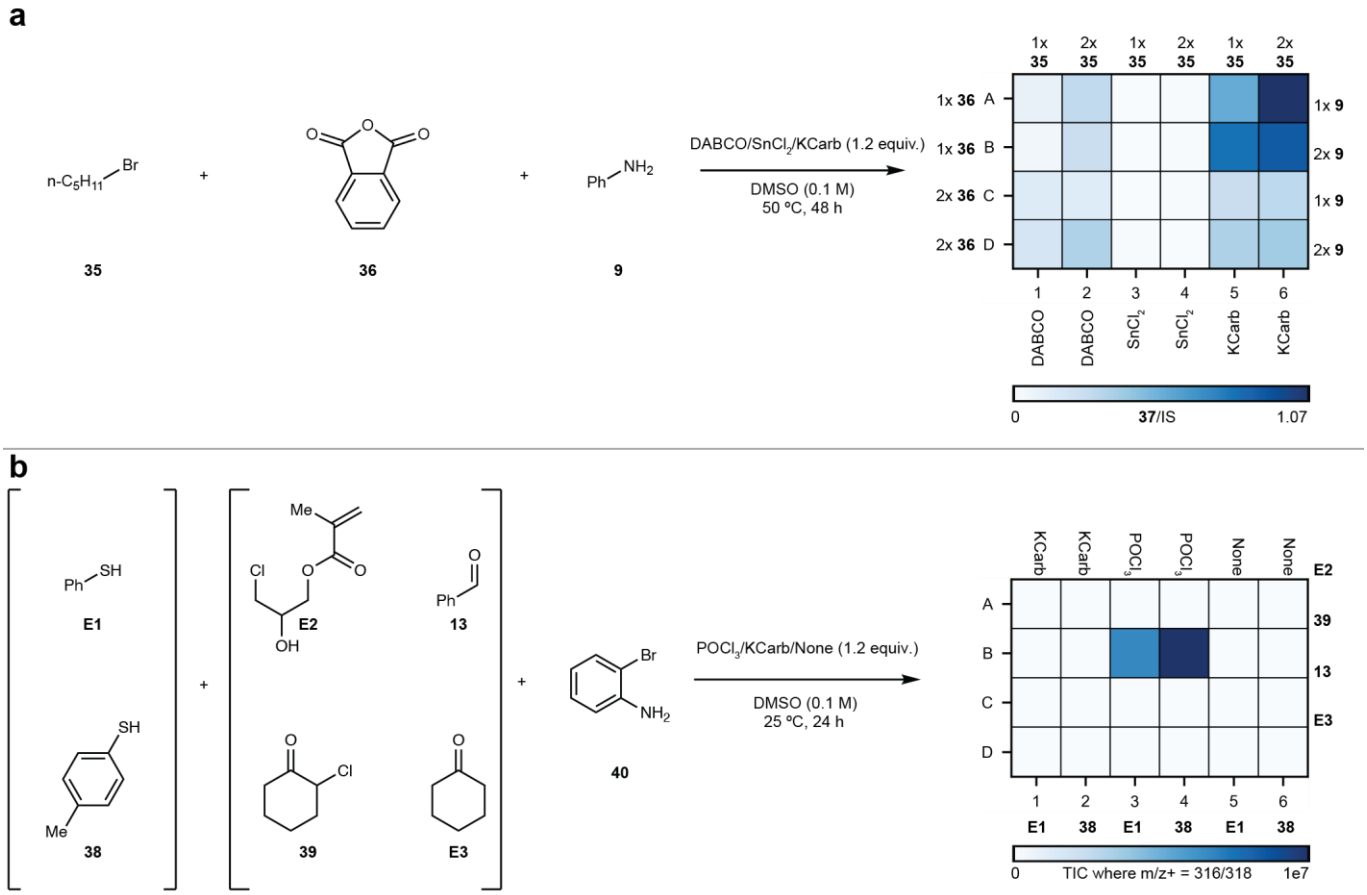
Extended Data Fig. 3 | Example enumerated mechanistic sequence hit. Passing hits can be visualised and inspected via interactive dashboard. Each hit is associated with a mechanistic sequence and additional metadata fully explaining the model's logic in forming the product and assessed metrics of value, with an emphasis on human readability.

https://doi.org/10.26434/chemrxiv-2024-qfjh9 **ORCID:** https://orcid.org/0000-0002-8225-6514 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0



603
604 **Extended Data Fig. 4 | Generated mechanistic sequences for novel reactivities identified through**
605 **workflow. a,** The virtual flask's proposed pathway that predicted the formation of product **37**. A nucleophilic
606 attack followed by hydrolysis opens the anhydride ring, generating a handle to capture bromopentane via SN2
607 and form a stable product. **b,** The virtual flask's hypothesis for the mechanistic sequence that forms **42**. Two
608 equivalents of the bromoaniline are proposed to capture the chlorosulfonium ion generated from Swern
609 oxidation-like activation of DMSO with $POCl_3$. Oxidation followed by formation of an iminium is then captured
610 via Mannich addition to form a quaternary centre.
611

https://doi.org/10.26434/chemrxiv-2024-qfjh9 **ORCID:** https://orcid.org/0000-0002-8225-6514 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0



Extended Data Fig. 5 | Assay Details for Fig. 3. a, The reaction condition assay ran that enabled the formation of desired product **37**. Pleasingly, the assay enabled strong control of the system via reaction conditions, deactivating the pathway entirely when using tin II chloride. These results enabled a faster understanding of the mechanistic network underlying the multicomponent reactivity. **b,** A multicomponent reaction assay that led to the unexpected product **42** using DMSO/$POCl_3$ as the third component as opposed to either of the intended thiols.

https://doi.org/10.26434/chemrxiv-2024-qfjh9 **ORCID:** https://orcid.org/0000-0002-8225-6514 Content not peer-reviewed by ChemRxiv. **License:** CC BY-NC-ND 4.0