

Automatic discovery of chemical reactions using imposed activation

Cyrille Lavigne,¹ Gabriel dos Passos Gomes,^{§1,2} Robert Pollice,^{§1,2} Alán Aspuru-Guzik^{*1,2,3,4}

¹ Department of Computer Science, University of Toronto, 214 College St., Toronto, Ontario M5T 3A1, Canada.

² Chemical Physics Theory Group, Department of Chemistry, University of Toronto, 80 St George St, Toronto, Ontario M5S 3H6, Canada.

³ Vector Institute for Artificial Intelligence, 661 University Ave Suite 710, Toronto, Ontario M5G 1M1, Canada.

⁴ Lebovic Fellow, Canadian Institute for Advanced Research (CIFAR), 661 University Ave, Toronto, Ontario M5G

§ These authors contributed equally to this work.

* Corresponding author: aspuru@utoronto.ca

Abstract

Computational power and quantum chemical methods have improved immensely since computers were first applied to the study of reactivity, but the *de novo* prediction of chemical reactions has remained challenging. We show that complex reactions can be efficiently and autonomously predicted using chemical activation imposed by simple geometrical constraints. Our approach is demonstrated on realistic and challenging chemistry, such as a triple cyclization cascade involved in the total synthesis of a natural product and several oxidative addition reactions of complex drug-like molecules. Notably and in contrast with traditional hand-guided computational chemistry calculations, our method requires minimal human involvement and no prior knowledge of products or mechanisms. Imposed activation can be a transformational tool to screen for chemical reactivity and mechanisms as well as to study byproduct formation and decomposition.

Main

Under typical conditions, chemical kinetics, with timescales of seconds to days, is at least fifteen orders of magnitude slower than the femtosecond timescales of quantum dynamics. Even the most facile reactions are vanishingly unlikely to occur spontaneously in large-scale, state-of-the-art computer simulations.¹ Consequently, the *de novo* prediction of chemical reactivity remains one of the most important challenges in computational chemistry.

Despite the unfavourable timescales, multiple methods have been developed to predict chemical reactions using dynamical simulations, potential energy surface exploration, and machine learning (ML). In dynamical approaches, time-dependent simulations are biased to increase the probability of rare events by artificial attractive interatomic forces²⁻⁵ or by performing simulations under high-energy conditions.^{6,7} Alternatively, reaction paths can be generated by “walking” the potential energy surface between products and reactants, probing only necessary dimensions of the full surface.⁸⁻¹⁰ Machine learning

models can predict chemical reactions based on learned information from the literature and databases (e.g., patent filings).^{11–15} Finally, heuristics approaches encode and formalize human-derived chemical transformation rules.^{16–18}

Nevertheless, current techniques have significant drawbacks that curtail their general use. Standard computational chemistry techniques require fairly accurate initial guess structures to converge to reasonable transition states. Importantly, this necessitates knowledge of the reaction mechanism itself. Systematic and exhaustive reaction pathway searches are computationally costly,¹⁹ and difficult to scale to larger molecules.^{20–22} Heuristics approaches can map extensive reaction networks²³ but face issues when confronted with non-conventional bonding and organometallic complexes.¹⁶

Although these issues could, in principle, be solved using ML and data mining, the available data (e.g., from patent filings and publications) is heavily clustered around molecular families thought to be of economic or academic importance²⁴ and is skewed towards reactions commonly performed.¹³ Data biases produce, for example, unnecessary sequences of protection and deprotection steps.¹⁴ ML approaches trained on currently available data can be thought of as predicting the actions of chemists rather than the actions of molecules.

Herein, we describe a simple computational approach that automatically and robustly explores chemical reaction pathways from knowledge only of the reactants and their reactive bonds. Specifically, we show that these pathways can be obtained by conformational exploration with a chemically activating constraint. Reactions of sizable molecules are found rapidly with only modest computational resources using state-of-the-art semi-empirical methods.²⁵ We demonstrate this approach using three case studies: (1) an epoxide-initiated stereoselective polyene tricyclization cascade with many side products, (2) a solvent-mediated Michael addition with a proton shuttle involving multiple water molecules, and (3) the late-stage oxidative addition of ten drug-like molecules with electronically unsaturated palladium phosphine complexes.

Results

Transition state theory describes chemical reactions as primarily occurring through the transition state, the highest energy saddle point (a critical point of order 1) that lies on the reaction coordinate connecting reactants and products.²⁶ If the barrier to products is a mountain range, the transition state is the highest point on the most accessible pass through this range.²⁷ Transition states are not readily found without a suitable initial geometry. The high dimensionality of the coordinate space of even moderately sized molecules precludes any brute force search,²⁸ and characterizing transition states requires costly second-derivative calculations.²⁹

Conformer generation,³⁰ the identification of local minima on the potential energy surface, suffers from the same dimensionality issues but is otherwise significantly simpler. Indeed, energy minimization always yields a local minimum, irrespective of starting structure, as a consequence of the variational principle.²⁶ Modern conformer generation methods can

rapidly and reliably find and identify many low-lying conformers starting from only a single reasonable chemical structure as input.^{31–35}

Here, we find reaction pathways and products in a fully automated fashion using conformer generation with applied constraints. A chemical coordinate (such as the length of a bond) is chosen as the activating coordinate q^\ddagger . Conformer generation is performed while constraining q^\ddagger to a value between the initial reactant structure and possible product structures to explore the orthogonal coordinates q_\perp^\ddagger . To continue with the mountain range analogy, this corresponds to a search for valleys, local energy minima in the constrained space amidst the range separating products from reactants. These valleys, found by conformational exploration, are possible reaction pathways. A subsequent relaxed scan of q^\ddagger pushes the molecule along discovered pathways, walking the valleys to the products. We note that neither approximate transition structures, nor the identity of reaction products, nor any training data are required in the search.

a Standard conformer search

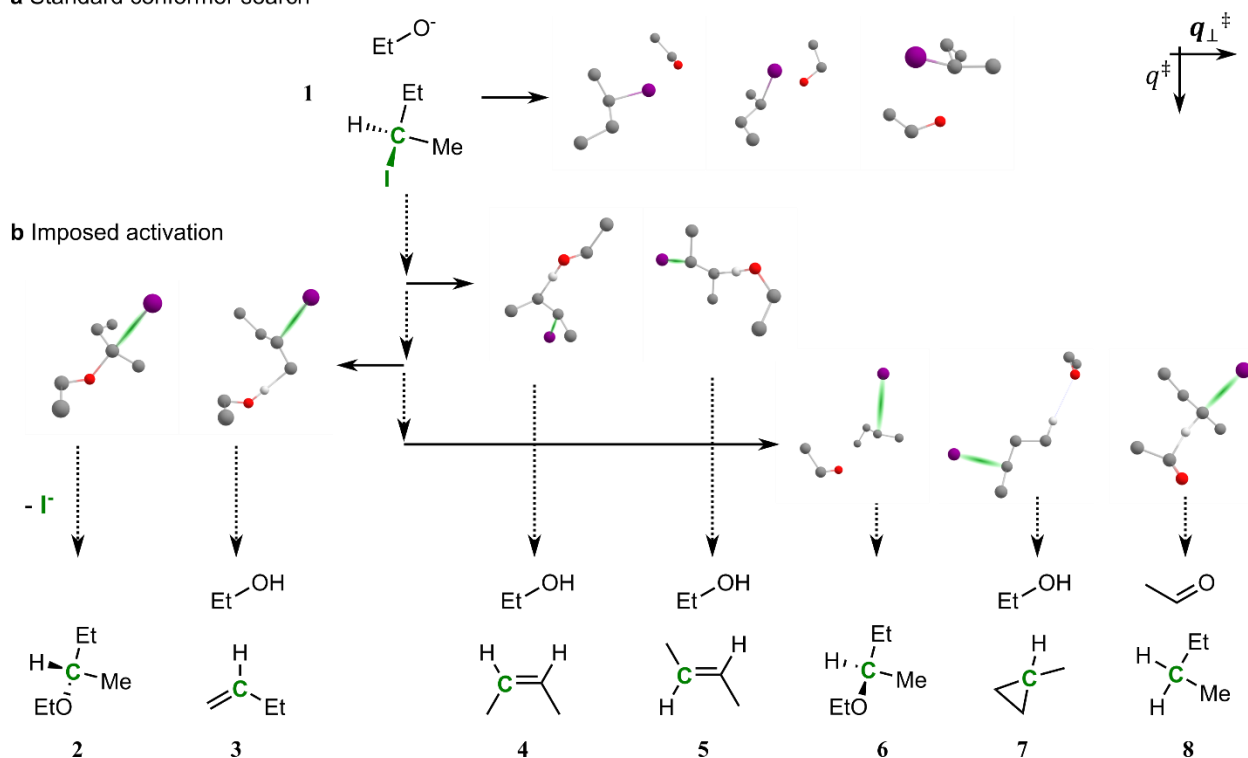


Figure 1: Diagram of reaction search by conformer exploration with imposed activation. **a**, Conformer search methods generate stable three-dimensional molecular structures, such as those shown for complex **1**, composed of a molecule of (*R*)-2-iodobutane and an ethoxide anion. **b**, Our reaction prediction methodology consists of constraining a specific activation coordinate q^\ddagger to out-of-equilibrium values (vertical dotted arrows) and performing searches for activated conformers with the other coordinates free (solid horizontal arrows). This is demonstrated here on **1**, using the carbon-iodine bond as q^\ddagger (green). Further increasing the carbon-iodine distance of the activated conformers yield reaction pathways to multiple products (compounds **2–8**). Non-reacting hydrogen atoms are omitted.

For example, consider the various elimination and substitution reactions of (*R*)-2-iodobutane and ethoxide anion. Generating ground-state conformers for this system yields many loosely bound complexes (Figure 1a). Reaction prediction from these structures is challenging, as they are quite far from any transition states, such as those shown in Figure 1b.

The traditional approach to obtaining a transition state for a given reaction requires setting up the reacting molecules in a configuration close to the transition state, but this requires a fairly good assessment of the reaction mechanism and transition state geometry.³⁶ Coordinate driving and related approaches circumvent this issue by pulling together (in relaxed scans) pairs of atoms to induce new bonds to form.^{20,21,37} For example, bringing together the ethoxide anion oxygen and the α -carbon of 2-iodobutane produces the S_N2 reaction. However, a fully systematic search of all possible bond-forming pairs scales quadratically with the number of atoms²¹ and generally can only be performed for small systems.⁵

The cornerstone of our approach is to activate the reactants using a single coordinate q^\ddagger that can produce many interesting reactions, thereby circumventing the scaling issues of searching for possible bond formation pairs systematically. In the case of the (*R*)-2-iodobutane molecule shown in Figure 1, bond dissociation energies dictate that the carbon-iodine bond is readily broken.³⁸ Stretching the carbon-iodine bond (increasing q^\ddagger) produces an activated electrophilic molecule. In this case, a constrained conformer search (in the orthogonal space q_\perp^\ddagger) yields geometries close to the S_N2 transition state, as the ethoxide oxygen stabilizes the complex when it lies close to the halogenated carbon. Further scanning q^\ddagger starting from one of these transition structures results in the S_N2 products (**2** in Figure 1).

Notably, other activated conformers corresponding to different reactions are found in q_\perp^\ddagger . Indeed, solely stretching the carbon-iodine bond of other activated structures also yields E2 and E1 elimination reactions (with Hofmann³⁹ product **3** and Zaitsev⁴⁰ products **4** and **5**), S_N1 substitution reactions (with products **2** and **6**), γ -elimination to methylcyclopropane **7** as well as a curious alkoxide hydride transfer⁴¹ forming **8**. The latter two reactions, which have significantly higher estimated activation energies, are unlikely to occur in this system but demonstrate the complex chemistry that can be obtained automatically from imposed activation by a single constrained coordinate.

Reactions from coordinate scans are highly dependent on starting geometries, especially when the scanned coordinate is not bond-forming. Indeed, for an iodine-carbon bond scan to alone generate a pathway to compound **7** by γ -elimination requires the ethoxide to be placed in very close proximity to the eliminated hydrogen. The diversity of products and paths discovered by the imposed activation approach comes at a cost: initial guess structures need to be relatively close to reactive pathways. Crucially, constrained conformer generation can produce these higher-energy, activated structures,^{31,42} and is

critical to making our approach practical. A diverse sample of similarly simple reactions is described in the SI. Below, we demonstrate this approach on some challenging examples.

Cyclization Cascade in the Synthesis of a Natural Product

A robust and automated reaction prediction approach can help to understand and rationalize complicated reactions with multiple outcomes. This is demonstrated here using the epoxide-initiated polycyclization of **9** to form **10** (Figure 2), an essential step in a short total synthesis of the fungal meroterpenoid berkeleyone A.⁴³

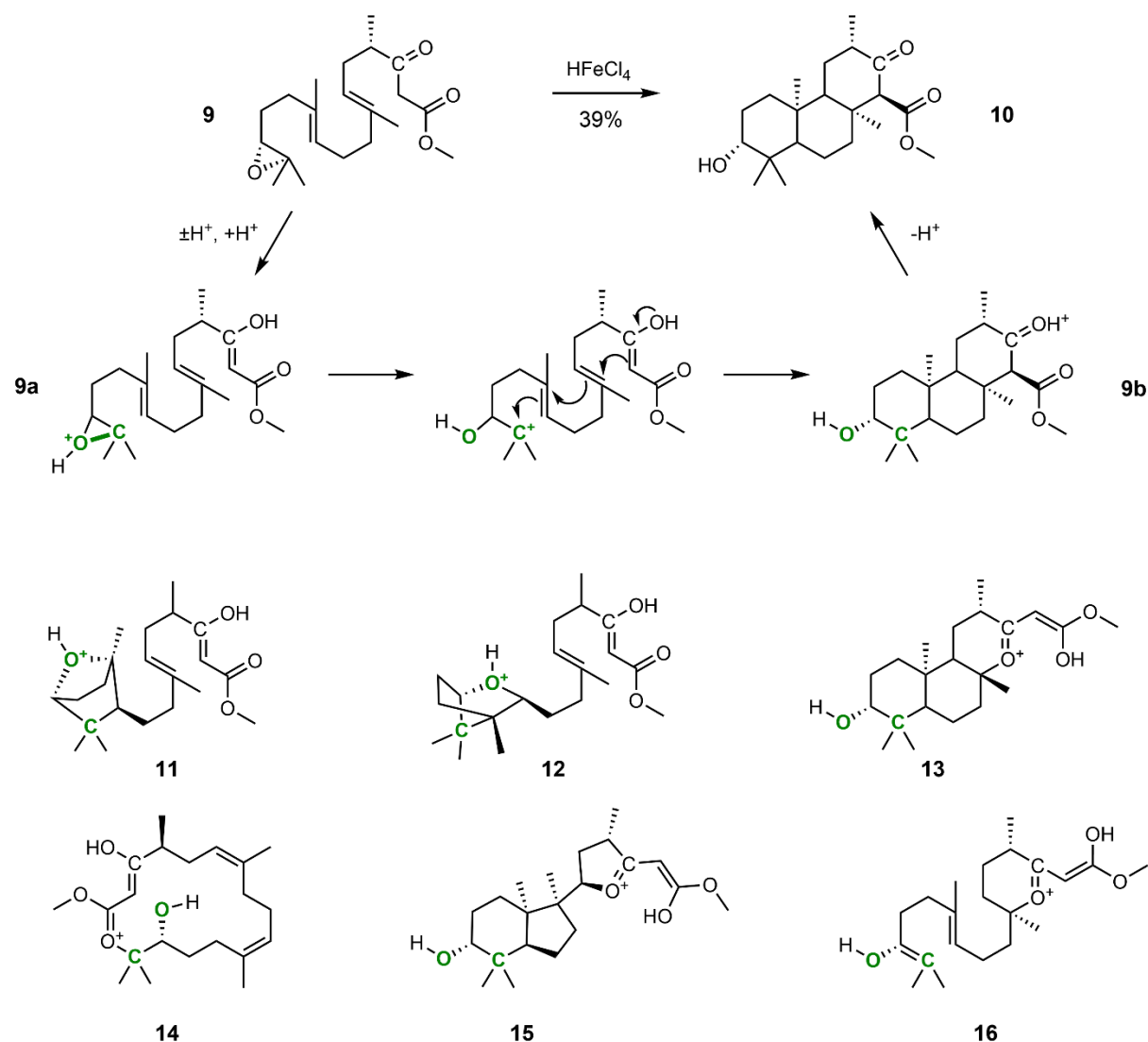


Figure 2 : Reaction mechanism and obtained side-products for an epoxide-initiated β -keto ester-terminated triple cyclization reaction. The polycyclization yielding **10** from **9** is an essential step in the total synthesis of berkeleyone A presented in ref. ⁴³. Here, a reaction search is performed starting from **9a** (details in the text), the protonated enol form of **9**, using the epoxide C-O bond of the tertiary carbon (green) as the reactive coordinate. Compounds **11** to **16** are example of side-products obtained in computations.

This triple cyclization was chosen as it is a complex reaction, with the potential to form many byproducts.^{44–47} Indeed, the 39% yield reported in ref. ⁴³ was only achieved after extensive optimization of reaction conditions. Previous attempts at similar reactions yielded a range of mono- and polycyclic, fused and bridged products from ether formation by the epoxide and cyclization at the β -keto ester oxygen instead of the desired carbon cyclization.⁴⁸ Here, we initiate the computation from **9a**, the reactive enol form of **9** protonated at the epoxide. Compound **9a** is activated by opening the epoxide on the tertiary carbon side using the carbon-oxygen distance as the constraint (highlighted in green in Figure 2). More than 200 distinct reaction pathways are discovered, including many tautomerizations and high energy processes as well as all experimentally reported modes of action.

A subset of thermodynamically favourable products with estimated activation energies below 35 kcal/mol is shown in Figure 2. Notably, **9b** is discovered by the reaction search from an epoxide-initiated triple cyclization cascade terminating at the β -keto ester and forming the desired product **10** by deprotonation. The approximate transition state for this reaction is shown in Figure 3, **a**, with the three bond-forming atoms highlighted. The cyclization cascade is barrier-free after this transition structure. **11** and **12** result from 1,2-addition of the epoxide to the nearest alkene, forming bridged bicyclic ethers.⁴⁴ Triple cyclization terminating at the β -keto ester oxygen forms **13**, a major experimental byproduct,⁴⁸ by a mechanism that differs from that forming **9b** only in the terminating atom (Figure 3, **b**). The nucleophilic opening of the epoxide by the same oxygen yields **14**, a 16-membered ring. Product **15** is obtained when the second step of the triple cyclization cascade is a five-membered, not six-membered, ring closure, as shown in Figure 3, **c**. Finally, the surprising product **16** is formed from the transfer of a proton from the epoxide (which forms an enol⁴⁵) to the far alkene followed by an oxygen cyclization (Figure 3, **c**).

The short duration and easy initialization of our computational experiments compensate for their unsystematic nature. All the above reactions were obtained from a single calculation targeting compound **9a**, which took only 6.3 hours on a single compute node. In this case, we note that success followed a failed search initiated from the less reactive keto form of **9a**. Similarly, computations starting from a tautomer of **9a** with protonation of the ester carbonyl instead of the ketone uncovers significantly fewer reactions and, specifically, fails to yield **9b**. In practice, the problem of sensitivity to the initial molecular identity is solved by iteratively trying new tautomeric structures, an approach enabled by the fast turnaround times of these calculations, and the ease of generating tautomers of a given structure using basic chemical rules.⁴⁹

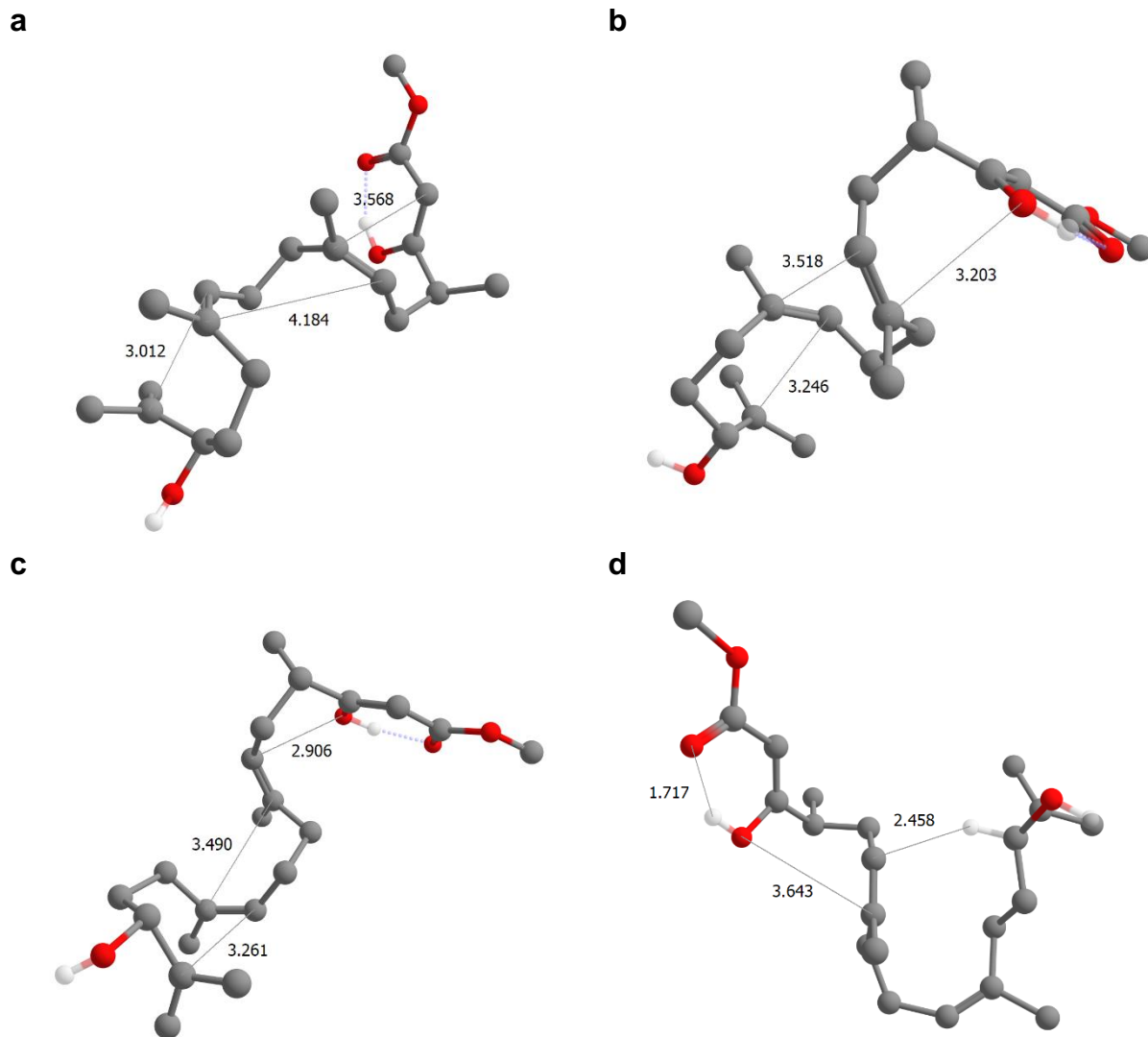


Figure 3: Selected transition structures forming triple cyclization products. a-d, Transition state guesses for the formation of **9b** (a), **13** (b), **15** (c) and **16** (d). Bond-forming atoms are connected by thin lines annotated with their interatomic distances. Non-reacting hydrogen atoms are omitted.

Water-Mediated Michael Addition

Automated reaction search is specifically needed when the reacting system consists of multiple loosely bound fragments with many possible arrangements. This is the case, for example, in solvent-mediated reactions, where specific solvation geometries play a major role in transition structures.⁵⁰ Setting up traditional transition state calculations for such systems is very time-consuming and error-prone, as it requires carefully arranging fragments in different and sometimes non-intuitive reactive geometries.⁵¹

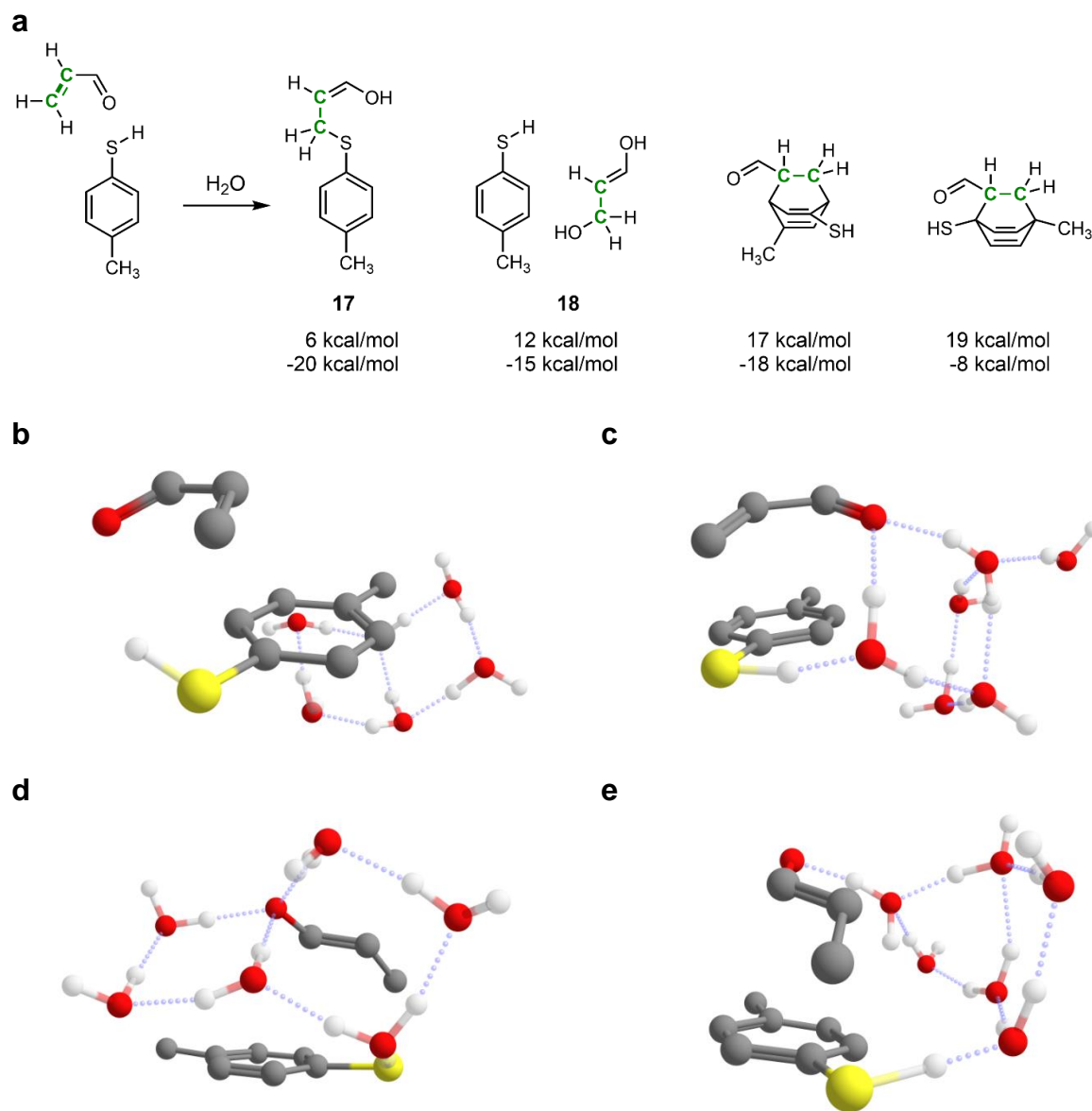


Figure 4: Water-catalyzed 1,4-addition reaction and some important side-products obtained by imposed activation. **a**, Reactions were obtained by stretching the carbon-carbon double bond of an acrolein molecule (shown in green) in the presence of *p*-toluenethiol and six explicit water molecules. Products are annotated with semi-empirically estimated activation energy (top) and reaction energy (bottom). **b-e**, Reaction pathways forming **17** include a direct proton transfer (**b**) and proton transfer shuttled by one (**c**), two (**d**), and four water molecules (**e**). Structures **b-c** are *s-cis* and structures **d-e** are *s-trans*, compact conformations.

Table 1. The 176 reaction trajectories yielding **17**, decomposed based on transition structure characteristics. For each entry, the estimated activation energies in kcal/mol (bold, <12 kcal/mol) and the number of occurrences is provided.

Acrolein isomer	TS conformation	Number of H ₂ O molecules in H ⁺ transfer chain				
		0	1	2	3	4
<i>s-cis</i>	compact	21.2 6	8.7 20	11.2 8	—	—
	expanded	—	15.6 17	13.6 5	17.1 1	—
<i>s-trans</i>	compact	—	27.4 1	5.8 49	8.5 14	20.6 1
	expanded	—	23.3 4	9.8 41	17.1 11	16.6 1

Estimated ΔE^\ddagger (kcal/mol) | no. of occurrences

Here, we demonstrate imposed activation on the water-mediated Michael addition of acrolein to *p*-methylthiophenol, shown in **Error! Reference source not found.**, an example inspired by a recent popular science article.⁵² Six explicit water molecules are added to the reacting system. The addition reaction forming **17** is obtained as both the thermodynamic product and the kinetic product upon stretching the carbon-carbon double bond of the acrolein. Byproducts include hydration of acrolein to 3-hydroxypropanal (**18**), a biologically important process,⁵³ and various Diels-Alder reactions between acrolein and *p*-methylthiophenol.

The Michael addition involves the transfer of a proton from the thiol to the carbonyl group of acrolein, which is facilitated by a chain of hydrogen-bonded water molecules. Acrolein has *s-cis* and *s-trans* isomers, and the transition state for this reaction has the acrolein double bond either lying along the aromatic ring (the compact conformation) or away from it (the expanded conformation). Amongst the possible combinations of isomers, conformers, and number of chained water molecules, our reaction search method identifies, in a single run of the algorithm, fourteen different pathways and distinct transition structures (Table 1), five of which are kinetically favoured over the formation of **18**.

The uncatalyzed reaction only occurs for the *s-cis* isomer, as it requires the carbonyl and thiol to be in proximity (Figure 4b). The larger distance required for the proton transfer in the *trans* isomer favours proton shuttling by two or three water molecules. In contrast, addition of the *cis* isomer is efficiently mediated by one water molecule (Figure 4c). The lowest activation energy is obtained for the *trans* isomer facilitated by two water molecules in the compact conformation (Figure 4d). The longest identified water chain comprises four water molecules (Figure 4e) with a total of five distinct proton transfers. In the expanded conformation, only *trans*-acrolein with a two-water molecule bridge is favourable.

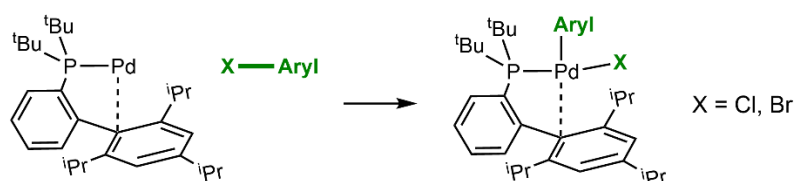
A significant saving in human labour is achieved using the automated search, at the cost of moderate computational resources. Indeed, obtaining sufficient sampling of all these transition structures took only 20 hours of wall-clock time on a mid-range workstation. This calculation required minimal human involvement as the search is entirely automatic,

even starting from a structure far from the identified reactive pathways. By contrast, the human labour needed to set up any one of those 14 transition state searches using traditional means is on the order of hours.

Oxidative Addition Complexes of Drug-Like Compounds

In addition to the exploration of chemical pathways, our reaction search methodology can also minimize the effort of setting up transition state calculations. We demonstrate this application here using calculations on the oxidative addition of a palladium catalyst to a set of drug-like molecules.⁵⁴ Forming oxidative addition complexes in this way provides a potentially powerful means of late-stage diversification in screening studies and avoids the still-inconsistent performance of catalytic Buchwald-Hartwig couplings of highly functionalized substrates.

(a)



(b)

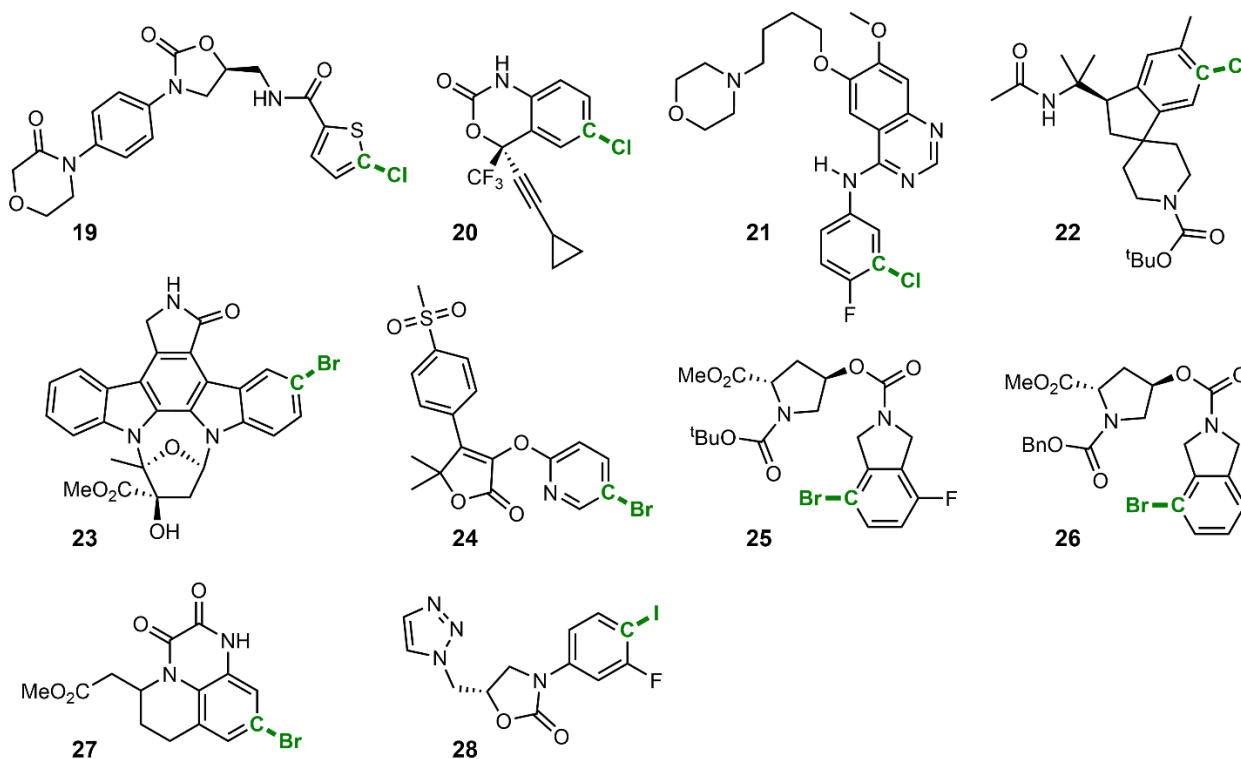


Figure 5: Formation of oxidative addition complexes of drug-like substrates. (a) The reaction studied here is the formation of complexes with ^tBuXPhosPd by oxidative addition. The aryl-halide distance is the activation coordinate (green). (b) Imposed activation reaction search is performed for the ten drug-like, functionalized molecules **19-28**.

Computational studies of organometallic catalysis generally include either complex ligands or complex substrates, but very rarely both, due to the required computational resources. Such studies are specifically challenging when including designer catalysts such as ^tBuXPhosPd (studied here and shown in Figure 5a) and functionalized ligands, due to the many possible coordination modes around the metal catalyst.⁵⁵

Here, we present transition state calculations enabled by our methodology for ten reactions exhibiting both catalyst and substrate complexity. Because the aims of this section differ from those of the previous two (i.e., finding transition states for a predetermined reaction as opposed to searching for new reactions), the calculations were set up slightly differently. In these computations, the activated catalyst was initially placed (programmatically) such that the Pd atom is ~5 angstroms away from the target halide; the halide-aryl bond is the reactive coordinate. Imposed activation reaction search was performed on this structure and the most accessible reaction pathway was selected. Finally, the lowest-energy reactant and product structures were re-optimized at a higher level of theory. Transition states were obtained by relaxing the automatically obtained approximate transition structures using DFT with the aryl-halide bond constrained, followed by a subsequent systematic transition state optimization.

The corresponding results are shown in Table 2. The data show two general trends. First, as expected, activation energies tend to decrease when changing the halogen from chlorine to bromine to iodine.^{56,57} Secondly, electron-rich (hetero)aromatic rings seem to lead to more facile oxidative addition. This is in contrast to the expected trends for bisligated palladium(0) catalysts but in line with the results obtained for monoligated palladium(0), as the pre-reactive complexes of monoligated palladium(0) and aryl halide tend to be more tightly bound for electron-poor (hetero)aromatic rings.^{57,58} Furthermore, oxidative addition products with electron-poor (hetero)aromatic rings tend to be more thermodynamically favourable, as previously demonstrated.^{58,59} Moreover, all the estimated oxidative addition barriers are consistent with facile room temperature reactions, as observed in the original study.⁵⁴ Looking at the contributions of London dispersion to the reaction energies, we find that larger Gibbs free energies of activation correlate with more repulsive dispersion contributions. This indicates that attractive interactions are lost during the oxidative addition, causing an overall higher barrier, and suggests that proper alignment of noncovalent interactions in the transition states is important in accelerating these reactions.^{60,61} Finally, we show in Table 2 that the computing demands of imposed activation reaction search at a semi-empirical level of theory are comparable to those of a DFT optimization of the reactant geometry, and thus well within the realm of current high-throughput screening capabilities.⁶²

Table 2. Summary of reaction searches and DFT calculations for the oxidative additions shown in Figure 5. The first four columns provide the substrate, the total number of atoms in the system, the reacting halide and the reacting aromatic system on the substrate. Gibbs free energies of activation ΔG^\ddagger and Gibbs free energies of reaction ΔG are obtained from PBE-D3/def2-SVP calculations. The last two columns compare the times required (in hours) for the reaction search and DFT optimizations of product, reactant, and transition state geometries.

Substrate	N_{atoms}	Halogen	Core aryl group	Reaction energies (kcal/mol)		Compute time (node hours)	
				ΔG^\ddagger	ΔG	Reaction search	DFT Optimizations
19	123	Cl	Thiophene	13.5	-12.2	10.6	4.5
20	106	Cl	Benzoxazinone	8.9	-16.7	4.8	5.1
21	131	Cl	Benzene	16.0	-6.7	10.3	3.7
22	141	Cl	Indane	4.4	-15.8	10.5	4.5
23	132	Br	Indole	1.1	-18.1	8.7	4.1
24	118	Br	Pyridine	12.2	-12.1	7.6	3.9
25	130	Br	Isoindoline	7.6	-9.6	9.6	4.1
26	131	Br	Isoindoline	8.0	-5.6	10.7	4.6
27	110	Br	Quinoxalinedione	4.7	-16.2	5.2	6.1
28	106	I	Benzene	6.0	-2.1	4.9	4.3

Discussion

Virtual reaction prediction could accelerate the search for novel catalysts, automatically flag detrimental reaction pathways (*e.g.*, solvolysis or formation of side products), and help discover new reaction mechanisms. However, chemical reactions are extremely rare events not spontaneously captured by *in silico* dynamics. The complexity of all but the smallest molecules limits full, systematic searches of all possible bonding patterns and bonding transformations.

We showed that conformer searches with imposed activation can identify chemical pathways and provide approximate transition state structures using modest computational resources. Leveraging some human-derived heuristics, that is, the selection of one chemical bond to undergo reaction, allows us to bypass the quadratic scaling of possible bond-forming processes encountered by more systematic approaches^{4,20,21} while still maintaining sufficient chemical freedom to discover new reaction products and mechanisms. In essence, imposed activation is analogous to starting a curly arrow mechanism by breaking one bond, then computationally completing the mechanism by systematically finding the most energetically feasible responses.

Imposed activation using semi-empirical approaches is sufficiently low-cost to be deployed in an exploratory fashion for medium- to high-throughput virtual screening, predicting and studying the mechanisms behind byproduct formation or catalyst decomposition, and exploring the role of explicit solvation in chemical reactions. Although the semi-empirical calculations used here are appropriate for qualitative discovery, they are not accurate enough for quantitative analysis.⁴² Yet the high degree of inherent parallelism present in our approach makes it scalable on distributed and cloud compute systems, which would compensate for the computational demands required by higher

levels of theory.⁶³ Efforts are ongoing in this direction. Additionally, we are currently investigating the use of machine learning methods (such as Gaussian process regression^{29,64,65}) to systematically and robustly refine the transition state guesses obtained using semi-empirical calculations at higher levels of theory.

Investigations are also underway to improve systematicity by, for example, automating the selection of activation coordinates using chemical and computational means. This kind of automation could enable the study of complex chemical reaction networks^{16,66} by applying imposed activation computations recursively over obtained products.²³ Finally, the inclusion of automated (de)protonation and tautomerization capabilities^{67,68} could increase the range of chemistry that can be studied systematically.

Methods

The results presented in this paper were obtained using the xTB package^{69,70} and a custom interfacing and data analysis code. Additional results for the Buchwald-Hartwig coupling reactions were computed using ORCA,⁷¹ as described below. Timing data is reported for a single dual-socket compute node with two 20-core, 2.4 GHz Intel Skylake processors.

Algorithm

The reaction search consists of the four steps below, followed by post-processing of obtained reaction trajectories.

1. The initial structure of the reactants is optimized, and the user selects a reactive coordinate q (*i.e.*, an interatomic distance, a bending angle, or a torsional angle). With the reactive coordinate fixed at its equilibrium value q_0 , an initial conformer search is performed to obtain a diverse set of structures. A relaxed scan of the reactive coordinate follows, starting with one of the initially generated conformers, to an intermediate value between that of reactants and products q_i .
2. A second, shorter and less exhaustive conformer search is undertaken with the reactive coordinate constrained at q_i to explore the local energy landscape.
3. All the structures thereby generated are scanned from $q = q_i$ to q_N to obtain new reaction products. A backward scan from q_i to q_0 yields a complete trajectory back to a reactant conformation.
4. Finally, structures corresponding to energy minima of the trajectory (potential reactants and products conformations) are optimized without constraints to ensure stability.

Steps 2 to 4 above are then repeated for different values of q_i to obtain an ensemble of reactive trajectories sampled from multiple transition structures. Specific numerical values of all parameters are given in the SI.

For generating conformers, the metadynamics module integrated in xTB is used with parameters based on those of the highly reliable algorithm of Grimme et al.³¹ In effect, molecular dynamics are performed with regular snapshots, and these simulations are augmented with a biasing potential $V_{\text{bias}} = \sum_i k \exp(-\alpha \Delta_i^2)$ where Δ_i is the root-mean-square displacement (RMSD) between the current structure and the i -th previous snapshot, and k and α are numerical parameters. The biasing potential strongly drives the discovery of new structures. The obtained structures are then optimized, with similar structures being removed based on RMSD, rotational constant and energy thresholds.³¹ Finally, structures exceeding a maximum energy threshold are discarded. Numerical values for all parameters are given in the Supplementary Information.

In the examples presented, the overall number of sampled trajectories varies between a few hundreds (such as the iodobutane example) and 10,000 (to obtain the results of Table 1), depending on the number of atoms and the chosen sampling parameters. The duration of metadynamics simulations, and thus the number of trajectories, is chosen proportional to the size of the system. A more thorough sampling is done for the 1,4-addition to sample many reactive geometries.

The trajectories are collected and parsed into reactions by transforming geometries of trajectory minima to canonical SMILES using Open Babel⁷² and identifying changes in the SMILES strings. For structures containing a metal complex, the metal atom is removed before conversion to eliminate spurious bond changes. The transition structure of a trajectory is taken to be the highest energy structure between reactants and products. The guess transition state for a given reaction is the lowest energy transition structure amongst all trajectories for this reaction, and the estimated activation energy is computed by subtracting the energy of the most stable identified reactant conformer from the energy of the approximate transition state.

Refinement of Trajectories for Oxidative Addition Complexes

Additional refinement of reaction trajectories was performed for the results shown in Table 2. The specific protocol described here broadly follows that of the computational study by Barder et al. involving a similar catalyst.⁷³ We note that our study involves more substrates and that those substrates are more complex. DFT calculations were performed at the PBE-D3/def2-SVP level of theory.

Approximate transition structures obtained from the reaction search were used to find transition states with Berny optimization at the semi-empirical level of theory employed in the search (GFN2-xTB). Here, performance is excellent, with every guess transition structure converging to a transition state. At this point, further transition state refinements were attempted using DFT, but GFN2-xTB was found to significantly overestimate the length of the aryl-halide bond in the transition state, hindering convergence and making the transition state poorly transferable to higher-level theory. Hence, a more sophisticated procedure was followed for transition state refinements. Specifically, a first transition state guess was obtained by geometry optimization of the xTB-obtained transition state, constraining the carbon-halogen bond. Unconstrained geometry optimization of the

resulting structures resulted in the respective oxidative addition products in all cases. Subsequently, a relaxed scan of the carbon-halogen bond to 0.8 times the initial bond length starting from the geometry obtained after constrained optimization was carried out. In addition, a partially relaxed scan of the palladium-halogen bond to 1.5 times the initial bond length was performed by constraining the carbon-halogen bond length. Relaxed geometry optimization of either the last point of the first scan or the minimum energy structure of the second scan, whichever was lower in energy, yielded the reactant structure. Next, nudged elastic band (NEB) calculations together with a subsequent transition state optimization (keywords NEB-TS or ZOOM-NEB-TS in Orca) was performed using the reactant as starting structure and the transition state guess from the constrained geometry optimization as product structure. Final optimizations of reactants, products and transition states were performed using exact initial Hessians to remove any residual imaginary frequencies. The transition states were verified by calculating both the forward and backward intrinsic reaction coordinates and comparing the resulting endpoint structures with the respective reactant and product structures.

Acknowledgements

We acknowledge the Defense Advanced Research Projects Agency (DARPA) under the Accelerated Molecular Discovery Program under Cooperative Agreement No. HR00111920027 dated August 1, 2019. The content of the information presented in this work does not necessarily reflect the position or the policy of the Government. G. P. G gratefully acknowledges the Natural Sciences and Engineering Research Council of Canada (NSERC) for the Banting Postdoctoral Fellowship. A. A.-G. thanks Anders G. Frøseth for his generous support. A. A.-G. also acknowledges the generous support of Natural Resources Canada and the Canada 150 Research Chairs program.

Author Contributions

C.L., G.P.G. and R.P. conceived the project based on an idea by C.L. C.L. designed and wrote the custom codes used in reaction discovery and data analysis, with scientific input from G.P.G. and R.P. DFT computations on oxidative addition complexes were performed by R.P. and C.L. C.L., G.P.G., R.P., and A.A.-G.. contributed to the analysis of generated data. All authors contributed to the writing of the manuscript.

Supporting Information

Additional results for varied chemical reactions, as well as values for every numerical parameter used in conformer generation and relaxed scans, are provided in the attached supplemental information.

References

1. Unke, O. T. & Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *J. Chem. Theory Comput.* **15**, 3678–3693 (2019).
2. Maeda, S. & Morokuma, K. Finding Reaction Pathways of Type $A + B \rightarrow X$: Toward Systematic Prediction of Reaction Mechanisms. *J. Chem. Theory Comput.* **7**, 2335–2345 (2011).
3. Sameera, W. M. C., Maeda, S. & Morokuma, K. Computational Catalysis Using the Artificial Force Induced Reaction Method. *Acc. Chem. Res.* **49**, 763–773 (2016).
4. Sameera, W. M. C., Sharma, A. K., Maeda, S. & Morokuma, K. Artificial Force Induced Reaction Method for Systematic Determination of Complex Reaction Mechanisms. *Chem. Rec.* **16**, 2349–2363 (2016).
5. Lee, C. W. *et al.* An Unexpected Ireland–Claisen Rearrangement Cascade During the Synthesis of the Tricyclic Core of Curcusone C: Mechanistic Elucidation by Trial-and-Error and Automatic Artificial Force-Induced Reaction (AFIR) Computations. *J. Am. Chem. Soc.* **141**, 6995–7004 (2019).
6. Wang, L.-P. *et al.* Discovering chemistry with an ab initio nanoreactor. *Nat. Chem.* **6**, 1044–1048 (2014).
7. Jara-Toro, R. A., Pino, G. A., Glowacki, D. R., Shannon, R. J. & Martínez-Núñez, E. Enhancing Automated Reaction Discovery with Boxed Molecular Dynamics in Energy Space. *ChemSystemsChem* **2**, e1900024 (2020).
8. Zimmerman, P. M. Growing string method with interpolation and optimization in internal coordinates: Method and examples. *J. Chem. Phys.* **138**, 184102 (2013).

9. Suleimanov, Y. V. & Green, W. H. Automated Discovery of Elementary Chemical Reaction Steps Using Freezing String and Berny Optimization Methods. *J. Chem. Theory Comput.* **11**, 4248–4259 (2015).
10. Zimmerman, P. M. Single-ended transition state finding with the growing string method. *J. Comput. Chem.* **36**, 601–611 (2015).
11. Schwaller, P., Gaudin, T., Lányi, D., Bekas, C. & Laino, T. “Found in Translation”: predicting outcomes of complex organic chemistry reactions using neural sequence-to-sequence models. *Chem. Sci.* **9**, 6091–6098 (2018).
12. Coley, C. W. *et al.* A graph-convolutional neural network model for the prediction of chemical reactivity. *Chem. Sci.* **10**, 370–377 (2019).
13. Schwaller, P. *et al.* Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction. *ACS Cent. Sci.* **5**, 1572–1583 (2019).
14. Schwaller, P. *et al.* Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy. *Chem. Sci.* **11**, 3316–3325 (2020).
15. Bajczyk, M. D., Dittwald, P., Wołos, A., Szymkuć, S. & Grzybowski, B. A. Discovery and Enumeration of Organic-Chemical and Biomimetic Reaction Cycles within the Network of Chemistry. *Angew. Chem. Int. Ed.* **57**, 2367–2371 (2018).
16. Simm, G. N., Vaucher, A. C. & Reiher, M. Exploration of Reaction Pathways and Chemical Transformation Networks. *J. Phys. Chem. A* **123**, 385–399 (2019).
17. Szymkuć, S. *et al.* Computer-Assisted Synthetic Planning: The End of the Beginning. *Angew. Chem. Int. Ed.* **55**, 5904–5937 (2016).

18. Rappoport, D. & Aspuru-Guzik, A. Predicting Feasible Organic Reaction Pathways Using Heuristically Aided Quantum Chemistry. *J. Chem. Theory Comput.* **15**, 4099–4112 (2019).
19. Maeda, S. & Harabuchi, Y. On Benchmarking of Automated Methods for Performing Exhaustive Reaction Path Search. *J. Chem. Theory Comput.* **15**, 2111–2115 (2019).
20. Fadrná, E. & Koča, J. Single-Coordinate-Driving Method Coupled with Simulated Annealing. An Efficient Tool To Search Conformational Space. *J. Phys. Chem. B* **101**, 7863–7868 (1997).
21. Yang, M., Zou, J., Wang, G. & Li, S. Automatic Reaction Pathway Search via Combined Molecular Dynamics and Coordinate Driving Method. *J. Phys. Chem. A* **121**, 1351–1361 (2017).
22. Koča, J. A mathematical model of the logical structure of chemistry. A bridge between theoretical and experimental chemistry and a general tool for computer-assisted molecular design. *Theor. Chim. Acta* **80**, 29–50 (1991).
23. Rappoport, D., Galvin, C. J., Zubarev, D. Yu. & Aspuru-Guzik, A. Complex Chemical Reaction Networks from Heuristics-Aided Quantum Chemistry. *J. Chem. Theory Comput.* **10**, 897–907 (2014).
24. Grambow, C. A., Pattanaik, L. & Green, W. H. Reactants, products, and transition states of elementary chemical reactions based on quantum chemistry. *Sci. Data* **7**, 137 (2020).
25. 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. *J. Chem. Theory Comput.* **15**, 1652–1671 (2019).
26. Schlegel, H. B. Geometry optimization. *WIREs Comput. Mol. Sci.* **1**, 790–809 (2011).
27. Field-Theodore, T. E. & Taylor, P. R. ALTRUISM: A Higher Calling. *J. Chem. Theory Comput.* (2020) doi:10.1021/acs.jctc.0c00388.
28. Bolhuis, P. G., Chandler, D., Dellago, C. & Geissler, P. L. TRANSITION PATH SAMPLING: Throwing Ropes Over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem.* **53**, 291–318 (2002).
29. Denzel, A. & Kästner, J. Hessian Matrix Update Scheme for Transition State Search Based on Gaussian Process Regression. *J. Chem. Theory Comput.* **16**, 5083–5089 (2020).
30. Ebejer, J.-P., Morris, G. M. & Deane, C. M. Freely Available Conformer Generation Methods: How Good Are They? *J. Chem. Inf. Model.* **52**, 1146–1158 (2012).
31. Grimme, S. Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *J. Chem. Theory Comput.* **15**, 2847–2862 (2019).
32. Hawkins, P. C. D., Skillman, A. G., Warren, G. L., Ellingson, B. A. & Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **50**, 572–584 (2010).
33. Miteva, M. A., Guyon, F. & Tufféry, P. Frog2: Efficient 3D conformation ensemble generator for small compounds. *Nucleic Acids Res.* **38**, W622–W627 (2010).

34. O'Boyle, N. M., Vandermeersch, T., Flynn, C. J., Maguire, A. R. & Hutchison, G. R. Confab - Systematic generation of diverse low-energy conformers. *J. Cheminformatics* **3**, 8 (2011).
35. Vainio, M. J. & Johnson, M. S. Generating Conformer Ensembles Using a Multiobjective Genetic Algorithm. *J. Chem. Inf. Model.* **47**, 2462–2474 (2007).
36. Young, D. C. *Computational chemistry: a practical guide for applying techniques to real world problems.* (2002).
37. Maeda, S. *et al.* Implementation and performance of the artificial force induced reaction method in the GRRM17 program. *J. Comput. Chem.* **39**, 233–251 (2018).
38. Blanksby, S. J. & Ellison, G. B. Bond Dissociation Energies of Organic Molecules. *Acc. Chem. Res.* **36**, 255–263 (2003).
39. Hofmann, A. W. V. & Clark, J. XIV. Researches into the molecular constitution of the organic bases. *Philos. Trans. R. Soc. Lond.* **141**, 357–398 (1851).
40. Saytzeff, A. Zur Kenntniss der Reihenfolge der Analgerung und Ausscheidung der Jodwasserstoffelemente in organischen Verbindungen. *Justus Liebigs Ann. Chem.* **179**, 296–301 (1875).
41. Tojo, G. & Fernández, M. Oxidations by Hydride Transfer from Metallic Alkoxide. in *Oxidation of Alcohols to Aldehydes and Ketones: A Guide to Current Common Practice* 255–279 (Springer US, 2006). doi:10.1007/0-387-25725-X_6.
42. Rasmussen, M. H. & Jensen, J. H. Fast and Automatic Estimation of Transition State Structures Using Tight Binding Quantum Chemical Calculations. *Chem. Sci.* (2020) doi:10.26434/chemrxiv.12600443.v1.

43. Elkin, M., Szewczyk, S. M., Scruse, A. C. & Newhouse, T. R. Total Synthesis of (\pm)-Berkeleyone A. *J. Am. Chem. Soc.* **139**, 1790–1793 (2017).
44. Corey, E. J. & Sodeoka, M. An effective system for epoxide-initiated cation-olefin cyclization. *Tetrahedron Lett.* **32**, 7005–7008 (1991).
45. Van Tamelen, E. E. Bioorganic chemistry: sterols and acrylic terpene terminal epoxides. *Acc. Chem. Res.* **1**, 111–120 (1968).
46. Van Tamelen, E. E. & McCormick, J. P. Terpene terminal epoxides. Mechanistic aspects of conversion to the bicyclic level. *J. Am. Chem. Soc.* **91**, 1847–1848 (1969).
47. Sharpless, K. B. & Van Tamelen, E. E. Terpene terminal epoxides. Skeletal rearrangement accompanying bicyclization of squalene 2,3-oxide. *J. Am. Chem. Soc.* **91**, 1848–1849 (1969).
48. Aggarwal, V. K., Bethel, P. A. & Giles, R. A formal synthesis of (+)-pyripyropene A using a biomimetic epoxy-olefin cyclisation: effect of epoxy alcohol/ether on cyclisation efficiency. *J. Chem. Soc. Perkin 1* 3315–3321 (1999) doi:10.1039/A906589J.
49. Dhaked, D. K., Ihlenfeldt, W.-D., Patel, H., Delannée, V. & Nicklaus, M. C. Toward a Comprehensive Treatment of Tautomerism in Chemoinformatics Including in InChI V2. *J. Chem. Inf. Model.* **60**, 1253–1275 (2020).
50. B. Sunoj, R. & Anand, M. Microsolvated transition state models for improved insight into chemical properties and reaction mechanisms. *Phys. Chem. Chem. Phys.* **14**, 12715–12736 (2012).

51. Lyngvi, E. & Schoenebeck, F. Oxidative addition transition states of Pd(0) complexes in polar solvent—a DFT study involving implicit and explicit solvation. *Tetrahedron* **69**, 5715–5718 (2013).
52. Rzepa, H. Choreographing a chemical ballet: a story of the mechanism of 1,4-Michael addition. | Henry Rzepa's Blog. <https://www.ch.imperial.ac.uk/rzepa/blog/?p=22153> (2020).
53. Engels, C. *et al.* Acrolein contributes strongly to antimicrobial and heterocyclic amine transformation activities of reuterin. *Sci. Rep.* **6**, (2016).
54. Uehling, M. R., King, R. P., Krska, S. W., Cernak, T. & Buchwald, S. L. Pharmaceutical diversification via palladium oxidative addition complexes. *Science* **363**, 405–408 (2019).
55. Melvin, P. R., Nova, A., Balcells, D., Hazari, N. & Tilset, M. DFT Investigation of Suzuki–Miyaura Reactions with Aryl Sulfamates Using a Dialkylbiarylphosphine-Ligated Palladium Catalyst. *Organometallics* **36**, 3664–3675 (2017).
56. Senn, H. M. & Ziegler, T. Oxidative Addition of Aryl Halides to Palladium(0) Complexes: A Density-Functional Study Including Solvation. *Organometallics* **23**, 2980–2988 (2004).
57. Anjali, B. A. & Suresh, C. H. Interpreting Oxidative Addition of Ph–X (X = CH₃, F, Cl, and Br) to Monoligated Pd(0) Catalysts Using Molecular Electrostatic Potential. *ACS Omega* **2**, 4196–4206 (2017).
58. Ahlquist, M. & Norrby, P.-O. Oxidative Addition of Aryl Chlorides to Monoligated Palladium(0): A DFT-SCRF Study. *Organometallics* **26**, 550–553 (2007).

59. Kégl, T. R., Kollár, L. & Kégl, T. DFT Study on the Oxidative Addition of 4-Substituted Iodobenzenes on Pd(0)-Phosphine Complexes. *Adv. Phys. Chem.* **2015**, 985268 (2015).
60. Wagner, J. P. & Schreiner, P. R. London Dispersion in Molecular Chemistry—Reconsidering Steric Effects. *Angew. Chem. Int. Ed.* **54**, 12274–12296 (2015).
61. Lyngvi, E., Sanhueza, I. A. & Schoenebeck, F. Dispersion Makes the Difference: Bisligated Transition States Found for the Oxidative Addition of Pd(PtBu₃)₂ to Ar-OSO₂R and Dispersion-Controlled Chemoselectivity in Reactions with Pd[P(iPr)(tBu)₂]₂. *Organometallics* **34**, 805–812 (2015).
62. Hachmann, J. *et al.* Lead candidates for high-performance organic photovoltaics from high-throughput quantum chemistry – the Harvard Clean Energy Project. *Energy Environ. Sci.* **7**, 698–704 (2014).
63. Grambow, C. A. *et al.* Unimolecular Reaction Pathways of a γ -Ketohydroperoxide from Combined Application of Automated Reaction Discovery Methods. *J. Am. Chem. Soc.* **140**, 1035–1048 (2018).
64. Denzel, A. & Kästner, J. Gaussian Process Regression for Transition State Search. *J. Chem. Theory Comput.* **14**, 5777–5786 (2018).
65. Koistinen, O.-P., Ásgeirsson, V., Vehtari, A. & Jónsson, H. Nudged Elastic Band Calculations Accelerated with Gaussian Process Regression Based on Inverse Interatomic Distances. *J. Chem. Theory Comput.* **15**, 6738–6751 (2019).
66. Proppe, J. & Reiher, M. Mechanism Deduction from Noisy Chemical Reaction Networks. *J. Chem. Theory Comput.* **15**, 357–370 (2019).

67. Pracht, P., Bauer, C. A. & Grimme, S. Automated and efficient quantum chemical determination and energetic ranking of molecular protonation sites. *J. Comput. Chem.* **38**, 2618–2631 (2017).
68. Grimme, S. A. & Reiher, M. The electrostatic potential as a descriptor for the protonation propensity in automated exploration of reaction mechanisms. *Faraday Discuss.* **220**, 443–463 (2019).
69. *xTB: Semiempirical Extended Tight-Binding Program Package*. (Lab of Stefan Grimme, 2020).
70. Bannwarth, C. *et al.* Extended tight-binding quantum chemistry methods. *WIREs Comput. Mol. Sci.* e01493 (2020) doi:10.1002/wcms.1493.
71. Neese, F. *ORCA (v. 6.2.1)*.
72. *Open Babel: The Open Source Chemistry Toolbox (v. 3.1.1)*.
73. Barder, T. E., Biscoe, M. R. & Buchwald, S. L. Structural Insights into Active Catalyst Structures and Oxidative Addition to (Biaryl)phosphine–Palladium Complexes via Density Functional Theory and Experimental Studies. *Organometallics* **26**, 2183–2192 (2007).

Supplemental information for automatic discovery of chemical reactions using imposed activation

Cyrille Lavigne,¹ Gabriel dos Passos Gomes,^{§1,2} Robert Pollice,^{§1,2} Alán Aspuru-Guzik^{*1,2,3,4}

¹ Department of Computer Science, University of Toronto, 214 College St., Toronto, Ontario M5T 3A1, Canada.

² Chemical Physics Theory Group, Department of Chemistry, University of Toronto, 80 St George St, Toronto, Ontario M5S 3H6, Canada.

³ Vector Institute for Artificial Intelligence, 661 University Ave Suite 710, Toronto, Ontario M5G 1M1, Canada.

⁴ Lebovic Fellow, Canadian Institute for Advanced Research (CIFAR), 661 University Ave, Toronto, Ontario M5G

[§] These authors contributed equally to this work.

** Corresponding author: aspuru@utoronto.ca*

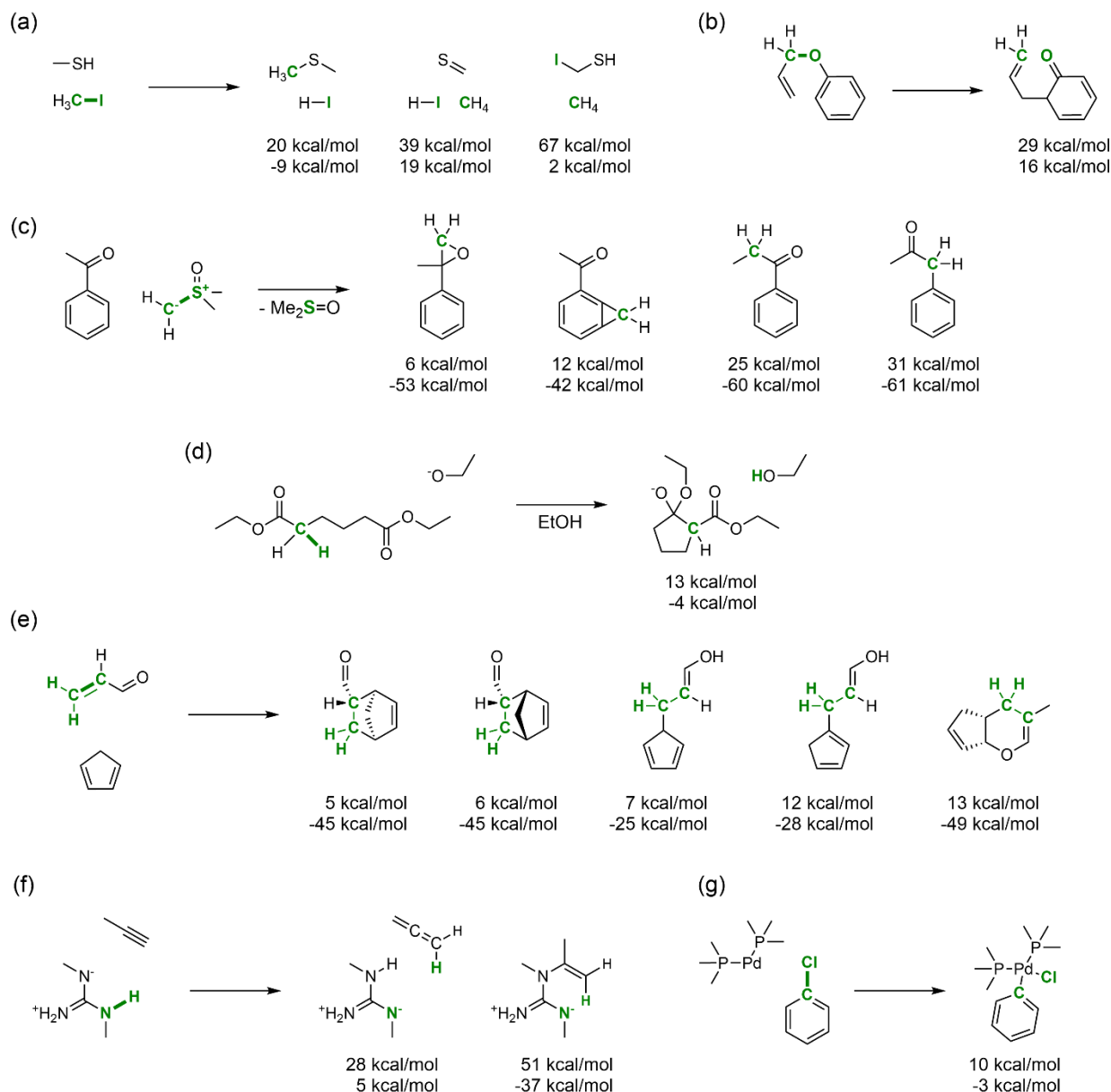


Figure S6: Selected set of diverse chemical reactions obtained by imposed activation. Representative products are shown in each case, annotated with estimated activation energy (top) and reaction energy (bottom) calculated at the semi-empirical GFN2-xTB level.¹ All results are obtained directly from the automated search without additional refinement. Activated bonds and constituent atoms are shown in green. In (e), an improper dihedral angle is activated instead of a bond length. Details of each reaction are in the text.

Additional set of diverse chemical reactions

Initial tests were performed using the small but chemically diverse set of simple reactions, the results of which we describe here. All the chemical reactions shown in Figure S6 are adequately captured with minimal reaction-specific parameter choices (Table S4). The relatively small size of the molecules involved allows this entire set of reactions to be run on a single compute node in less than 5 hours.

The S_N2 reaction of methanethiol and iodomethane (Figure S6a) is readily obtained by stretching the carbon-iodine bond. Some high-energy side products include thioformaldehyde formation (by transfer hydrogenolysis) and iodomethanethiol (by a stepwise σ -bond metathesis). The aromatic Claisen² rearrangement (Figure S6b) is found as the sole product of stretching the allyl ether bond. Energies of the two expected transition structures, six-membered rings with chair and boat conformations, respectively, are well reproduced.³

In Figure S6c, the Johnson-Corey-Chaikovsky reaction⁴ is obtained from stretching the methylene-sulfur bond on the ylide. The expected epoxide has by far the lowest estimated activation energy. Interesting side products include addition to the aromatic ring as well as homologations by methylene insertion into a C–H bond and by insertion into the aryl-carbon bond. This last reaction has been reported previously for benzophenones⁵, with a hypothesized mechanism matching the one found here. The stretching of the sulfur-methylide (C₁ in Figure S7a) bond yields the reactive intermediate shown in Figure S7b. Insertion of C₁ between carbons marked C₂ and C₃ occurs through a four-membered cyclic transition state, displacing the sulfoxide and forming the homologation product (Figure S7c).

The Dieckmann condensation⁶ in Figure S6d is found readily upon dissociation of a carbon-hydrogen bond in the presence of an ethoxide molecule. The reactants (an ester, an ethoxide molecule and two molecules of ethanol) are initially arranged without regard to the expected transition structure. The assembly is optimized to obtain the initial conformer in Figure S7d. The explicit solvent is required here to stabilize the ethoxide. The reaction search finds the Dieckmann condensation product in Figure S7e, with the standard⁷ two-step mechanism of formation of an enolate ion followed by a nucleophilic attack. Approximate transition structures are shown in Figure S7f-g. It should be noted that the initial input structure is highly dissimilar to both the structures of the transition states and the products, illustrating that our method allows handling explicit solvent molecules without prior spatial placement by the user.

For the Diels-Alder reaction, activation can be imposed in two ways: by stretching the dienophile double bond to the length of a single bond or by pyramidalizing one of the dienophile carbons through an improper dihedral constraint applied to the two ene carbons and terminal hydrogen atoms. Both approaches yield the same products, with the pyramidalization resulting in significantly better transition structures and energies due to the asymmetric dienophile and resulting asynchronous transition state⁸. The results for this reaction are shown in Figure S6e with the atoms of the dihedral constraint marked in green. Notably, both the endo- and the exo-products are obtained, though the activation energy is approximated too poorly to reproduce experimental selectivities. Observed side reactions include Michael additions as well as an inverse electron-demand Diels-Alder reaction, with the acrolein acting as the diene and the cyclopentadiene acting as the dienophile.

The final two examples are taken from catalysis and organometallic chemistry, respectively: a base-catalysed alkyne-allene isomerization and an oxidative addition of palladium bisphosphine into a carbon-chlorine bond (Figure S6f and g). In the first case, the N-H bond on the catalyst is stretched to obtain the isomerization, and the reaction pathway and transition structure are near identical to those obtained by traditional means.⁹ The oxidative addition of Pd in Figure S6g is readily found by increasing the Ph-Cl bond distance, showing that imposed activation can be applied to organometallic reactions using extended density-functional tight-binding theory.¹⁰

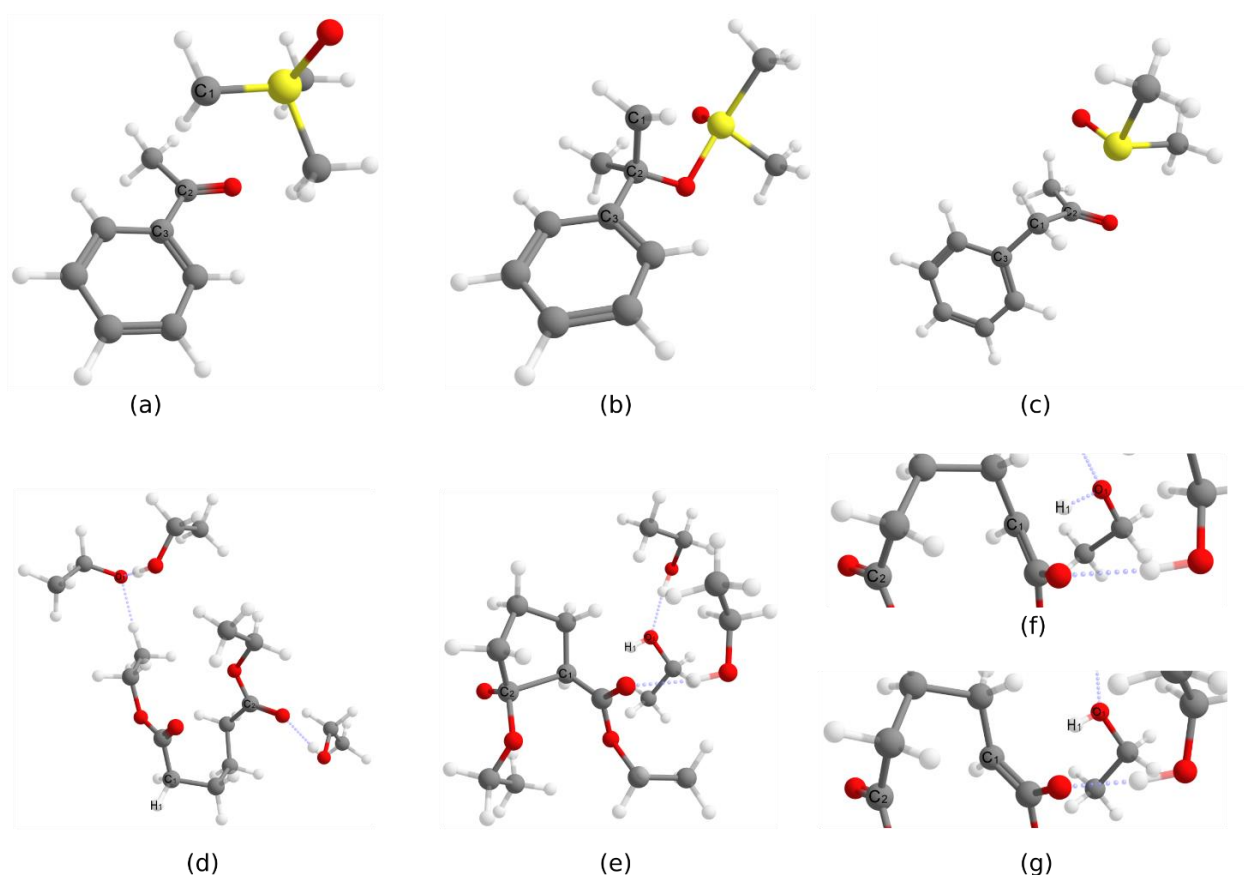


Figure S7: Detailed structures for Johnson-Corey-Chaikovsky methylene insertion and Dieckmann condensation reactions. (a-c) Reactants (a), reactive intermediate (b), and product structures (c) obtained for the homologation reaction by methylene insertion in Figure S6c. (d-g) Input structure (d) and products (e) for the Dieckmann condensation in Figure S6d and transition structures of the two steps, formation of an enolate (f) and nucleophilic attack (g), as obtained from a single relaxed scan.

Algorithmic details and numerical parameters.

Table S3 lists default numerical parameters used for the computations described in this article. Unless otherwise noted, these values are used throughout. Reaction-specific parameters (specifically, those that differ from defaults) are shown in Table S4. Here, we described in more details the implemented algorithm.

All runs are initialized by performing a metadynamics propagation (MTD1 in Table S3) filtering all the resulting structures and selecting the bottom 10% on the basis of energy. This is done with the activation coordinate q^\ddagger constrained to its initial value.

Relaxed scans are performed from the initial value of q_i^\ddagger to its final value q_f^\ddagger over a uniformly discretized grid of values. Metadynamics searches for activated conformers (MTD2) are performed for points $j = M_L$ to M_U of the scans with q^\ddagger constrained to its value at that point, denoted q_j^\ddagger . Each metadynamics search is initialized from one of the structures obtained in the initialization.

Structures obtained from metadynamics propagation at point j are optimized with q^\ddagger constrained to q_j^\ddagger and screened for duplicates. Those structures with energy below two thresholds, a local threshold $E < \min_j E + \Delta_l$ (where the minimum is over structures from the same metadynamics propagation) and a global threshold $E < E_0 + \Delta_g$ (where E_0 is the energy of the reactant conformer) are selected as starting points for reaction scans. The reaction scans are relaxed scans from q_j^\ddagger to q_f^\ddagger and to q_i^\ddagger , yielding new product and reactant structures.

Finally, all reaction scans are searched for potential stable structures, which are local energy minima on the q^\ddagger axis. Those stable points are optimized without any constraining potentials. The overall scan is then analyzed for reactions, as described in the methods section.

Table S3 Default parameters used in computations.

xtb parameters	Parametrization	GFN2
	Electronic temperature	300K
	Cavity potential ¹	Logfermi, spherical with diameter $d_{\text{reactants}} + 4\text{\AA}$
Metadynamics propagation (MTD1)	Propagation time	0.5 ps per atom
	Bias potential ²	$k = 0.20 E_H$, $\alpha = 0.2/\text{\AA}^2$ Structure saved every 100 fs Biasing by previous 10 saved structures
	Timestep ³	5 fs
	SHAKE constraints ³	All atoms
	Structures kept	Bottom 10% in energy after filtering
Relaxed scans	Number of scan points between q_i^\ddagger and q_f^\ddagger	$N_S = 50$
	Constraint force constant ⁴	$k = \text{computed or } 1 \text{ kcal/mol}$
	Optimization tolerance	“normal” $\Delta E = 5 \times 10^{-6} E_H$, max grad = $10^{-3} E_H/\alpha$
Metadynamics propagation (MTD2)	Propagation time	0.1 ps per atom
	Bias potential ⁵	$k = 0.20 E_H$, $\alpha = 0.8/\text{\AA}^2$ $k = 0.20 E_H$, $\alpha = 0.2/\text{\AA}^2$ $k = 0.05 E_H$, $\alpha = 0.8/\text{\AA}^2$ $k = 0.05 E_H$, $\alpha = 0.2/\text{\AA}^2$ Structure saved every 100 fs Biasing by previous 10 saved structures
	Timestep	2 fs
	SHAKE constraints	None
	Performed at...	... every point between $M_L = 0$ and $M_U = 0.5 \times N_S$
Conformer filtering ⁶	Optimization tolerance	“tight” $\Delta E = 10^{-6} E_H$, max grad = $8 \times 10^{-4} E_H/\alpha$
	Similarity thresholds	RMSD = 0.4 Å Energy = 1.0 kcal/mol
	Local threshold Δ_l	12 kcal / mol
	Global threshold Δ_g	60 kcal/mol above reactants

¹ For metadynamics propagations, a spherical constraining potential¹¹ is added to stop molecules in the reacting system from separating. $d_{\text{reactants}}$ is the maximum atom pair distance of the reactant structure.

² The metadynamics potential has the following form $V_{\text{bias}} = \sum_i k \exp(-\alpha \Delta_i^2)$ where the sum is over contributing structures and Δ_i is the RMSD between contributing structure i and the current structure.

³ If the initial metadynamics propagation fails to converge, it is restarted with a timestep of 2 fs and no SHAKE constraints.

⁴ For bond stretches, the force constant is computed from an estimated bond force constant, computed from a quadratic fit of the energy from a five-point relaxed scan around the equilibrium bond length.

⁵ Four metadynamics propagations are done,¹¹ one for each pair of values k, α .

⁶ Structures from all metadynamics calculations are first optimized and then duplicated structures are removed based on similarity, followed by elimination using energy thresholds.

Table S4 Parameters used for specific reactions described in this article.

	Activation coordinate scans	Parameters changed from Table S1
Reactions of 2-iodobutane	$q_0^\ddagger \rightarrow 6.0 \text{ \AA}$	GBSA solvent model for THF $N_S = 100$
Cyclization cascade	$1.5 \text{ \AA} \rightarrow 3.0 \text{ \AA}$	$N_S = 100$ $M_U = 0.3 \times N_S$
Acrolein water-mediated 1,4-addition	$q_0^\ddagger \rightarrow 1.6 \text{ \AA}$	$N_S = 100$ $M_U = 0.3 \times N_S$ MTD2 time per atom = 0.5 ps
Oxidative additions of drug candidates	$q_0^\ddagger \rightarrow 3.5 \text{ \AA}$	$N_S = 200$ $M_L = 0.06 \times N_S, M_U = 0.12 \times N_S$
Reactions in Figure S1	(a) $q_0^\ddagger \rightarrow 6.0 \text{ \AA}$	GBSA solvent model for methanol $M_L = 0.30 \times N_S, M_U = 0.80 \times N_S$
	(b) $q_0^\ddagger \rightarrow 4.0 \text{ \AA}$	
	(c) $q_0^\ddagger \rightarrow 4.0 \text{ \AA}$	
	(d) $q_0^\ddagger \rightarrow 3.5 \text{ \AA}$	
	(e) $q_0^\ddagger \rightarrow 120^\circ$	
	(f) $q_0^\ddagger \rightarrow 3.0 \text{ \AA}$	
	(g) $q_0^\ddagger \rightarrow 3.5 \text{ \AA}$	

q_0^\ddagger denotes the equilibrium value of the activation coordinate.

References

1. 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. *J. Chem. Theory Comput.* **15**, 1652–1671 (2019).
2. Claisen, L. Über Umlagerung von Phenol-allyläthern in C-Allyl-phenole. *Berichte Dtsch. Chem. Ges.* **45**, 3157–3166 (1912).
3. Yamabe, S., Okumoto, S. & Hayashi, T. Transition Structures for the Aromatic Claisen Rearrangements by the Molecular Orbital Method. *J. Org. Chem.* **61**, 6218–6226 (1996).
4. Corey, E. J. & Chaykovsky, M. Dimethyloxosulfonium Methylide ((CH₃)₂SOCH₂) and Dimethylsulfonium Methylide ((CH₃)₂SCH₂). Formation and Application to Organic Synthesis. *J. Am. Chem. Soc.* **87**, 1353–1364 (1965).
5. Chittimalla, S. K., Chang, T.-C., Liu, T.-C., Hsieh, H.-P. & Liao, C.-C. Reactions of 2-hydroxybenzophenones with Corey–Chaykovsky reagent. *Tetrahedron* **64**, 2586–2595 (2008).
6. Dieckmann, W. Zur Kenntniss der Ringbildung aus Kohlenstoffketten. *Berichte Dtsch. Chem. Ges.* **27**, 102–103 (1894).
7. Schaefer, J. P. & Bloomfield, J. J. The Dieckmann Condensation (Including the Thorpe-Ziegler Condensation). in *Organic Reactions* 1–203 (John Wiley & Sons, 2011). doi:10.1002/0471264180.or015.01.
8. Loncharich, R. J., Brown, F. K. & Houk, K. N. Transition structures of the Diels-Alder reaction of butadiene with acrolein. *J. Org. Chem.* **54**, 1129–1134 (1989).

9. dos Passos Gomes, G., Morrison, A. E., Dudley, G. B. & Alabugin, I. V. Optimizing Amine-Mediated Alkyne–Allene Isomerization to Improve Benzannulation Cascades: Synergy between Theory and Experiments. *Eur. J. Org. Chem.* **2019**, 512–518 (2019).
10. Bursch, M., Neugebauer, H. & Grimme, S. Structure Optimisation of Large Transition-Metal Complexes with Extended Tight-Binding Methods. *Angew. Chem.* **131**, 11195–11204 (2019).
11. Grimme, S. Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations. *J. Chem. Theory Comput.* **15**, 2847–2862 (2019).