Predicting Feasible Organic Reaction Pathways Using Heuristically Aided Quantum Chemistry

Dmitrij Rappoport* and Alán Aspuru-Guzik*

Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA

E-mail: dmitrij@rappoport.org; aspuru@chemistry.harvard.edu

Abstract

Studying organic reaction mechanisms using quantum chemical methods requires from the researcher an extensive knowledge of both organic chemistry and first-principles computation. The need for empirical knowledge arises because any reasonably complete exploration of the potential energy surfaces (PES) of organic reactions is computationally prohibitive. We have previously introduced the Heuristically-Aided Quantum Chemistry (HAQC) approach to modeling complex chemical reactions, which abstracts the empirical knowledge in terms of chemical heuristics—simple rules guiding the PES exploration—and combines them with structure optimizations using quantum chemical methods. The HAQC approach makes use of heuristic kinetic criteria for selecting reaction paths that are not only plausible, that is, consistent with the empirical rules of organic reactivity, but also feasible under the reaction conditions. In this work, we develop heuristic kinetic feasilibity criteria, which correctly predict feasible reaction pathways for a wide range of simple polar (substitutions, additions, and eliminations) and pericyclic organic reactions (cyclizations, sigmatropic shifts, and cycloadditions). In contrast to knowledge-based reaction mechanism prediction methods, the same kinetic heuristics are successful in classifying reaction pathways as feasible or infeasible

1

across this diverse set of reaction mechanisms. We discuss the energy profiles of HAQC and their potential applications in machine learning of chemical reactivity.

1 Introduction

The understanding of the existing and the design of new reaction pathways in organic chemistry build on a large body of empirical knowledge in organic reaction mechanisms accumulated over the course of more than a century. The success of physical organic chemistry and, more recently, quantum chemical computation created a theoretical foundation of organic reactivity and also provided practical guidance for developing synthetic routes.[1–4] In parallel to these advances, the last decade has seen a shift in focus to complex biological and chemical systems.[5–8] Complex chemical reactions, which consist of networks of concurrent and interlinked reaction paths, are increasingly recognized as the rule rather than the exception in chemistry.[5,9–12] These systems present considerable challenges for quantum chemical modeling due to the involved topology of the reactive potential energy surface (PES) with numerous energy minima connected by transition states.

In our previous work, we have developed the Heuristically Aided Quantum Chemistry (HAQC) approach to modeling complex chemical reactions,[13] which leverages the existing empirical rules of polar organic reactions as powerful heuristics for selecting *plausible* reactive transformations, followed by structure optimizations using quantum chemical methods. By means of the HAQC methodology, we were able to reproduce the key products and reaction paths of the formose reaction and to explore the structure and properties of the corresponding reaction network.[13] However, further improvements are necessary to the HAQC methodology in selecting reaction paths, which are not only plausible, that is, consistent with the empirical rules of organic reactivity, but are also thermodynamically and kinetically *feasible* under the reaction conditions, broadly defined. While the reaction thermodynamics can be directly estimated from the molecule energies computed using quantum chemical meth-

ods, the formal derivation of kinetic parameters requires the knowledge of transition state energies, which is computationally expensive or impractical in complex chemical reactions. While methodological improvements in transition-state computations have greatly increased their applicability to large systems,[14–18] they will likely remain a computational bottleneck in studying of complex chemical reactions for some time. In this work we explore an alternative approach to boosting the predictive power of the HAQC procedure, which takes advantage of heuristic kinetic feasibility criteria. We define formal requirements that these heuristic criteria must broadly fulfill and develop several simple proposals, the performance of which we assess using a set of well-characterized organic reactions.

Predictions of reaction pathways encompass two related but distinct tasks: finding one or all feasible pathways leading from a given reactant to a given product (*single-source, single-product* reaction prediction), and identifying all products, for which at least one feasible pathway exists from a given reactant (*single-source, multi-product* reaction prediction). In both cases, the results are dependent on the reaction conditions such as solvent, temperature, presence of catalysts, etc. All these factors also determine the relative yields of the feasible reaction products. For the purposes of this work, we wish to abstract from the specific reaction conditions and consider a generalized feasibility of reaction pathways in *typical* reaction conditions, while admitting a certain degree of fuzziness under this definition. Furthermore, we distinguish only between feasible and infeasible pathways, reducing the task at hand to a simple binary classification problem. As we show in this work, this type of binary classification is achievable and is in line with the accumulated empirical knowledge of organic reaction mechanisms. Moreover, we are able to develop a consensus heuristic kinetic criterion, which applies across broad class of polar organic reactions within the HAQC approach. This result is highly encouraging and raises the prospect for a robust and inexpensive general predictor of organic reactivity, complementing *ab initio* transition-state computations. Potential applications of the HAQC methodology to organic synthesis and complex chemical reactions, such as the ones arising in the context of origins of life, are the primary motivator for this

and future work.

This paper is structured as follows. We review the HAQC approach and introduce the relevant definitions in Section 2 and briefly describe our implementation of the HAQC scheme in Section 3. Section 4 contains a study of heuristic kinetic criteria for feasibility classification in the single- source, single- product problem for a series of well-characterized organic reactions including aliphatic additions, substitutions, and eliminations, aromatic substitution, and pericyclic reactions. The results are discussed in Section 5, and the conclusions of this work in Section 6.

2   Chemical Heuristics

The network representation affords a radically simplified but powerful discrete model of a complex chemical reaction that gives insight into its structure, properties, and dynamics in an efficient way. In this and the preceding work, we employ the transition network (TN) representation, which may be thought of as a discretized version of the reactive PES. The network nodes in the TN representation describe *flasks*—closed collections of molecules, whose stoichiometry is kept constant throughout the reaction network—while the directed network edges are stoichiometry-preserving reactive transformations. A series of directed network edges connecting a pair of (not necessarily adjacent) flasks comprises a reaction pathway or chemical mechanism. In the HAQC approach, each transformation implements one of the predefined transformation heuristics, which are derived from empirical rules chemists use to write reasonable mechanisms for reactions. This set of transformation heuristics, together with the HAQC approach, defines the scope of the chemical reaction network, *i.e.*, the kinds of chemical reactions that the TN model is capable of describing, as well as the granularity of the description. Organic synthesis planning has a long tradition of using empirical rules to encode known synthetic reactions in the manner of an expert system, beginning with the seminal work on the LHASA system by Corey and Wipke,[19] and including works by Ugi

and Dugunji,[20,21] Jorgensen,[22] and many others, most recently by Grzybowski and co workers.[12,23] For the purposes of this work, we employ an alternative approach and utilize the set of rules describing heteropolar breaking and formation of single, double, and triple bonds as elementary transformation heuristics. (Table 1) Our motivation for using this much simpler rule set this approach is to broaden the scope of the chemical reactivity we can describe and to reduce knowledge bias. At the same time, we rely on energy-based heuristic criteria for choosing thermodynamically and kinetically feasible transformations.

The set of elementary transformations is chosen to be reversible, that is, each transformation appears both in the forward and the backward direction, and stoichiometry-preserving. While the choice of bond dissociations and associations as elementary transformation steps is hardly surprising, the polarization and depolarization of multiple bonds deserve an additional comment. By writing double and triple bonds in their polarized forms, we do not mean to imply that these charge-separated structures are distinct intermediates in the reaction mechanisms or distinct energy minima on the corresponding reactive PES. Depending on the molecular structure and the reaction mechanism, this might be or might not be the case. Regardless, we consider (de)polarization steps as purely fictitious constructs, which are formally energy-neutral and provide us with a convenient notation for expressing the reactivity of double and triple bonds. Finally, we point out that all polar reactions involving the elements C, H, O with normal polarity as well as halogens in their oxidation states of 0 and -1 can be derived from these elementary transformation heuristics by composition. The extension to, $e.\,g.$, reactions of nitrogen containing compounds, reactions with inverted polarity, or different types of bonds, such as O–O, is straightforward but is not pursued in this paper.

While the elementary transformation heuristics effectively narrow down all conceivable reaction mechanisms to the subset of those consistent with the pre-defined rules of chemical reactivity (plausible reaction paths), they do not rule out high-energy intermediates, $e.g.$, strained small rings, or high-barrier reaction paths, $e.g.$, breaking of unactivated C–H and C–

Table 1: Elementary transformation steps employed in this work (X = F, Cl, Br, I).

Bond dissociation	Bond association
C−H → C⁻ ⋯ H⁺	C⁻ ⋯ H⁺ → C−H
C−C → C⁺ ⋯ C⁻	C⁺ ⋯ C⁻ → C−C
C−O → C⁺ ⋯ O⁻	C⁺ ⋯ O⁻ → C−O
C−X → C⁺ ⋯ X⁻	C⁺ ⋯ X⁻ → C−X
C−X → C⁻ ⋯ X⁺	C⁻ ⋯ X⁺ → C−X
O−H → O⁻ ⋯ H⁺	O⁻ ⋯ H⁺ → O−H
X−H → X⁻ ⋯ H⁺	X⁻ ⋯ H⁺ → X−H
X−X → X⁻ ⋯ X⁺	X⁻ ⋯ X⁺ → X−X
Polarization	**Depolarization**
C=O → C⁺−O⁻	C⁺−O⁻ → C=O
C=C → C⁺−C⁻	C⁺−C⁻ → C=C
C≡C → C⁺=C⁻	C⁺=C⁻ → C≡C

C bonds. This is the price we have to pay for the broader scope of organic reactions covered by the elementary transformation heuristics. As a consequence, it is necessary to further select for reaction paths that are not only plausible but also feasible under the reaction conditions, as judged by thermodynamic and kinetic criteria.

The thermodynamic and kinetic criteria developed in the previous work showed promise in minimizing the number of infeasible reaction paths.[13] The thermodynamic criterion can be taken to be the energy difference $\Delta E_{K \to L} = E_L - E_K$ between the initial flask K and the final flask L and is thus straightforward to obtain from the results of quantum chemical structure optimizations. By contrast, the kinetic criterion must be approximated by heuristic expressions. One of the two heuristic kinetic criteria considered previously is *climb* $W_{c,K \to L}$, which yields the energy of the highest point in the energy profile of reaction path relative to the initial flask. This criterion follows from the idea of Hammond's postulate that allows us to take the energy of the highest-energy intermediate flask to approximate the activation barrier of the rate-limiting step.[1,24] The second kinetic criterion is *arc* $W_{a,K \to L}$, which modifies the climb criterion as to penalize longer reaction paths by adding a constant penalty factor α for each step. In both cases, higher values of the heuristic kinetic criteria indicate a less favorable

reaction path. The feasible reaction paths are thus determined by defining thermodynamic and kinetic cutoff criteria values such that all reaction paths with $\Delta E_{K \to L} \leq \Delta E_{\max}$ and $W_{K \to L} \leq W_{\max}$ are considered as feasible.[13]

The focus of the present work is on generalizing this successful strategy to other organic reactions and on refining our capability for discriminating between feasible and infeasible reaction paths. To aid the criterion selection, we first define a set of expectations that a useful heuristic kinetic criterion should fulfill in keeping with the qualitative picture of transition state theory, Hammond's postulate, and computational considerations. A heuristic kinetic criterion $W_{K \to L}$ for the N-step reaction path $K \to L = \{K = K_0, K_1, \ldots, K_N = L\}$ thus should possess the following properties:

1. *Non-Negativity.* $W_{K \to L} \geq 0$.

2. *Additivity.* $W_{K \to L \to M} = W_{K \to L} + W_{L \to M}$ for consecutive reaction paths $K \to L$ and $L \to M$.

3. *Hammond's Postulate.* $W_{K \to L}$ should approximate the activation energy of the reaction path $K \to L$ (up to a constant factor) and should thus be most sensitive to energy maximum of the energy profile along the reaction path $K \to L$, corresponding to the highest-energy intermediate flask.

4. *Length Penalty.* $W_{K \to L}$ should increase roughly linearly with the number of steps N between K and L.

5. *Detailed Balance.* The heuristic kinetic criteria for the forward and the backward reaction should be related to each other via the energy difference between the initial and the final flasks,

$$W_{K \to L} - W_{L \to K} = \Delta E_{K \to L} = E_L - E_K \,. \tag{1}$$

The above properties allow us to interpret the heuristic kinetic criterion $W_{K \to L}$ as the

analog of the energy factor for the reaction path $K \to L$ in the Boltzmann distribution of all plausible reaction paths such that its rate is given by

$$V_{K \to L} \propto \exp(-\beta \, W_{K \to L}) \,, \tag{2}$$

where $\beta = 1/(k_B T)$ with Boltzmann constant k_B and absolute temperature T. Additivity of $W_{K \to L}$ ensures that multiplication of the probabilities for each step of the reaction pathway generates the probability of the entire pathway. The length penalty makes longer reaction paths relatively less likely and qualitatively describes a branching of reaction pathways at each elementary step. Hammond's postulate provides an approximation for the transition state energy of a reaction pathway by calculation of the energy of the highest energy intermediate. Finally, if the heuristic kinetic criterion $W_{K \to L}$ satisfies the detailed balance condition in Eq. 1, the rates of the forward and backward reactions computed according to Eq. 2 obey the detailed balance in a stationary state.

The heuristic kinetic criteria introduced in our previous work[13] are climb and arc. The *climb* criterion is

$$W_{c, K \to L} = \sum_{i=0}^{N} \max(E_{K_{i+1}} - E_{K_i}, 0) \,. \tag{3}$$

This climb kinetic criterion gives the highest activation barrier between the highest-energy intermediate and the initial reactant and thus generalizes the intuitive notion of the activation energy to the multi-step case. The climb criterion also fulfills the above requirements 1–3. However, the discriminating power of the climb criterion is modest, since it fails to account for the length of the reaction pathway and thus gives preference to long, low-barrier reaction mechanisms. In addition, $W_{c, K \to L} = 0$ for all reactions $K \to L$ with linear or concave energy profiles. The *arc* kinetic criterion introduces a constant penalty α per reaction step,

$$W_{a, K \to L} = \sum_{i=0}^{N} \left[(E_{K_{i+1}} - E_{K_i})^2 + \alpha^2 \right]^{1/2} \,, \tag{4}$$

where we can consider $\alpha \geq 0$ as an adjustable parameter. $\alpha = 0$ is equivalent to the climb criterion, while for $\alpha > 0$, the value of the arc criterion increases approximately linearly with the number of steps, thus additionally fulfilling the requirement 4 listed above. In the limit $\alpha \to \infty$, the arc criterion favors the pathways $K \to L$ with the smallest number of steps. The arc criterion performs relatively well for the reaction pathways involved in the formose reaction and predicts the mechanisms of the enolization, hemiacetal formation, and aldol addition in agreement with empirical knowledge. Moreover, the key products and reaction paths of the formose reaction are correctly reproduced by using the arc criterion with $\alpha = 1$ eV.[13] At the same time, the arc criterion penalizes every elementary transformation equally, even the formally energy-neutral (de)polarization steps (Table 1). In order to prevent the latter, we define the modified kinetic arc criterion *karc* $W_{k, K \to L}$ by

$$W_{k, K \to L} = \sum_{i=0}^{N} \left[(E_{K_{i+1}} - E_{K_i})^2 + \alpha^2 \left(\frac{1 - \exp(-\kappa(E_{K_{i+1}} - E_{K_i})^2)}{1 + \exp(-\kappa(E_{K_{i+1}} - E_{K_i})^2)} \right) \right]^{1/2}, \qquad (5)$$

where $\alpha \geq 0$ and $\kappa \geq 0$ are adjustable parameters. This functional form attenuates the penalty penalty α per reaction step depending on the corresponding energy change. In the limit of large energy changes, the attenuating term tends to 1, and the karc criterion increases approximately linearly with the number of steps. At the same time, energy-neutral steps $(E_{K_{i+1}} = E_{K_i})$, in particular the fictitious opening and closing of multiple bonds, do not contribute to the karc criterion. The parameter κ adjusts the rate of transition between the limiting cases. $\kappa = 0$ corresponds to the climb criterion, while $\kappa \to \infty$ recovers the arc criterion.

3   Building Large Chemical Reaction Networks

Automatic construction (or discovery) of reaction mechanisms is a fundamental task in kinetic modeling,[11,25–37] *ab initio* simulation of reaction dynamics,[38–41] and synthesis planning.[12,19,21,23,42–44] The HAQC approach[13] seeks the middle ground between database-driven

empirical methods[12,19,21,23,43] and *ab initio* PES exploration.[14–18,45,46] The basic unit of data used in constructing reaction mechanisms by HAQC is the flask, which refers to a collection of discrete molecules, subject to the constraint that only flasks with equal total stoichiometries may be transformed into one another. The constituent molecules of a flask are represented by their structural formulas and are considered as independent of each other. As a consequence, the network model containing flasks as its nodes and their mutual transformations as edges is well-defined as a discretized model of the reactive PES, while the corresponding energy evaluations involve only structure optimizations of the individual molecules and are cheaply and easily performed. At the same time, the mapping between molecules and flasks is of a many-to-many type and requires careful data synchronization. Another trade–off used in the HAQC approach is between the structural formulas and the corresponding three-dimensional molecular configurations used for energy evaluations. The transformation rules are formulated in the language of structural formulas in terms of bond breaking and bond making (Table 1), whereas quantum chemical structure optimizations have no notion of localized bonds and the optimized structures need to be validated with respect to the atomic connectivities and potentially rejected if rearrangements or fragmentations occur.

Due to the hybrid structure of the HAQC approach, the problem of building chemical reaction networks is divisible into a series of interlocking simple tasks, which can be efficiently executed in parallel fashion. Data throughput and synchronization thus become primary concerns and need to be optimized for in an efficient HAQC implementation. These requirements are quite unlike the situation in traditional quantum chemical computation, which is typically constrained by CPU, memory, and disk space needs. The open-source prototype Python implementation of the HAQC approach[47] is described in Section S1 of the Supporting Information.

4   Simple Reaction Mechanisms

In order to assess the predictive capabilities of the arc and karc heuristics defined in Sec. 2 and to optimize their adjustable parameters, we applied them to a set of 14 simple organic reactions (Table 2) representing a variety of polar reactions (substitutions, additions, and eliminations) as well as pericyclic reactions (6π electrocyclization, Claisen rearrangement, Diels–Alder and ene reactions). The appeal of the simple reaction mechanisms is that only one major is expected for each of them, and an abundance of experimental and theoretical data on their mechanisms is available for comparison with our single-source, single-product reaction predictions using HAQC. For simplicity we limit ourselves to reactions that contain only the compounds of C, H, O, and halogens here.

For each of the reactions of the simple reaction set, we created a model of the reactants (initial flask) and generated the complete reaction network beginning with it by repeatedly applying the elementary transformation rules (Table 1) until no new flasks could be made. All structure optimizations were performed using the PM7 semiempirical method as implemented in the MOPAC program[48] and the solvent effects were represented by the conductor-like solvation model (COSMO)[49] with the effective dielectric constant $\varepsilon = 78.4$. The energy of the proton was computed assuming a neutral aqueous solution (pH = 7).

The complete network of each simple reaction is the exhaustive description of the corresponding reactive system within the framework of HAQC. It contains all *plausible* reactive pathways between the reactant and the product flasks consistent with the defined elementary transformation rules but ignoring their kinetic feasibility (effectively in the infinite-temperature limit). Due to the presence of loops in the reaction network, the path length between a pair of nodes is generally unbounded, and a reasonable maximum path length N^{max} needs to be defined for plausible pathways. Of the plausible reaction pathways, we identify the (typically one or few) empirically *feasible* reaction pathways by comparing with the accepted reaction mechanisms from the organic chemistry literature. The statistics of the plausible (n_{p}) and empirically feasible (n_{f}) for the simple reaction mechanisms discussed

in this paper are given in Table 2.

We assess the usefulness of the arc and karc heuristic kinetic feasibility criteria based on their ability to distinguish the empirically feasible reaction pathways within the constructed reaction network from those pathways, which are plausible but infeasible based on the existing literature evidence. In order to facilitate the comparison between the reaction pathways modeled by HAQC and literature reaction mechanisms, two observations should be made. Firstly, the HAQC procedure by design divides reaction steps into individual bond breaking and bond making events (elementary transformations). For many reactions, several distinct orderings of elementary transformations are plausible (as defined above) and consistent with the accepted reaction mechanism. Moreover, both arc and karc criteria are invariant with respect to the ordering of the energy changes (Eq. 4 and Eq. 5, respectively). In this case, we denote the number of the plausible reaction pathways, which map onto the same reaction mechanism, as the *multiplicity* of this reaction pathway.

Secondly, we note that not all elementary transformations change the flask energy; indeed, the (de)polarization steps (Table 1) are considered energy-neutral within the HAQC procedure. To maintain consistency with these definitions, we view all flasks connected to each other by (de)polarization as an equivalence class and assign the energy of the flask with the most saturated valences (lowest total absolute charge) to all its members. While this approach in not the only possible one for dealing with (de)polarization steps, it is the simplest and helps prevent numerical noise.

Before turning to the individual discussions of the simple reaction mechanisms, we summarize the overall performance of the arc and karc heuristics for classifying the reaction pathways as feasible or infeasible. Fig. 1 gives an overview of the classification results for the reactions of Table 2 using the arc heuristic kinetic feasibility criterion W_a with $\alpha = 1$ eV. For each reaction mechanism, the plausible pathways with up to N^{max} elementary transformations are depicted as vertical lines in the order of increasing W_a values. Pathways consistent with the empirical knowledge of the respective reaction mechanism (empirically feasible) are

shown in blue, those deviating from the empirical reaction mechanisms are shown in red. We refer to the following Subsections 4.1–4.13 for detailed discussions.

As Fig. 1 illustrates, even the simple arc heuristic succeeds in correctly ranking the reaction pathways by feasibility for 10 out of 14 reaction mechanisms (except diol formation (DI), esterification (ES), Diels–Alder (DA), and ene (EN) reactions), assigning lower values to the empirically feasible reaction pathways. As a visual guide, the interval between the feasible pathway with the highest W_a value and the infeasible pathway with the lowest W_a value (*feasibility gap*, by analogy with the orbital gap of molecular orbital theory) is highlighted by gray shading in Fig. 1. The correct ranking provides for a simple binary classification of reaction pathways as feasible/infeasible using a threshold value $\overline{W_a}$ such that pathways with $W_a \leq \overline{W_a}$ are classified as feasible and those with $W_a > \overline{W_a}$ are considered infeasible. Any choice of $\overline{W_a}$ from within the (positive) feasibility gap will result in a correct classification for a given reaction mechanism. We further observe from Fig. 1 that the feasibility gaps for different reaction mechanisms vary significantly with respect to their location and width on the W_a scale. This is perhaps to be expected, given that the reaction mechanisms considered in this work occur at different temperatures, a fact we purposely ignored in our broad definition of kinetic feasibility. The best consensus threshold value across the set of reaction mechanisms considered here is $\overline{W_a} = 6$ eV (as indicated by a heavy vertical line in Fig. 1).

The karc heuristic criterion W_k with $\alpha = 1$ eV and $\kappa = 5$ eV^{-2} improves upon the arc heuristic by reducing the length penalty for elementary transformations with small energy changes. In particular, the fictitious (de)polarization of multiple bonds is not penalized (see Sec. 2 for details). As a result, the values of the karc heuristic are shifted downward compared to the arc heuristic with the same α parameter (see Fig. 2). With this correction, the ranking performance of the karc heuristic improves: The empirically feasible reaction pathways are consistently assigned smaller W_k values than infeasible reaction pathways for each of the 14 reaction mechanisms discussed below. Put differently, the karc heuristic allows for a binary

feasible/infeasible classification of all reaction mechanisms considered in this work, when considered separately. This finding leads us to assume that the karc heuristic correctly picks up on basic features of energy profiles contributing to the kinetic feasibility. The consensus threshold value of $\overline{W_k} = 6$ eV (shown as a heavy vertical line in Fig. 2) predicts that 9 out of 14 reactions (except bromine addition (BA), aromatic substitution (SA), epoxide hydrolysis (EP), Diels–Alder (DA) and ene (EN) reactions) should have at least one feasible reaction pathway under a comparable set of conditions. The comparisons of arc and karc heuristic criteria using other values of the α and κ parameters are given in Section S2 of the Supporting Information. We return to the discussion of the implications of these results in Sec. 5.

Table 2: Reactants, products, plausible (n_p) and empirically feasible (n_f) pathways, and maximum reactive pathway lengths (N^max) of the simple reaction set.
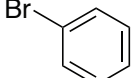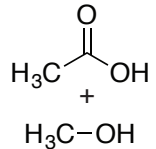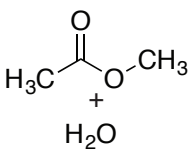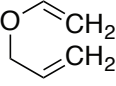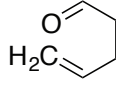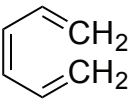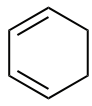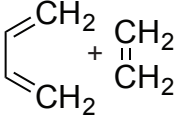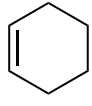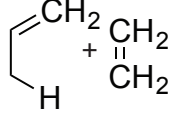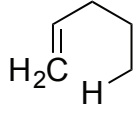
	Reactants	Products	n_f	n_p	N^max
Tautomerizaton (TA)	H_2C⁀OH	H_3C⁀O	2	1	8
Diol formation (DI)	$H_2C{=}O + H_2O$	HO⁀OH	38	7	8
S_N1 substitution (S1)	H_3C–(CH_3)(CH_3)–$Br + H_2O$	H_3C–(CH_3)(CH_3)–$OH + HBr$	94	5	6
S_N2 substitution (S2)	H_3C⁀$Br + H_2O$	H_3C⁀$OH + HBr$	102	5	6
Br_2 addition (BA)	$H_2C{=}CH_2 + Br{-}Br$	Br⁀Br	14	2	6
HBr addition (HA)	$H_3C{-}C({=}CH_2){-}CH_3 + HBr$	H_3C–(CH_3)(CH_3)–Br	28	3	6
$E1$ elimination (E1)	H_3C–(CH_3)(CH_3)–Br	$H_3C{-}C({=}CH_2){-}CH_3 + HBr$	28	3	6
S_EAr substitution (SA)	⬡ $+ Br{-}Br$	Br–⬡ $+ HBr$	96	4	6
Epoxide hydrolysis (EP)	▷O $+ H_2O$	HO⁀OH	61	2	6

14

Table 2: Reactants, products, plausible (n_p) and empirically feasible pathways (n_f), and maximum reactive pathway lengths (N_{max}) of the simple reaction set. (*cont.*)

Reaction	Reactants	Products	n_f	n_p	N^{max}
Esterification (ES)			523	9	6
Claisen rearrangement (CL)			264	108	6
6π electrocyclization (6C)			12241	56	8
Diels–Alder reaction (DA)			5532	248	8
Ene reaction (EN)			10443	138	8

4.1 Enol tautomerization

The keto–enol tautomerism is one of the best studied organic reaction mechanisms, with no shortage of experimental and computational studies.[50–53] The one-step tautomerization in the gas phase is governed by the Woodward–Hoffmann rules[54–56] and thus proceeds via an antarafacial 1,3-H shift with a highly strained and energetically unfavorable transition state.[53,57,58] The computed free energy of reaction and free energy of activation for the tautomerization of vinyl alcohol to acetaldehyde in the gas phase are -11.54 kcal/mol and 56.62 kcal/mol, respectively (using the Gaussian 2(MP2) method).[58]

In water, the activation barrier is considerably lower due to acid/base catalysis by the solvent. The acid-catalyzed tautomerization consists of a fast and reversible protonation of the
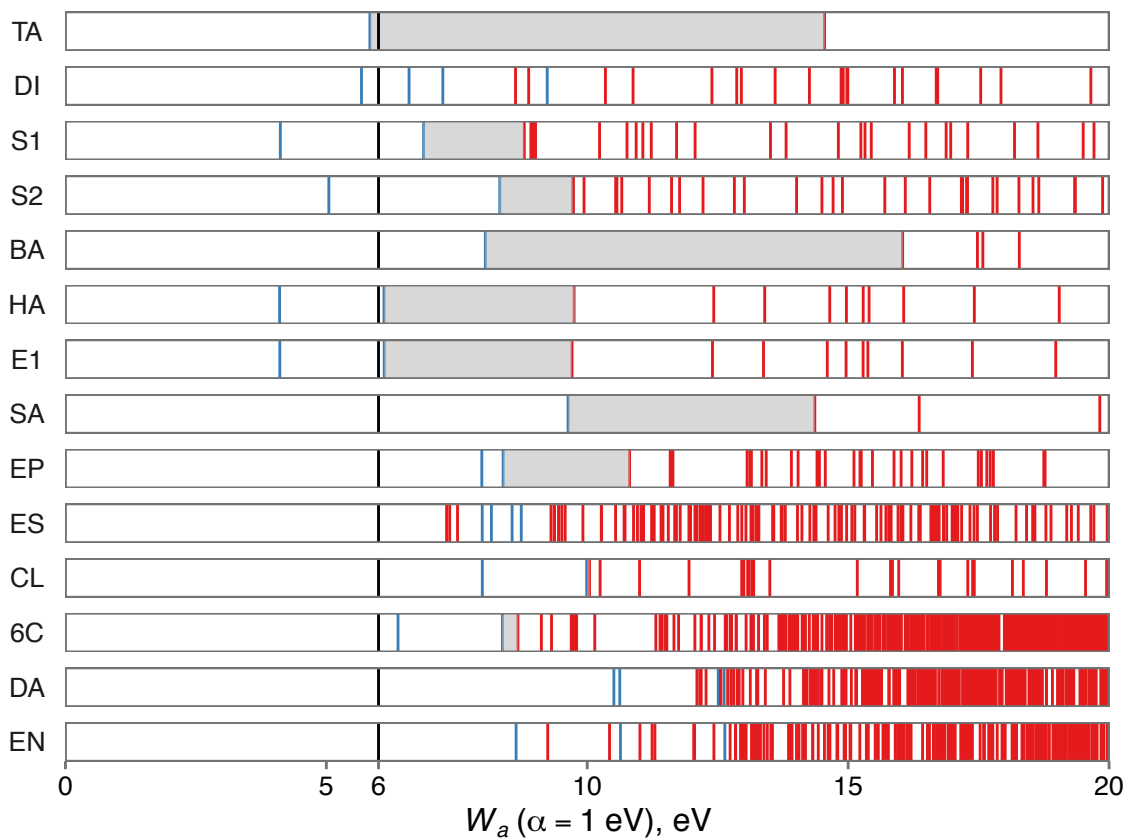
Figure 1: Classification of reaction pathways as feasible/infeasible using the arc heuristic with $\alpha = 1$ eV for simple reaction mechanisms. Empirically feasible reaction pathways are marked by blue lines, infeasible pathways are marked by red lines. The interval between the feasible reaction pathway with the highest W_a value and the infeasible reaction pathways with the lowest W_a value (feasibility gap) is highlighted by gray shading. $\overline{W_a} = 6$ eV estimates the consensus threshold for binary feasible/infeasible classification. See Table 2 for reaction abbreviations.

Figure 2: Classification of reaction pathways as feasible/infeasible using the karc heuristic with $\alpha = 1$ eV, $\kappa = 5$ eV^{-2} for simple reaction mechanisms. $\overline{W_k} = 6$ eV estimates the consensus threshold for binary feasible/infeasible classification See Table 2 for reaction abbreviations and Fig. 1 for additional details.
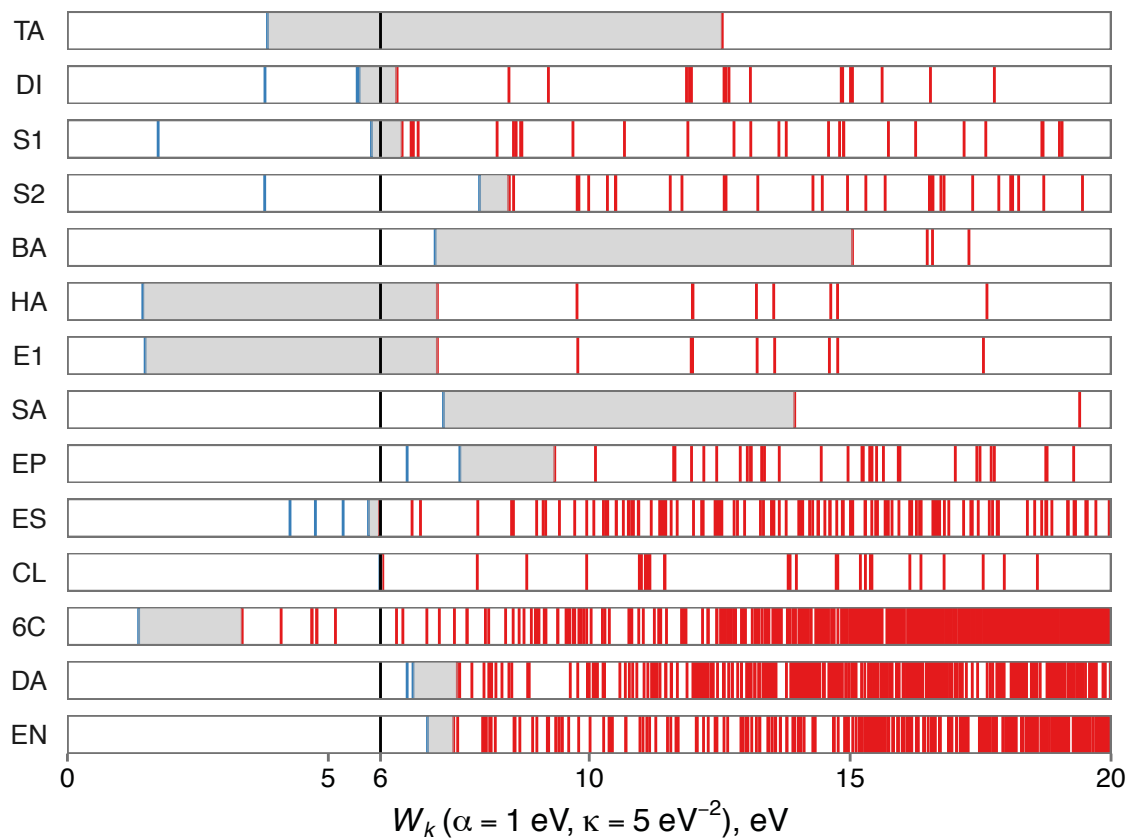
vinyl alcohol at the α-carbon atom, followed by the slower and rate-determining deprotonation of the conjugated acid of acetaldehyde.[50–52,59] Under base catalysis conditions, the enol group in vinyl alcohol is reversibly deprotonated, while the re-protonation at the α-carbon atom is rate-determining.[50–52] The experimental free energy of tautomerization of vinyl alcohol in water is -8.5 kcal/mol,[59,60] while the activation energy for the acid-catalyzed process was determined as 11.82 kcal/mol.[61] The computed energy of tautomerization was found to be only weakly affected by solvation (-9.6 kcal/mol in the gas phase, -8.0 kcal/mol in water using density functional theory (DFT) with the B3LYP exchange–correlation functional and 6-311++g(3df,3pd) basis sets).[62] In contrast, the activation energy of the tautomerization is lowered from 78.4 kcal/mol for the gas-phase mechanism to 41.8 kcal/mol when two explicit water molecules are included in the reaction mechanism.[62]

The feasible tautomerization pathway TA–0 (Fig. 3) mirrors the concerted gas-phase mechanism. The first three steps of the predicted pathway are fictitious (de)polarization steps. The proton shift from the oxygen to the α-carbon atom corresponds to the rate-determining step (Fig. 5, left), although the HAQC procedure by construction does not include stereochemical information and makes no distinction between suprafacial and antarafacial 1,3-H shift.[54,55]

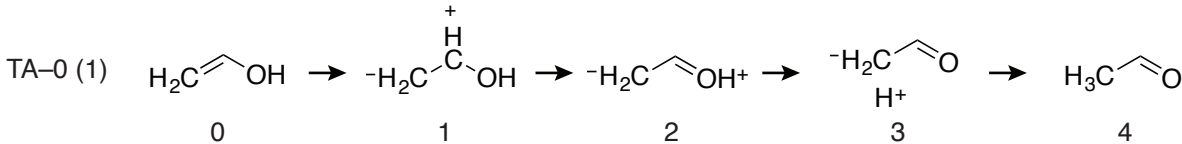Figure 3: Feasible tautomerization (TA) reaction pathway. Path multiplicity is given in parentheses.

4.2 Geminal diol formation

The addition of nucleophiles, such as water, to the carbonyl group is characteristic of aldehyde and ketone reactivity.[63,64] In formaldehyde CH_2O, the hydration equilibrium in water lies predominantly on the side of the hydrate $CH_2(OH)_2$, with the hydration equilibrium

constant at 298 K, $K_{eq} = 1300$.[65,66] The mechanisms of the water addition to formaldehyde are significantly different in the gas phase and in solution, with the solvent playing an active catalytical role.[63,65] The gas-phase mechanism is concerted with a simultaneous proton transfer and C–O bond formation taking place along the reaction coordinate.[67,68] This process has a high activation barrier of 35.2 kcal/mol (with B3LYP exchange–correlation functional and aug-cc-pVTZ basis sets).[69] In the presence of additional water molecules, a cooperative hydration mechanism with an expanded cyclic transition state can be attained, drastically lowering the activation energy of hydration.[70]

The diol formation in solution is subject to general acid/base catalysis.[63,65,67,71] In the presence of acids, the slow and reversible first step consists of the concerted attack by the water molecule on the carbonyl C atom and the simultaneous protonation of the carbonyl O atom by the acid catalyst. The following deprotonation step is fast and reversibly yields the hydrate.[63,71] Similarly, the rate-determining first step of the base-catalyzed water addition includes the bond formation between the water O atom and the carbonyl C atom, while the base simultaneously removes a proton from the water molecule. The fast proton exchange between the base and the reaction intermediate concludes the reaction mechanism.[63,71]

Three feasible pathways of water addition to formaldehyde are predicted by the HAQC procedure (Fig. 4). The pathways DI–0 and DI–1 correspond to the concerted water addition mechanism and differ only in the order of the elementary transformations. In the DI–0 pathway, the C–O bond formation precedes the O–H bond dissociation, while their sequence is reversed in the DI–1 pathway. In accordance with the acid- and base-catalyzed mechanisms, the C–O bond formation is part of the rate-determining step. The third feasible pathway, DI–2, describes an alternative mechanism of diol formation, in which formaldehyde is first reversibly protonated and subsequently reacts with the water nucleophile. This mechanism was considered by Jencks and co-workers and ruled out based on experimental evidence.[63,71]
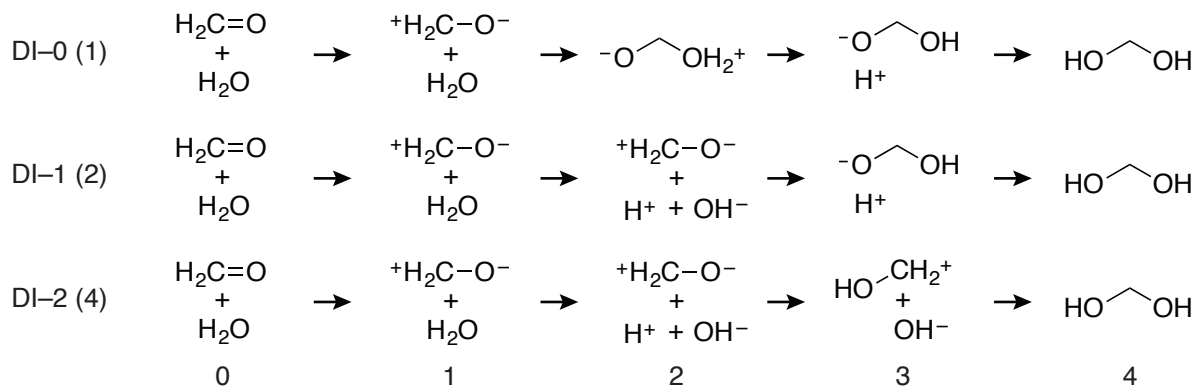
Figure 4: Feasible diol formation (DI) reaction pathways. Path multiplicities are given in parentheses.
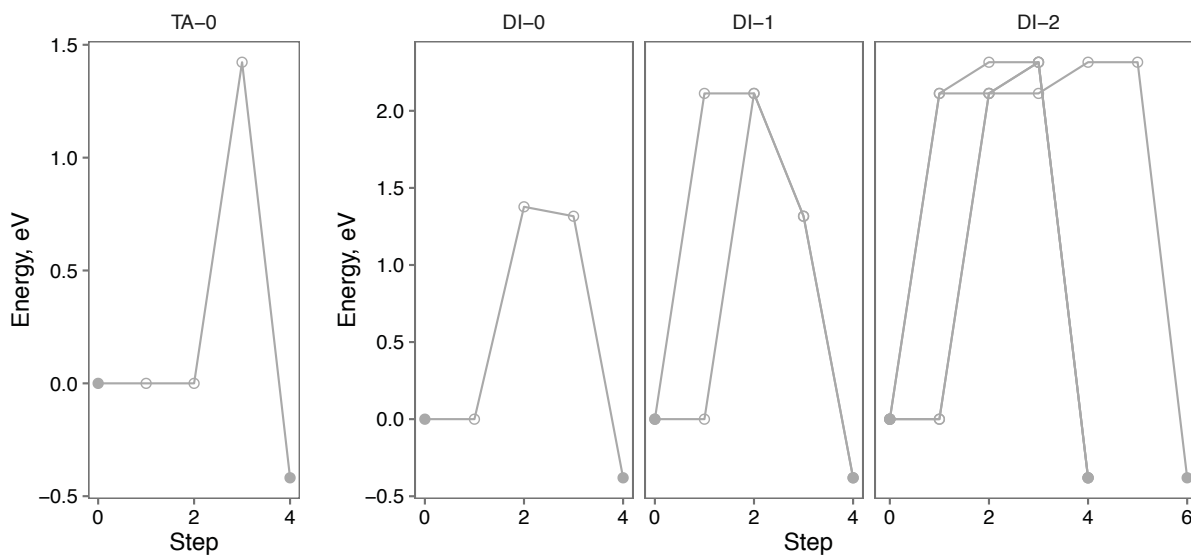

Figure 5: Energy profiles of feasible tautomerization (TA, left) and diol formation (DI, right) reaction pathways

20

4.3 S_N1 and S_N2 nucleophilic substitution

Nucleophilic substitution at the saturated carbon center is known to proceed in a bimolecular (S_N2) or unimolecular (S_N1) fashion, dependent on the structure of the substrate, the nucleophilicity of the reagent, the solvent, and many other factors.[4,72–77] The classical S_N2 mechanism is expected in primary substrates, *e.g.*, primary alkyl halides, in which the nucleophilic attack displaces the leaving group in a concerted step. At the other end of the mechanistic spectrum, the S_N1 mechanism is typical of tertiary substrates, which undergo reversible ionization to the corresponding carbocations in the rate-determining step, followed by a fast capture by the nucleophile. The relevance of ion-pair intermediates and the nucleophilic assistance by the solvent have been the subject of considerable debate and depend on the details of the reaction.[78–86]

Although the S_N2 mechanism is also operative in the gas phase, the latter is an imperfect model for the solution reaction.[87–90] The PES of the gas-phase S_N2 reaction features two deep energy minima, which describe the reactant and product complexes (bound by ion–dipole interactions), which are separated by a relatively low barrier.[88,91–93] The reaction barrier of the classical solution S_N2 mechanism is, to a significant degree, due to desolvation of the nucleophile.[4,94,95] In the best-studied gas-phase S_N2 reaction, the chloride self-exchange, Cl^- + $CH_3Cl \rightarrow ClCH_3$ + Cl^-, the ion–dipole complex has an experimental binding enthalpy of 10.5 kcal/mol, while the effective activation enthalpy with respect to the reactants is only 2.9 kcal/mol.[93] The corresponding solution reaction in water has an experimental activation enthalpy of 26.3 kcal/mol.[96]

The HAQC formulation, by construction, does not distinguish between the concerted S_N2 and the stepwise S_N1 mechanisms. In both ethyl bromide (S2–0) and tert-butyl bromide hydrolysis (S1–0) reactions, the C–Br bond is broken before the C–O bond is formed (Figs. 6 and 7). However, the two pathways differ in the height of the energy maxima relative to their reaction energies (Fig. 8). The experimental free activation enthalpy for the hydrolysis of ethyl bromide in water is 26 kcal/mol, while the free reaction enthalpy is

-6.4 kcal/mol.[97] The hydrolysis of tert-butyl bromide has an experimental free activation enthalpy of 18 kcal/mol and a free reaction enthalpy of -9.3 kcal/mol.[97] It is thus reasonable to posit that the degree of concertedness in the sequence of steps predicted by HAQC depends on the ratio of the energy maximum to the total energy change. In fact, this suggestion is closely related to the concept of *enforced concertedness* of Jencks.[79] The alternative substitution mechanisms (S1–1 and S2–1) are dominated by the dissociation of water, which overwhelms the details of the substitution mechanism (Fig. 8).
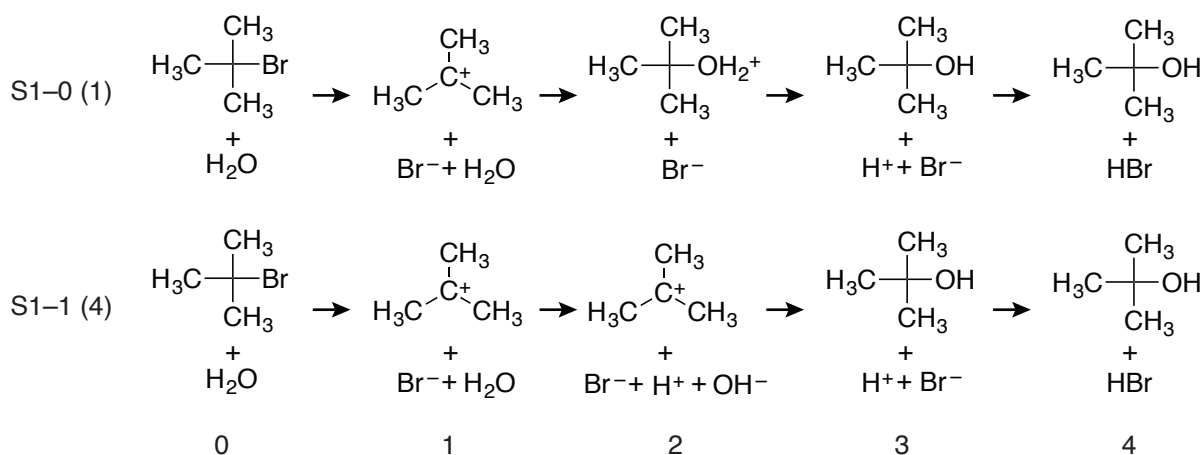

Figure 6: Feasible S_N1 substitution (S1) reaction pathways. Path multiplicities are given in parentheses.
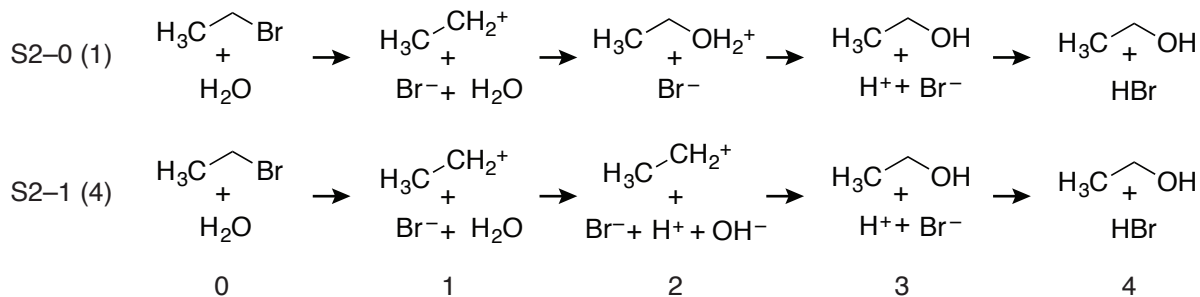

Figure 7: Feasible S_N2 substitution (S2) reaction pathways. Path multiplicities are given in parentheses.
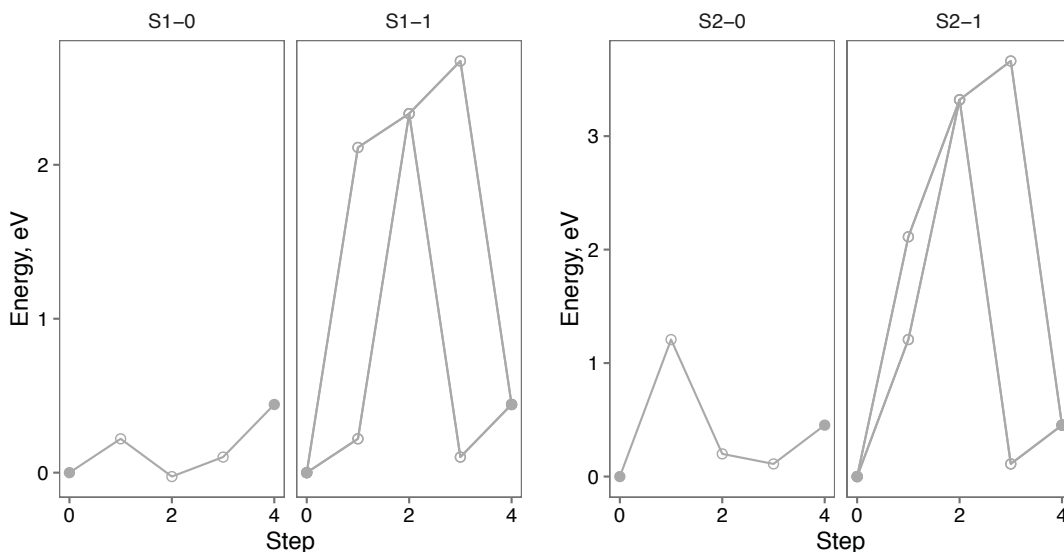
Figure 8: Energy profiles of feasible S_N1 (S1, left) and S_N2 (S2, right) substitution reaction pathways

4.4 Electrophilic bromine addition to alkenes

The addition of molecular bromine to alkenes is a fast reaction at room temperature in water with a surprisingly complex reaction mechanism.[98–107] The key step of the bromination reaction is the electrophilic attack by the Br_2 molecule that yields a cyclic bromonium ion as an ionic intermediate. The subsequent backside addition of the bromide counterion is fast and results in the formation of a *trans* dibromo product. Due to the charge separation in the rate-determining step, the bromination reaction is unknown in the gas phase and is driven by the interaction with the solvent.[106] The experimental activation energy of bromination of aliphatic alkenes in water is 8–11 kcal/mol.,[99,106] while the theoretical estimate for the activation energy of the gas-phase reaction is 64.2 kcal/mol using the Gaussian 3(MP2) method.[108] By using an implicit solvation model, the reduction in the activation barrier could be explained in terms of the dielectric stabilization by the solvent. The computed free activation enthalpy for the reaction between ethene and Br_2 was 64.5 kcal/mol in the gas phase and 8.2 kcal/mol in water (with MP2 method and CEP-121(aug) basis sets).[109] The experimental reaction enthalpy of bromination of ethene is -28.9 kcal/mol.[108]

The experimentally observed reaction rates of bromination often decrease with temperature, which indicates the presence of a pre-equilibrium, which is attributed to the formation of the π-complex between the alkene and Br_2.[106,110,111] The experimental reaction enthalpy for the π-complex formation is -2.8 kcal/mol (for hex-1-ene),[99] while the computed complexation energy is -1.6 kcal/mol with the MP2 method and 6-311+G** basis sets.[112] Kinetic studies have also shown that more than one molecule of bromine may be involved in the rate-determining addition step, resulting in both bimolecular (Ad_E2) and termolecular (Ad_E3) reaction kinetics.[98,104,105,107]

The only feasible reaction pathway for bromine addition to ethene (BA–0) is consistent with the established reaction mechanism (Fig. 9). The energy maximum is attained by heterolytic dissociation of the Br_2 molecule; the resulting the electrophilic reagent Br^+ adds to the ethene molecule in the subsequent step (Fig. 10). Due to the restriction in the elementary transformations disallowing divalent halogens (Table 1), the ionic intermediate is not a cyclic bromonium ion but an isomeric open-chain carbenium ion. The latter structure was identified as unstable with respect to the bromonium ion using configuration interaction with single and double excitations (CISD) and DZ+d basis sets.[113] The incorporation of additional elementary transformation heuristics to allow divalent bromine and thus the cyclic bromonium ion as a reaction intermediate is a straightforward extension of this work.

4.5   Addition of HBr to alkenes

The addition of HBr to 2-methylpropene both in gas phase and in solution yields tert-butyl bromide in accordance with Markovnikov's rule.[77] Protic solvents favor a polar mechanism, in which the protonation of the alkene molecule constitutes the rate-determining step.[101,103,104,114] The resulting tert-butyl cation captures a bromide ion and yields the reaction product. On the other hand, the gas-phase reaction is likely to proceed via a cyclic transition state with some degree of charge separation, which directs the regioselectivity in Markovnikov's sense.[25,115–118] The gas-phase Markovnikov addition of HBr to

2-methylpropene is exothermic with reaction enthalpy -16.8 kcal/mol and an activation energy of 23.2 kcal/mol.[115] The activation energy is well reproduced by calculations using the MP2 method and 6-311+G(3df,2p) basis sets: the computed value is 26.5 kcal/mol for the Markovnikov pathway and 38.2 kcal/mol for the anti-Markovnikov addition.[118]

The HA–0 pathway is the exact representation of the polar addition mechanism consisting of the sequential addition of a proton and a bromide ion to the alkene molecule (Fig. 9). However, the energy profile does not conform to the typical barrier-like shape but goes solely downhill instead (Fig. 10). This energy profile points to a very fast reaction in solution and is consistent with the computed value for the protonation energy of ethene in water of -234 kcal/mol, which was obtained with DFT using the B3LYP exchange –correlation functional and 6-31G(d,p) basis sets.[119] In contrast, the experimental estimates for the energetics of the protonation of 2-methylpropene to the tert-butyl cation in water yield a vastly different picture, with the free reaction enthalpy for protonation of 16.5 kcal/mol and the corresponding free activation enthalpy of 22.1 kcal/mol.[97] This qualitative discrepancy is likely due to the specific solvation effects in water and deserves further study.

4.6 Elimination of HBr from alkyl bromides

The elimination of HBr from tert-butyl bromide[74,77,120–129] is the reverse process of the addition reaction discussed in the previous section and should thus contain exactly the same steps in the inverted order, according to the principle of microscopic reversibility. The gas-phase elimination is consequently a concerted process with partial charge separation in the transition state. The activation energy of the endothermic gas-phase elimination of HBr from tert-butyl bromide is 39 kcal/mol.[130]

Eliminations from alkyl halides in solution are usually base-catalyzed and exhibit a wide range of reaction mechanisms, depending on the substrate, the leaving group, and the base.[74,77,121,124,128,129] The bimolecular mechanism ($E2$) involves the abstraction of a proton by the base, the dissociation of the halide ion, and the for formation of the C=C bond

in a concerted step. On the other hand, the unimolecular mechanism ($E1$) includes a slow bond breaking step, which produces a carbenium ion intermediate, and a fast deprotonation step. Stabilization of the carbenium ion results in the reaction mechanism with a more pronounced $E1$ character. The experimental free reaction enthalpy for the elimination of HBr from tert-butyl bromide in water is 2.9 kcal/mol.,[131,132] while the experimental free activation enthalpy for the elimination using ethoxide as base is 23.6 kcal/mol.[132]

The HAQC procedure is consistent with the principle of microscopic reversibility if the set of elementary transformations is invariant with respect to exchanging reactants and products (Table 1) and if the heuristic kinetic criterion obeys detailed balance. Therefore, the E1–0 pathway is the reverse of the HA –0 pathway and describes the polar $E1$ elimination mechanism (Figs. 9 and 10).
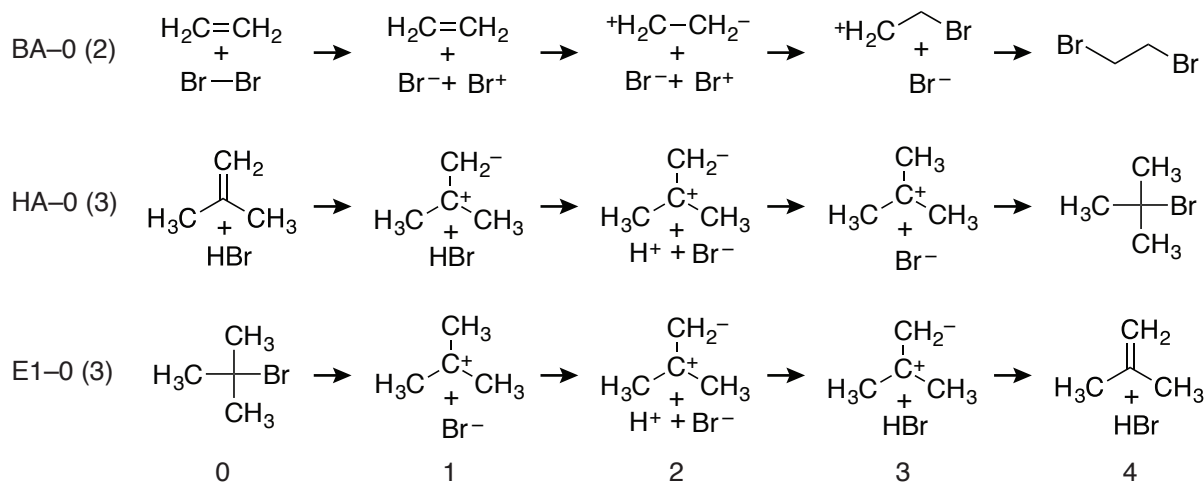
Figure 9: Feasible Br_2 addition (BA), HBr addition (HA), and $E1$ elimination (E1) reaction pathways. Path multiplicities are given in parentheses.

4.7 Electrophilic aromatic substitution

The reaction mechanism of electrophilic aromatic substitution (S_EAr) seemed firmly established for decades[74,77,102,133–138] before being recently called into question by a series of detailed computational studies.[139–141] In its traditional formulation, the S_EAr mecha-
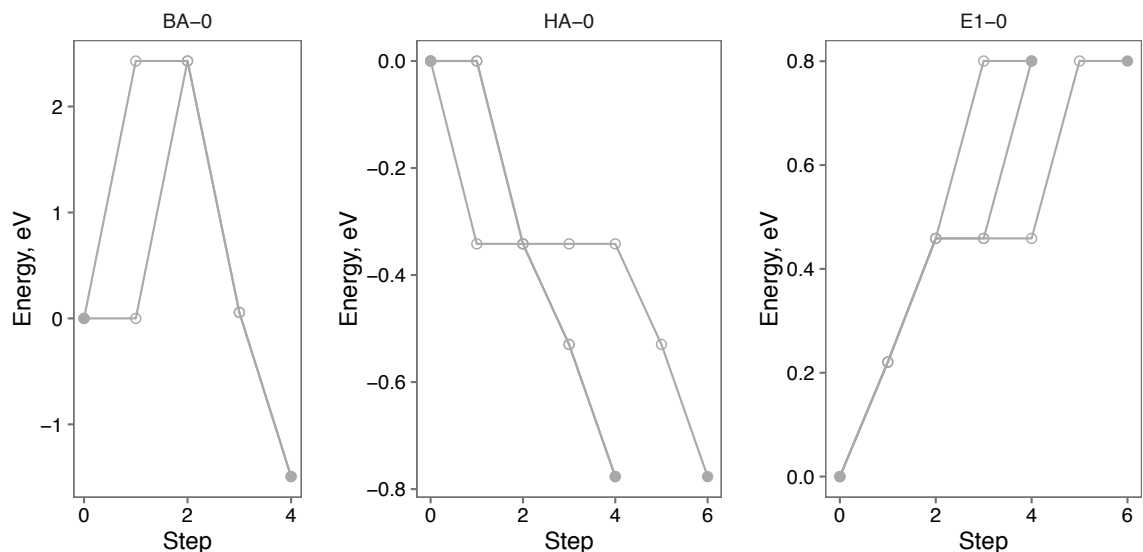
Figure 10: Energy profiles of feasible Br_2 addition (BA, left), HBr addition (HA, center), and $E1$ elimination (E1, right) reaction pathways

nism is a two-step process involving the arenium ion (σ-complex) as the crucial intermediate.[74,102,134,137,138,142] The arenium ion is reversibly produced in the rate-determining step and subsequently loses a proton to give the substitution product. The electrophilic attack is preceded by the equilibrium formation of the less strongly bound π-complex between the electrophile and the aromatic substrate.[143,144]

Bromination of benzene is an exothermic process with the reaction enthalpy of -17.2 kcal/mol and a high activation energy of 64.7 kcal/mol in the absence of catalysts. These values were computed using the MP2 method and 6-311++G** basis sets.[145] Unexpectedly, the computed barriers for alternative reaction mechanisms were found to be substantially lower in recent computational studies by Schleyer, Schaefer, and co-workers.[139–141] The barrier for the direct (one-step) substitution was computed as 41.8 kcal/mol using the B2PLYP double-hybrid exchange–correlation functional and 6-311+G(2d,2p) bases sets, while several addition–elimination pathways have activation energies < 40 kcal/mol.[139] It should be noted, however, that brominations of unactivated aromatics are typically carried out in the presence of Lewis acid catalysts. In any case, electrophilic aromatic substitution might be mechanistically richer than traditionally assumed.[146]

The pathway SA–0 (Fig. 11) describes essentially the S_EAr mechanism. Similar to the bromination of ethene (Sec. 4.4), the heterolysis of the Br_2 molecule produces the maximum of the energy profile (Fig. 13). The Br^+ electrophile adds to the benzene molecule and forms the σ-complex, which is then deprotonated to yield the product, bromobenzene. The addition–elimination bromination pathways were not considered as feasible for the purposes of this work.
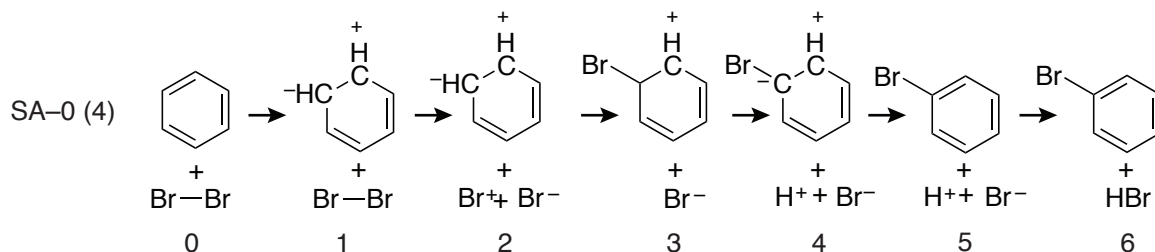
Figure 11: Feasible S_EAr substitution (SA) reaction pathway. Path multiplicity is given in parentheses.

4.8 Epoxide hydrolysis

The opening of the epoxide ring by nucleophiles is analogous to nucleophilic substitution at saturated carbon centers, save for the effect of the ring strain. The mechanistic alternatives are the bimolecular S_N2-like mechanism, also known as $A2$, and the unimolecular mechanism similar to S_N1, also called $A1$.[74,147,148] The base-catalyzed hydrolysis follows the bimolecular $A2$ mechanism, in which the nucleophilic attack and the epoxide ring opening occur in a single concerted step.[147,149,150] In contrast, the stepwise $A1$ mechanism is found in acid-catalyzed hydrolysis. The first step, which is rate-determining, involves the opening of the protonated epoxide ring. The resulting carbenium ion reacts with the nucleophile in the fast second step.[147,149,151–153] The experimental reaction enthalpy for the hydrolysis of ethene oxide is -46.0 kcal/mol.[154] The activation energy for the acid-catalyzed hydrolysis was determined as 19.0 kcal/mol.[152] The computed activation energies for the acid-catalyzed and base-catalyzed hydrolysis are 18.9 kcal/mol and 10.0 kcal/mol, respectively, using DFT with the B3LYP

exchange–correlation functional and 6-311+G(d,p) basis sets.[150]

The pathways of epoxide hydrolysis predicted as feasible (EP–0 and EP–1) (Fig. 12) are consistent with experimentally known $A1$ mechanism. The energy maximum of the EP–0 pathway corresponds to the ring opening in the protonated ethene oxide (Fig. 13), in line with experimental results. The C–O bond breaking in the neutral ethene oxide (EP–1) generates an energetically unfavorable zwitterion with the positive charge on a primary carbon center, which might enforce the concertedness of the reaction mechanism ($A2$), similar to the nucleophilic substitution mechanisms discussed above (Sec. 4.3).
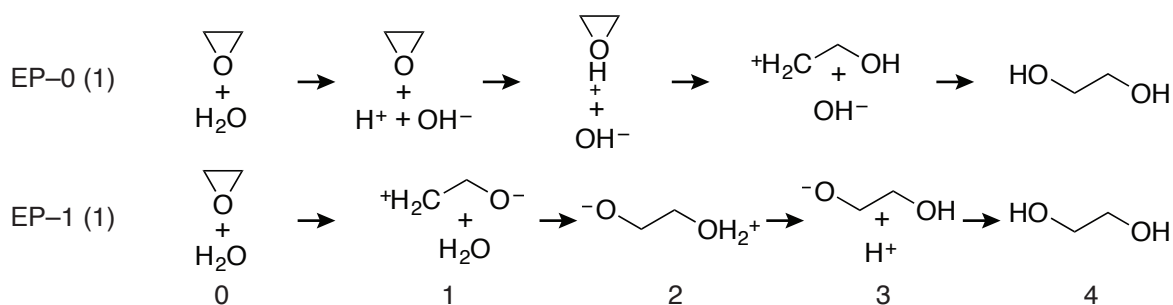

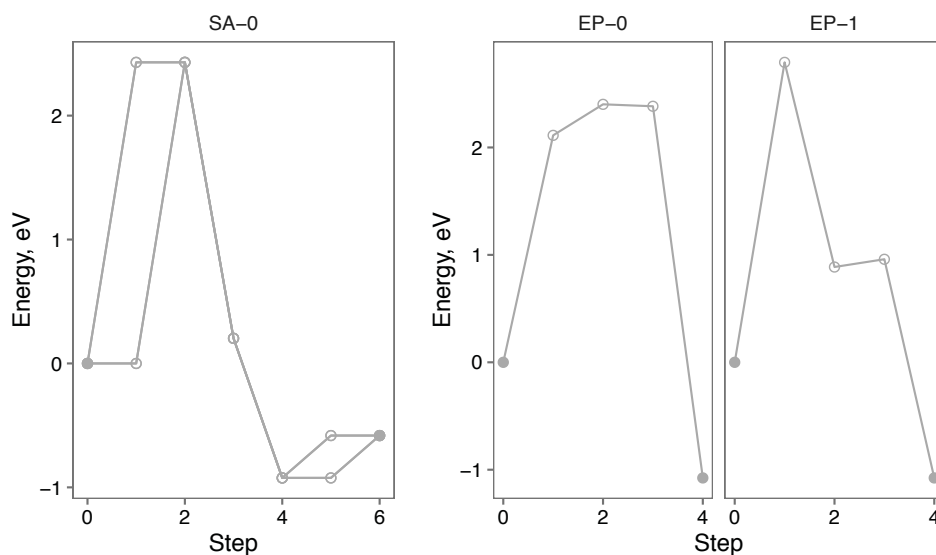Figure 12: Feasible epoxide hydrolysis (EP) reaction pathways. Path multiplicities are given in parentheses.

Figure 13: Energy profiles of feasible S_EAr substitution (SA, left) and epoxide hydrolysis (EP, right) reaction pathways

29

4.9 Esterification

Esterification and its reverse reaction, ester hydrolysis, offer a study in rich mechanistic variation.[74,155–160] The traditional three-factor classification of Ingold and co-workers characterizes the reaction mechanisms in terms of acidic (A) or basic (B) conditions, the breaking of the acyl–O (AC) or alkyl–O (AL) bond, and the molecularity of the reaction.[74,161] While ester hydrolysis is well known under both acidic and basic conditions, only acid-catalyzed esterification is possible due to the formation of the inert carboxylate anions in basis solution. The predominant mechanism of esterification in aliphatic carboxylic acids is the two-step $A_{AC}2$ mechanism involving a tetrahedral acid–alcohol adduct as the reaction intermediate.[74,155–157,159,161,162] The first step of the $A_{AC}2$ mechanism—the formation of the tetrahedral intermediate—is rate-determining. A common alternative is the $A_{AL}2$ mechanism, which amounts to S_N2 nucleophilic substitution at the alcohol carbon atom.[74,159,161,163]

Esterification reactions are usually close to energy-neutral and thus reversible. The free enthalpy of the reaction of acetic acid with methanol is +1.9 kcal/mol.[164] The formation of the tetrahedral intermediate is strongly endergonic with 16.9 kcal/mol and has a significant free activation enthalpy of 34.0 kcal/mol for the uncatalyzed process, or 23.4 kcal/mol in the presence of an acidic catalyst.[164] The decomposition of the tetrahedral intermediate has a free activation enthalpy of 18.2 kcal/mol uncatalyzed, or 10.7 kcal/mol under acidic catalysis.[164] The sizable negative activation entropy of the esterification reaction (-22.5 cal/(mol·K) for the formation of ethyl acetate[157,165]) is indicative of the $A_{AC}2$ addition–elimination mechanism.

Numerous computational studies addressed the energetics of esterification and ester hydrolysis under various conditions.[166–171] The computed energy change for the acid-catalyzed reaction of acetic acid with methanol in water is -8.6 kcal/mol using the MP2 method and 6-31+G(d,p) basis sets for structure optimizations and the larger 6-311+G(d,p) basis sets for the single-point energy calculations.[171] The tetrahedral intermediate is 4.4 kcal/mol above the reactants. The formation of the tetrahedral intermediate has the highest bar-

rier of 9.7 kcal/mol, while the activation energy for the decomposition of the tetrahedral intermediate is merely 2.7 kcal/mol.

The feasible esterification pathways predicted by HAQC (ES–0 through ES–3) all describe the canonical $A_{AC}2$ mechanism (Fig. 14) and proceed via a high-energy tetrahedral intermediate (Fig. 15). The ES–0 pathway consists of the nucleophilic attack by a neutral methanol molecule and the subsequent elimination of a water molecule. The other feasible pathways include the attack by the methoxide anion (ES–1), the elimination of hydroxide anion (ES–2), or both (ES–3).
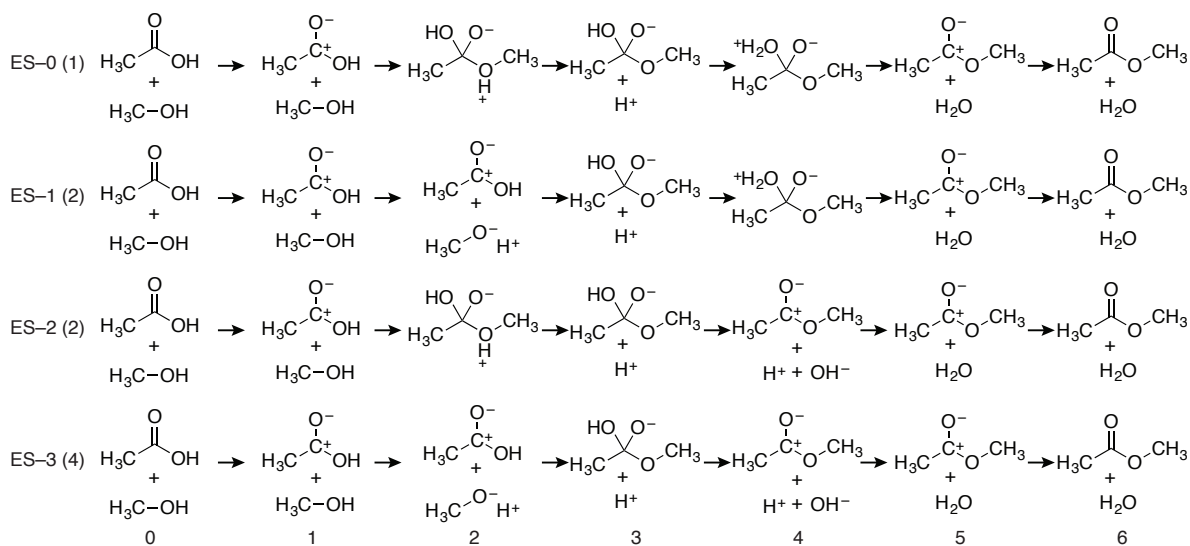
Figure 14: Feasible esterification (ES) reaction pathways. Path multiplicities are given in parentheses.

4.10 Claisen rearrangement

We begin our discussion of pericyclic reactions with the Claisen rearrangement of vinyl allyl ether to 4-pentenal.[77,172–178] Typical features of pericyclic reactions are the absence of observable intermediates and relative insensitivity to acid–base catalysis. A concerted cyclic reaction mechanism was proposed for the Claisen rearrangement early on;[172,173] and the Woodward–Hoffmann principle of the conservation of symmetry provided a qualitative theoretical framework for the Claisen rearrangement and related reactions.[54–56,179] The Claisen
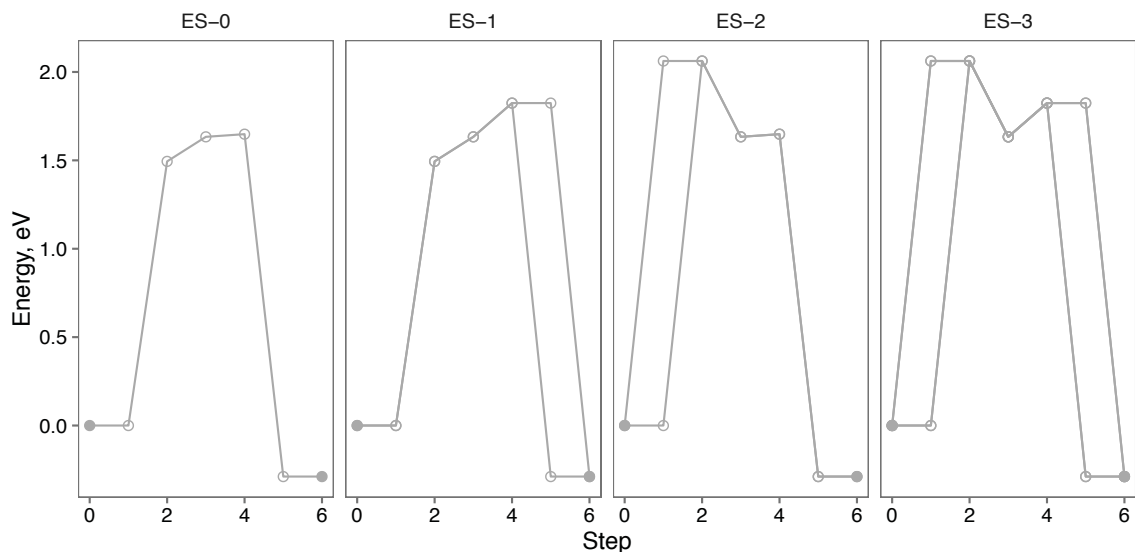
Figure 15: Energy profiles of feasible esterification (ES) reaction pathways

rearrangement is classified as a [3,3]-sigmatropic reaction and is thus thermally allowed with a favorable *supra,supra*-topology. The transition state assumes a chair conformation and is subject to aromatic stabilization.[180–183] The experimental activation enthalpy for the Claisen rearrangement of allyl vinyl ether is 29.8 kcal/mol[180,184] in gas phase and 25.4 kcal/mol in a non-polar solvent (dibutyl ether).[185] Density functional calculations with the B3LYP functional and 6-31+G** basis sets yield the activation energy of 26.1 kcal/mol,[183] the composite Gaussian-3 method gives 29.4 kcal/mol.[186]

The Claisen rearrangement is appreciably exothermic with an experimental reaction enthalpy of −17.1 kcal/mol for the allyl vinyl ether.[187] The reaction rate is increased by 2–3 orders of magnitude in aqueous solutions.[177,188–192] The lowering of the activation barrier by −3.5 kcal/mol in water is due to hydrogen bonding[189–191] and indicates that the transition state is more polar than the reactant. The reaction mechanism is nonsynchronous with the C–O bond breaking having progressed further in the transition state than the C–C bond formation.[180,192–194]

The HAQC procedure was not specifically designed to reproduce the concerted mechanisms of pericyclic reactions involving multiple simultaneous bond breaking and bond making

events. Nevertheless, the predicted reaction profile (Fig. 17) has the correct single-barrier shape. The energy maximum corresponds to the polar dissociation of the C–O bond, followed by the recombination of the charged fragments to yield the product (Fig. 16).

4.11 6π electrocyclizaton

The cyclization of hexa-1,3,5-triene to cyclohexa-1,3-diene is a thermally allowed disrotatory process according to the Woodward–Hoffmann rules.[55,56,77,195–200] The concerted mechanism of the 6π electrocyclization reaction has been extensively studied by quantum chemical methods.[182,183,186,201–205] The experimental reaction enthalpy of -15.2 kcal/mol[198] is in good agreement with the computed reaction energy of -17.9 kcal/mol using the MP2 method and 6-311++G(3df,3pd) basis sets.[186] However, DFT calculations using the B3LYP exchange–correlation functional and 6-311++G(3df,3pd) basis sets underestimate the reaction energy, -9.3 kcal/mol.[186] The experimental activation enthalpy for the disrotatory mechanism is 29 kcal/mol, while that for the forbidden conrotatory process is estimated to be only about 5 kcal/mol higher.[198,203] The computed activation enthalpy is 28.8 kcal/mol the complete basis set (CBS-QB3) method and 30.8 kcal/mol with B3LYP and 6-311++G(2d,p) basis sets.[205]

The energy profile of the 6π electrocyclizaton reaction produced by the HAQC procedure corresponds to a concerted barrierless process (Fig. 17), at variance with the available experimental and computational results. This is in part due to the lack of an energetic penalty for charge separation in double bonds. Since we consider the (de)polarization steps as formally energy-neutral, the only evident energy change describes the exothermic ring closure (Fig. 16). This finding suggests that an empirical penalty for charge separation might be appropriate for (de)polarization elementary steps and warrants further study.
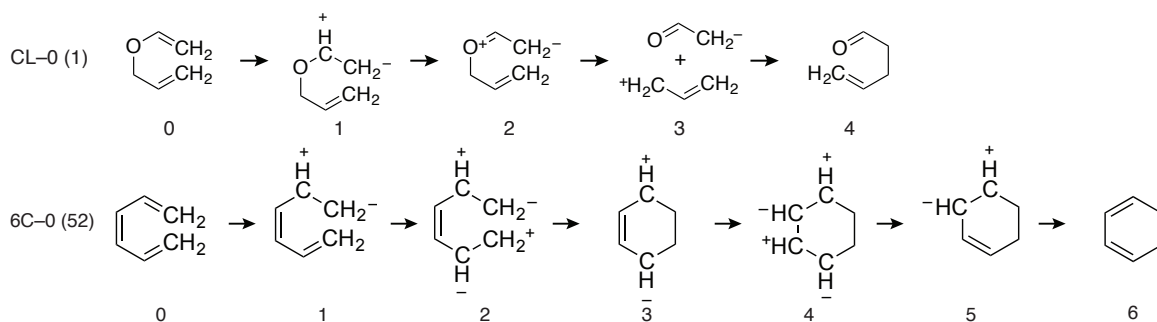
Figure 16: Feasible Claisen rearrangement (CL) and 6π electrocyclizaton (6C) reaction pathways. Path multiplicities are given in parentheses.
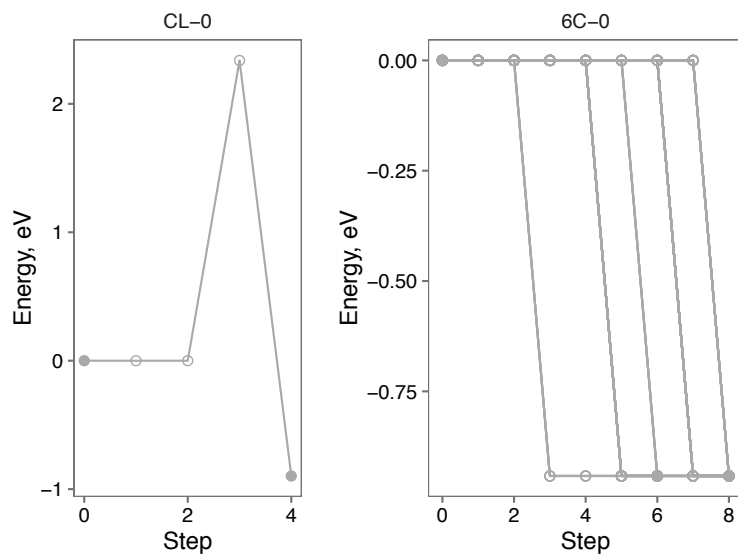

Figure 17: Energy profiles of feasible Claisen rearrangement (CL, left) and 6π electrocyclizaton (6C, right) reaction pathways

34

4.12   Diels–Alder reaction

The Diels–Alder reaction is at once the canonical example for a thermal pericyclic reaction and an invaluable synthetic tool in the construction of cyclic systems.[4,77,206–217] The prototype reaction of 1,3-butadiene with ethene requires high-temperature conditions and is mostly of theoretical interest. The principle of conservation of symmetry[55,56,218] predicts that the product cyclohexene is formed in a one-step thermal process. Many quantum chemical studies at different levels of theory confirmed the existence of the C_s-symmetric transition state in the Diels–Alder reaction, in accordance with the Woodward–Hoffmann principle.[4,182,212,219–223] The experimental reaction energy of the cycloaddition between 1,3-butadiene and ethene is reported as -37.6 kcal/mol;[4] however, the computed reaction energies show considerable scatter as a function of level of theory and basis sets used.[4,186,205,223–226]

The computed reaction energy with DFT using B3LYP exchange–correlation functional and 6-31* basis sets is -36.9 kcal/mol,[186] in good agreement with the experimental value. Meanwhile, adding diffuse basis functions to the basis sets, counterintuitively, makes the agreement with the experiment worse: the reaction energy is predicted as -27.5 kcal/mol with B3LYP and 6-311++G(2df,2p) basis sets.[4] The CBS-Q3 method gives the reaction enthalpy as -38.3 kcal/mol.[205] The experimental activation energy of 27.5 kcal/mol for the reaction of 1,3-butadiene with ethene[211,227] is well reproduced by DFT methods (24.8 kcal/mol with B3LYP exchange–correlation functional and 6-31G* basis sets,[223,224] 28.3 kcal/mol with B3LYP and large 6-311++G(3df,3pd) basis sets[186]). The predicted activation energy using the CBS-QB3 method is 22.9 kcal/mol,[205] in agreement with other high-level correlated methods.[186,228,229]

The predictions of the HAQC methodology are consistent with the concerted reaction mechanism of the Diels–Alder reaction . Again, the concerted mechanism is split into a sequence of elementary transformations (bond breaking and bond making, Fig. 18). We observe a single energy maximum in the energy profile, associated with first C–C bond formation, while the second step follows without an additional barrier (Fig. 19). The two

predicted reaction mechanisms (DA–0 and DA–1) are near-degenerate with respect to the modified kinetic arc criterion W_k and differ only in the order of the bond formation steps.
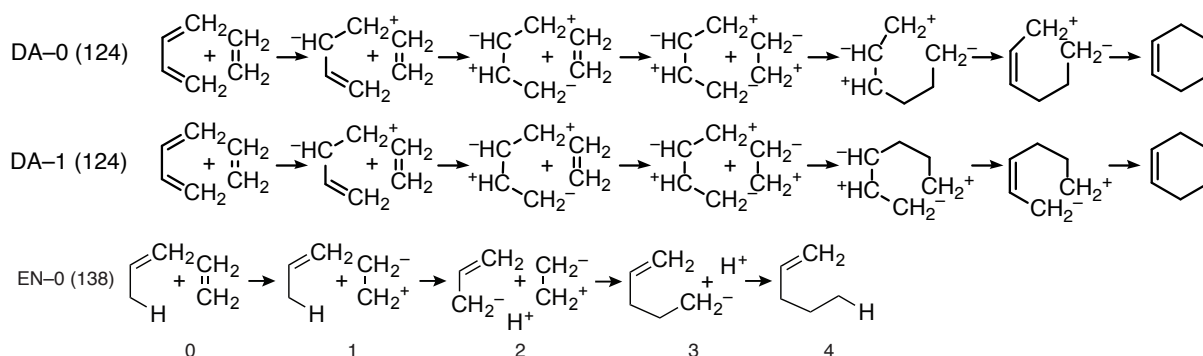
Figure 18: Feasible Diels–Alder (DA) and ene (EN) reaction pathways. Path multiplicities are given in parentheses.
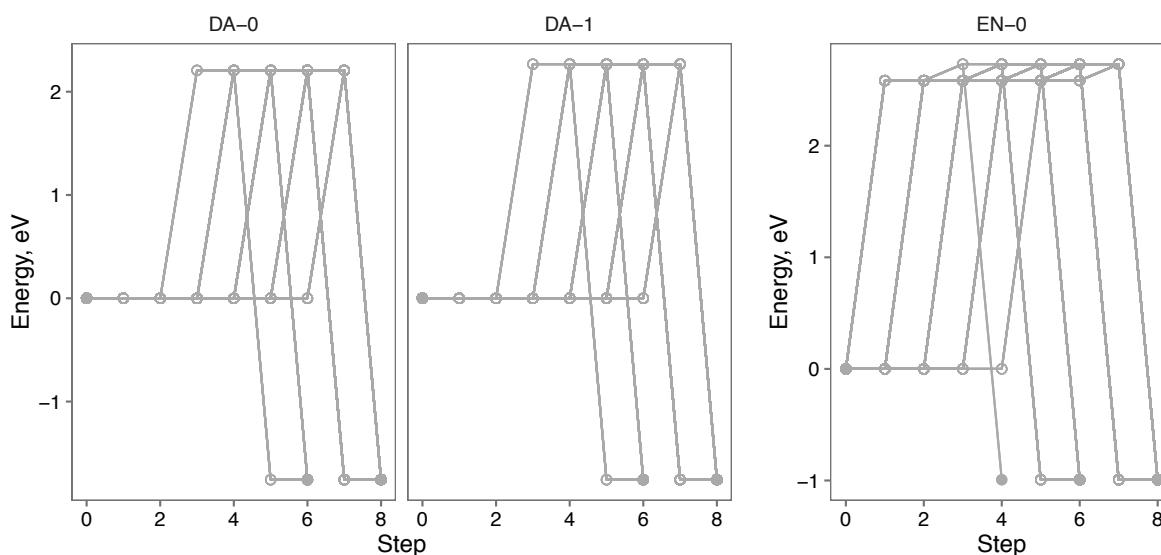
Figure 19: Energy profiles of feasible Diels–Alder (DA, left) and ene (EN, right) reaction pathways

4.13 Ene reaction

The Alder ene reaction.[216,230,231] consists in the addition of an alkene containing allyl hydrogen atoms to the C=C bond of a second alkene molecule in a manner formally similar to the Diels –Alder reaction. Computational results have shown that the reaction mechanism of the

36

ene reaction is concerted, however, the formation of the C–C bond precedes that of the C–H bond.[182,216,221] The formation of 1-pentene by thermal pyrolysis of mixtures of propene and ethene is exothermic with the reaction enthalpy of -21.8 kcal/mol at 700 K.[232,233] The computed reaction energy of the ene reaction between propene and ethene is -21.0 kcal/mol with DFT using the B3LYP exchange–correlation functional and 6-31G* basis sets.[234] The MP2 reaction energy using 6-31G* basis sets is -24.5 kcal/mol.[234] The experimental activation energy for this reaction was reported as 57.4 kcal/mol[232] but is at variance with quantum chemical predictions.[233–235] DFT using the B3LYP functional and 6-31G* basis sets predicts the activation energy for the reaction of propene with ethene to be 33.3 kcal/mol,[234] with a similar value of 31.5 kcal/mol yielded by the MP2 calculations using 6-31G* basis sets.[234]

The feasible pathway of the ene reaction (EN–0 in Fig. 18) reflects the concerted reaction mechanism consisting of the C–C bond formation followed by the C–H bond formation.[182,216,221] These two steps are observable as separate energy changes in the energy profile (Fig. 19).

5   Discussion

## 5.1   Energy Profiles and Their Interpretation

Having analyzed a diverse set of 14 polar and pericyclic reaction mechanisms in the preceding Subsections 4.1–4.13, we are now in the position to make some tentative generalizations. The most immediately commonality of the feasible reaction pathways considered so far is that their energy profiles have the intuitively expected single-barrier shape. Among polar reactions, we find only two exceptions: the HBr addition and elimination reactions, in which the energy profiles are entirely non-increasing and non-decreasing, respectively. For polar reactions, which feature high-energy charged intermediates, these results are straightforward to explain by Hammond's postulate. But perhaps more surprisingly, we find that the energy profiles of pericyclic reactions considered in this work are also described as a single barrier

crossing with the HAQC approach. The 6π electrocyclization reaction is an exception, however, having an strictly downhill energy profile. By design, the HAQC approach tends to err on the side of being too granular and to represent any elementary reaction—as defined by the reaction kinetics—as a sequence of one or more elementary transformations of Table 1. As we have seen in Subsection 4.3, this approach erases the much-studied qualitative distinction between concerted (S_N2) and stepwise (S_N1) reaction mechanisms, reducing it to a quantitative difference in the relative height of the energy barrier (Fig. 8). In this case, the model suggested by the HAQC approach is in line with the accepted notion of a S_N2–S_N1 mechanistic continuum.[4,72–77] More generally, however, we wish to be able to identify concerted elementary reactions from the sequences of elementary transformations generated by the HAQC procedure. Borrowing on Jencks' concept of "enforced concertedness",[79,236,237] one promising approach to treating concerted reaction mechanisms within the HAQC framework could be to use the barrier height or the curvature of the energy profile as a heuristic for merging adjacent elementary transformations into one concerted reaction step.

While the HAQC energy profiles are useful conceptual devices and seem to correctly capture some patterns of chemical reactivity, they should not be overinterpreted. It is perhaps best to think of them as taking an intermediate position between the qualitative "arrow pushing" reaction diagrams and the complex realities of multidimensional reactive PES. Importantly, the HAQC energy profiles should not be taken as the literal traces of reactive trajectories in energy space. We highlight 3 fundamental differences: The points of the energy profile are discrete, the connection between them are nothing more than visual guides. Moreover, each of the points corresponds to a (local) energy minimum on the PES, even if it appears to be a maximum of the energy profile. While the relationship between transition-state structures and high-energy reaction intermediates are implied by the Hammond postulate, it is not quantitative. Finally, the abscissa of the energy profile is not a true reaction coordinate, recording the overall changes in molecular structure. Instead, it follows discrete steps of bond making and bond breaking with each bond transformation event being

weighted equally. While this simple approach is a useful first step, it neglects the differences in bond distances and bond strengths between bonds of different types. Improved variants of the simple HAQC energy profiles could take the latter into account by using abstract measure of molecular change[238] or molecular similarity[239,240] as their abscissa instead on the number of steps.

5.2   Reaction Mechanism Prediction

Taking into consideration only the narrow goal of predicting feasible reaction pathways, we have pursued a somewhat circuitous and inefficient approach in this work. We have first constructed an exhaustive transition network (TN) model of plausible elementary transformations (see Section S3 of the Supporting Information for details of the reaction network models), only to filter it down to the feasible reaction pathways using thermodynamic and heuristic kinetic feasibility criteria. Only between 50 % (for tautomerization) and 0.5 % (for 6π electrocyclization) of all plausible reaction pathways in Table 2 are empirically feasible. Our approach is justified because we are additionally interested in comparing and optimizing kinetic heuristics. However, given a fixed set of feasibility criteria, we are free to choose a much more efficient strategy based on local search.[241,242] The formal properties of the kinetic heuristics, specifically non-negativity and additivity (see Section 2), are designed to make them applicable in the context of local search or network-flow optimization algorithms.[243] The simplest local-search variant of the HAQC approach might use a global kinetic feasibility threshold $\overline{W_k}$ along with an additional local kinetic feasibility threshold $\overline{w_k} = \overline{W_k}/N^{\max} + \delta w_k$, where $N^{\max}$ is the maximum path length, and δw_k is a local slack parameter. Within this local- threshold approach, one would in every step only accept elementary transformations $K_i \to K_{i+1}$ ($0 \le i \le N^{\max} - 1$), for which $W_{K_i \to K_{i+1}} \le \overline{w_k}$ (local threshold criterion) and $W_{K \to K_{i+1}} \le \overline{W_k}$ (global threshold criterion).

Path search algorithms are exceptionally versatile and widely used.[242,244] The breadth-first search algorithm[244] is suitable for finding all reaction pathways with up to $N^{\max}$ steps

between the given source and the target flasks (single-source, single-product reaction prediction), which is the focus of this work. But with only small modifications, this approach can be adapted to solve different problems in reaction mechanism prediction. We will only give a brief enumeration here. The feasibility of a particular reaction, given the source and target flasks, translates into finding the path with the smallest value of the kinetic feasibility heuristic and can be readily solved by a shortest-path algorithm. Breadth-first search can also be extended to solve the single-source, multi-product reaction problem. The only significant difference to this work is that we need to take into consideration both kinetic and thermodynamic feasibility of the predicted reaction pathways.[13] Provided that a suitable calibration exists, it should even be possible to estimate overall reaction rates and product compositions using Eq. 2.

5.3   Reaction Networks and Their Properties

While local search is likely the most efficient approach to reaction mechanism prediction, the full reaction networks in the TN representation are worthy of a closer examination in their own right. We recall that main difference between the TN model and the interaction network (IN) representation, which plays a central role in kinetic modeling[11,25–37] and is usually the implied meaning of the term *reaction network*,[7,9,12,23,27] is that network nodes in the TN representation correspond to collections of molecules (flasks), while network edges of the TN representation denote stoichiometry-preserving transformations between flasks (see Section 2 and Ref. 13 for details). The latter definition has a number of practically useful consequences. Firstly, the reaction networks in the TN representation are large but fundamentally finite, meaning that the exhaustive generation of the reaction network by the HAQC procedure is guaranteed to terminate after a finite number of steps. The network size in the TN representation increases exponentially with the number of atoms, making it impractical to generate the entire network for large enough systems. Nevertheless, as we have demonstrated in this work, reaction networks in the TN representation remain of manageable

size for some systems relevant to chemical reactivity (up to 8 non-hydrogen atoms, $< 2 \cdot 10^5$ nodes, $< 10^6$ edges, see Section S3 of the Supporting Information).

Secondly, the formal properties of the reaction networks in the TN representation influence the efficiency of many network algorithms.[244] As a simple example, we note that the reaction networks in the TN representation are relatively sparse, *i.e.*, each network node is connected to only a small fraction of the overall network. Sparse directed networks are characterized by small values of *network density*

$$\rho = \frac{N_{\mathrm{e}}}{N_{\mathrm{n}} \left(N_{\mathrm{n}} - 1\right)} \, , \tag{6}$$

where N_{n} and N_{e} are the numbers of network nodes and edges in the network, respectively. We find that, with the exception of the very small network of the diol formation reaction, $\rho < 0.1$ for all reaction networks considered in this work. Moreover, the network density ρ is expected to decrease even further for larger systems, as the number of adjacent edges for a given node (plausible elementary transformations) increases roughly linearly with the number of atoms, while at the same time the number of network nodes grows exponentially. This finding is encouraging because network search algorithms are known to perform better for sparse networks.[244]

Finally, the reaction networks in the TN representation form a well-defined hierarchy (partially ordered set), in which the network of the stoichiometry $\mathrm{C}_{n_{\mathrm{C}}}\mathrm{H}_{n_{\mathrm{H}}}\mathrm{O}_{n_{\mathrm{O}}}...$ is contained within the $\mathrm{C}_{n'_{\mathrm{C}}}\mathrm{H}_{n'_{\mathrm{H}}}\mathrm{O}_{n'_{\mathrm{O}}}...$ network if $n_{\mathrm{C}} \leq n'_{\mathrm{C}}$, $n_{\mathrm{H}} \leq n'_{\mathrm{H}}$, $n_{\mathrm{O}} \leq n'_{\mathrm{O}}$, ... The relationship between the analogous reaction pathways in the smaller and the larger network expresses in a straightforward way the effect of a *catalyst*, namely, a compound, whose addition to the chemical system improves the kinetic feasibility of some desirable reaction pathway.

6 Conclusions

At present most reaction mechanism predictions fall squarely into one of two disjoint categories: the maximally empirical approach based on the existing mechanistic and synthetic data[1,19–23,74,77] and the the maximally non-empirical approach relying PES explorations by quantum chemical methods.[11,14–18,33–37] In this work, we are exploring a middle ground, combining a series of broadly applicable empirical rules and heuristics with quantum chemical structure optimizations. The resulting HAQC framework is conceptually simple, general, extensible, and consistent with empirical understanding of organic chemistry. Instead of relying on the ever-growing and often proprietary compilation of reaction data, it makes use of a small number of reaction-independent heuristics (elementary transformations and reaction feasibility criteria). The ability to find a consensus set of heuristics correctly predicting a wide range of polar and pericyclic reactions make HAQC a promising proposition for discovering novel reaction pathways across organic chemistry.

However, much room for improvement remains in the level of computational treatment used thus far within the HAQC approach. In all calculations performed in this work, we excluded any conformational data or explicit solvent effects from the energy computations and restricted ourselves to the inexpensive PM7 semiempirical method for reasons of simplicity. Improvements along these lines can be incorporated at some additional computational cost but without fundamental changes to the HAQC approach. Additionally, the parallel implementation and the local-search variant of the HAQC approach help to limit their impact on the overall computational effort.

On the more fundamental level, one of the most important shortcomings of the HAQC approach is the tight coupling between the structural formulas, on which elementary transformation rules operate, and the three-dimensional structural representations used in quantum chemical structure optimizations. While the structural formulas have the advantage of being intimately familiar to any chemist, alternative discretized representations of molecular structure might be more suitable for expressing elementary transformation rules. We have

previously explored some discrete encodings of three-dimensional molecular structures based on space-filling curves.[245] These or some other improvements will be necessary to incorporate other kinds of interactions, such as metal–ligand interactions, hydrogen bonds and other weak interactions, and to express the corresponding reaction semantics. Fast progress in this direction is not to be expected but the success of some three-dimensional molecular structure descriptors[246] is encouraging.

We should also note that we have barely scratched the surface of the various ways, in which machine learning techniques might be utilized to improve the reaction feasibility criteria. The arc and karc heuristics were largely motivated by transition-state theory[38–41] but we are equally well justified in deriving kinetic heuristics by regression against experimental or computed barrier heights. Combining discrete features such as types of elementary transformations (see Table 1) with the computed HAQC enegy profiles using regression trees[247] is yet another promising avenue to explore. In similar vein, the criteria for combining elementary transformations into concerted reactions steps might be more easily obtained by machine learning rather than the empirical reasoning offered in Section 5.1.

We close by expressing the hope that chemical reaction networks might become more useful as a conceptual framework for both understanding the patterns of chemical reactivity and for predicting new reactions with moderate computational effort. It is notable that one of early classic texts on graph theory references chemical structures as graphs.[248] We would like to argue that a deeper examination of chemical reaction graphs will similarly enrich our understanding of chemical reactivity.

Acknowledgement

References

(1) Lowry, T. H.; Richardson, K. S. *Mechanism and Theory in Organic Chemistry*; Harper Row: New York, 1987.

(2) Isaacs, N. *Physical Organic Chemistry*, 2nd ed.; Longman: Harlow UK, 1996.

(3) Anslyn, E. V.; Dougherty, D. A. *Modern Physical Organic Chemistry*; University Science Books: Sausalito CA, 2005.

(4) Bachrach, S. M. *Computational Organic Chemistry*; Wiley: Hoboken NJ, 2014.

(5) Whitesides, G. M.; Ismagilov, R. F. Complexity in Chemistry. *Science* **1999**, *284*, 89–92.

(6) Ludlow, R. F.; Otto, S. Systems chemistry. *Chem. Soc. Rev.* **2007**, *37*, 101–108.

(7) Nitschke, J. R. Systems chemistry: Molecular networks come of age. *Nature* **2009**, *462*, 736–738.

(8) Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits*; Chapman & Hall: Boca Raton FL, 2007.

(9) Jeong, H.; Tombor, B.; Oltvai, Z. N.; Barabási, A.-L. The large-scale organization of metabolic networks. *Nature* **2000**, *407*, 651–654.

(10) Szostak, J. W. Origins of life: Systems chemistry on early Earth. *Nature* **2009**, *459*, 171–172.

(11) Green Jr., W. H. Predictive Kinetics: A New Approach for the 21st Century. *Adv. Chem. Eng.* **2007**, *32*, 1–313.

(12) Gothard, C. M.; Soh, S.; Gothard, N. A.; Kowalczyk, B.; Wei, Y.; Baytekin, B.; Grzy- bowski, B. A. Rewiring chemistry: Algorithmic discovery and experimental validation of one-pot reactions in the network of organic chemistry. *Angew. Chem. Int. Ed.* **2012**, *51*, 7922–7927.

(13) Rappoport, D.; Galvin, C. J.; Zubarev, D. Y.; Aspuru-Guzik, A. Complex Chemi- cal Reaction Networks from Heuristics-Aided Quantum Chemistry. *J. Chem. Theory Comput.* **2014**, *10*, 897–907.

(14) Bernardi, F.; Robb, M. A. In *Ab Initio Methods in Quantum Chemistry. Part I*; Lawley, K. P., Ed.; Advances in Chemical Physics; Wiley: Hoboken, NJ, 1987; Vol. 67; Chapter 3, pp 155–248.

(15) Schlegel, H. B. In *Ab Initio Methods in Quantum Chemistry. Part I*; Lawley, K. P., Rice, S. A., Eds.; Advances in Chemical Physics; Wiley: Hoboken, NJ, 1987; Vol. 67; Chapter 4, pp 249–286.

(16) Henkelman, G.; Jóhannesson, G.; Jónsson, H. In *Theoretical Methods in Condensed Phase Chemistry*; Schwartz, S. D., Ed.; Kluwer: Dordrecht Netherlands, 2002; Chap- ter 10, pp 269–302.

(17) Hratchian, H. P.; Schlegel, H. B. In *Theory and Applications of Computational Chem- istry*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier: Ams- terdam, 2005; Chapter 10, pp 195–249.

(18) Schlegel, H. B. Geometry optimization. *WIREs Comput. Mol. Sci.* **2011**, *1*, 790–809.

(19) Corey, E. J.; Wipke, W. T. Computer-Assisted Design of Complex Organic Syntheses. *Science* **1969**, *166*, 178–192.

(20) Dugundji, J.; Ugi, I. *Computers in Chemistry*; Topics in Current Chemistry; Springer: Berlin, Heidelberg, 1973; Vol. 39; pp 19–64.

(21) Ugi, I.; Bauer, J.; Bley, K.; Dengler, A.; Dietz, A.; Fontain, E.; Gruber, B.; Herges, R.; Knauer, M.; Reitsam, K.; Stein, N. Computer-Assisted Solution of Chemical Problems—The Historical Development and the Present State of the Art of a New Discipline of Chemistry. *Angew. Chem. Int. Ed.* **1993**, *32*, 201–227.

(22) Jorgensen, W. L.; Laird, E. R.; Gushurst, A. J. CAMEO: A program for the logical prediction of the products of organic reactions. *Pure Appl. Chem.* **1990**, *62*, 1921–1932.

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

(24) Hammond, G. S. A Correlation of Reaction Rates. *J. Am. Chem. Soc.* **1955**, *77*, 334–338.

(25) Benson, S. W. *Thermochemical Kinetics*, 2nd ed.; Methods for the Estimation of Thermochemical Data and Rate Parameters; Wiley: New York, 1976.

(26) Feinberg, M. In *Dynamics and Modelling of Reactive Systems*; Stewart, W. E., Ray, W. H., Conley, C. C., Eds.; Academic Press: New York, 1980; Chapter 3, pp 59–130.

(27) Temkin, O. N.; Zeigarnik, A. V.; Bonchev, D. G. *Chemical Reaction Networks: A Graph-Theoretical Approach*; CRC Press: Boca Raton FL, 1996.

(28) Côme, G.-M. M. *Gas-Phase Thermal Reactions*; Springer Netherlands: Dordrecht, 2001.

(29) Helfferich, F. G. *Kinetics of Multistep Reactions*, 2nd ed.; Comprehensive Chemical Kinetics; Elsevier: Amsterdam, 2004; Vol. 40.

(30) Swihart, M. T. In *Modeling of Chemical Reactions*; Carr, R. W., Ed.; Comprehensive Chemical Kinetics; Elsevier, 2007; Vol. 42; Chapter 5, pp 185–242.

(31) Marin, G. B.; Yablonsky, G. S. *Kinetics of Chemical Reactions*; Wiley-VCH: Weinheim Germany, 2011.

(32) Zielinski, D. C.; Palsson, B. Ø. In *Systems Metabolic Engineering*; Wittmann, C., Lee, S. Y., Eds.; Springer Netherlands: Dordrecht, 2012; pp 25–55.

(33) Broadbelt, L. J.; Stark, S. M.; Klein, M. T. Computer generated reaction networks: on-the-fly calculation of species properties using computational quantum chemistry. *Chem. Eng. Sci.* **1994**, *49*, 4991–5010.

(34) Susnow, R. G.; Dean, A. M.; Green, W. H.; Peczak, P.; Broadbelt, L. J. Rate-Based Construction of Kinetic Models for Complex Systems. *J. Phys. Chem. A* **1997**, *101*, 3731–3740.

(35) Green Jr, W. H.; Moore, C. B.; Polik, W. F. Transition States and Rate Constants for Unimolecular Reactions. *Annu. Rev. Phys. Chem.* **1992**, *43*, 591–626.

(36) Sumathi, R.; Green Jr, W. H. A priori rate constants for kinetic modeling. *Theor. Chem. Acc.* **2002**, *108*, 187–213.

(37) Gao, C. W.; Allen, J. W.; Green, W. H.; West, R. H. Reaction Mechanism Generator: Automatic construction of chemical kinetic mechanisms. *Comput. Phys. Comm.* **2016**, *203*, 212–225.

(38) Billing, G. D.; Mikkelsen, K. V. *Introduction to Molecular Dynamics and Chemical Kinetics*; Wiley: New York, 1996.

(39) Truhlar, D. G.; Garrett, B. C.; Klippenstein, S. J. Current Status of Transition-State Theory. *J. Phys. Chem.* **1996**, *100*, 12771–12800.

(40) Steinfeld, J. I.; Francisco, J. S.; Hase, W. L. *Chemical Kinetics and Dynamics*; Prentice Hall: Upper Saddle River, NJ, 1998.

(41) Levine, R. D. *Molecular Reaction Dynamics*; Cambridge University Press: Cambridge UK, 2009.

(42) Dengler, A.; Fontain, E.; Knauer, M.; Stein, N.; Ugi, I. Competing concepts in CAOS. *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 262–269.

(43) Ugi, I., Ed. *Computer Chemistry*; Topics in Current Chemistry; Springer: Berlin, Heidelberg, 1993; Vol. 166.

(44) Hoffmann, R. W. *Elements of Synthesis Planning*; Springer: Berlin, Heidelberg, 2009.

(45) Fischer, T. H.; Almlöf, J. E. General methods for geometry and wave function optimization. *J. Phys. Chem.* **1992**, *96*, 9768–9774.

(46) Wales, D. J. *Energy Landscapes*; Cambridge University Press: Cambridge UK, 2003.

(47) Rappoport, D. Colibri is your lightweight and gregarious chemistry explorer, v.0.9.7, `https://bitbucket.org/rappoport/colibri`. 2017.

(48) Stewart, J. J. P. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.* **2013**, *19*, 1–32.

(49) Klamt, A.; Schüürmann, G. COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799–805.

(50) Bell, R. P. *The Proton in Chemistry*; Chapman & Hall: London, 1973.

(51) Toullec, J. In *Enols*; Gold, V., Bethell, D., Eds.; Academic Press: London, 1982; Vol. 18; Chapter 1, pp 1–77.

(52) Keeffe, J. R.; Kresge, A. J. In *Enols*; Rappoport, Z., Ed.; Wiley: Chichester UK, 1990; Chapter 7, pp 399–480.

(53) Apeloig, Y. In *Enols*; Rappoport, Z., Ed.; Wiley: Chichester UK, 1990; Chapter 1, pp 1–74.

(54) Woodward, R. B.; Hoffmann, R. Selection Rules for Sigmatropic Reactions. *J. Am. Chem. Soc.* **1965**, *87*, 2511–2513.

(55) Woodward, R. B.; Hoffmann, R. The Conservation of Orbital Symmetry. *Angew. Chem. Int. Ed.* **1969**, *8*, 781–853.

(56) Hoffmann, R.; Woodward, R. B. Conservation of orbital symmetry. *Acc. Chem. Res.* **1968**, *1*, 17–22.

(57) Bouma, W. J.; Vincent, M. A.; Radom, L. Ab initio molecular orbital studies of sigmatropic rearrangements. *Int. J. Quantum Chem.* **1978**, *14*, 767–777.

(58) Chuang, C.-H.; Lien, M.-H. Computational Study on the Effects of Substituents and Functional Groups in the Isomerization of 1- and 2-Substituted Propenes, Acetaldimines, and Aldehydes. *Eur. J. Org. Chem.* **2004**, *2004*, 1432–1443.

(59) Chiang, Y.; Hojatti, M.; Keeffe, J. R.; Kresge, A. J.; Schepp, N. P.; Wirz, J. Vinyl alcohol. Generation and decay kinetics in aqueous solution and determination of the tautomerization equilibrium constant and acid dissociation constants of the aldehyde and enol forms. *J. Am. Chem. Soc.* **1987**, *109*, 4000–4009.

(60) Guthrie, J. P. In *Enols*; Rappoport, Z., Ed.; Wiley: Chichester UK, 1990; Chapter 2, pp 75–93.

(61) Capon, B.; Zucco, C. Simple enols. 2. Kinetics and mechanism of the ketonization of vinyl alcohol. *J. Am. Chem. Soc.* **1982**, *104*, 7567–7572.

(62) González-Rivas, N.; Cedillo, A. Solvent effects on the energetic parameters and chemical reactivity in the keto–enol tautomeric equilibrium of substituted carbonyl compounds. *Comp. Theor. Chem.* **2012**, *994*, 47–53.

(63) Jencks, W. P. In *Mechanism and Catalysis of Simple Carbonyl Group Reactions*; Cohen, S. G., Streitwieser Jr., A., Taft, R. W., Eds.; Interscience: New York, 1964; Vol. 2; Chapter 2, pp 63–128.

(64) Ogata, Y.; Kawasaki, A. In *The Carbonyl Group*; Zabicky, J., Ed.; Wiley: Chichester UK, 1970; Vol. 2; Chapter 1, pp 1–69.

(65) Bell, R. P. In *The Reversible Hydration of Carbonyl Compounds*; Gold, V., Ed.; Academic Press: London, 1966; Vol. 4; Chapter 1, pp 1–29.

(66) Winkelman, J. G. M.; Voorwinde, O. K.; Ottens, M.; Beenackers, A. A. C. M.; Janssen, L. P. B. M. Kinetics and chemical equilibrium of the hydration of formaldehyde. *Chem. Eng. Sci.* **2002**, *57*, 4067–4076.

(67) Williams, I. H.; Maggiora, G. M.; Schowen, R. L. Theoretical models for mechanism and catalysis in carbonyl addition. *J. Am. Chem. Soc.* **1980**, *102*, 7831–7839.

(68) Spangler, D.; Williams, I. H.; Maggiora, G. M. Determination and characterization of a transition structure for water–formaldehyde addition. *J. Comput. Chem.* **1983**, *4*, 524–541.

(69) Ignatyev, I.; Montejo, M.; Ortega, P. G. R.; González, J. J. L. Effect of substituents and hydrogen bonding on barrier heights in dehydration reactions of carbon and silicon geminal diols. *Phys. Chem. Chem. Phys.* **2011**, *13*, 18507–18515.

(70) Wolfe, S.; Kim, C.-K.; Yang, K.; Weinberg, N.; Shi, Z. Hydration of the Carbonyl Group. A Theoretical Study of the Cooperative Mechanism. *J. Am. Chem. Soc.* **1995**, *117*, 4240–4260.

(71) Funderburk, L. H.; Aldwin, L.; Jencks, W. P. Mechanisms of general acid and base catalysis of the reactions of water and alcohols with formaldehyde. *J. Am. Chem. Soc.* **1978**, *100*, 5444–5459.

(72) Streitwieser, A. *Solvolytic Displacement Reactions*; McGraw-Hill: New York, 1962.

(73) Bunton, C. A. *Nucleophilic Substitution at a Saturated Carbon Atom*; Elsevier: Amsterdam, 1963.

(74) Ingold, C. K. *Structure and Mechanism in Organic Chemistry*, 2nd ed.; Cornell University Press: Ithaca NY, 1969.

(75) de la Mare, P. B. D.; Swedlund, B. E. In *The Carbon–Halogen Bond*; Patai, S., Ed.; Wiley: Chichester UK, 1973; Chapter 7, pp 407–548.

(76) Hartshorn, S. R. *Aliphatic Nucleophilic Substitution*; Cambridge University Press: Cambridge UK, 1973.

(77) Smith, M. B. *March's Advanced Organic Chemistry. Reactions, Mechanisms, and Structure*, 7th ed.; Wiley: Hoboken NJ, 2013.

(78) Gold, V. Border-line mechanisms in nucleophilic displacement reactions. *J. Chem. Soc.* **1956**, 4633–4637.

(79) Jencks, W. P. When is an intermediate not an intermediate? Enforced mechanisms of general acid-base, catalyzed, carbocation, carbanion, and ligand exchange reaction. *Acc. Chem. Res.* **1980**, *13*, 161–169.

(80) Winstein, S.; Appel, B.; Baker, R.; Diaz, A. Ion pairs in solvolysis and exchange. *Chem. Soc. (London), Spec. Publ.* **1965**, *19*, 109–30.

(81) Sneen, R. A. Substitution at a saturated carbon atom. XVII. Organic ion pairs as intermediates in nucleophilic substitution and elimination reactions. *Acc. Chem. Res.* **1973**, *6*, 46–53.

(82) Harris, J. M. In *Progress in Physical Organic Chemistry*; Streitwieser Jr., A., Taft, R. W., Eds.; Interscience: New York, 1974; Vol. 11; Chapter 2, pp 89–173.

(83) Amyes, T. L.; Toteva, M. M.; Richard, J. P. In *Reactive Intermediate Chemistry*; Moss, R. A., Platz, M. S., Jones, M., Eds.; Wiley: Hoboken NJ, 2005; Chapter 2, pp 41–68.

(84) McLennan, D. J. A case for the concerted SN2 mechanism of nucleophilic aliphatic substitution. *Acc. Chem. Res.* **1976**, *9*, 281–287.

(85) Bentley, T. W.; v R Schleyer, P. The SN2–SN1 spectrum. 1. Role of nucleophilic solvent assistance and nucleophilically solvated ion pair intermediates in solvolyses of primary and secondary arenesulfonates. *J. Am. Chem. Soc.* **1976**, *98*, 7658–7666.

(86) Bentley, T. W.; von Ragué Schleyer, P. In *Advances in Physical Organic Chemistry*; Gold, V., Bethell, D., Eds.; Academic Press: London, 1977; Vol. 14; Chapter 1, pp 1–67.

(87) Brauman, J. I.; Olmstead, W. N.; Lieder, C. A. Gas-phase nucleophilic displacement reactions. *J. Am. Chem. Soc.* **1974**, *96*, 4030–4031.

(88) Olmstead, W. N.; Brauman, J. I. Gas-phase nucleophilic displacement reactions. *J. Am. Chem. Soc.* **1977**, *99*, 4219–4228.

(89) Riveros, J. M.; José, S. M.; Takashima, K. In *Gas-phase Nucleophilic Displacement Reactions*; Gold, V., Bethell, D., Eds.; Academic Press: London, 1985; Vol. 21; Chapter 4, pp 197–240.

(90) Laerdahl, J. K.; Uggerud, E. Gas phase nucleophilic substitution. *Int. J. Mass Spectrom.* **2002**, *214*, 277–314.

(91) Chabinyc, M. L.; Craig, S. L.; Regan, C. K.; Brauman, J. I. Gas-Phase Ionic Reactions:

Dynamics and Mechanism of Nucleophilic Displacements. *Science* **1998**, *279*, 1882–1886.

(92) Shaik, S. S.; Schlegel, H. B.; Wolfe, S. *Theoretical Aspects of Physical Organic Chemistry: The SN2 Mechanism*; Wiley: New York, 1992.

(93) Parthiban, S.; de Oliveira, G.; Martin, J. M. L. Benchmark ab Initio Energy Profiles for the Gas-Phase SN2 Reactions Y- + $CH_3X \rightarrow CH_3Y + X^-$ (X,Y = F,Cl,Br). Validation of Hybrid DFT Methods. *J. Phys. Chem. A* **2001**, *105*, 895–904.

(94) Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. SN2 reaction profiles in the gas phase and aqueous solution. *J. Am. Chem. Soc.* **1984**, *106*, 3049–3050.

(95) Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. Theoretical examination of the SN2 reaction involving chloride ion and methyl chloride in the gas phase and aqueous solution. *J. Am. Chem. Soc.* **1985**, *107*, 154–163.

(96) Abraham, M. H.; McLennan, D. J. Comments on the putative ion pair mechanism for hydrolysis of methyl halides and methyl perchlorate. *J. Chem. Soc., Perkin Trans. 2* **1977**, 873–879.

(97) Abraham, M. H. Thermodynamic parameters for ionisation and dissociation of alkyl halides in water and nonaqueous solvents. Comments on the ion-pair mechanism of nucleophilic substitution. *J. Chem. Soc., Perkin Trans. 2* **1973**, 1893–1899.

(98) Fahey, R. C. In *Topics in Stereochemistry*; Allinger, N. L., Eliel, E. L., Eds.; Wiley: Hoboken NJ, 1968; Vol. 3; Chapter 4, pp 237–342.

(99) Sergeev, G. B.; Serguchev, Y. A.; Smirnov, V. V. Molecular Complexes in the Liquid-Phase Halogenation of Unsaturated Compounds. *Russ. Chem. Rev.* **1973**, *42*, 697–712.

(100) Freeman, F. Possible criteria for distinguishing between cyclic and acyclic activated

complexes and among cyclic activated complexes in addition reactions. *Chem. Rev.* **1975**, *75*, 439–490.

(101) de la Mare, P. B. D.; Bolton, R. *Electrophilic Additions to Unsaturated Systems*; Elsevier: Amsterdam, 1966.

(102) de la Mare, P. B. D. *Electrophilic Halogenation: Reaction Pathways Involving Attack by Electrophilic Halogens on Unsaturated Compounds*; Cambridge University Press: London, 1976.

(103) Schmid, G. H.; Garratt, D. G. In *Double-Bonded Functional Groups*; Patai, S., Ed.; Wiley: Chichester UK, 1977; Vol. 2; Chapter 9, pp 725–912.

(104) de la Mare, P. B. D.; Bolton, R. *Electrophilic Additions to Unsaturated Systems*, 2nd ed.; Elsevier: Amsterdam, 1982.

(105) Schmid, G. H. In *Double-Bonded Functional Groups*; Patai, S., Ed.; Wiley: Chichester UK, 1989; Vol. 2, Part 1; Chapter 11, pp 679–731.

(106) Ruasse, M.-F. In *Electrophilic Bromination of Carbon–Carbon Double Bonds: Structure, Solvent and Mechanism*; Bethell, D., Ed.; Academic Press: London, 1993; Vol. 28; Chapter 5, pp 207–291.

(107) V'yunov, K. A.; Ginak, A. I. The Mechanism of the Electrophilic Addition of Halogens to a Multiple Bond. *Russ. Chem. Rev.* **1981**, *50*, 151–163.

(108) Islam, S. M.; Poirier, R. A. New Insights into the Bromination Reaction for a Series of Alkenes A Computational Study. *J. Phys. Chem. A* **2007**, *111*, 13218–13232.

(109) Cammi, R.; Mennucci, B.; Pomelli, C.; Cappelli, C.; Corni, S.; Frediani, L.; Trucks, G. W.; Frisch, M. J. Second-order Møller–Plesset second derivatives for the polarizable continuum model: theoretical bases and application to solvent effects in electrophilic bromination of ethylene. *Theor. Chem. Acc.* **2004**, *111*, 66–77.

(110) Fukuzumi, S.; Kochi, J. K. Transition-state barrier for electrophilic reactions. Solvation of charge-transfer ion pairs as the unifying factor in alkene addition and aromatic substitution with bromine. *J. Am. Chem. Soc.* **1982**, *104*, 7599–7609.

(111) Bellucci, G.; Bianchini, R.; Ambrosetti, R. Direct evidence for bromine-olefin charge-transfer complexes as essential intermediates of the fast ionic addition of bromine to cyclohexene. *J. Am. Chem. Soc.* **1985**, *107*, 2464–2471.

(112) Lenoir, D.; Chiappe, C. What is the Nature of the First-Formed Intermediates in the Electrophilic Halogenation of Alkenes, Alkynes, and Allenes? *Chem. Eur. J.* **2003**, *9*, 1036–1044.

(113) Hamilton, T. P.; Schaefer, H. F. Structure and energetics of C2H4Br+: ethylene-bromonium ion vs. bromoethyl cations. *J. Am. Chem. Soc.* **1990**, *112*, 8260–8265.

(114) Bolton, R. In *Addition and Elimination Reactions of Aliphatic Compounds*; Bamford, C. H., Tipper, C. F. H., Eds.; Comprehensive Chemical Kinetics; Elsevier: Amsterdam, 1973; Vol. 9; Chapter 1, pp 1–86.

(115) Benson, S. W.; Bose, A. N. Structural Aspects of the Kinetics of Four-Center Reactions in the Vapor Phase. *J. Chem. Phys.* **1963**, *39*, 3463–3473.

(116) Sergeev, G. B.; Stepanov, N. F.; Leensov, I. A.; Smirnov, V. V.; Pupyshev, V. I.; Tyurina, L. A.; Mashyanov, M. N. Molecular mechanism of hydrogen bromide addition to olefins. *Tetrahedron* **1982**, *38*, 2585–2589.

(117) Sergeev, G. B.; Smirnov, V. V.; Rostovshchikova, T. N. Hydrochlorination of Unsaturated Compounds. *Russ. Chem. Rev.* **1983**, *52*, 259–274.

(118) Yang, Z.-Z.; Ding, Y.-L.; Zhao, D.-X. Insight into Markovnikov Reactions of Alkenes in Terms of Ab Initio and Molecular Face Theory. *ChemPhysChem* **2008**, *9*, 2379–2389.

(119) Bagno, A.; Modena, G. Solvent Effect on the Protonation of Acetylene and Ethylene–Continuum Solvent Quantum Chemical Calculations. *Eur. J. Org. Chem.* **1999**, *1999*, 2893–2897.

(120) Hughes, E. D.; Ingold, C. K. The mechanism and kinetics of elimination reactions. *Trans. Faraday Soc.* **1941**, *37*, 657–685.

(121) Bunnett, J. F. In *Survey of Progress in Chemistry*; Scott, A. F., Ed.; Academic Press: New York, 1969; Vol. 5; Chapter 2, pp 53–93.

(122) Maccoll, A. In *Advances in Physical Organic Chemistry*; Gold, V., Ed.; Academic Press: London, 1965; Vol. 3; Chapter 2, pp 91–122.

(123) Maccoll, A. Heterolysis and the Pyrolysis of Alkyl Halides in the Gas Phase. *Chem. Rev.* **1969**, *69*, 33–60.

(124) O'Ferrall, R. A. M. In *The Carbon–Halogen Bond*; Patai, S., Ed.; Wiley: Chichester UK, 1973; Chapter 9, pp 609–675.

(125) Cockerill, A. F. In *Addition and Elimination Reactions of Aliphatic Compounds*; Bamford, C. H., Tipper, C. F. H., Eds.; Comprehensive Chemical Kinetics; Elsevier: Amsterdam, 1973; Vol. 9; Chapter 3, pp 163–372.

(126) Cockerill, A. F.; Harrison, R. G. In *Double-Bonded Functional Groups: Part 1*; Patai, S., Ed.; Wiley: Chichester UK, 1977; Chapter 4, pp 149–329.

(127) Saunders, W. H.; Cockerill, A. F. *Mechanisms of Elimination Reactions*; Wiley: New York, 1973.

(128) Baciocchi, E. In *Halides, Pseudo-Halides and Azides: Part 2*; Patai, S., Rappoport, Z., Eds.; Wiley: Chichester UK, 1983; Chapter 23, pp 1173–1227.

(129) Gandler, J. R. In *Double-Bonded Functional Groups*; Patai, S., Ed.; Wiley: Chichester UK, 1989; Vol. 2 Part 1; Chapter 12, pp 733–797.

(130) Benson, S. W.; Haugen, G. R. A Simple, Self-Consistent Electrostatic Model for Quantitative Prediction of the Activation Energies of Four-Center Reactions. *J. Am. Chem. Soc.* **1965**, *87*, 4036–4044.

(131) Guthrie, J. P. Concertedness and E2 elimination reactions: prediction of transition state position and reaction rates using two-dimensional reaction surfaces based on quadratic and quartic approximations. *Can. J. Chem.* **1990**, *68*, 1643–1652.

(132) Guthrie, J. P.; Leandro, L.; Pitchko, V. Elimination reactions—Calculation of rate constants from equilibrium constants and distortion energies by means of no barrier theory. *Can. J. Chem.* **2005**, *83*, 1654–1666.

(133) de la Mare, P. B. D.; Ridd, J. H. *Aromatic Substitution. Nitration and Halogenation*; Butterworth: London, 1959.

(134) Berliner, E. In *Progress in Physical Organic Chemistry*; Streitwieser Jr., A., Taft, R. W., Eds.; Wiley: Hoboken NJ, 1964; Chapter 5, pp 253–321.

(135) Norman, R. O. C.; Taylor, R. *Electrophilic Substitution in Benzenoid Compounds*; Elsevier: Amsterdam, 1965.

(136) Stock, L. M. *Aromatic Substitution Reactions*; Prentice–Hall: Englewood Cliffs NJ, 1968.

(137) Taylor, R. In *Comprehensive Chemical Kinetics*; Bamford, C. H., Tipper, C. F. H., Eds.; Elsevier: Amsterdam, 1972; Vol. 13; Chapter 1, pp 1–406.

(138) Taylor, R. *Electrophilic Aromatic Substitution*; Wiley: Chichester UK, 1990.

(139) Kong, J.; Galabov, B.; Koleva, G.; Zou, J.-J.; Schaefer, H. F.; von Ragué Schleyer, P. The Inherent Competition between Addition and Substitution Reactions of Br2 with Benzene and Arenes. *Angew. Chem. Int. Ed.* **2011**, *50*, 6809–6813.

(140) Galabov, B.; Koleva, G.; Simova, S.; Hadjieva, B.; Schaefer, H. F.; von Ragué Schleyer, P. Arenium ions are not obligatory intermediates in electrophilic aromatic substitution. *Proc. Natl. Acad. Sci.* **2014**, *111*, 10067–10072.

(141) Galabov, B.; Nalbantova, D.; von Ragué Schleyer, P.; Schaefer, H. F. Electrophilic Aromatic Substitution: New Insights into an Old Class of Reactions. *Acc. Chem. Res.* **2016**, *49*, 1191–1199.

(142) Brouwer, D. M.; Mackor, E. L.; MacLean, C. In *Carbonium Ions*; Olah, G. A., von Ragué Schleyer, P., Eds.; Interscience: New York, 1970; Vol. 2; Chapter 20, pp 837–898.

(143) Dewar, M. J. S.; Marchand, A. P. Physical Organic Chemistry: π-Complexes as Intermediates in Organic Reactions. *Annu. Rev. Phys. Chem.* **1965**, *16*, 321–346.

(144) Banthorpe, D. V. π Complexes as reaction intermediates. *Chem. Rev.* **1970**, *70*, 295–322.

(145) Smith, W. B. Ab initio studies of the bromination of benzene. *J. Phys. Org. Chem.* **2003**, *16*, 34–39.

(146) de la Mare, P. B. D. Pathways in electrophilic aromatic substitutions. Cyclohexadienes and related compounds as intermediates in halogenation. *Acc. Chem. Res.* **1974**, *7*, 361–368.

(147) Parker, R. E.; Isaacs, N. S. Mechanisms of Epoxide Reactions. *Chem. Rev.* **1959**, *59*, 737–799.

(148) Whalen, D. L. In *Advances in Physical Organic Chemistry*; Richard, J. P., Ed.; Elsevier: Amsterdam, 2005; Vol. 40; Chapter 6, pp 247–298.

(149) Long, F. A.; Pritchard, J. G. Hydrolysis of Substituted Ethylene Oxides in H_2O^{18} Solutions. *J. Am. Chem. Soc.* **1956**, *78*, 2663–2667.

(150) Lundin, A.; Panas, I.; Ahlberg, E. A Mechanistic Investigation of Ethylene Oxide Hydrolysis to Ethanediol. *J. Phys. Chem. A* **2007**, *111*, 9087–9092.

(151) Pritchard, J. G.; Long, F. A. Kinetics and Mechanism of the Acid-catalyzed Hydrolysis of Substituted Ethylene Oxides. *J. Am. Chem. Soc.* **1956**, *78*, 2667–2670.

(152) Long, F. A.; Pritchard, J. G.; Stafford, F. E. Entropies of Activation and Mechanism for the Acid-Catalyzed Hydrolysis of Ethylene Oxide and its Derivatives. *J. Am. Chem. Soc.* **1957**, *79*, 2362–2364.

(153) Ford, G. P.; Smith, C. T. Gas-phase hydrolysis of protonated oxirane. Ab initio and semiempirical molecular orbital calculations. *J. Am. Chem. Soc.* **1987**, *109*, 1325–1331.

(154) *Ullmann's Encyclopedia of Industrial Chemistry*, 7th ed.; Wiley-VCH: Weinheim Germany, 2011.

(155) Bender, M. L. Mechanisms of Catalysis of Nucleophilic Reactions of Carboxylic Acid Derivatives. *Chem. Rev.* **1960**, *60*, 53–113.

(156) Johnson, S. L. In *Advances in Physical Organic Chemistry*; Gold, V., Ed.; Academic Press: London, 1967; Vol. 5; Chapter 5, pp 237–330.

(157) Euranto, E. K. In *Carboxylic Acids and Esters*; Patai, S., Ed.; Wiley: Chichester UK, 1969; Chapter 11, pp 505–588.

(158) Jencks, W. P. *Catalysis in Chemistry and Enzymology*; McGraw-Hill: New York, 1969.

(159) Kirby, A. J. In *Comprehensive Chemical Kinetics*; Bamford, C. H., Tipper, C. F. H., Eds.; Elsevier: Amsterdam, 1972; Vol. 10; Chapter 2, pp 57–207.

(160) Talbot, R. J. E. In *Comprehensive Chemical Kinetics*; Bamford, C. H., Tipper, C. F. H., Eds.; Elsevier: Amsterdam, 1972; Vol. 10; Chapter 3, pp 209–293.

(161) Day, J. N. E.; Ingold, C. K. Mechanism and kinetics of carboxylic ester hydrolysis and carboxyl esterification. *Trans. Faraday Soc.* **1941**, *37*, 686–705.

(162) Bender, M. L. Oxygen Exchange as Evidence for the Existence of an Intermediate in Ester Hydrolysis. *J. Am. Chem. Soc.* **1951**, *73*, 1626–1629.

(163) Davies, A. G.; Kenyon, J. Alkyl–oxygen heterolysis in carboxylic esters and related compounds. *Q. Rev., Chem. Soc.* **1955**, *9*, 203–228.

(164) Guthrie, J. P. Hydration of carboxylic acids and esters. Evaluation of the free energy change for addition of water to acetic and formic acids and their methyl esters. *J. Am. Chem. Soc.* **1973**, *95*, 6999–7003.

(165) Schaleger, L. L.; Long, F. A. In *Advances in Physical Organic Chemistry*; Gold, V., Ed.; Academic Press: London, 1963; Vol. 1; Chapter 1, pp 1–33.

(166) Hæffner, F.; Hu, C.-H.; Brinck, T.; Norin, T. The catalytic effect of water in basic hydrolysis of methyl acetate: a theoretical study. *J. Mol. Struct. (THEOCHEM)* **1999**, *459*, 85–93.

(167) Zhan, C.-G.; Landry, D. W.; Ornstein, R. L. Theoretical Studies of Fundamental Pathways for Alkaline Hydrolysis of Carboxylic Acid Esters in Gas Phase. *J. Am. Chem. Soc.* **2000**, *122*, 1522–1530.

(168) Zhan, C.-G.; Landry, D. W.; Ornstein, R. L. Reaction Pathways and Energy Barriers for Alkaline Hydrolysis of Carboxylic Acid Esters in Water Studied by a Hybrid Supermolecule-Polarizable Continuum Approach. *J. Am. Chem. Soc.* **2000**, *122*, 2621–2627.

(169) Pliego Jr, J. R.; Riveros, J. M. A Theoretical Analysis of the Free-Energy Profile of the Different Pathways in the Alkaline Hydrolysis of Methyl Formate In Aqueous Solution. *Chem. Eur. J.* **2002**, *8*, 1945–1953.

(170) Takano, Y.; Houk, K. N. Benchmarking the Conductor-like Polarizable Continuum Model (CPCM) for Aqueous Solvation Free Energies of Neutral and Ionic Organic Molecules. *J. Chem. Theory Comput.* **2005**, *1*, 70–77.

(171) Hori, K.; Ikenaga, Y.; Arata, K.; Takahashi, T.; Kasai, K.; Noguchi, Y.; Sumimoto, M.; Yamamoto, H. Theoretical study on the reaction mechanism for the hydrolysis of esters and amides under acidic conditions. *Tetrahedron* **2007**, *63*, 1264–1269.

(172) Tarbell, D. S. The Claisen Rearrangement. *Chem. Rev.* **1940**, *27*, 495–546.

(173) Tarbell, D. S. In *Organic Reactions*; Adams, R., Ed.; Wiley: Hoboken NJ, 1944; Chapter 1, pp 1–48.

(174) Rhoads, S. J. In *Molecular Rearrangements*; de Mayo, P., Ed.; Interscience: New York, 1963; Vol. 1; Chapter 11, pp 655–706.

(175) Rhoads, S. J.; Raulins, N. R. In *Organic Reactions*; Dauben, W. G., Ed.; Wiley: Hoboken NJ, 1975; Vol. 22; Chapter 1, pp 1–252.

(176) Ziegler, F. E. The thermal, aliphatic Claisen rearrangement. *Chem. Rev.* **1988**, *88*, 1423–1452.

(177) Ganem, B. The Mechanism of the Claisen Rearrangement: Déjà Vu All Over Again. *Angew. Chem. Int. Ed.* **1996**, *35*, 936–945.

(178) Rehbein, J.; Hiersemann, M. In *The Claisen Rearrangement: Methods and Applications*; Hiersemann, M., Nubbemeyer, U., Eds.; Wiley–VCH: Weinheim Germany, 2007; Chapter 11, pp 525–557.

(179) Hoffmann, R.; Woodward, R. B. Orbital Symmetries and Orientational Effects in a Sigmatropic Reaction. *J. Am. Chem. Soc.* **1965**, *87*, 4389–4390.

(180) Vance, R. L.; Rondan, N. G.; Houk, K. N.; Jensen, F.; Borden, W. T.; Komornicki, A.; Wimmer, E. Transition structures for the Claisen rearrangement. *J. Am. Chem. Soc.* **1988**, *110*, 2314–2315.

(181) Dewar, M. J. S.; Jie, C. Mechanism of the Claisen rearrangement of allyl vinyl ethers. *J. Am. Chem. Soc.* **1989**, *111*, 511–519.

(182) Houk, K. N.; Li, Y.; Evanseck, J. D. Transition Structures of Hydrocarbon Pericyclic Reactions. *Angew. Chem. Int. Ed.* **1992**, *31*, 682–708.

(183) Wiest, O.; Houk, K. N. In *Density Functional Theory IV*; Nalewajski, R. F., Ed.; Topics in Current Chemistry; Springer: Berlin, Heidelberg, 1996; Vol. 183; Chapter 1, pp 1–24.

(184) Schuler, F. W.; Murphy, G. W. The Kinetics of the Rearrangement of Vinyl Allyl Ether. *J. Am. Chem. Soc.* **1950**, *72*, 3155–3159.

(185) Burrows, C. J.; Carpenter, B. K. Substituent effects on the aliphatic Claisen rearrangement. 1. Synthesis and rearrangement of cyano-substituted allyl vinyl ethers. *J. Am. Chem. Soc.* **1981**, *103*, 6983–6984.

(186) Goumans, T. P. M.; Ehlers, A. W.; Lammertsma, K.; Würthwein, E.-U.; Grimme, S. Improved Reaction and Activation Energies of [4+2] Cycloadditions, [3,3] Sigmatropic Rearrangements and Electrocyclizations with the Spin-Component-Scaled MP2 Method. *Chem. Eur. J.* **2004**, *10*, 6468–6475.

(187) Dewar, M. J. S.; Healy, E. F. MNDO study of the Claisen rearrangement. *J. Am. Chem. Soc.* **1984**, *106*, 7127–7131.

(188) Cramer, C. J.; Truhlar, D. G. What causes aqueous acceleration of the Claisen rearrangement? *J. Am. Chem. Soc.* **1992**, *114*, 8794–8799.

(189) Severance, D. L.; Jorgensen, W. L. Effects of hydration on the Claisen rearrangement of allyl vinyl ether from computer simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10966–10968.

(190) Jorgensen, W. L.; Blake, J. F.; Lim, D.; Severance, D. L. Investigation of solvent effects on pericyclic reactions by computer simulations. *Faraday Trans.* **1994**, *90*, 1727–1732.

(191) Davidson, M. M.; Hillier, I. H. Aqueous Acceleration of the Claisen Rearrangement of Allyl Vinyl Ether: A Hybrid, Explicit Solvent, and Continuum Model. *J. Phys. Chem.* **1995**, *99*, 6748–6751.

(192) Gajewski, J. J. The Claisen Rearrangement. Response to Solvents and Substituents: The Case for Both Hydrophobic and Hydrogen Bond Acceleration in Water and for a Variable Transition State. *Acc. Chem. Res.* **1997**, *30*, 219–225.

(193) Meyer, M. P.; DelMonte, A. J.; Singleton, D. A. Reinvestigation of the Isotope Effects for the Claisen and Aromatic Claisen Rearrangements. The Nature of the Claisen Transition States. *J. Am. Chem. Soc.* **1999**, *121*, 10865–10874.

(194) Gajewski, J. J.; Conrad, N. D. Aliphatic Claisen rearrangement transition state structure from secondary α-deuterium isotope effects. *J. Am. Chem. Soc.* **1979**, *101*, 2747–2748.

(195) Woodward, R. B.; Hoffmann, R. Stereochemistry of Electrocyclic Reactions. *J. Am. Chem. Soc.* **1965**, *87*, 395–397.

(196) Longuet-Higgins, H. C.; Abrahamson, E. W. The Electronic Mechanism of Electrocyclic Reactions. *J. Am. Chem. Soc.* **1965**, *87*, 2045–2046.

(197) Miller, S. I. In *Advances in Physical Organic Chemistry*; Gold, V., Ed.; Academic Press: London, 1968; Vol. 6; Chapter 4, pp 185–332.

(198) Marvell, E. N. *Thermal Electrocyclic Reactions*; Academic Press: New York, 1980.

(199) Boyd, G. V. In *The Chemistry of Dienes and Polyenes*; Rappoport, Z., Ed.; Wiley: Chichester UK, 1997; Vol. 1; Chapter 11, pp 507–546.

(200) Bakulev, V. A. 1,6-Electrocyclic reactions. *Russ. Chem. Rev.* **2007**, *64*, 99–124.

(201) Marvell, E. N. The cis-hexatriene electrocyclization: Attempted calculation of the transition state geometry. *Tetrahedron* **1973**, *29*, 3791–3796.

(202) Komornicki, A.; McIver, J. W. Structure of transition states. III. MINDO/2 study of the cyclization of 1,3,5-hexatriene to 1,3-cyclohexadiene. *J. Am. Chem. Soc.* **1974**, *96*, 5798–5800.

(203) Evanseck, J. D.; Thomas IV, B. E.; Spellmeyer, D. C.; Houk, K. N. Transition Structures of Thermally Allowed Disrotatory Electrocyclizations. The Prediction of Stereoselective Substituent Effects in Six-Electron Pericyclic Reactions. *J. Org. Chem.* **1995**, *60*, 7134–7141.

(204) Rodríguez-Otero, J. Study of the Electrocyclization of (Z)-Hexa-1,3,5-triene and Its Heterosubstituted Analogues Based on Ab Initio and DFT Calculations. *J. Org. Chem.* **1999**, *64*, 6842–6848.

(205) Guner, V.; Khuong, K. S.; Leach, A. G.; Lee, P. S.; Bartberger, M. D.; Houk, K. N. A Standard Set of Pericyclic Reactions of Hydrocarbons for the Benchmarking of Computational Methods: The Performance of ab Initio, Density Functional, CASSCF, CASPT2, and CBS-QB3 Methods for the Prediction of Activation Barriers, Reaction Energetics, and Transition State Geometries. *J. Phys. Chem. A* **2003**, *107*, 11445–11459.

(206) Huisgen, R.; Grashey, R.; Sauer, J. In *The Chemistry of Alkenes*; Patai, S., Ed.; Wiley: Chichester UK, 1964; Vol. 1; Chapter 11, pp 739–953.

(207) Wassermann, A. *Diels–Alder reactions: Organic Background and Physicochemical Aspects*; Elsevier: Amsterdam, 1965.

(208) Sauer, J. Diels–Alder reactions II: The Reaction Mechanism. *Angew. Chem. Int. Ed.* **1967**, *6*, 16–33.

(209) Seltzer, S. In *Advances in Alicyclic Chemistry*; Hart, H., Karabatsos, G. J., Eds.; Elsevier: New York, 1968; Vol. 2; Chapter 1, pp 1–57.

(210) Herndon, W. C. Theory of cycloaddition reactions. *Chem. Rev.* **1972**, *72*, 157–179.

(211) Beltrame, P. In *Comprehensive Chemical Kinetics*; Bamford, C. H., Tipper, C. F. H., Eds.; Elsevier: Amsterdam, 1973; Vol. 9; Chapter 2, pp 87–162.

(212) Houk, K. N. *Theoretical and Experimental Insights into Cycloaddition Reactions*; Topics in Current Chemistry; Springer: Berlin, Heidelberg, 1979; Vol. 79; Chapter 1, pp 1–40.

(213) Babichev, S. S.; Kovtunenko, V. A.; Voitenko, Z. V.; Tyltin, A. K. Application of Quantum-chemical Methods to the Diels–Alder Reaction. *Russ. Chem. Rev.* **1988**, *57*, 397–405.

(214) Sauer, J.; Sustmann, R. Mechanistic Aspects of Diels–Alder Reactions: A Critical Survey. *Angew. Chem. Int. Ed.* **1980**, *19*, 779–807.

(215) Bach, R. D.; McDouall, J. J. W.; Schlegel, H. B.; Wolber, G. J. Electronic factors influencing the activation barrier of the Diels–Alder reaction. An ab initio study. *J. Org. Chem.* **1989**, *54*, 2931–2935.

(216) Carruthers, W., Ed. *Cycloaddition Reactions in Organic Synthesis*; Pergamon Press: Oxford UK, 1990.

(217) Beusker, P. H.; Scheeren, H. W. In *The Chemistry of Dienes and Polyenes*; Rappoport, Z., Ed.; Wiley: Chichester UK, 2003; Vol. 2; pp 329–479.

(218) Hoffmann, R.; Woodward, R. B. Selection Rules for Concerted Cycloaddition Reactions. *J. Am. Chem. Soc.* **1965**, *87*, 2046–2048.

(219) Townshend, R. E.; Ramunni, G.; Segal, G.; Hehre, W. J.; Salem, L. Organic transition states. V. The Diels–Alder reaction. *J. Am. Chem. Soc.* **1976**, *98*, 2190–2198.

(220) Houk, K. N.; Lin, Y. T.; Brown, F. K. Evidence for the concerted mechanism of the Diels–Alder reaction of butadiene with ethylene. *J. Am. Chem. Soc.* **1986**, *108*, 554–556.

(221) Borden, W. T.; Loncharich, R. J.; Houk, K. N. Synchronicity in Multibond Reactions. *Annu. Rev. Phys. Chem.* **1988**, *39*, 213–236.

(222) Houk, K. N.; Gonzalez, J.; Li, Y. Pericyclic Reaction Transition States: Passions and Punctilios, 1935-1995. *Acc. Chem. Res.* **1995**, *28*, 81–90.

(223) Goldstein, E.; Beno, B.; Houk, K. N. Density Functional Theory Prediction of the Relative Energies and Isotope Effects for the Concerted and Stepwise Mechanisms of the Diels–Alder Reaction of Butadiene and Ethylene. *J. Am. Chem. Soc.* **1996**, *118*, 6036–6043.

(224) Wiest, O.; Montiel, D. C.; Houk, K. N. Quantum Mechanical Methods and the Interpretation and Prediction of Pericyclic Reaction Mechanisms. *J. Phys. Chem. A* **1997**, *101*, 8378–8388.

(225) Barone, V.; Arnaud, R. Study of prototypical Diels–Alder reactions by a hybrid density functional/Hartree–Fock approach. *Chem. Phys. Lett.* **1996**, *251*, 393–399.

(226) Barone, V.; Arnaud, R. Diels–Alder reactions: An assessment of quantum chemical procedures. *J Chem Phys* **1997**, *106*, 8727–8732.

(227) Rowley, D.; Steiner, H. Kinetics of diene reactions at high temperatures. *Discuss. Faraday Soc.* **1951**, *10*, 198.

(228) Lischka, H.; Ventura, E.; Dallos, M. The Diels–Alder Reaction of Ethene and 1,3-Butadiene: An Extended Multireference ab initio Investigation. *ChemPhysChem* **2004**, *5*, 1365–1371.

(229) Karton, A.; Goerigk, L. Accurate reaction barrier heights of pericyclic reactions: Surprisingly large deviations for the CBS-QB3 composite method and their consequences in DFT benchmark studies. *J. Comput. Chem.* **2015**, *36*, 622–632.

(230) Hoffmann, H. M. R. The Ene Reaction. *Angew. Chem. Int. Ed.* **1969**, *8*, 556–577.

(231) Boyd, G. V. In *Double-Bonded Functional Groups*; Patai, S., Ed.; Wiley: Chichester UK, 1989; Vol. 2 Part 1; Chapter 8, pp 477–525.

(232) Richard, C.; Back, M. H. Ene reactions of olefins, Part II. The addition of ethylene to propylene and to isobutene and the addition of propylene to propylene. *Int. J. Chem. Kinet.* **1978**, *10*, 389–405.

(233) Pranata, J. The ene reaction: Comparison of results of Hartree–Fock, Møller–Plesset, CASSCF, and DFT calculations. *Int. J. Quantum Chem.* **1997**, *62*, 509–514.

(234) Deng, Q.; Thomas, B. E.; Houk, K. N.; Dowd, P. Transition Structures of the Ene Reactions of Cyclopropene. *J. Am. Chem. Soc.* **1997**, *119*, 6902–6908.

(235) Loncharich, R. J.; Houk, K. N. Transition structures of ene reactions of ethylene and formaldehyde with propene. *J. Am. Chem. Soc.* **1987**, *109*, 6947–6952.

(236) Critchlow, J. E. Prediction of transition state configuration in concerted reactions from the energy requirements of the separate processes. *J. Chem. Soc., Faraday Trans. 1* **1972**, *68*, 1774–19.

(237) Jencks, W. P. Ingold Lecture. How does a reaction choose its mechanism? *Chem. Soc. Rev.* **1981**, *10*, 345.

(238) Fukui, K. The Path of Chemical Reactions–The IRC Approach. *Acc. Chem. Res.* **1981**, *14*, 363–368.

(239) Carbó-Dorca, R., Gironés, X., Mezey, P. G., Eds. *Fundamentals of Molecular Similarity*; Springer: New York, 2001.

(240) Maggiora, G. M.; Shanmugasundaram, V. In *Molecular Modeling of Proteins*; Bajorath, J., Ed.; Humana Press: Totowa NJ, 2011; Chapter 2, pp 39–100.

(241) Kleinberg, J. M. Navigation in a Small World. *Nature* **2000**, *406*, 845–845.

(242) Edelkamp, S.; Schroedl, S. *Heuristic Search. Theory and Applications*; Morgan Kaufmann: Waltham MA, 2012.

(243) Ahuja, R. K.; Magnanti, T. L.; Orlin, J. B. *Network Flows: Theory, Algorithms, and Applications*; Prentice Hall: Englewood Cliffs NJ, 1993.

(244) Newman, M. E. J. *Networks: An Introduction*; Oxford University Press: Oxford UK, 2009.

(245) Jasrasaria, D.; Pyzer-Knapp, E. O.; Rappoport, D.; Aspuru-Guzik, A. Space-Filling Curves as a Novel Crystal Structure Representation for Machine Learning Models. arXiv:1608.05747. arXiv.org e-Print archive, 2016.

(246) Faulon, J.-L., Bender, A., Eds. *Handbook of Chemoinformatics Algorithms*; Chapman & Hall/CRC: Boca Raton FL, 2010.

(247) Breiman, L.; Friedman, J. H.; Olshen, R. A.; Stone, C. J. *Classification And Regression Trees*; Wadsworth & Brooks/Cole: Monterey CA, 1984.

(248) Pólya, G.; Read, R. C. *Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds*; Springer: New York, 1987.

Supporting Information Available

Distributed Implementation of the HAQC Procedure; Reaction Network Models; Heuristic Kinetic Feasibility Criteria.

Supplementary Information

Predicting Feasible Organic Reaction Pathways Using Heuristically-Aided Quantum Chemistry

Dmitrij Rappoport[*] and Alán Aspuru-Guzik[*]

Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA

E-mail: dmitrij@rappoport.org; aspuru@chemistry.harvard.edu

S1   Distributed Implementation of the HAQC Procedure

The HAQC approach as implemented in the prototype Python code named Colibri is briefly described in this section.[1] The Colibri code comprises a suite of independent processes that operate in parallel and interact through a common MongoDB[2] document store for the Molecule and Flask objects. (Fig. S1) Each Molecule object is behaves like a finite-state machine with the following states: FORMULA has only the structural formula of the molecule, represented by a canonical SMILES string;[3,4] CONFIGURATION additionally contains three-dimensional atomic coordinates from an empirical structure builder or force-field optimization; and GEOMETRY is obtained as a result of a quantum chemical structure optimization and includes an associated molecular energy. Similarly, the Flask object has the following states: INITIAL holds only the references to the constituent Molecule objects; the Flask object is converted to RUNNING status after submitting the Molecule objects

for structure optimization; upon successfully collecting the molecular energies and evaluating the total flask energy, the Flask object becomes REACTIVE and reaches the terminal UNREACTIVE state when the elementary reactive transformations are applied. Each state transition is implemented as a separate continuously running task, which executes whenever documents with the matching state are available in the data store and waits for data otherwise. Synchronization between concurrent tasks is implemented using locking. The main advantage of this microservice architecture compared to a monolithic application is that the individual tasks remain independent and may be launched or shut down based on data volume. Moreover, scaling out to a distributed system and fine-grained process control are straightforward in the microservice architecture.

The most basic Colibri configuration comprises five types of tasks, not including data import and export, which are denoted by solid arrows in Fig. S1. BuildTask generates a three-dimensional molecular model from the structural formula using the RDKit library[5] (FORMULA → CONFIGURATION), and the MOPAC semiempirical quantum chemistry program[6] is utilized in GeometryTask for structure optimizations (CONFIGURATION → GEOMETRY) for the purposes of this work. However, other structure builders and quantum chemistry codes are readily integrated with the Colibri implementation. The FlaskMapperTask and FlaskReducerTask implement the asynchronous coupling between the Flask and Molecule objects (empty arrows in Fig. S1). The FlaskMapperTask inserts new Molecule objects into the data store based on the INITIAL Flask objects. Its counterpart is the FlaskReducerTask, which waits for the molecular energies of all constituent molecules in a RUNNING Flask to become available and subsequently computes the total flask energy. Finally, the FlaskReactionTask generates a number of new children Flask objects (each having INITIAL state) from a REACTIVE flask according to elementary transformation rules and inserts them into the data store along the parent Flask object, which changes its state to UNREACTIVE.

Once initialized by one or more starting Flask objects with INITIAL state, the Colibri

program executes autonomously and iteratively constructs the reaction network in the TN representation until all Flask objects in the data store reach their terminal states or until it is manually stopped. In addition to the normal UNREACTIVE state, additional terminal states UNAVAILABLE, ERROR, and INVALID indicate various error conditions. The complete network is then exported to a Neo4J graph store[7] or as a GraphML graph file[8,9] for analysis and visualization. The Colibri source code is released under the permissive Apache 2 license along with this paper.[1]
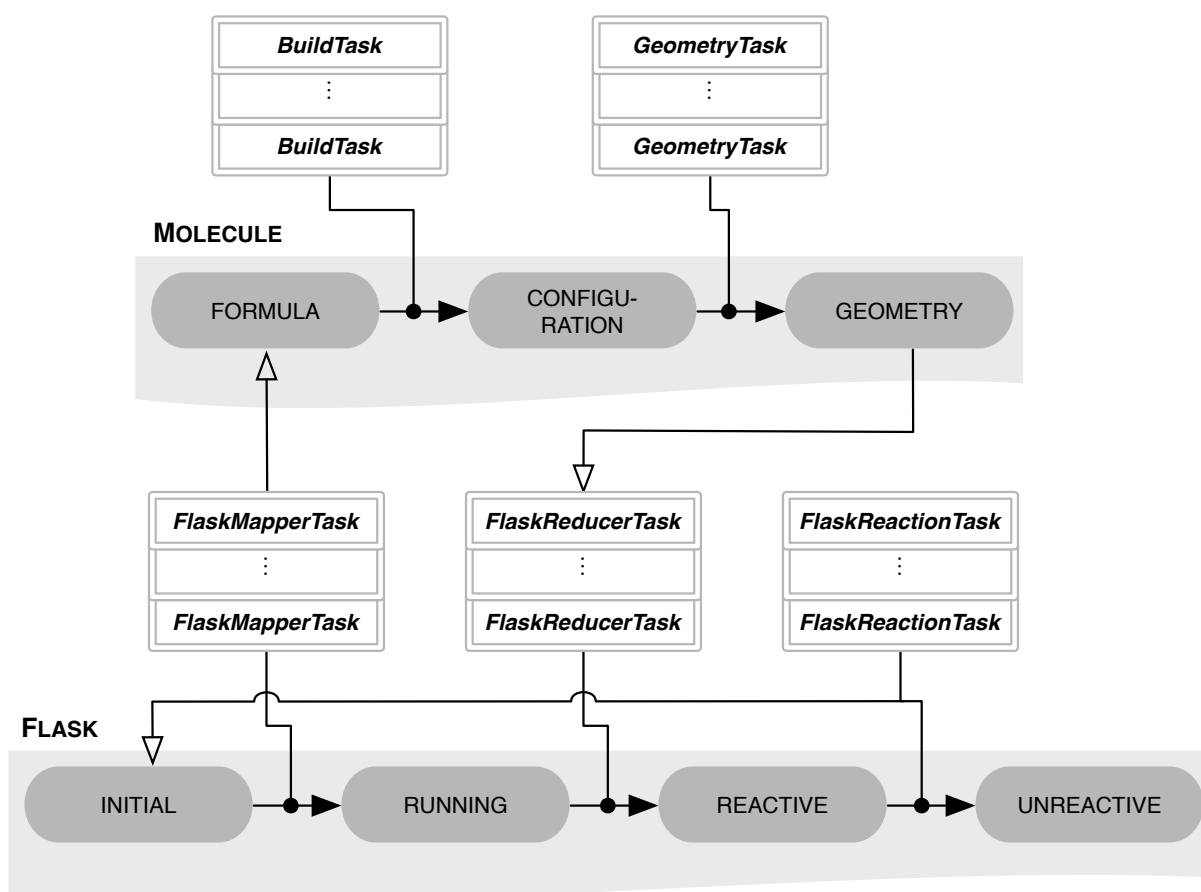
Figure S1: Simplified scheme of the Colibri code. The Molecule and Flask objects are shown hatched, their states are in dark gray. Tasks executing in parallel are drawn in white. State changes are indicated by solid arrows, additional data streams are shown by empty arrows. Data import, export, synchronization, process control, and error handling are omitted for clarity.

S2   Heuristic Kinetic Feasibility Criteria

We show the performance of arc and karc heuristic kinetic feasibility criteria with the parameters $\alpha = 5, 10$ eV and $\kappa = 1, 5, 10$ eV^{-2}. See Section 4 of the main text for definitions and discussion.
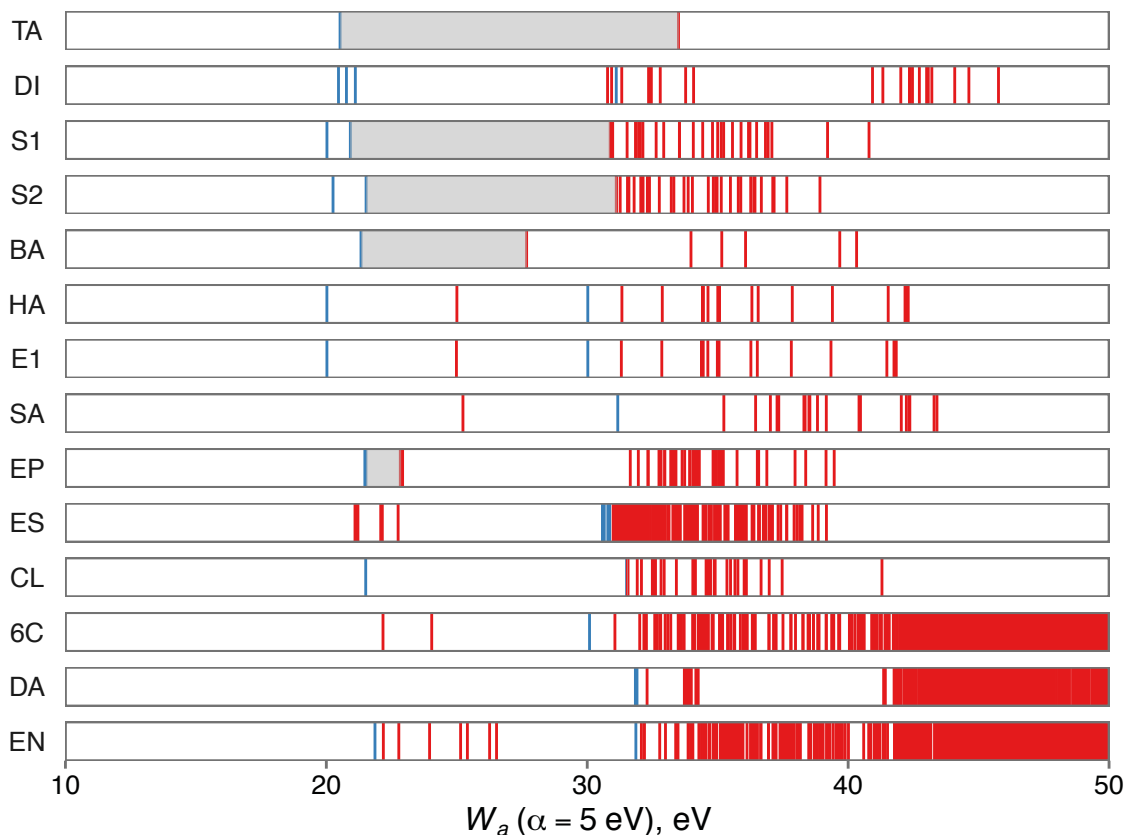
Figure S2: Classification of reaction pathways as feasible/infeasible using the arc heuristic with $\alpha = 1$ eV for simple reaction mechanisms. Empirically feasible reaction pathways are marked by blue lines, infeasible pathways are marked by red lines. The interval between the feasible reaction pathway with the highest W_a value and the infeasible reaction pathways with the lowest W_a value (feasibility gap) is highlighted by gray shading.
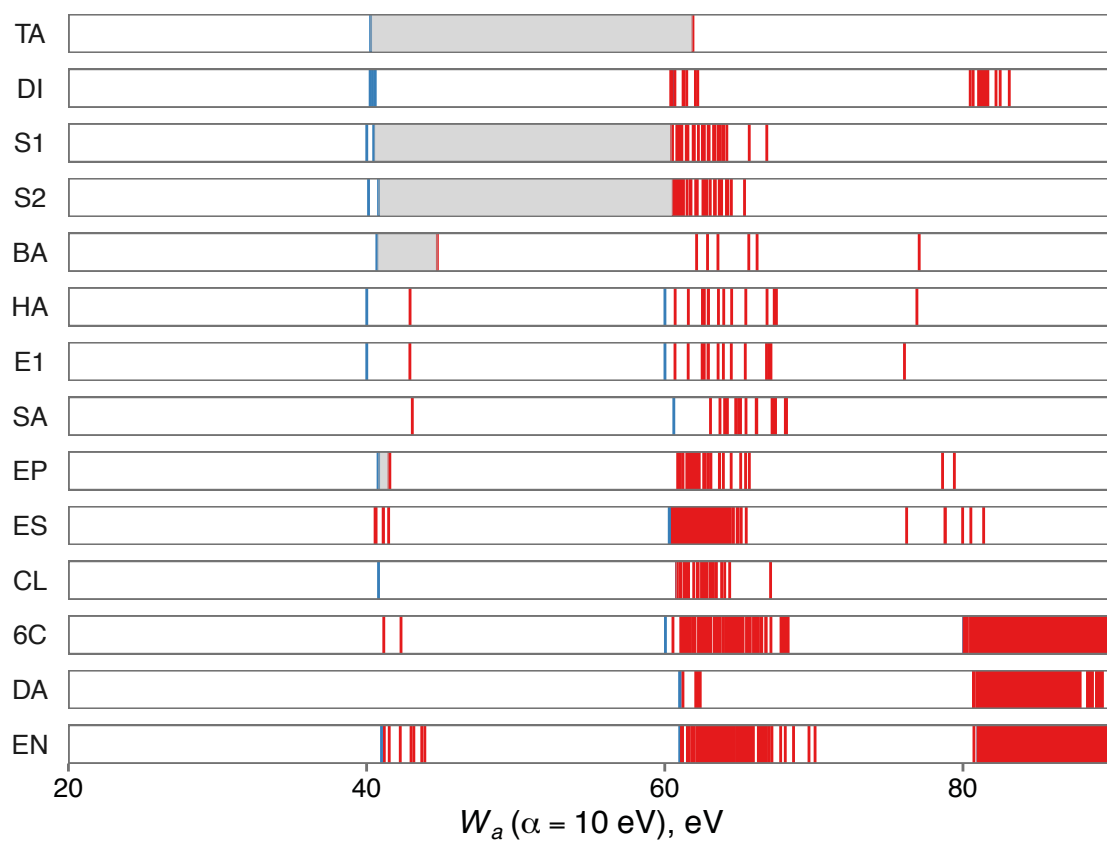
Figure S3: Classification of reaction pathways as feasible/infeasible using the arc heuristic with $\alpha = 10$ eV for simple reaction mechanisms. See Fig. S2 for details.
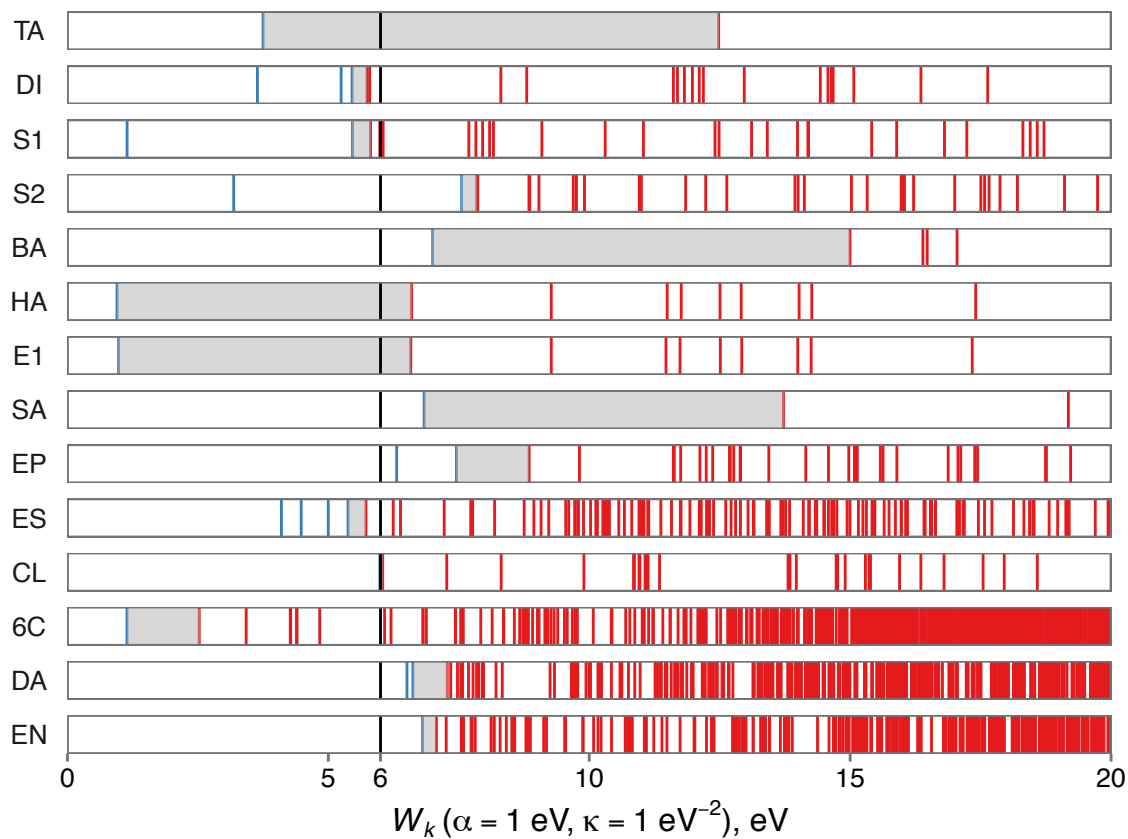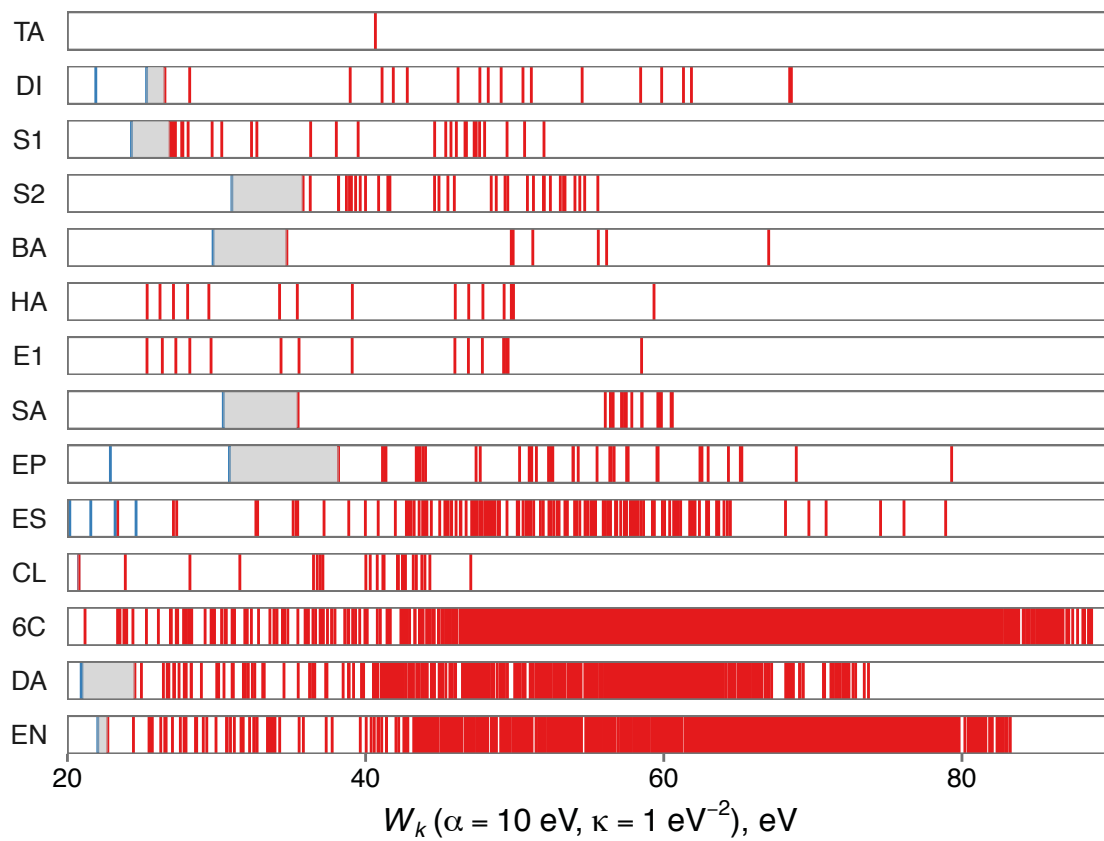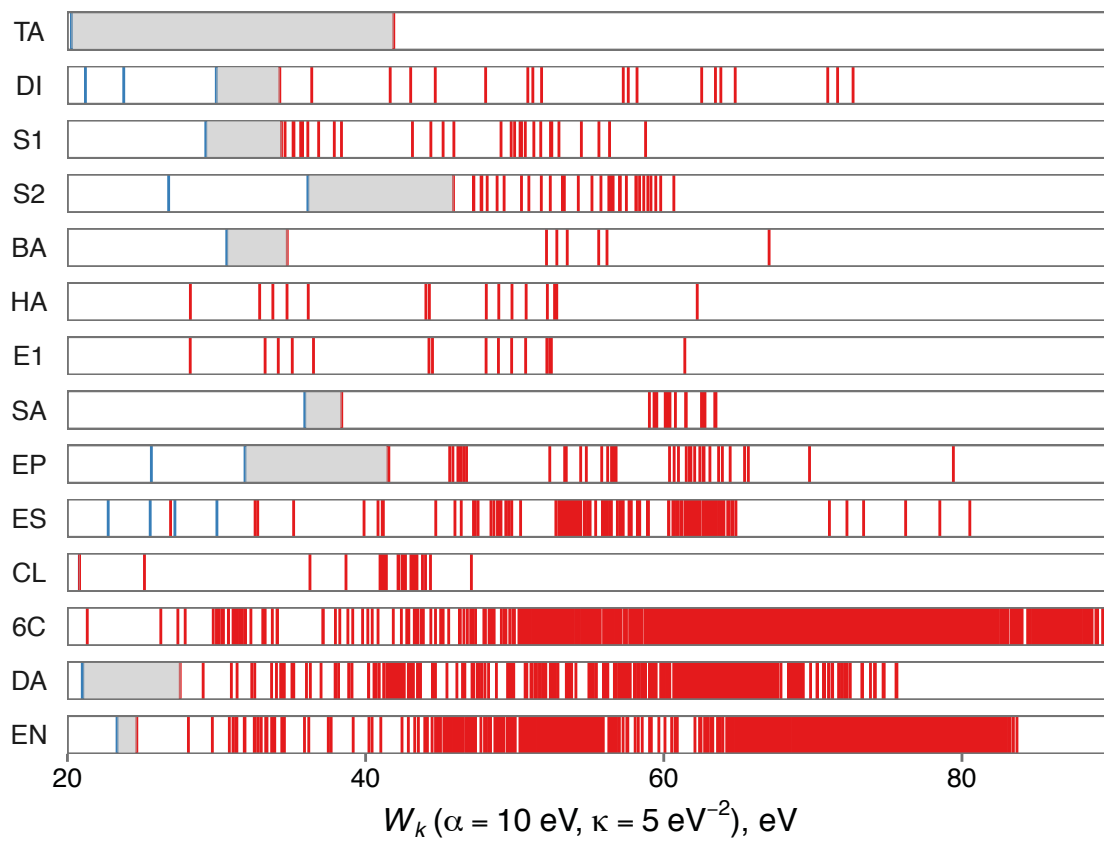
Figure S4: Classification of reaction pathways as feasible/infeasible using the karc heuristic with $\alpha = 1$ eV, $\kappa = 1$ eV^{-2} for simple reaction mechanisms. $\overline{W_k} = 6$ eV estimates the consensus threshold for binary feasible/infeasible classification See Fig. S2 for additional details.
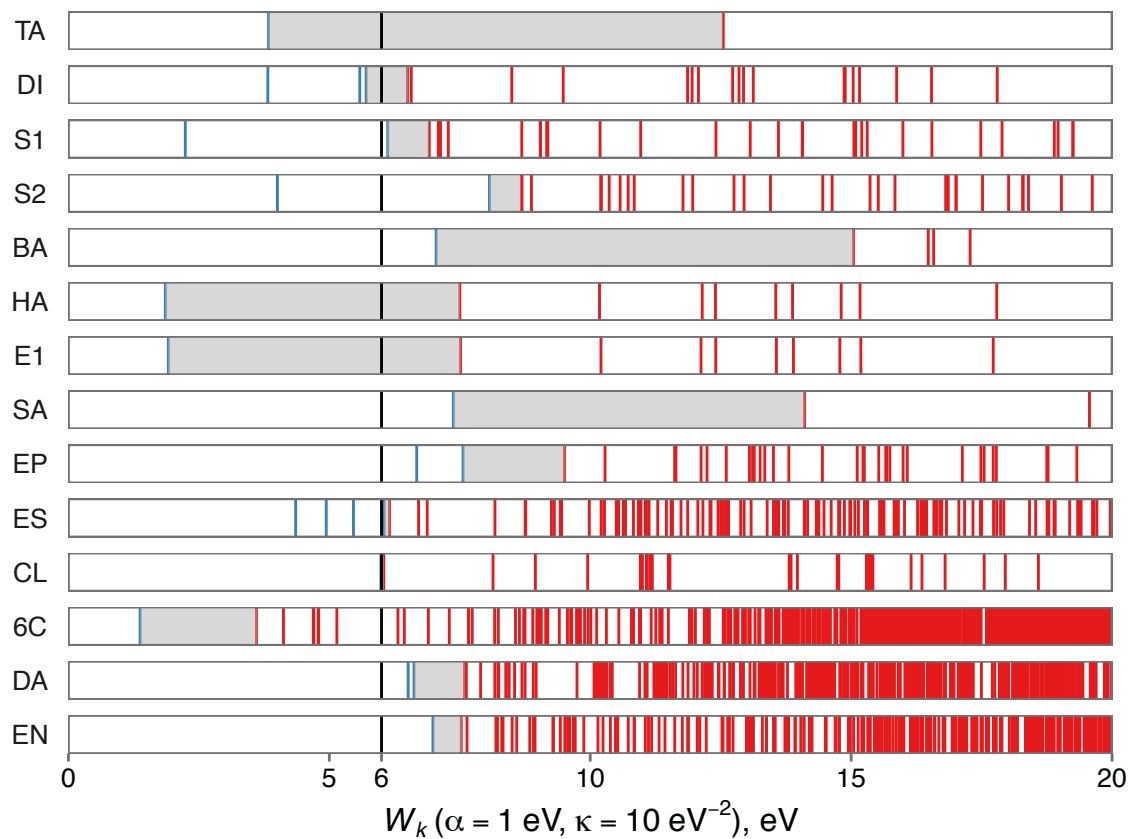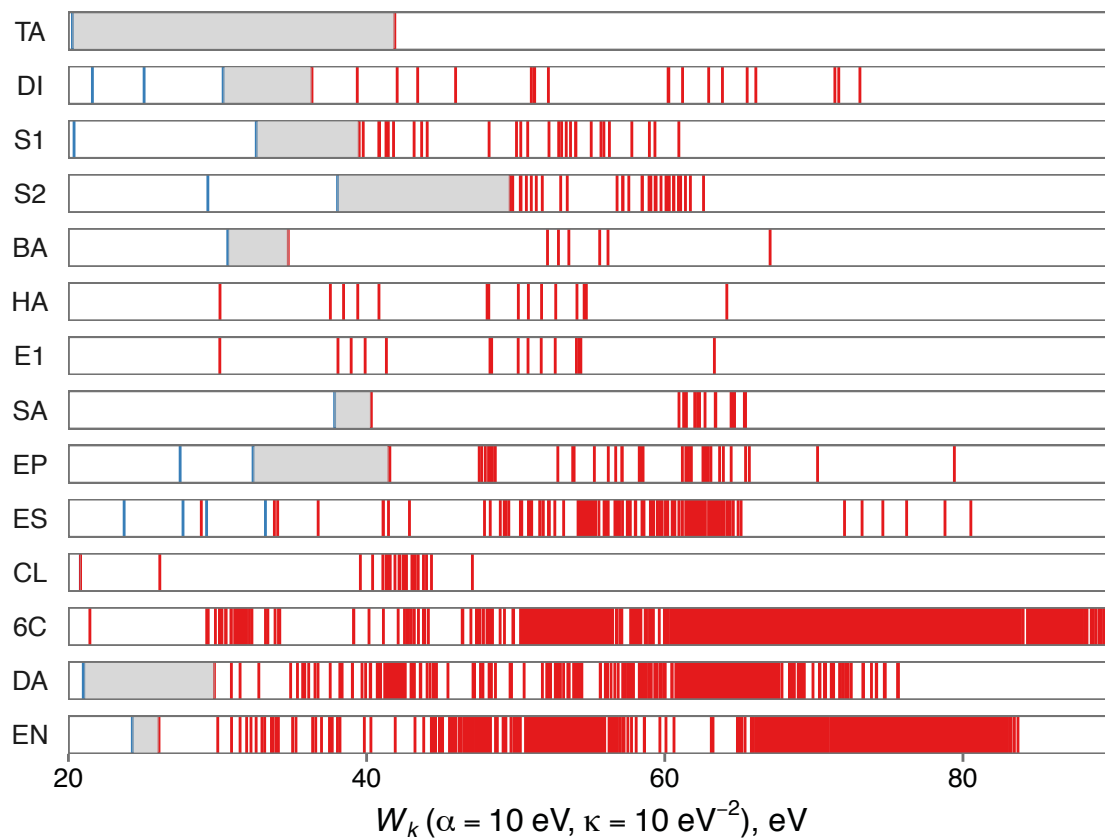
Figure S5: Classification of reaction pathways as feasible/infeasible using the karc heuristic with $\alpha = 1$ eV, $\kappa = 10$ eV^{-2} for simple reaction mechanisms. $\overline{W_k} = 6$ eV estimates the consensus threshold for binary feasible/infeasible classification See Fig. S2 for additional details.
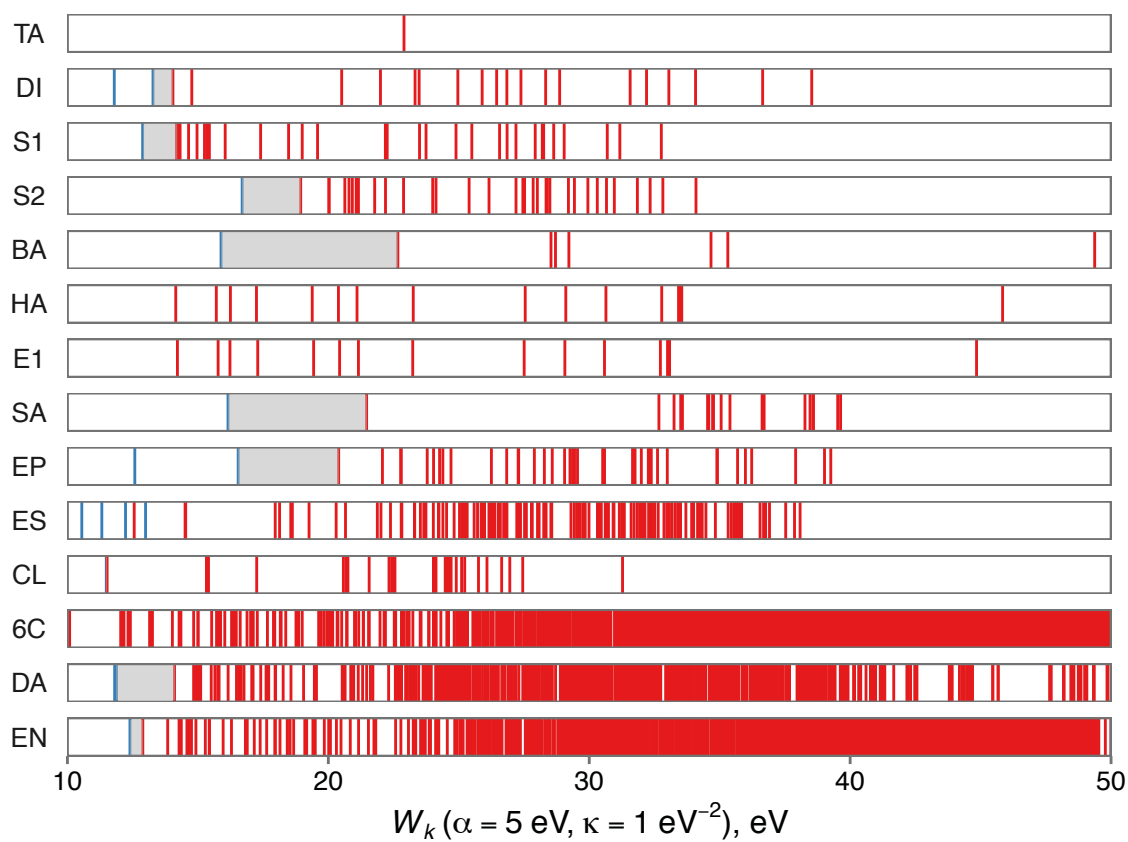
Figure S6: Classification of reaction pathways as feasible/infeasible using the karc heuristic with $\alpha = 5$ eV, $\kappa = 1$ eV^{-2} for simple reaction mechanisms. See Fig. S2 for additional details.
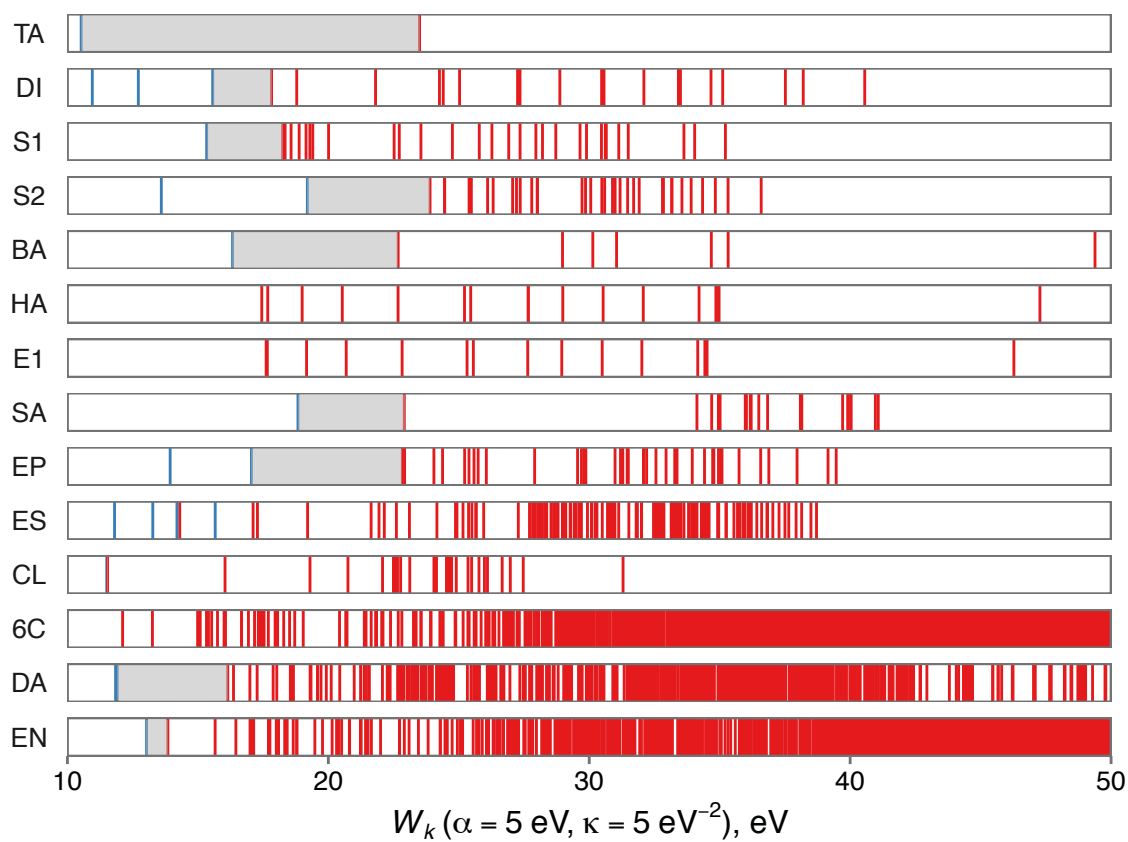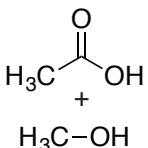
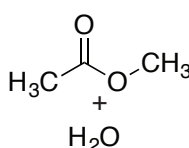Figure S7: Classification of reaction pathways as feasible/infeasible using the karc heuristic with $\alpha = 5$ eV, $\kappa = 5$ eV^{-2} for simple reaction mechanisms. See Fig. S2 for additional details.
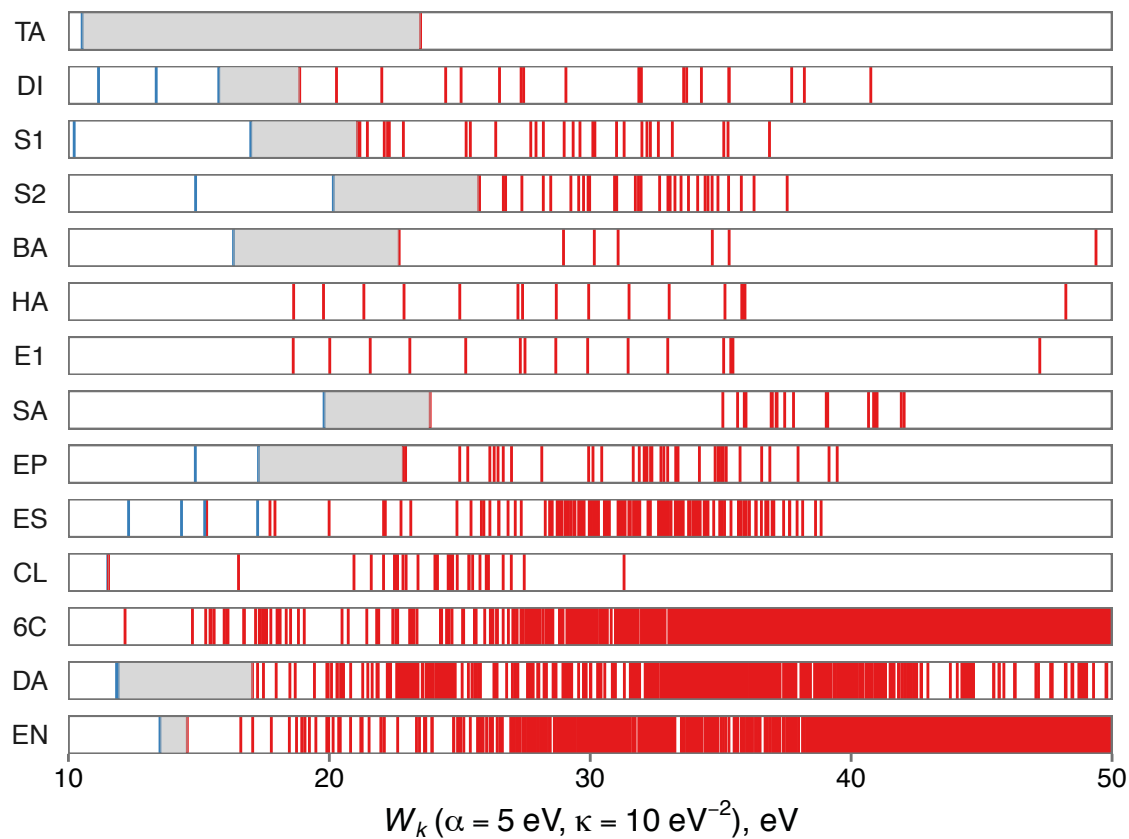
Figure S8: Classification of reaction pathways as feasible/infeasible using the karc heuristic with $\alpha = 5$ eV, $\kappa = 10$ eV^{-2} for simple reaction mechanisms. See Fig. S2 for additional details.
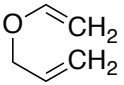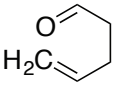
Figure S9: Classification of reaction pathways as feasible/infeasible using the karc heuristic with $\alpha = 10$ eV, $\kappa = 1$ eV^{-2} for simple reaction mechanisms. See Fig. S2 for additional details.

Figure S10: Classification of reaction pathways as feasible/infeasible using the karc heuristic with $\alpha = 10$ eV, $\kappa = 5$ eV^{-2} for simple reaction mechanisms. See Fig. S2 for additional details.

Figure S11: Classification of reaction pathways as feasible/infeasible using the karc heuristic with $\alpha = 10$ eV, $\kappa = 10$ eV^{-2} for simple reaction mechanisms. See Fig. S2 for additional details.
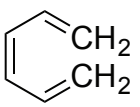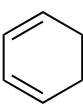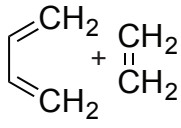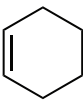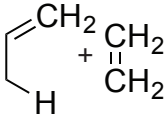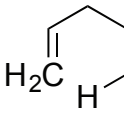
S3 Reaction Network Models

All reaction networks were constructed using the Colibri[1] program. Reactive transformations and 3D structure generation used the RDKit[5] library. Quantum chemical structure optimizations were carried out by the MOPAC program[6] at the PM7 semiempirical level using the COSMO implicit solvation model for water ($\varepsilon = 78.4$).[10] The energy of the proton in solution was computed for pH = 7. The network construction procedure was executed until all Flasks were in the UNREACTIVE state or were marked as invalid. The details of the reaction network models are shown in Table S1.

Table S1: Reaction network models generated in this work

	Reactants	Products	Nodes	Edges
Tautomerizaton (TA)	H₂C=CHOH	H₃C—CHO	42	102
Diol formation (DI)	H₂C=O + H₂O	HO–CH₂–OH	18	50
S_N1 substitution (S1)	(CH₃)₃C—Br + H₂O	(CH₃)₃C—OH + HBr	5359	32684
S_N2 substitution (S2)	H₃C–CH₂–Br + H₂O	H₃C–CH₂–OH + HBr	139	572
Br₂ addition (BA)	H₂C=CH₂ + Br—Br	Br–CH₂–CH₂–Br	53	168
HBr addition (HA)	(CH₃)₂C=CH₂ + HBr	(CH₃)₃C—Br	591	2698
$E1$ elimination (E1)	(CH₃)₃C—Br	(CH₃)₂C=CH₂ + HBr	591	2698
S_EAr substitution (SA)	C₆H₆ + Br—Br	C₆H₅Br + HBr	115999	570169
Epoxide hydrolysis (EP)	epoxide + H₂O	HO–CH₂–CH₂–OH	193	730

Table S1: Reaction network models generated in this work (*cont.*)

	Reactants	Products	Nodes	Edges
Esterification (ES)			6532	36239
Claisen rearrangement (CL)*			37504	264522
6π electrocyclization (6C)*			14860	100506
Diels–Alder reaction (DA)*			11114	78432
Ene reaction (EN)*			1033	5786

* Up to 4 atom charges allowed

References

(1) Rappoport, D. Colibri is your lightweight and gregarious chemistry explorer, v.0.9.7, `https://bitbucket.org/rappoport/colibri`. 2017.

(2) MongoDB, Inc., MongoDB, version 3.0. 2016.

(3) Weininger, D.; Weininger, A.; Weininger, J. L. SMILES. 2. Algorithm for generation of unique SMILES notation. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 97–101.

(4) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36.

(5) Landrum, G. RDKit: Open-Source Cheminformatics, version 2015.09.2.

(6) Stewart, J. J. P. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.* **2013**, *19*, 1–32.

(7) Neo Technology Inc., Neo4J, version 3.2. 2016.

(8) GraphML Working Group, GraphML Specification.

(9) Brandes, U.; Eiglsperger, M.; Herman, I.; Himsolt, M.; Marshall, M. S. In *GraphML Progress Report Structural Layer Proposal*; Mutzel, P., Jünger, M., Leipert, S., Eds.; Springer: Berlin Heidelberg, 2002; pp 501–512.

(10) Klamt, A.; Schüürmann, G. COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799–805.