Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Digital Discovery*, 2026, **5**, 1372

Automated reaction transition state search for bimolecular liquid-phase reactions using internal coordinates: a test case for neutral hydrolysis

Leen Fahoum [a] and Alon Grinberg Dana *[ab]

Transition-state (TS) identification for bimolecular liquid-phase reactions is notoriously sensitive to the initial spatial arrangement of reactants, making automated searches difficult, especially in solvation where conformational effects dominate barrier heights. We address this gap with a fully automated, heuristic framework integrated into the Automated Reaction Calculator (ARC) software tool that generates TS guesses for neutral hydrolysis, as a model reaction type, by positioning water using atom-centered internal-coordinate rules derived from representative DFT-optimized cases. The approach is parameterized for three hydrolysis families: carbonyl-based (esters, amides, and acyl halides), ethers, and nitriles, and operates from reactant/product SMILES alone. Validation across 91 diverse reactions shows that chemically guided internal-coordinate placement yields relatively high success rates under SMD-water conditions: 96.9% for carbonyl-based substrates, 86.2% for nitriles, and 72.4% for ethers, consistent with the greater conformational variability and weaker intrinsic directionality of ether substrates. An ablation study highlights that small, targeted reactant-dihedral adjustments and $\pm\varphi$ sign-sampling are essential to robustly align the water nucleophile, while the electronegativity-based neighbor ranking primarily fine-tunes local orientation. By automating the classically manual step of water placement and orientation, and producing chemically faithful geometry initializations, this framework enables scalable, high-throughput TS searches for neutral hydrolysis reactions. It provides a practical foundation for mechanistic studies and kinetic modeling in condensed-phase organic chemistry. The methodology is readily extensible to additional solvents and catalytic regimes, and to other bimolecular liquid-phase reactions where directed fragment placement is the key bottleneck.

Received 14th November 2025
Accepted 26th February 2026

DOI: 10.1039/d5dd00506j

rsc.li/digitaldiscovery

1. Introduction

Hydrolysis and its reverse reaction, condensation, are fundamental processes in many reactive liquid-phase systems, with relevance for (retro)synthesis[1,2] and degradation.[3–7] Hydrolysis typically involves the cleavage of a chemical bond through the addition of a water molecule, while condensation forms a bond by releasing water. These reactions play a pivotal role in biochemistry, pharmaceuticals, food science, and environmental chemistry, domains that represent the current frontiers in predictive chemical kinetic modeling.[8]

Hydrolysis can proceed under neutral, acidic, or basic conditions, each with distinct mechanistic characteristics. Acidic hydrolysis accelerates bond cleavage by protonating key functional groups, while base-catalyzed hydrolysis improves the nucleophilic attack to facilitate bond scission. Hydrolysis under neutral conditions is generally slower and often requires elevated temperatures.[9,10] For this reason, acid- and base-catalyzed hydrolysis has been studied more extensively. However, neutral hydrolysis remains highly relevant in systems where extreme pH conditions are undesirable or impractical, such as buffered aqueous media, prebiotic chemistry, specific drug formulations, and environmentally mild synthetic protocols. The study of hydrolysis under neutral conditions provides insight into intrinsic reaction pathways free of external catalytic effects, which is essential for understanding the baseline reaction behavior and capturing transient intermediates that may be obscured under catalyzed conditions.

Recent research demonstrates that hydrolysis significantly affects drug stability even under neutral conditions. For instance, hydrolysis of peptide bonds in antibody-based drugs has been observed during long-term storage at neutral pH, leading to structural degradation and potential immunogenicity.[11] In particular, IgG1 antibodies incubated in neutral aqueous media undergo hydrolysis in the hinge region without enzymatic assistance,[11] highlighting that even relatively slow, uncatalyzed neutral hydrolysis reactions can significantly reduce therapeutic effectiveness over time.

[a]*Wolfson Department of Chemical Engineering, Technion – Israel Institute of Technology, Haifa 3200003, Israel. E-mail: alon@technion.ac.il*
[b]*Grand Technion Energy Program (GTEP), Technion – Israel Institute of Technology, Haifa 3200003, Israel*

More broadly, hydrolysis represents a major degradation pathway for pharmaceuticals, affecting the stability and shelf-life of drug products, especially if the active pharmaceutical ingredient (API) contains an ester or an amide group. Drug formulation strategies must consider hydrolytic stability, employing moisture-resistant packaging and appropriate excipients to minimize degradation.[12] Furthermore, hydrolysis reactions play a key role in prodrug activation, where ester hydrolysis is commonly used to release the API at the target site, while the formation of amide bonds is a fundamental step in the construction of peptides and small-molecule drugs.[13]

Understanding hydrolysis in the context of synthesis and degradation is important for modern drug design and synthesis, from the design of stable pharmaceuticals and drug performance optimization to prediction of degradation pathways.[7,14] To address these challenges, predictive chemical kinetic models are essential, as they facilitate our understanding of reaction mechanisms and enable design and optimization of reactive systems. Such models can be instrumental in predicting drug shelf life[7] and validating proposed synthetic pathways.

Refining these quantitative models by computing first-principles rate coefficients requires identification of proper reaction transition state (TS) configurations in 3D space. While conformer search of stable reactants is comparatively routine, even in solution, locating TSs remains the bottleneck for barrier-height estimation, *e.g.*, *via* quantum chemical computations. Neutral hydrolysis reactions present unique computational challenges due to the significant impact of molecular conformations on reaction energetics. Conformational changes in TSs can lead to substantial activation energy variations, with rate coefficient differences of up to 350×.

Traditional string-based methods for TS searches[15–18] are computationally intensive and require significant manual intervention in the initial steps of mapping the reactant and product atoms in 3D and orienting the reacting fragments. Several automated methods have been developed to address these challenges. KinBot,[19] for example, identifies unimolecular gas-phase reaction families and generates TS guesses (TSGs) based on family-specific heuristics, iteratively altering its geometry toward TSGs for reaction classes such as intramolecular hydrogen migration, Diels–Alder, and cycloadditions. AutoTST[20] employs decision trees to suggest TSGs for specific gas-phase reactions by analyzing reactive centers from predefined chemical rules. Graph Convolutional Networks (GCN)[21] apply machine-learning to propose TSGs geometry for isomerization reactions.

Beyond targeted search algorithms, the field of automated chemical discovery has expanded to include global exploration methods that identify novel reaction pathways and products without prior mechanistic knowledge. Approaches such as the *ab initio* nanoreactor[22] and metadynamics-based simulations[23,24] allow for the blind discovery of molecules and mechanisms. Some methods utilize atomic connectivity rules,[25] stochastic surface walking,[26] basin confinement with temperature-accelerated dynamics,[27] or kinetics-guided exploration[28] to build extensive reaction networks and predict reactive events, while others use artificial force-induced reactions (AFIR)[29,30] to map potential energy surfaces. Complementary to these discovery-first methods are emerging path-search tools that leverage machine learning (ML), such as iteratively trained neural network potentials,[31] learned analytical Hessians,[32] and geometric flow matching,[33] to predict TS geometries with reduced computational cost.

The challenge the present work focuses on when searching for a TSG is correctly positioning and orienting two distinct molecular fragments (*e.g.*, the substrate and water) in a complex energetic landscape, a step that is often trivial or non-existent in unimolecular cases. Established methods such as QST2/QST3 (Gaussian),[15] GSM,[16] autodE,[34] and AFIR[29,30] can locate bimolecular TSs but require pre-aligned reactant/product geometries, 3D mapped reactant and product atoms (currently done manually), or parameter tuning. The current methods therefore cannot chemically guide a water molecule placement (Table 1). Despite these advances, to our knowledge no tool automatically generates TS structures for condensed-phase neutral hydrolysis, including chemically informed placement and orientation of the water nucleophile, let alone more complex bimolecular degradation pathways. As a result, computational studies of hydrolysis involving esters or other functional groups[35,36] still

Table 1 Comparison of existing automated TS-search frameworks

Framework	Condensed-phase/ bimolecular	Water placement	User preparation	Main limitation
QST2/QST3 (Gaussian)	Yes	Manual alignment of reactants and products	3D mapped reactant and product geometries required	Needs pre-aligned endpoints
GSM (Growing String Method)	Yes	No defined placement of nucleophile	3D mapped reactant and product geometries required	Computationally intensive
autodE	Yes (PCM/SMD)	Random orientation sampling	Reactant SMILES input; conformers may need review	Limited control of water direction; performs only a 2D atom-mapping
AFIR	Yes	Bias-force placement approach	User defines reactive fragments and bias parameters	May generate non-specific paths
Present work	Yes (SMD)	Chemically guided internal-coordinate placement	Fully automatic (reactants + products SMILES only)	Currently implemented only for neutral hydrolysis

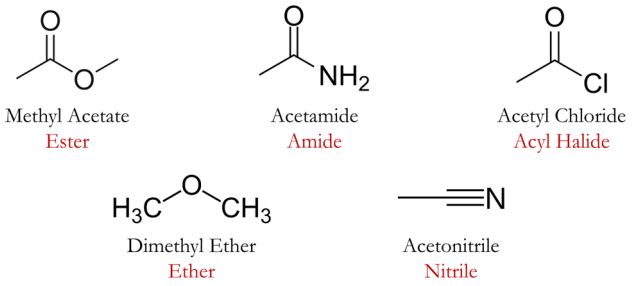

Fig. 1 Example molecules representing the five hydrolysis classes (ester, amide, acyl halide, ether, and nitrile), each belonging to one of the reaction families defined in this work.

depend heavily on manual TS searches for refining predictive chemical kinetic models, often requiring manual expert-guided orientation of the water molecule(s) near the reactive site. This manual trial-and-error workflow limits the scalability of chemical kinetic modeling for liquid-phase synthesis and degradation.

The present work fills this gap by providing a targeted, high-efficiency alternative for established reaction families, specifically addressing the chemically guided orientation of bimolecular fragments in solution, a challenge that remains a bottleneck for scalable kinetic modeling. Here we introduce a heuristic internal-coordinate algorithm, implemented in the Automated Reaction Calculator (ARC) software tool,[37] that automatically positions the water nucleophile for a single-molecule hydrolysis event to generate TSGs for condensed-phase bimolecular reactions using an implicit solvation correction. Our approach, to our knowledge, is the first to automate this process for such systems, using a labeled-graph representation of the reactants to assemble reacting complexes by adding atoms in 3D *via* internal coordinates to yield high-quality initial TSGs suitable for subsequent refinement, *e.g.*, with implicit solvent models. We implement and validate this methodology for neutral hydrolysis, demonstrating its success in orienting water near a variety of reactive sites, which are referred to as classes in this work, including esters, amides, acyl halides, ethers, and nitriles (Fig. 1). By populating

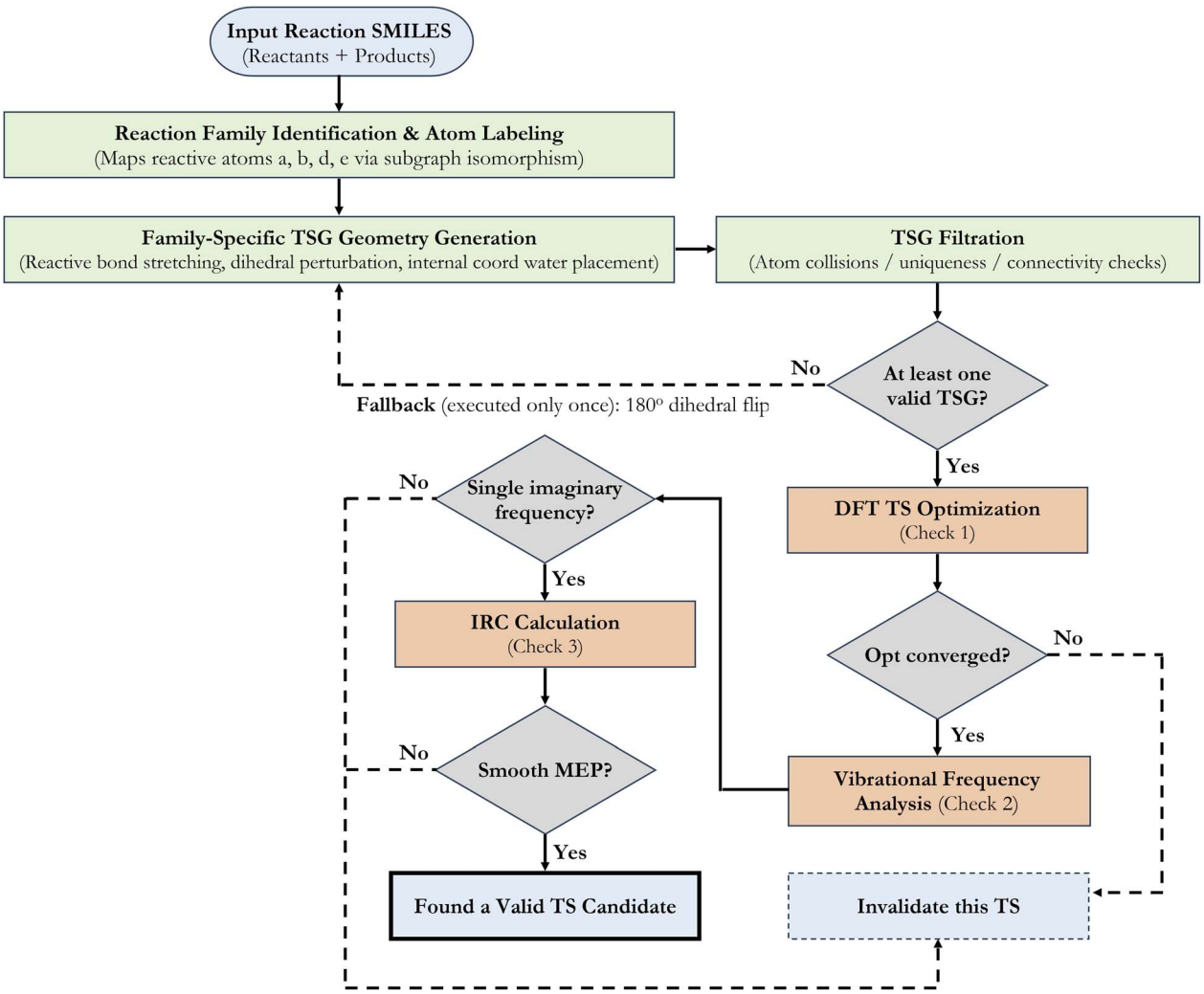

Fig. 2 An integrated workflow scheme for automated TS generation and validation. SMILES: simplified molecular-input line-entry system;[38] TSG: transition state guess; DFT: density functional theory; IRC: intrinsic reaction coordinate; MEP: minimum energy path.

key internal degrees of freedom, the algorithm yields high-quality initializations for TS searches. This establishes a robust and chemically intuitive framework that can be extended to other critical reactions, such as polymer degradation in fuel cells, where general-purpose approaches are insufficient.

2. Methodology

2.1. Integrated workflow overview

Fig. 2 summarizes the workflow developed in this work for automated transition-state guess (TSG) generation and validation for neutral hydrolysis reactions. The workflow consists of two main stages: (i) chemically guided TSG generation and (ii) quantum-mechanical (QM) optimization and validation.

The workflow begins from hydrolysis reaction SMILES (reactants and products). The reaction family is identified, and the reactive atoms (a, b, d, e) are assigned using subgraph isomorphism. An initial reactive geometry is then prepared by applying a family-dependent elongation of the bond undergoing hydrolysis. To capture local conformational variability, multiple reactant geometry variants are generated by systematic dihedral perturbations.

For each reactant geometry variant, a reactive water molecule is introduced using sequential internal-coordinate placement. The oxygen and hydrogen atoms (o, h_1, h_2) are positioned using family-specific distance, bond angle, and dihedral angle definitions designed to pre-organize the nucleophilic attack geometry. The resulting TSGs are filtered to remove nonphysical structures based on atomic collision checks and connectivity validation. If no valid TSG is obtained, a single deterministic fallback step involving a 180° inversion of the reaction-center dihedral angle is applied before repeating the generation cycle.

Once at least one valid TSG is identified, the workflow proceeds to QM optimization and validation. TS optimization is performed at the Density Functional Theory (DFT) level, followed by vibrational frequency analysis to confirm the presence of a single imaginary mode. Intrinsic reaction coordinate (IRC) calculations and connectivity checks are then used to verify that the TS connects the intended reactants and products. Successful completion of these steps yields a validated TS structure.

2.2. Classification of hydrolysis reactions

Reaction classes (Fig. 1) were initially studied *via* manual TS searches followed by DFT optimization to extract reaction class-specific structural parameters. Reaction classes with similar TS geometries and parameter values were grouped into a broader family.

The carbonyl-based hydrolysis family (esters, amides, and acyl halides) proceeds *via* nucleophilic attack at the electrophilic carbonyl carbon (Fig. 3). The ether hydrolysis family (Fig. 4A)[39] serves as a high-barrier model reaction that proceeds negligibly in the absence of acid, base, or catalytic surfaces, due to the strength of the carbon–oxygen bond and the poor nucleophilicity of water.[40,41] The nitrile hydrolysis family follows a three-step mechanism; our work focuses on the rate-determining first step (step a in Fig. 4B), forming an imidic acid intermediate.[42–45] Although nitrile hydrolysis under neutral conditions is slow, experimental studies at elevated temperatures have confirmed that both aliphatic and aromatic nitriles can hydrolyze without acid or base catalysis, following the same stepwise sequence.[44,45]

For multi-functional molecules (Fig. 5), the algorithm identifies all potential reactive sites. By default, it resolves site ambiguity by matching the user-supplied products through subgraph isomorphism, ensuring the water molecule is

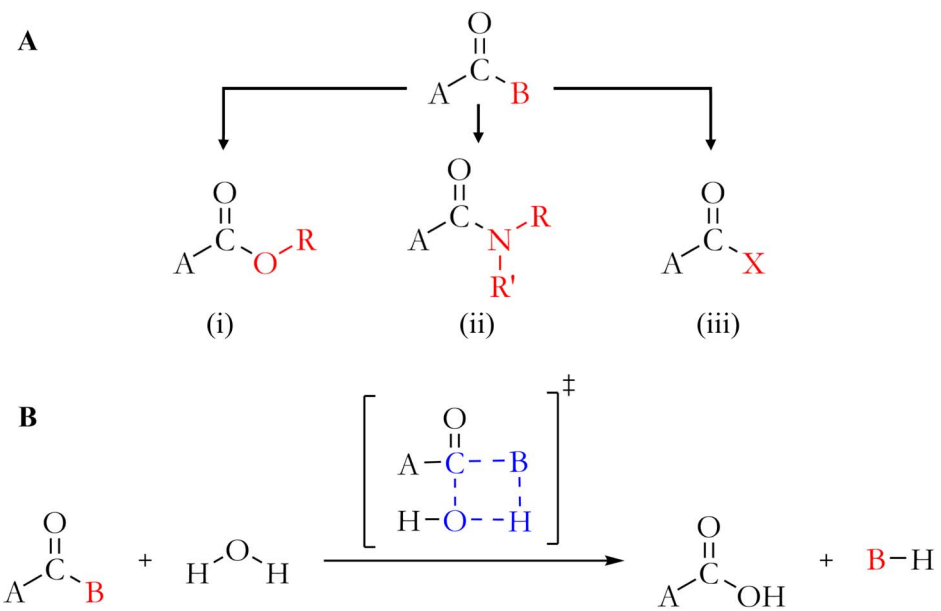

Fig. 3 A general hydrolysis scheme of the carbonyl-based hydrolysis family. (A) The carbonyl-based family classes: (i) esters, (ii) amides, (iii) acyl halides. (B) The carbonyl-based family hydrolysis as modeled here. B represents the electrophilic center substituent (O–R, R–N–R′, or X), X represents a halogen atom (Cl, Br, F). R and R′ denote generic substituents.

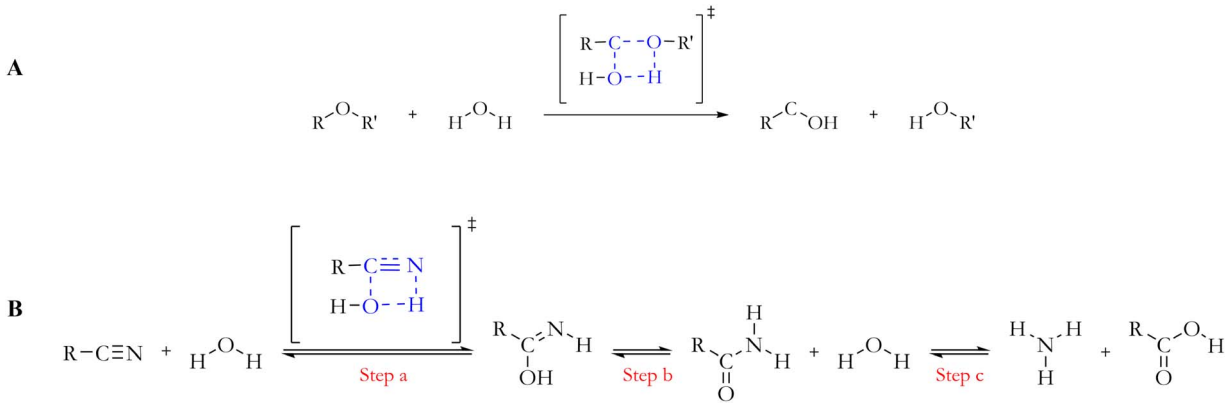

Fig. 4 A general hydrolysis scheme for non-carbonyl-based classes. (A) Ether hydrolysis family as modeled in this work. (B) General nitrile hydrolysis multi-step mechanism, highlighting the first step modeled in this work (step a). R, and R′ denote generic substituents.

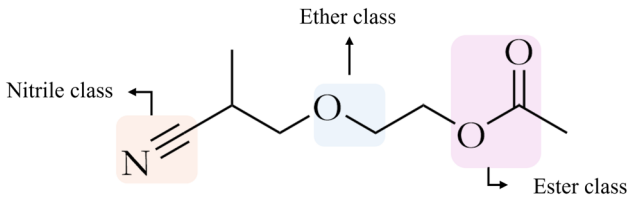

Fig. 5 An example of a molecule containing multiple hydrolysis classes. Note that the ester group is also connected *via* an ether group.

correctly oriented toward the specific bond undergoing cleavage. This mapping procedure resolves potential ambiguities when multiple reactive sites are present, which could lead to distinct chemical products, by comparing the 2D connectivity of the user-specified products with that of all possible products of the bimolecular reaction. This graph-isomorphism-based method results in the identification of the unique bond undergoing cleavage. For instance, in a molecule containing both an ester and an amide group, providing the products of ester cleavage (a carboxylic acid and an alcohol) enables the isomorphism check to identify the unique bond undergoing hydrolysis as the C–O bond in the ester, ensuring that the water molecule is oriented specifically for the ester site rather than the competing amide group. Another possible operational mode is in the context of automated kinetic model generation tools,[46,47] where all possible hydrolysis sites can be explored simultaneously if the user does not specify specific desired products.

2.3. Chemical perception and atom labeling

Following reaction family identification, to generate a TSG geometry the algorithm identifies the key atoms involved in the reaction zone. We defined the following labels for atoms that play a significant role in the hydrolysis reaction zone of the TS, designated as atoms *a*, *b*, *d*, *e* for the molecule undergoing hydrolysis, and *o*, h_1, h_2 for the water molecule (Fig. 6).

Atom *a* represents the site where the nucleophilic attack occurs. For instance, in carbonyl-based and ether hydrolysis, atom *a* is typically the carbon in the carbonyl group or in the C–O bond, respectively. Atom *b* is the atom connected to atom *a* by the bond that is cleaved during the hydrolysis reaction, *e.g.*, in ether hydrolysis it is the oxygen attached to the carbon that undergoes the nucleophilic attack. Atom *e* is the most electronegative neighbor of atom *a*, excluding atom *b*. Atom *d* is the second most electronegative neighbor of atom *a*, if the former exists. We avoid using the label *c* to prevent confusion with the carbon element. Atom *o* is the oxygen atom in the attacking water molecule. Atoms h_1 and h_2 are the reactive and non-reactive hydrogen atoms in the water molecule, respectively (Fig. 6).

The identification of the key atoms in a given molecule follows a systematic procedure. First, the algorithm determines the relevant reaction family, as discussed above. Each supported hydrolysis family is implemented in ARC[37] using a graph representation of the generic reacting group for each family along with a "recipe" for breaking and forming bonds that

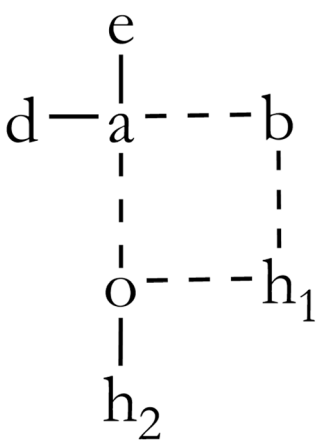

Fig. 6 Atom labeling of the hydrolysis reaction zone. Atoms *a* and *b* represent the bond undergoing hydrolysis, atoms *d*, and *e* are atom *a* neighbors, used for orienting the water molecule. Atom *o* represents the water's oxygen, atom h_1 is the hydrogen being ruptured from the water, and h_2 is the hydrogen atom that remains with the original water's oxygen atom, forming an OH group on the hydrolysis product. Broken lines represent reactive chemical bonds.

convert reactants into products. This reaction family implementation closely follows the convention implemented in the Reaction Mechanism Generator (RMG) software,[46,47] yet lacks kinetic data since it is only used to identify and label reacting atoms rather than to estimate rate coefficients.

The algorithm assigns labels to the four reacting atoms (a, b, o, and h_1) using subgroup isomorphism checks. To resolve e and d labeling ambiguity in complex scenarios, the algorithm ranks atoms attached to a based on their effective electronegativity. This concept is introduced to prioritize atoms involved in multiple bonds (*e.g.*, a carbonyl oxygen). The algorithm ranks atoms directly attached to atom a (excluding atom b), as defined in eqn (1). Here, χ_i^{eff} is the effective electronegativity of neighbor atom i, χ_i is the intrinsic electronegativity value of atom i according to the Pauling scale,[48] and $\text{BO}_{a,i}$ is the bond order between atoms a and i. In cases where two neighboring atoms have identical χ_i^{eff} values, the tie is resolved by comparing the sum of the χ_i^{eff} of the next-level neighboring atoms, with appropriately adjusted atom indices in eqn (1). The neighbor with the highest χ_i^{eff} is designated as atom e, establishing the primary orientation around atom a, while all additional neighboring atoms are collected in a sorted list (by decreasing χ_i^{eff} values) that represents potential atoms for the d label.

$$\chi_i^{\text{eff}} = \chi_i \cdot \text{BO}_{a,i} \tag{1}$$

The highest-ranked atom in this list is labeled by default as atom d, while the others are retained as potential alternatives for generating a more diverse set of TSGs if initial attempts fail. This approach allows flexibility in later stages of TSG generation by providing multiple options to define geometrical parameters. In nitrile hydrolysis, atom a often only has two neighbors, namely b and e. As a result, no d atom is defined in the TSG generation for this reaction family.

The atom labeling provides essential and uniform reference points for positioning and orienting the attacking water molecule in 3D. It facilitates positioning atoms o, h_1, and h_2 relative to atoms a, b, d, and e. Different examples for neighboring atoms labeling are provided in Fig. S1.

2.4. The internal-coordinate atom placement engine

The TSG generation engine relies on an existing ARC[37] module for representing internal coordinates ("Z-matrices"), including conversion between Cartesian and internal coordinates ("folding": cartesian → internal; "unfolding": internal → Cartesian). The unfolding implementation follows the *SN-NeRF* algorithm described by Parsons *et al.*[49]

In the present framework, ARC was enhanced with the ability to append a new atom X to an existing Cartesian structure $\{\mathbf{r}_i\}_{i=1}^{N}$, using the Z-matrix module. The new atom is placed to satisfy three internal constraints relative to specified reference atoms: a bond length $\rho = \|\mathbf{r}_N - \mathbf{r}_X\|$ to atom N, a bond angle between three atoms, $\theta = \measuredangle M-N-X$, and a dihedral angle involving a four-atom torsion, $\phi = \measuredangle L-M-N-X$ (Fig. 7). The three reference tuples of atom indices in the given Cartesian coordinates, (N), (M, N), and (L, M, N), may be chosen independently. This flexibility allows each internal DOF to be anchored to the most chemically meaningful atoms, even if they are not sequentially bonded, providing more intuitive control over the placement than a standard Z-matrix construction.

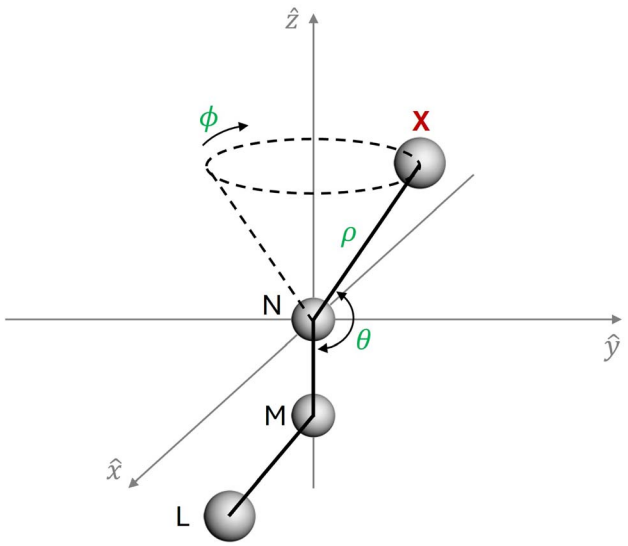

Fig. 7 Internal-coordinate specification for placing atom X. ρ: the N–X distance; θ: the M–N–X bond angle; ϕ: the L–M–N–X dihedral angle. In the present implementation, each degree of freedom may reference an independent set of atoms.

Let $\mathbf{r}_i = (x_i, y_i, z_i)$ denote the Cartesian coordinates of atom i. Given ρ, θ, ϕ, we place a new atom X at $\mathbf{r}_X = (x_X, y_X, z_X)$ so that the computed values ρ_{calc}, θ_{calc}, ϕ_{calc} match the independent targets.

The distance residual is

$$g_\rho(\mathbf{r}_X) = \rho_{\text{calc}} - \rho = \|\mathbf{r}_X - \mathbf{r}_N\| - \rho.$$

For the angle defined by M–N–X, with $\mathbf{NM} = \mathbf{r}_M - \mathbf{r}_N$ and $\mathbf{NX} = \mathbf{r}_X - \mathbf{r}_N$, the residual is

$$\theta_{\text{calc}} = a\tan 2(\|\mathbf{NM} \times \mathbf{NX}\|, \mathbf{NM}\cdot\mathbf{NX}), \quad g_\theta(\mathbf{r}_X) = \text{wrap}(\theta_{\text{calc}} - \theta),$$

where $a\tan 2$ is the quadrant-aware arctangent and wrap$(\cdot)$ maps angles to $(-\pi, \pi]$.

For the dihedral defined by L–M–N–X, define the plane normals

$$\mathbf{n}_1 = (\mathbf{r}_L - \mathbf{r}_M) \times (\mathbf{r}_N - \mathbf{r}_M), \quad \mathbf{n}_2 = (\mathbf{r}_X - \mathbf{r}_N) \times (\mathbf{r}_M - \mathbf{r}_N).$$

Let $\hat{\mathbf{u}}_{MN} = (\mathbf{r}_N - \mathbf{r}_M)/\|\mathbf{r}_N - \mathbf{r}_M\|$. Then

$$c = \frac{\mathbf{n}_1 \cdot \mathbf{n}_2}{\|\mathbf{n}_1\|\|\mathbf{n}_2\|}, \quad s = \hat{u}_{\text{MN}} \cdot \frac{\mathbf{n}_1 \times \mathbf{n}_2}{\|\mathbf{n}_1\|\|\mathbf{n}_2\|}, \quad \phi_{\text{calc}} = a\tan 2(s, c).$$

and the dihedral residual to minimize is

$$g_\phi(\mathbf{r}_X) = \text{wrap}(\phi_{\text{calc}} - \phi).$$

Finally, we determine $\mathbf{r}_X$ by minimizing the scaled sum of squares

$$F(\mathbf{r}_X) = \left(\frac{g_\rho(\mathbf{r}_X)}{\rho}\right)^2 + (g_\theta(\mathbf{r}_X))^2 + (g_\phi(\mathbf{r}_X))^2,$$

where the distance residual is scaled by its target value to form a dimensionless objective function. This equal weighting of normalized terms ensures that all three geometric constraints contribute balanced penalties during local optimization. This scheme was found to be numerically robust for appending a new atom X to an existing Cartesian structure.

We employ a sequential fallback strategy: a short list of heuristic initializations (ρ–θ directed, midpoint, perpendicular, bond-length-randomized, randomly perturbed, and shifted) is generated, and a local optimization of $F(\mathbf{r}_X)$ is attempted from each initialization in turn. An initialization is accepted immediately if the absolute errors

$$\varepsilon_\rho = |g_\rho| < 0.01\ \text{Å}, \quad \varepsilon_\theta = |g_\theta| < 0.1\ \text{rad}, \quad \varepsilon_\phi = |g_\phi| < 0.1\ \text{rad}.$$

If none meet all tolerances but at least one optimization converges, we select the candidate with the smallest $F(\mathbf{r}_X)$. On success, atom X is appended to the original Cartesian coordinate list at its optimized position $\mathbf{r}_X$. Following atom placement and TSG construction, newly generated structures are screened for geometric redundancy to avoid unnecessary downstream quantum chemical optimizations; the structural merging and screening procedure is described in detail in Section S2 of the SI.

2.5. Heuristic strategies for neutral hydrolysis

2.5.1. TSG generation strategy. Following reaction family identification and atom labeling, several TSGs are generated per reaction. ARC[37] strictly preserves a consistent atom order in its dual representations of each chemical species: a connectivity graph object (similar to RMG's Molecule object[46]) and a 3D Cartesian representation. This feature of ARC allows the translation from subgraph-isomorphism-based atom labels to actual 3D positioning of atoms. The computational workflow begins with an RDKit-based conformational search[50] followed by DFT optimization with an implicit solvation correction.[51] TSGs are then constructed by positioning a water molecule using family-specific internal coordinate parameters natively integrated into the ARC heuristics TS search adapter. These governing parameters were derived from approximately 10 DFT-optimized TSs per hydrolysis reaction class. By extracting heuristics from converged stationary points rather than initial manual guesses, the framework ensures that the hardcoded rules reflect the true electronic structure requirements of the nucleophilic approach. Consequently, the user's workflow remains fully automated, as the software independently applies these pre-defined rules to any new SMILES input without requiring manual TS fragment orientation.

The approach described here generates multiple structural variants for each reactive configuration. Each generated TSG undergoes a series of validation checks to remove redundant, unstable, or chemically unreasonable geometries by avoiding atomic collisions, verifying water molecule connectivity, and removing non-unique structures (see Section S2 in the SI).

To facilitate cleavage, the a–b bond is systematically stretched based on the identified family-specific parameters (Tables S1–S5). The algorithm achieves robust conformational diversity and overcomes steric hindrances through a hierarchical sampling strategy. While carbonyl-based reactions require rotations to accommodate the transition from planar sp^2 toward a partially tetrahedral configuration, ether hydrolysis requires alignment opposite the leaving group to ensure proper orbital overlap. In contrast, nitriles typically preserve their linear sp geometry, with dihedral modifications implemented only as a fallback if initial placements fail.

Core dihedrals involving atoms a, b, e, or d are identified from the internal coordinates and undergo a systematic scan from 15° to 55° in 10° intervals. This process displaces angles away from unstable eclipsed or near-planar boundary values, such as 0° or 180°, while maintaining the overall molecular framework. If no valid TSG passes the filtration step, the algorithm rotates the selected dihedrals by 180° to explore the complementary side of the local torsional landscape. By sequentially modifying additional dihedrals, the framework broadens the accessible conformational space until a chemically valid initialization is generated.

2.5.2. Water molecule positioning and orientation. The three atoms of the attacking water molecule (o, h_1, and h_2) are added consecutively following the preparation of the main reactant geometry. Each atom is positioned sequentially by specifying bond distance, bond angle, and dihedral angle based on one of two tracks determined by the presence of neighbor atom d.

Atom o is first placed using three parameters, as illustrated in step A1 (Fig. 8). The distance r_1 represents the forming a–o bond, which is critical for capturing the progress of the nucleophilic attack. The b–a–o angle (α_1), is defined using the leaving group (b) and the electrophilic center (a) as reference points to establish the nucleophilic approach trajectory. This angle controls the approach of atom o relative to the bond being broken, ensuring proper orbital overlap for bond formation while avoiding unfavorable steric interactions. The dihedral angle e–d–a–o (φ_1), is important for avoiding steric interference with the most electronegative neighbors around the reaction center, ensuring that the water molecule approaches the reactant from an accessible direction.

The reactive hydrogen (h_1) in water is positioned relative to the newly placed oxygen atom, as shown in Fig. 8, step B. The o–h_1 distance, labeled as r_2, is slightly elongated relative to the equilibrium O–H bond length to facilitate the proton transfer. The a–o–h_1 angle (α_2) uses the newly formed a–o bond as a reference to position h_2 for optimal interaction with the leaving group. This arrangement, combined with the b–a–o–h_1 dihedral angle (φ_2), specifically pre-organizes the 4-center (4-membered ring-like) geometry characteristic of the concerted neutral hydrolysis mechanism. The dihedral configuration (φ_2) establishes the spatial relationship between atoms h_1 and b, ensuring reasonable positioning for the proton transfer step. By sampling multiple dihedral orientations in a single run, the algorithm provides high-quality initializations that allow the

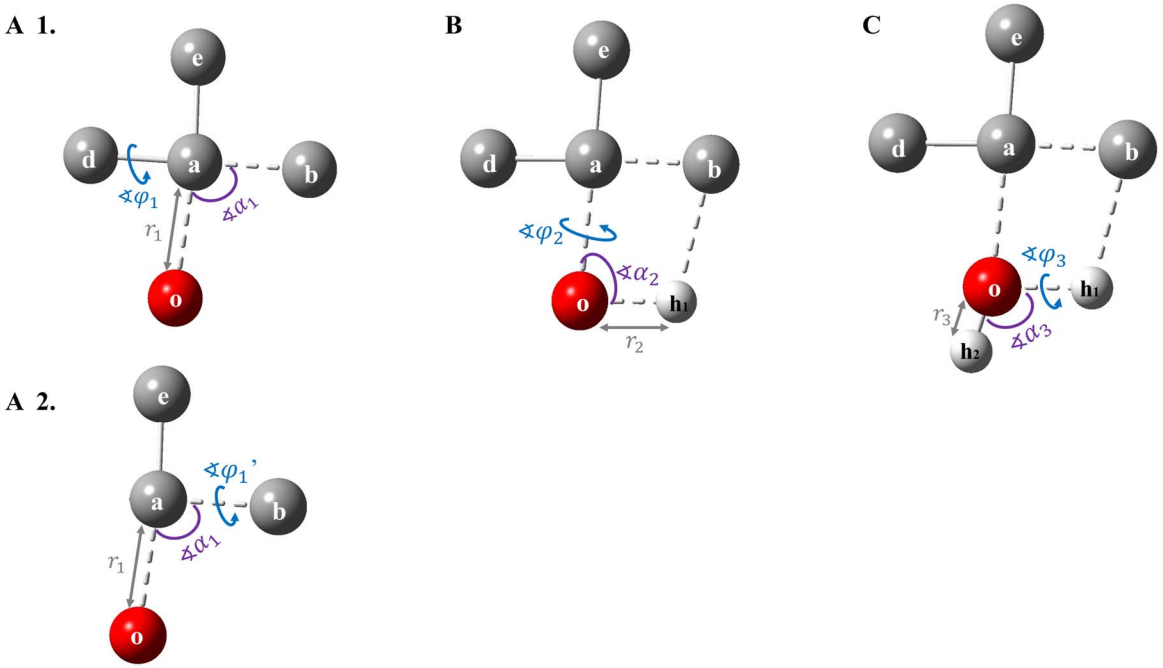

Fig. 8 The three sequential steps for positioning the water molecule using internal-coordinate heuristics. (A1 and A2) Placement of the water oxygen (atom *o*) relative to the substrate electrophilic center; step A1 is utilized when neighbor atom *d* is present, whereas A2 serves as the fallback when *d* is absent. (B) Placement of the reactive hydrogen (h_1) to pre-organize the 4-center transition state geometry characteristic of the concerted neutral hydrolysis mechanism. (C) Placement of the nonreactive hydrogen (h_2) to complete the transition state initialization.

DFT optimization to converge to this concerted saddle point across diverse chemical systems.

Finally, the nonreactive atom h_2 is positioned as illustrated in Fig. 8, step C. The o–h_2 distance (r_3) is close to the equilibrium O–H bond length. Similarly, the h_1–o–h_2 bond angle (α_3) is defined as slightly larger than water's equilibrium angle, 104.45°. The dihedral angle a–o–h_1–h_2 (φ_3) ensures the proper orientation of the water molecule, typically to balance the overall dipole moment in the TS.

In the absence of atom *d*, where other than the leaving group (atom *b*) atom *a* has only one neighbor, atom *e*. This typically indicates a higher bond order between atoms *a* and *b* (such as a C≡N bond). Above, atom *d* participates in the definition of the *o* atom dihedral angle (φ_1). In the present case, this dihedral is defined as the *e*–*b*–*a*–*o* angle (φ_1'), Fig. 8 step A 2. This dihedral angle uses both the most electronegative neighbor (*e*) and the leaving group (*b*) as reference points to define the relative location of atom *o*.

2.6. Quantum chemical computations and systematic validation protocol

Ab initio calculations were performed using the open-source Automated Rate Calculator (ARC) software tool.[37] ARC was used for preparing, scheduling, and post-processing quantum-chemical jobs. Quantum-chemical calculations were executed using the Gaussian 09 software suite. The initial set of TS geometries was generated through an exhaustive manual search. This manually curated set was used to derive the reaction-family-specific heuristics (*i.e.*, the target internal coordinates), as discussed in Section 3, and to benchmark its performance.

Initial TS geometry optimizations were performed at the CBS-QB3 (ref. 52) composite level of theory (utilizing B3LYP/CBSB7 (ref. 53 and 54)) followed by a vibrational-frequency calculation at the same level to obtain zero-point vibrational energy. The CBS-QB3 method was chosen for initial optimizations due to its established record of providing reliable and relatively robust geometries and energies for a wide range of organic reactions at a manageable computational cost. Additional electronic-structure calculations, including geometry optimizations, vibrational frequency analyses, single-point energy calculations, and intrinsic reaction coordinate (IRC)[55] calculations, were performed using the ωB97X-D[56] functional with the jul-cc-pVTZ basis set.[57] Solvation effects were modeled with the implicit Solvation Model based on Density (SMD)[51] with water as the solvent. While the internal-coordinate water placement algorithm is independent of the chosen solvation model, all validations reported in this work employed SMD. Other implicit or explicit solvation treatments are compatible within ARC but were not benchmarked here. For the high-throughput validation of the TSG engine, subsequent hindered rotor scans were omitted to reduce computational cost, as the primary objective was to confirm the successful and efficient location of the correct TS structure.

All TSGs generated by the heuristics adapter were internally screened within ARC for automated quality control after generation to eliminate redundant or unphysical geometries, such as structures containing atomic overlaps, before geometry

optimization. Each optimized structure, whether generated automatically or constructed manually, was then subjected to a systematic validation protocol to ensure its physical and mechanistic correctness. First, the algorithm was required to produce at least one TSG that successfully converged during DFT optimization, demonstrating its ability to generate chemically reasonable geometries that correspond to stationary points on the potential energy surface. Second, the optimized TS was examined for thermodynamic consistency by verifying that its electronic energy (E_0) was higher than that of both the reactant and product wells, as expected for an energy barrier between two minima. Third, harmonic frequency analysis was performed to confirm the presence of exactly one imaginary vibrational mode, indicative of a first-order saddle point. Fourth, the normal-mode displacement (NMD) associated with this imaginary frequency was visually inspected to ensure that the atomic motions corresponded to the intended bond-breaking and bond-forming events characteristic of the hydrolysis step. Finally, intrinsic reaction coordinate (IRC) calculations were carried out to confirm that the forward and reverse trajectories connected smoothly to the designated products and reactants, respectively.[58] IRC calculations were attempted for all TSs, but in cases where the calculation was as numerically unstable or computationally prohibitive, and the NMD clearly represented the expected bond-forming and bond-breaking motions, the TS was still considered validated.

2.7. Workflow example

The TSG generation procedure is demonstrated for a model hydrolysis reaction of methyl acetate, as illustrated in Fig. 9. After identifying the reaction as belonging to the carbonyl-based hydrolysis family (Fig. 3), reactant atoms are appropriately labeled (Fig. 9B). The labeling of atoms *e* and *d* follows the respective χ_i^{eff} values (eqn (2) and (3)). In this example, the carbonyl oxygen has the highest χ_i^{eff} value and is therefore assigned as atom *e*. The methyl carbon is trivially labeled as *d*.

$$\chi_{\text{O}}^{\text{eff}} = 3.44 \times 2 = 6.88 \quad \text{(carbonyl oxygen)} \tag{2}$$

$$\chi_{\text{C}}^{\text{eff}} = 2.55 \times 1 = 2.55 \quad \text{(methyl carbon)} \tag{3}$$

The *a*–*b* bond in the lowest-energy 3D conformer of methyl acetate in water is stretched by $\sim x1.3$ of its original length (Fig. 9C), in this example to 1.72 Å (Table S6). Next, the algorithm explores several placement branches. In the first branch, water is added to the stretched but otherwise unmodified conformer; the resulting TSG is shown from two viewpoints in Fig. 9D1.1 and 1.2. In the other branch, near-planar dihedrals around the reaction center are perturbed by series of angular offsets (here, the dihedral φ in Fig. 9D2.0), after which the water atoms are placed again shown from two viewpoints in Fig. 9D2.1 and 2.2. This branch strategy systematically samples both sterically hindered and accessible approach geometries to maximize the chance of generating a viable TSG.

The generated variants then undergo filtration according to the predefined criteria. As shown in Fig. 9D1.2, the variant generated without dihedral adjustments was eliminated during this step due to a collision between atoms *o* and one of the hydrogen atoms attached to atom *d*. The dihedral adjustment, shown here as an out-of-plane motion of atom *e* (Fig. 9C and D2.0), resulted in a valid TSG (Fig. 9D2.2). To assess the accuracy of the generated TSG, its geometry was compared to the respective DFT optimized structure using internal coordinates. The optimized TS in the methyl acetate example (Fig. 9E) showed a high degree of agreement with the initial guess with mean similarities of 95.75% for bonds, 97.74% for angles, and

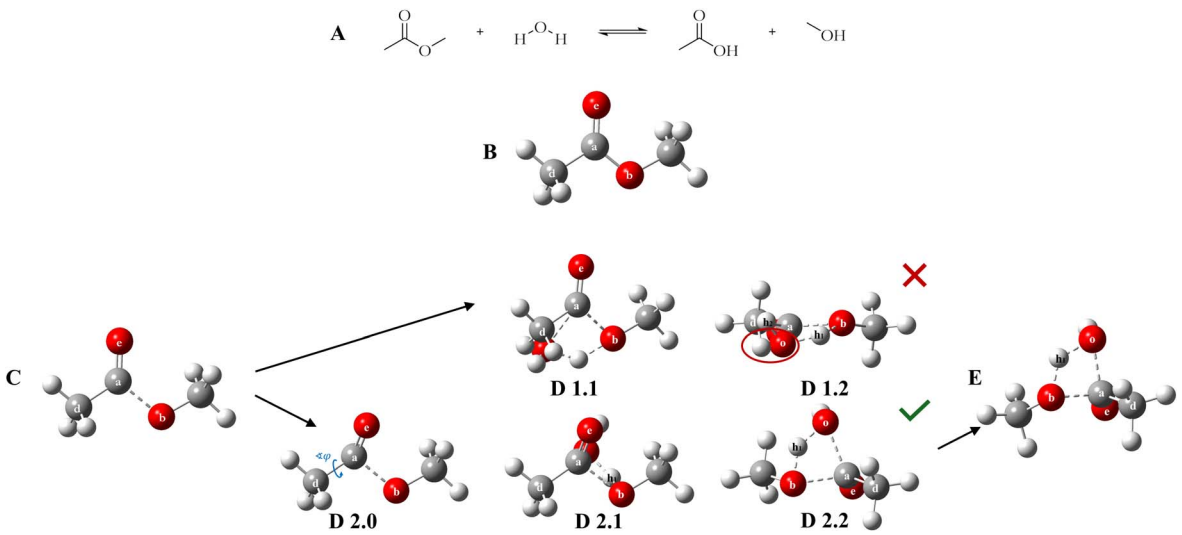

Fig. 9 A TSG generation workflow example for methyl acetate hydrolysis. (A) Reaction family identification. (B) Reactant atom label assignment. (C) *a*–*b* bond stretch. (D) Orienting the water molecule. Panels D 1.1 and D 1.2 display two viewpoints of the same TSG (unmodified conformer), with D 1.2 highlighting an atom collision. In contrast, D 2.1 and D 2.2 show two viewpoints of the TSG that resulted from the first dihedral-modified geometry. (E) The DFT optimized TS structure.

98.35% for dihedral angles, yielding an overall geometric similarity of approximately 97.28% (Table S6).

3. Results and discussion

3.1. Manual TS parameter determination

An initial set of 10 representative reactants was generated for each class (Fig. 3, and 4). For the acyl halide class, only chloro-substituted derivatives were included in the initial set because chlorides lie in the mid-range of halide leaving group strength, making them a practical baseline. The impact of this assumption on other halides (Br, F) that differ in bond strength[59] is examined below (Section 3.2).

The dataset used for the study of the transition state (TS) geometries comprised a wide range of molecules for each class. This was done to ensure that the chosen parameter values would be applicable across different molecular motifs rather than being tuned to a single molecular genre. To confirm coverage over a wide chemical space, structural diversity was quantified using Morgan circular fingerprints (ECFP)[50,60] and the Tanimoto similarity coefficient.[61] These fingerprints encode atom-centered environments within a two-bond radius into a fixed-length binary vector, enabling robust comparison of substructural features.

Analysis of within-class Tanimoto similarity distributions (Fig. 10) shows that all classes include both closely related and structurally distinct molecules. The median within-class similarity values ranged from 0.13 for ethers and amides, to 0.24 for acyl halides, indicating an overall high structural diversity. Ethers displayed the widest spread of similarity values, consistent with their structural flexibility.

To visualize the diversity of the dataset in physicochemical property space, Principal Component Analysis (PCA) was performed on nine molecular descriptors (Fig. 11). The broad and overlapping distribution across hydrolysis families provides a secondary confirmation that the benchmark set spans a wide property space. Detailed descriptor definitions and loading analyses are provided in Section S3, while complete descriptor Tables are provided in Tables S7 and S8 and datasets are available in Databases S1 and S2 of the SI.

For each molecule, the water was positioned using internal coordinates defined relative to the atoms in the other reactant as described in Section 2.5.2 and illustrated in Fig. 8. Here, we report the mean and standard deviation of the nine geometric features extracted from each optimized TS (Table 2). The complete set of parameter values extracted from every study case job, alongside the median [IQR] values for each class is provided in Tables S1–S5.

During TS optimization for hydrolysis reactions, two of the dihedral angles used to position the water molecule, specifically φ_1 and φ_3 (Fig. 8), which were used to position the oxygen atom (atom o) and the non-reactive hydrogen atom (atom h_1) in the water, respectively, showed consistent magnitudes within each class, yet their signs varied across TSs. In many cases only a specific combination of signs across both dihedrals produces a physically reasonable TS geometry, whereas the opposite sign leads to less favorable geometry. An example that illustrates the effect of dihedral sign combinations on water placement is given Fig. S2. While certain sign combinations yielded successful TSs for specific substrates, these trends are coordinate-convention and conformer dependent rather than mechanistically universal. Therefore, the mean dihedral value was calculated from absolute magnitudes, and both the $+\varphi$ and

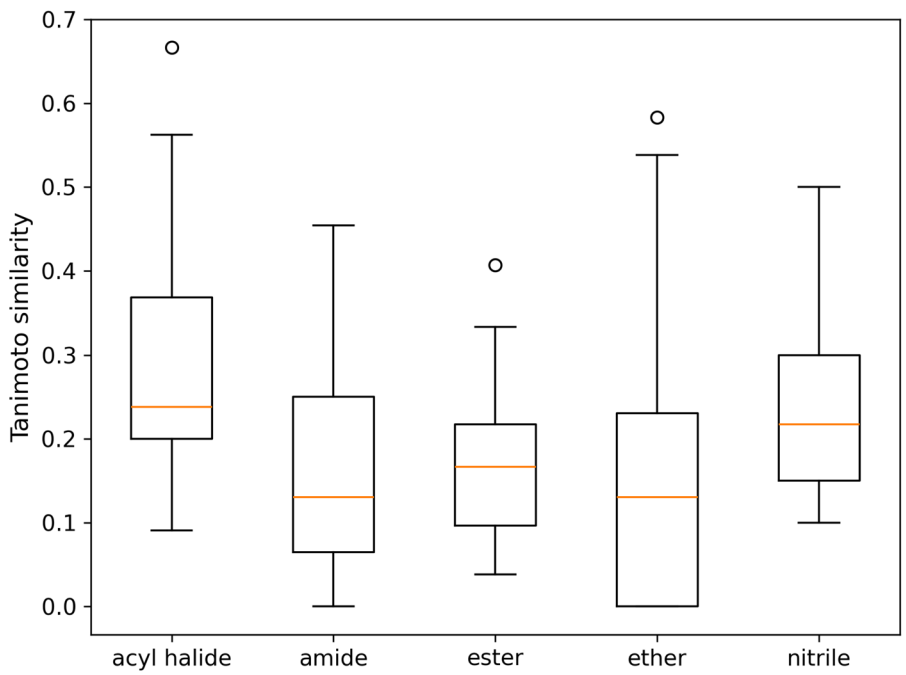

Fig. 10 Within-class similarity distributions of pairwise Tanimoto similarity values (Morgan fingerprints, radius = 2) for the study cases jobs. Circles represent outlier similarity values.

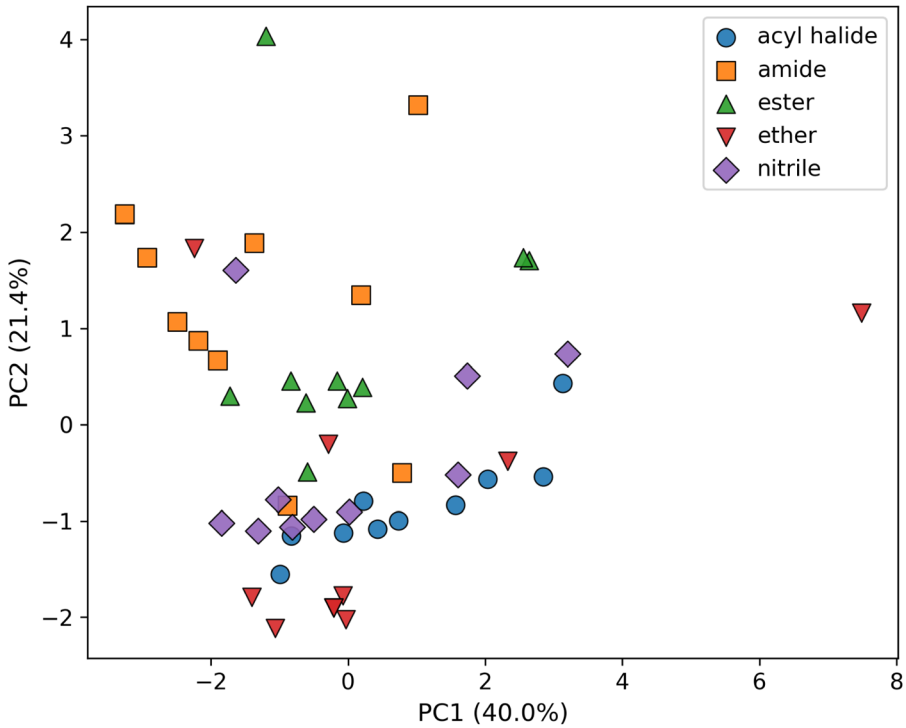

Fig. 11 PCA for the study cases jobs of nine standardized RDKit descriptors. PC1 (40.0% reflects molecular size and hydrophobicity; PC2 (21.4%) contrasts polarity and H-bonding with aromaticity for all studied classes.

Table 2 Mean and standard deviation of the internal coordinates used to position the water molecule in the TSG for all different classes. Bond lengths are in Å, angles are in degrees

Parameter	Esters	Amides	Acyl halides	Ethers	Nitriles
r_1	1.83 ± 0.09	1.91 ± 0.08	1.79 ± 0.07	2.14 ± 0.14	1.83 ± 0.03
α_1	76.18 ± 3.44	80.54 ± 2.06	75.43 ± 1.62	66.29 ± 3.84	97.23 ± 1.18
φ_1	142.59 ± 4.32	127.05 ± 2.10	152.24 ± 2.32	101.78 ± 13.34	174.29 ± 1.50
r_2	1.21 ± 0.05	1.39 ± 0.05	1.04 ± 0.01	1.10 ± 0.04	1.32 ± 0.01
α_2	73.29 ± 4.09	66.54 ± 1.94	88.43 ± 1.16	72.80 ± 0.90	57.89 ± 0.62
φ_2	1.57 ± 3.23	2.40 ± 5.90	0.95 ± 2.45	0.30 ± 2.49	−0.02 ± 1.88
r_3	0.97 ± 0.00	0.96 ± 0.00	0.97 ± 0.00	0.96 ± 0.00	0.97 ± 0.00
α_3	110.42 ± 1.98	120.03 ± 5.85	106.15 ± 1.01	107.20 ± 2.98	114.28 ± 0.72
φ_3	105.19 ± 3.86	97.59 ± 3.90	106.59 ± 3.45	102.19 ± 1.74	103.77 ± 0.70
a–b stretch[a]	1.34 ± 0.06	1.17 ± 0.02	1.41 ± 0.04	1.55 ± 0.10	1.04 ± 0.00

[a] The *a–b* bond stretch values represent a multiplicative factor relative to the optimized reactant bond length (*e.g.*, 1.34 corresponds to a 34% increase in bond length).

$-\varphi$ variants were considered when generating TSGs. For the third dihedral, φ_2 which was used to position the reactive hydrogen (atom h_2) in the water molecule, the optimized values across the TSs were consistently small and centered near 0°. In this range, the sign inversion reflects only minor fluctuations in a common geometry rather than distinct mirror image structures, so the signed mean was used directly.[62]

To improve statistical robustness and leverage mechanistic similarities, esters, amides, and acyl halides were grouped into a single carbonyl-based hydrolysis family. This grouping is chemically justified as all three classes undergo hydrolysis through a nucleophilic attack at the electrophilic carbonyl carbon, proceeding through a partially tetrahedral TS characteristic of the concerted 4-center mechanism observed under neutral conditions. The averaged geometric parameters derived from pooling the optimization results of these three classes are presented in Table 3.

Analysis of parameter distributions reveals both the strengths and limitations of the family-based approach. The consolidation of carbonyl-based classes demonstrates mixed results; while the oxygen positioning distance r_1 shows improved precision compared to individual class variations, other parameters such as φ_1 exhibit increased standard deviation after combination. This indicates that φ_1 is class sensitive, although the resulting variance remains within acceptable ranges for generating initial guesses. As expected, the most

Table 3 Mean and standard deviation of the internal coordinates used to position the water molecule of the carbonyl-based classes considered together. Bond lengths are in Å, angles are in degrees

Parameter	Carbonyl-based family
r_1	1.85 ± 0.09
α_1	77.39 ± 3.33
φ_1	140.63 ± 10.96
r_2	1.21 ± 0.15
α_2	76.09 ± 9.66
φ_2	1.64 ± 4.03
r_3	0.97 ± 0.00
α_3	112.20 ± 6.86
φ_3	103.12 ± 5.40
a–b stretch[a]	1.31 ± 0.11

[a] The *a–b* bond stretch value represents a multiplicative factor relative to the optimized reactant bond length.

consistent parameter across all functional groups is r_3, representing the non-reactive o–h_2 bond length in the approaching water molecule. In contrast, the o–h_1 bond distance, r_2, shows significant variations; carbonyl-based hydrolysis family exhibits intermediate approach distances (1.21 ± 0.15 Å), and ether hydrolysis family require the shortest distances (1.10 ± 0.04 Å).

The dihedral angles reveal important insights into water molecule orientation preferences. The values of φ_2 remain consistently small in all families (ranging from −12° to +2.4°), indicating a genuine chemical preference that likely minimizes steric interactions while improving the initial geometry guess of the approach. Notably, the ether hydrolysis family exhibits larger standard deviations specifically in the dihedral angle φ_1. This larger deviation suggests a lower geometric consistency and may indicate potential convergence challenges.

Despite these variations, the overall parameter distributions fall within acceptable ranges for automated TSG search. Bond lengths maintain standard deviations below ± 0.1 Å, and most angular parameters remain within ± 10°, meeting the established criteria for reliable initial guess generation.

3.2. Validation

To assess the reliability and accuracy of the automated TSG generation, comprehensive validation studies were conducted across all hydrolysis reaction families. Each validation job tested whether our approach could consistently locate a valid TSG for a given hydrolysis reaction.

To ensure diversity and capture different kinds of molecules undergoing hydrolysis across the defined classes, both the Tanimoto similarity coefficient and the PCA based on physicochemical descriptors were performed using the same definitions and methodological specifications applied to the case studies. These analyses provide a complementary confirmation of dataset breadth. Analysis of within-class Tanimoto similarity values shows that acyl halides and amides exhibit the highest internal molecular similarity, with median similarities of ∼0.26–0.27, while esters and nitriles lie in an intermediate range (∼0.22–0.24). In contrast, ethers display the lowest median similarity (∼0.14) and the widest overall spread, indicating substantially higher structural diversity within this class. PCA further supports these findings, confirming that all five classes span a broad and overlapping chemical property space. Complete similarity data for all classes are provided in Fig. S3, S4, Tables S9, S10, and Databases S3 and S4.

Within the carbonyl-based hydrolysis family, a total of 33 jobs were executed: 9 for the ester class, 12 for the amide class, and 12 for the acyl halide class. In addition, 30 jobs were performed for the ether hydrolysis family and 30 for the nitrile hydrolysis family.

A validation job was considered successful only when a converged TS satisfied the validation checks described in Section 2.6. This comprehensive framework confirmed that the automated TSG generation reliably identifies correct TSs across all defined hydrolysis families. In addition to validating the correctness of the identified TSGs, we report the number of TSGs generated by the heuristics adapter.

Table 4 presents the success rate (defined as the percentage of jobs with at least one converged TS that met the validation criteria), as well as the average number of generated TSGs for each reaction family. In Section S1 a complete summary of all jobs is provided, including detailed evaluations against each of the five validation criteria along with the number of TSGs generated (Tables S11–S13).

The results demonstrate that the carbonyl-based hydrolysis family achieved a remarkably high success rate of 96.9%, with an average of ∼6 generated TSGs. This strong performance correlates with the low variation observed during parameterization, and also arises from the highly polarized C=O bond, which defines a clear electrophilic carbon center and provides a directional π orbital that naturally guides the approach of the water molecule. As a result, the generated geometries align well with the electronic structure of the reactive site, producing stable and convergent TSs during optimization. Consequently, a relatively small number of TSGs were needed to undergo DFT optimization to yield a converged TS, with an overall high success rate (Table 4).

The nitrile hydrolysis family (as defined here, corresponding to step (a) in Fig. 4B) also exhibited high performance, with a success rate of 86.2% and ∼5 generated TSGs, consistent with the linear and strongly polarized carbon–nitrogen triple bond that creates a well-defined attack direction for the water molecule.

The lowest success rate in this study, 72.4%, was obtained for the ether hydrolysis family, consistent with the larger geometric variability in the parametrization step, particularly in the φ_1 dihedral. Because ethers lack a polarized π system and possess

Table 4 Validation success rates and the average number of the generated TSGs across reaction families

Family	Success rate(%)	Average number of generated TSGs
Carbonyl-based	96.9	5.8
Ether	72.4	14.3
Nitrile	86.2	4.9

two freely rotatable σ bonds around oxygen, the local environment is more flexible and less directive. As a result, significantly more TSGs are needed on average, and many converge into non-reactive minima during the TS optimization, where the water molecule is stabilized through hydrogen bonding to the substrate rather than attacking the reactive center.

While string methods like NEB[18] or GSM[16] offer robust path refinement, they are not fully automated for bimolecular systems, as they require manual 3D alignment and atom mapping. In contrast, our workflow operates from SMILES alone. Although CPU cost scales exponentially with system size, our method typically requires less than 10 independent TS DFT optimizations per reaction, on average (Table 4). Since these guesses are independent, they can be executed in parallel with zero communication overhead, significantly reducing wall-clock time compared to coupled reaction-path algorithms.

3.3. Ablation study of geometric heuristics

To assess the relative importance of the geometric heuristics used in the TSG generation algorithm, we performed a targeted ablation study in which three key components were individually removed: (i) dihedral modifications around the reaction center, (ii) the $\pm\varphi$ dihedral sign-sampling procedure, and (iii) the electronegativity-based d/e neighbor ranking. Each modification was tested on representative reactions from all hydrolysis families, and a case was considered successful only if at least one generated TSG converged to a validated TS. Full methodological details and per-reaction results are reported in Section S4, and Section S1 (Table S14).

The ablation results reveal that the geometric heuristics act cooperatively rather than independently. Removing either the dihedral-modification step or the $\pm\varphi$ sign sampling led to near-complete failure across all families, demonstrating that explicit control over torsional accessibility and attack direction is essential for placing the water molecule in a chemically meaningful orientation. In contrast, removing the d/e electronegativity ranking had a smaller, structure-dependent effect, with many carbonyl-based systems still converging successfully, while ethers showed greater sensitivity. Overall, this analysis confirms that dihedral modification and $\pm\varphi$ sampling are indispensable for robust TS generation, whereas the d/e ranking primarily improves consistency and reproducibility in asymmetric environments.

4. Conclusions

This work paves the way for transition state (TS) search of bimolecular liquid-phase reactions. By combining atom-labeled graph templates with atom-centered internal-coordinate placement and a lightweight local optimization, we automated the most failure-prone step in condensed-phase TS discovery: positioning and orienting two reacting fragments in 3D. While the underlying internal-coordinate rules were originally derived from a set of chemically-guided DFT-optimized case studies, they are now natively encoded within the Automated Reaction Calculator (ARC) framework to enable a fully automated operation. This architecture allows the software to generate high-quality TS guesses (TSGs) for neutral hydrolysis and validate them under a standardized protocol without user intervention.

Across 91 diverse reactions spanning three hydrolysis families, the method achieved relatively high success rates: 96.9% for carbonyl-based reactions (esters, amides, and acyl halides), 86.2% for nitriles, and 72.4% for ethers, while requiring only a modest number of TSGs per case. These outcomes reflect the electronic "guidance" available at polarized or linear centers (C═O and C≡N), and they clarify why ethers remain more challenging.

An ablation study demonstrates that two ingredients are essential for robust performance in solution: (i) small, targeted dihedral adjustments around the reaction center before water placement, and (ii) explicit $\pm\varphi$ sign sampling for dihedrals that control the approach of the water molecule from either side of the reactive plane of the substrate. Removing either step causes near-universal failures, whereas the electronegativity-based ranking of neighboring atoms (d/e) mainly fine-tunes orientation and improves reproducibility and robustness across asymmetric environments.

Although family parameters were extracted from CBS-QB3 case studies, they transferred well to ωB97X-D/jul-cc-pVTZ with SMD water for validation, indicating that the heuristics capture geometric, not method-specific, features of nucleophilic approach. By replacing manual trial-and-error water placement with these reproducible encoded internal-coordinate rules, the procedure turns condensed-phase TSG initialization into an automated, scalable step that can feed high-throughput kinetics workflows in pharmaceutical degradation and broader liquid-phase reactivity. This suggested approach bypasses the 3D orientation bottleneck that has historically required extensive expert intervention. With an average computational requirement of only ~10 TS DFT optimizations in parallel per reaction, the workflow is well-suited for high-throughput kinetics in pharmaceutical degradation and broader liquid-phase reactivity.

Looking ahead, the same internal-coordinate atom-addition strategy is readily extensible to alternative solvation treatments and microsolvation, where additional explicit water molecules can be positioned sequentially. The method could therefore be extended to address additional challenging reaction families, including acid- or base-catalyzed hydrolysis and other multi-fragment pathways relevant to polymer degradation and complex condensed-phase chemistry. This modular approach allows for the automated construction of complex, multi-fragment TSs that have historically required extensive manual manipulation. The suggested chemically informed internal-coordinate placement approach provides a practical foundation for automated TS discovery in solution and opens a path to general, high-throughput TS search for bimolecular liquid-phase reactions.

Conflicts of interest

There are no conflicts of interest to declare.

Data availability

The datasets supporting this article have been uploaded as part of the supplementary information (SI). Supplementary information: Section S1: figures and tables – visual logic and ranking rules for the *e* and *d* neighbor atom assignments (Fig. S1), the effect of dihedral sign combinations on water placement in TSG generation (Fig. S2), Tanimoto similarity matrices and PCA structural diversity analysis for the validation set (Fig. S3 and S4), hydrolysis TS parameters manually extracted from the optimized geometries (Tables S1–S5), comparison between the initial TS guess and the optimized TS geometry for methyl acetate hydrolysis (Table S6), within-class and between-class similarity statistics (Tables S7–S10), validation job summaries (Tables S11–S13), detailed per-reactant results of the ablation study (Table S14); Section S2: structural merging and redundancy screening – pre-optimization coordinate comparison, handling overlapping methods, selection of unique guesses; Section S3: detailed PCA descriptor definitions and methodological specifics; Section S4: ablation study of geometric heuristics; Section S5: datasets – molecular fingerprints, descriptor matrices, and PCA scores, quantum chemical validation output files, example ARC YAML file input for validation job; Section S6: references. See DOI: https://doi.org/10.1039/d5dd00506j.

The code for the ARC repository can be found at https://github.com/ReactionMechanismGenerator/ARC with DOI: https://doi.org/10.5281/zenodo.3356849.

Acknowledgements

This work was supported in part by The Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through the German-Israeli Project Cooperation (DIP) project number 6493/1-1, The Israel Science Foundation (ISF) under Grant No. 185/23, and the Stephen and Nancy Grand Technion Energy Program (GTEP). L. F. acknowledges financial support from the VATAT Scholarship (Planning and Budgeting Committee, Council for Higher Education, Israel).

References

1 R. Porta, M. Benaglia and A. Puglisi, Flow Chemistry: Recent Developments in the Synthesis of Pharmaceutical Products, *Org. Process Res. Dev.*, 2016, **20**, 2–25.
2 C. W. Coley, R. Barzilay, T. S. Jaakkola, W. H. Green and K. F. Jensen, Prediction of Organic Reaction Outcomes Using Machine Learning, *ACS Cent. Sci.*, 2017, **3**, 434–443.
3 S. Baertschi, K. Alsante, and R. Reed, *Pharmaceutical Stress Testing: Predicting Drug Degradation, Second Edition*, CRC Press, 2011.
4 M. Li, *Organic chemistry of drug degradation*, Royal Society of Chemistry, 2012.
5 D. Zhou, W. Porter, and G. Zhang, Chapter 5 – Drug Stability and Degradation Studies, in *Developing Solid Oral Dosage Forms*, ed. Y. Qiu, Y. Chen, G. G. Zhang, L. Yu, and R. V. Mantri, Academic Press, Boston, 2nd edn, 2017, pp. 113–149.
6 K. C. Waterman and R. C. Adami, Accelerated aging: Prediction of chemical stability of pharmaceuticals, *Int. J. Pharm.*, 2005, **293**, 101–125.
7 H. Wu, A. Grinberg Dana, D. S. Ranasinghe, F. C. I. Pickard, G. P. F. Wood, T. Zelesky, G. W. Sluggett, J. Mustakis and W. H. Green, Kinetic Modeling of API Oxidation: (2) Imipramine Stress Testing, *Mol. Pharm.*, 2022, **19**, 1526–1539.
8 A. Grinberg Dana, K. M. Van Geem, C. Cavallotti and W. H. Green, Predictive Chemical Kinetic Modeling: Where We Succeed, Where We Struggle, and What Comes Next, *ACS Eng. Au*, 2026, **6**, 1–19.
9 J. R. Campanelli, M. Kamal and D. Cooper, A kinetic study of the hydrolytic degradation of polyethylene terephthalate at high temperatures, *J. Appl. Polym. Sci.*, 1993, **48**, 443–451.
10 J. Krug, P. Popelier and R. Bader, Theoretical study of neutral and of acid and base-promoted hydrolysis of formamide, *J. Phys. Chem.*, 1992, **96**, 7604–7616.
11 M. Akbarian and S.-H. Chen, Instability challenges and stabilization strategies of pharmaceutical proteins, *Pharmaceutics*, 2022, **14**, 2533.
12 M. E. Aulton, and K. Taylor, *Aulton's pharmaceutics: the design and manufacture of medicines*, Elsevier Health Sciences, 2013.
13 D. G. Brown and J. Bostrom, Analysis of past and present synthetic methodologies on medicinal chemistry: where have all the new reactions gone? Miniperspective, *J. Med. Chem.*, 2016, **59**, 4443–4458.
14 A. Grinberg Dana, H. Wu, D. S. Ranasinghe, F. C. I. Pickard, G. P. F. Wood, T. Zelesky, G. W. Sluggett, J. Mustakis and W. H. Green, Kinetic Modeling of API Oxidation: (1) The AIBN/H_2O/CH_3OH Radical "Soup", *Mol. Pharm.*, 2021, **18**, 3037–3049.
15 C. Peng and H. Bernhard Schlegel, Combining Synchronous Transit and Quasi-Newton Methods to Find Transition States, *Isr. J. Chem.*, 1993, **33**, 449–454.
16 P. Zimmerman, Reliable transition state searches integrated with the growing string method, *J. Chem. Theory Comput.*, 2013, **9**, 3043–3050.
17 S. Mallikarjun Sharada, P. M. Zimmerman, A. T. Bell and M. Head-Gordon, Automated transition state searches without evaluating the Hessian, *J. Chem. Theory Comput.*, 2012, **8**, 5166–5174.
18 P. Maragakis, S. A. Andreev, Y. Brumer, D. R. Reichman and E. Kaxiras, Adaptive nudged elastic band approach for transition state calculation, *J. Chem. Phys.*, 2002, **117**, 4651–4658.
19 R. Van de Vijver and J. Zádor, KinBot: Automated stationary point search on potential energy surfaces, *Comput. Phys. Commun.*, 2020, **248**, 106947.
20 P. L. Bhoorasingh, B. L. Slakman, F. Seyedzadeh Khanshan, J. Y. Cain and R. H. West, Automated transition state theory calculations for high-throughput kinetics, *J. Phys. Chem. A*, 2017, **121**, 6896–6904.
21 S. Gong, Y. Wang, Y. Tian, L. Wang and G. Liu, Rapid enthalpy prediction of transition states using molecular graph convolutional network, *AIChE J.*, 2023, **69**, e17269.

22 L.-P. Wang, A. Titov, R. McGibbon, F. Liu, V. Pande and T. Martínez, Discovering chemistry with an ab initio nanoreactor, *Nat. Chem.*, 2014, **6**, 1044–1048.
23 Y. Zhang, C. Xu and Z. Lan, Automated Exploration of Reaction Networks and Mechanisms Based on Metadynamics Nanoreactor Simulations, *J. Chem. Theory Comput.*, 2023, **19**, 8718–8731.
24 U. Raucci, V. Rizzi and M. D. Parrinello, Sample, and Refine: Exploring Chemistry with Enhanced Sampling Techniques, *J. Phys. Chem. Lett.*, 2022, **13**, 1424–1430.
25 P. M. Zimmerman, Automated discovery of chemically reasonable elementary reaction steps, *J. Comput. Chem.*, 2013, **34**, 1385–1392.
26 X.-J. Zhang and Z.-P. Liu, Reaction sampling and reactivity prediction using the stochastic surface walking method, *Phys. Chem. Chem. Phys.*, 2015, **17**, 2757–2769.
27 L. Krep, I. S. Roy, W. Kopp, F. Schmalz, C. Huang and K. Leonhard, Efficient Reaction Space Exploration with ChemTraYzer-TAD, *J. Chem. Inf. Model.*, 2022, **62**, 890–902.
28 M. Woulfe and B. M. Savoie, Chemical Reaction Networks from Scratch with Reaction Prediction and Kinetics-Guided Exploration, *J. Chem. Theory Comput.*, 2025, **21**, 1276–1291.
29 S. Maeda, Y. Harabuchi, M. Takagi, T. Taketsugu and K. Morokuma, Artificial force induced reaction (AFIR) method for exploring quantum chemical potential energy surfaces, *Chem. Rec.*, 2016, **16**, 2232–2248.
30 Y. Sumiya, Y. Tabata and S. Maeda, Understanding the Acetalization Reaction Based on its Reaction Path Network, *ChemSystemsChem*, 2020, **2**, e1900022.
31 R. Staub, Y. Harabuchi, C. Seraphim, A. Varnek and S. Maeda, An Accurate and Efficient Reaction Path Search with Iteratively Trained Neural Network Potential: Answering the Passerini Mechanism Controversy, *J. Chem. Theory Comput.*, 2026, **22**, 422–440.
32 E. C.-Y. Yuan, A. Kumar, X. Guan, E. Hermes, A. Rosen, J. Zádor, T. Head-Gordon and S. Blau, Analytical ab initio hessian from a deep learning potential for transition state optimization, *Nat. Chem.*, 2024, **15**, 8865.
33 L. Galustian, K. Mark, J. Karwounopoulos, M. P.-P. Kovar and E. Heid, GoFlow: Efficient Transition State Geometry Prediction with Flow Matching and E(3)-Equivariant Neural Networks, *Digital Discovery*, 2025, **4**, 3492–3501.
34 T. A. Young, J. J. Silcock, A. J. Sterling and F. Duarte, autodE: automated calculation of reaction energy profiles—application to organic and organometallic reactions, *Angew. Chem.*, 2021, **133**, 4312–4320.
35 J. I. Mujika, J. M. Mercero and X. Lopez, Water-promoted hydrolysis of a highly twisted amide: Rate acceleration caused by the twist of the amide bond, *J. Am. Chem. Soc.*, 2005, **127**, 4445–4453.
36 H. Gunaydin and K. N. Houk, Revisiting the mechanism of neutral hydrolysis of esters, *J. Am. Chem. Soc.*, 2008, **130**, 15232–15233.
37 A. Grinberg Dana, D. Ranasinghe, H. Wu, C. Grambow, X. Dong, M. Johnson, M. Goldman, M. Liu, and W. Green, *ARC – Automated Rate Calculator*, version 1.1.0, DOI: 10.5281/zenodo.3356849, https://github.com/ReactionMechanismGenerator/ARC, 2019.
38 D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, *J. Chem. Inf. Comput. Sci.*, 1988, **28**, 31–36.
39 M. Riahinezhad, M. Hallman and J.-F. Masson, Critical Review of Polymeric Building Envelope Materials: Degradation, Durability and Service Life Prediction, *Buildings*, 2021, **11**, 299.
40 R. L. Burwell Jr, The Cleavage of Ethers, *Chem. Rev.*, 1954, **54**, 615–685.
41 B. C. Ranu and S. Bhar, Dealkylation of ethers. A review, *Org. Prep. Proced. Int.*, 1996, **28**, 371–409.
42 T. J. Ahmed, S. M. M. Knapp and D. R. Tyler, Frontiers in catalytic nitrile hydration: Nitrile and cyanohydrin hydration catalyzed by homogeneous organometallic complexes, *Coord. Chem. Rev.*, 2011, **255**, 949–974.
43 M. Kobayashi and S. Shimizu, Nitrile hydrolases, *Curr. Opin. Chem. Biol.*, 2000, **4**, 95–102.
44 B. Izzo, C. L. Harrell and M. T. Klein, Nitrile reaction in high-temperature water: Kinetics and mechanism, *AIChE J.*, 1997, **43**, 2048–2058.
45 A. Krämer, S. Mittelstädt, and H. Vogel, Hydrolysis of nitriles in supercritical water. *Chemical Engineering & Technology: Industrial Chemistry-Plant Equipment-Process Engineering-Biotechnology*, 1999, vol. 22, pp. 494–500.
46 M. Liu, A. Grinberg Dana, M. S. Johnson, M. J. Goldman, A. Jocher, A. M. Payne, C. A. Grambow, K. Han, N. W. Yee, E. J. Mazeau, *et al.*, Reaction mechanism generator v3. 0: advances in automatic mechanism generation, *J. Chem. Inf. Model.*, 2021, **61**, 2686–2696.
47 M. S. Johnson, X. Dong, A. Grinberg Dana, Y. Chung, D. Farina Jr, R. J. Gillis, M. Liu, N. W. Yee, K. Blondal, E. Mazeau, *et al.*, RMG database for chemical property prediction, *J. Chem. Inf. Model.*, 2022, **62**, 4906–4915.
48 National Center for Biotechnology Information, *Electronegativity in the Periodic Table of Elements*, https://pubchem.ncbi.nlm.nih.gov/periodic-table/electronegativity, 2025, Accessed: October 12, 2025.
49 J. Parsons, J. B. Holmes, J. M. Rojas, J. Tsai and C. E. M. Strauss, Practical conversion from torsion space to Cartesian space for in silico protein synthesis, *J. Comput. Chem.*, 2005, **26**, 1063–1068.
50 RDKit: Open-source cheminformatics, https://www.rdkit.org, Accessed: November 5, 2025.
51 A. V. Marenich, C. J. Cramer and D. G. Truhlar, Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions, *J. Phys. Chem. B*, 2009, **113**, 6378–6396.
52 Jr, J. A. Montgomery, M. J. Frisch, J. W. Ochterski and G. A. Petersson, A complete basis set model chemistry. VI. Use of density functional geometries and frequencies, *J. Chem. Phys.*, 1999, **110**, 2822–2827.
53 A. D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior, *Phys. Rev. A:At., Mol., Opt. Phys.*, 1988, **38**, 3098.

54 C. Lee, W. Yang and R. Parr, Accurate and simple analytic representation of the electron-gas correlation energy, *Phys. Rev. B:Condens. Matter Mater. Phys.*, 1988, **37**, 785–789.
55 K. Ishida, K. Morokuma and A. Komornicki, The intrinsic reaction coordinate. An ab initio calculation for HNC → HCN and H-+ CH4 → CH4+ H-, *J. Chem. Phys.*, 1977, **66**, 2153–2156.
56 J.-D. Chai and M. Head-Gordon, Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615–6620.
57 K. A. Peterson, D. Figgen, M. Dolg and H. Stoll, Energy-consistent relativistic pseudopotentials and correlation consistent basis sets for the 4d elements Y–Pd, *J. Chem. Phys.*, 2007, **126**, 124101.
58 H. P. Hratchian, and H. B. Schlegel, Finding minima, transition states, and following reaction pathways on ab initio potential energy surfaces, in *Theory and applications of computational chemistry*, Elsevier, 2005, pp. 195–249.
59 F. A. Carey, and R. J. Sundberg, *Advanced organic chemistry: part A: structure and mechanisms*, Springer, 2000.
60 D. Rogers and M. Hahn, Extended-connectivity fingerprints, *J. Chem. Inf. Model.*, 2010, **50**, 742–754.
61 D. Bajusz, A. Rácz and K. Héberger, Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations?, *J. Cheminf.*, 2015, **7**, 20.
62 N. R. Pillsbury and T. S. Zwier, Conformational isomerization of 5-phenyl-1-pentene probed by SEP-population transfer spectroscopy, *J. Phys. Chem. A*, 2009, **113**, 126–134.