Accurate transition state generation with an object-aware equivariant elementary reaction diffusion model

Chenru Duan[1, 2, *], Yuanqi Du[3], Haojun Jia[1, 2], and Heather J. Kulik[1, 2]

[1]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, 02139
[2]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, 02139
[3]Department of Computer Science, Cornell University, Ithaca, NY, 14850
[*]Corresponding to: duanchenru@gmail.com

Abstract

Transition state (TS) search is key in chemistry for elucidating reaction mechanisms and exploring reaction networks. The search for accurate 3D TS structures, however, requires numerous computationally intensive quantum chemistry calculations due to the complexity of potential energy surfaces. Here, we developed an object-aware SE(3) equivariant diffusion model that satisfies all physical symmetries and constraints for generating sets of structures – reactant, TS, and product – in an elementary reaction. Provided reactant and product, this model generates a TS structure in seconds instead of hours required when performing quantum chemistry-based optimizations. The generated TS structures achieve a median of 0.08 Å root mean square deviation compared to the true TS. With a confidence scoring model for uncertainty quantification, we approach an accuracy required for reaction rate estimation (2.6 kcal/mol) by only performing quantum chemistry-based optimizations on 14% of the most challenging reactions. We envision the proposed approach useful in constructing large reaction networks with unknown mechanisms.

Introduction

Breaking down complex chemical reactions into their constituent elementary reactions is key for understanding reaction mechanisms and designing processes that favor target reaction pathways.[1–3] Due to the transient nature of the intermediate and transition state (TS) involved in these elementary reactions, it is difficult to isolate and characterize these structures experimentally. Instead, high throughput quantum chemistry computation, e.g., with density functional theory (DFT),[4] provides valuable insights on potential reaction mechanisms by constructing comprehensive reaction networks.[2,5] These networks are established by either iteratively enumerating potential elementary reactions on-the-fly given existing species[6,7] or propagating biased *ab initio* molecular dynamics followed by elementary reaction refinement.[8–10] Both approaches, however, require a tremendous number of quantum chemistry calculations due to the large number of species potentially involved in a chemical reaction.[11–13]

Among all DFT energy evaluations, the overwhelming majority comes from locating an accurate TS structure solely based on reactant and product information.[2,3] Nonetheless, obtaining these TS structures is vital for estimating reaction rates and determining dominant reaction pathways in a reaction network. Conventional TS search algorithms (e.g., nudged elastic band[14]) are computationally intensive and notorious for their difficulty in convergence,[15] yielding

Preprint. Under review.

low success rates and wasting significant computational resources.[16] Recently, there has been growing interest in exploring the use of machine learning techniques for TS search. This includes ideas that formulate TS search as a 2D graph-to-structure conversion problem,[17] a "shooting game" solved by reinforcement learning,[18] generative tasks addressed alternately by graph neural networks (GNN),[19] a generative-adversarial network[20] and a combination of gated recurrent neural network and transformer,[21] and using an ML potential as a surrogate for DFT during TS optimizations.[22] However, these approaches do not respect all the physical symmetries in describing an elementary reaction and require further reconstruction and optimization to obtain the final 3D TS structure. In addition, they are still far from reaching the high precision required (i.e., 3 kcal/mol, corresponding to a change of one order of magnitude in reaction rate at 300 °C) for estimating a TS barrier height in lieu of DFT evaluation.[3]

Diffusion models[23–25] have recently been adapted in physical science problems, such as generating organic molecules[26] and their conformations, protein-ligand docking,[27] and structural-based drug design.[28] There, an SE(3) equivariant GNN is used as the scoring function to preserve the required permutation, transition, and rotation symmetry for a 3D object (e.g., molecule or protein) in the Euclidean space, which works ideally for systems that contain only one single object.[29–31] However, there are many scenarios in chemistry and materials science where the desired system consists of multiple objects, for which the relative positioning does not influence the system itself. This includes the design of compounds with multiple building blocks (e.g., metal organic frameworks[32]), pairs of molecules that have similar chemistry but demonstrate distinct properties (e.g., the well-known activity cliff in protein binding[33]), and chemical processes that involve multiple distinct structures such as in chemical reactions.[2] Existing diffusion models with SE(3) equivariant GNNs are problematic for modeling these systems as they do not respect all symmetries and constraints for describing these systems.

In this work, we developed a general procedure to adapt an SE(3) equivariant neural network to preserve all desired symmetries and constraints on systems that consist of multiple objects. We demonstrated this "object-aware" SE(3) GNN for generating sets of 3D molecules in elementary reactions under the diffusion model framework, which we refer to as OA-ReactDiff. In particular, we focused on TS search, an essential but computationally demanding step for estimating reaction barrier heights, rates, and exploring reaction networks. With OA-ReactDiff, the predicted TS structures are highly similar to the true TS structures with an average root mean square deviation (RMSD) of 0.18 Å within 6 seconds on a single GPU. We further built a recommender based on confidence ranking to select among samples generated by OA-ReactDiff, which reduced the average RMSD to 0.13 Å. Using the self-confidence score of OA-ReactDiff for uncertainty quantification, we obtain an MAE 2.6 kcal/mol on barrier height by only performing 14% of the DFT-based optimizations for the most challenging systems for the model, approaching the accuracy required (i.e., 3 kcal/mol) for exploring reaction networks with unknown mechanisms.[3] The excellent

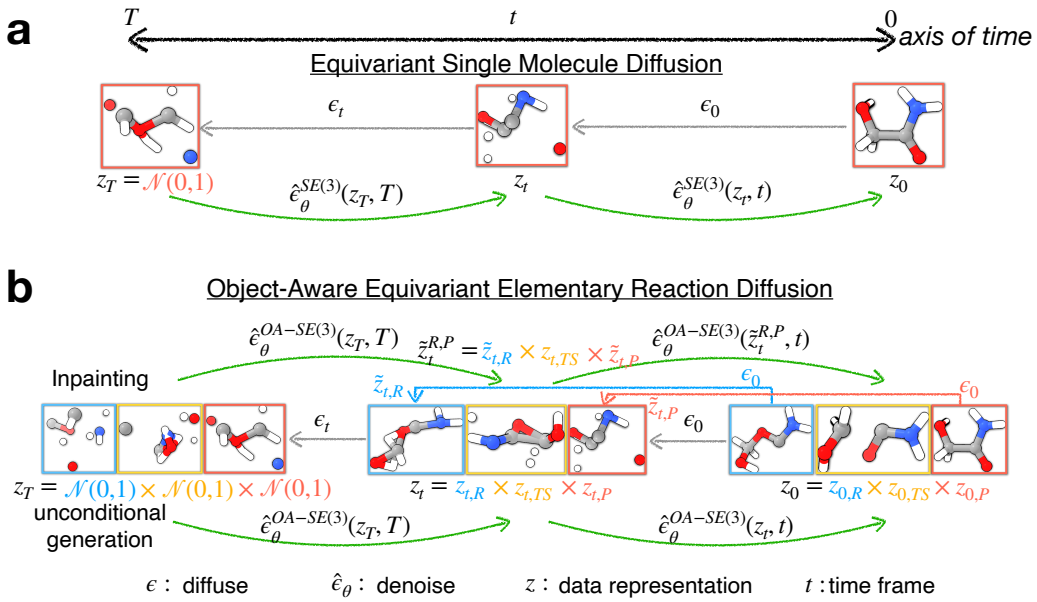

Fig. 1 | Overview of equivariant diffusion models for generative molecular system sampling. a. Equivariant diffusion model (EDM) for single molecule generation. In the diffusion process (gray arrows), the data distribution of molecules at time t (z_t) diffuses by a Gaussian noise (ϵ_t) from $t = 0$ to $t = T$ until z_T becomes a normal distribution. During the generation process (green arrows), a molecule is sampled from a normal distribution at $t = T$ after which an SE(3) equivariant denoising network ($\hat{\epsilon}_\theta^{SE(3)}$) is applied iteratively on the sample to remove noise, finally recovering the original data distribution at $t = 0$. **b.** Object-aware elementary reaction diffusion model (OA-ReactDiff), which generates a system as a joint distribution of multiple molecules that form a elementary reaction with reactant (R, blue), transition state structure (TS, orange), and product (P, red). Its diffusion process resembles that in EDM, while the sampling process requires that the denoising network respects object-level SE(3) equivariance (bottom). For conditional generation where part of the system is known *a priori*, a combined scheme named inpainting, where diffusing on known parts (e.g, R and P) and denoising on unknown parts (e.g., TS), will be used for recovering the original conditional joint distribution (top). Atoms are colored as follows: gray for C, blue for N, red for O, and white for H.

accuracy of generated 3D structures and reaction barrier estimate achieved by OA-ReactDiff provides the possibility of accelerating and even circumventing expensive quantum chemistry calculations normally required for TS search.

Results

Overview of OA-ReactDiff. A diffusion model contains two processes.[23–25] In the forward (i.e., diffusion) pass, Gaussian noise is continuously added to the original data distribution, which, over time, becomes an approximately normal distribution (Fig. 1a). In the reverse (i.e., sampling) pass, a random sample is drawn from the normal distribution, after which a denoising neural network is iteratively applied to remove noise, recovering the original data distribution. This denoising network is trained to predict the noise added to the original data distribution (see *Equivariant diffusion models*). Since a 3D molecule or macromolecule fulfills permutational, translational, and rotational symmetry, the denoising graph neural network (GNN) used in chemistry application requires SE(3) equivariance (Fig. 2a). For a molecule represented by atom types (i.e., scalars) and their Cartesian coordinates (i.e.,

vectors), as one applies an SE(3) transformation (e.g., rotation), the predicted noise on atom types should be the same while that on coordinates should undergo the same transformation.

Despite the success of SE(3) GNN-based equivariant diffusion models (EDM) in many chemistry applications, they inherently lack the symmetries required for systems containing multiple objects (e.g., molecules) whose interactions are independent of their coordinates in the 3D Euclidean space. An elementary reaction, as our system of interest, consists of three objects: reactant, TS, and product. If an SE(3) transformation (i.e., rotation) is applied on one of these object (e.g., reactant), the description of this elementary reaction should stay invariant/equivariant.[2,6] In addition, for a reactant or product that has multiple fragments, SE(3) transformations on individual fragments should also have no influence on the elementary reaction. A vanilla SE(3) GNN, however, would take these object-wise SE(3) transformations as if the entire system undergoes a non-SE(3) transformation and would yield non-equivariant results, breaking the symmetry required to predict the noise on atom types and Cartesian coordinates in EDMs (Fig. 2a).

There, we model an elementary reaction as a joint distribution of the 3D structures of the reactant, TS, and product (Fig. 1b). The diffusion process is essentially the same as the vanilla EDM, where independent Gaussian noise is added to reactant, TS, and product until they become independent normal distributions. In the denoising process, however, an object-aware SE(3) equivariant GNN is used to preserve correct physical symmetries and constraints in an elementary reaction (Supplementary Text S1). We consider two denoising schemes. One is unconditional generation where reactant, TS, and product are all sampled from the normal distribution, which can be used to generate new elementary reactions from scratch (Fig. 1b). In chemistry, however, many important applications are targeted for conditional generation, where some information of an elementary reaction of interest is known *a priori*. For example, in double-ended TS search, the 3D structure of both reactant and product is known, and the task is to find the unique corresponding TS structure. For these conditional generation tasks, we applied the inpainting scheme, which models the joint distribution of the reactant, TS, and product where the unknown objects are inpainted during the inference time.[34] In TS search, specifically, it combines distributions from the diffused reactant an product (i.e., known parts) and denoised TS structure (i.e, unknown parts) at each step before proceeding to the following denoising step (Fig. 1b, see *Inpainting for conditional generation*).

In an elementary reaction, any non-SE(3) transformation on a single object (e.g., reactant) should simultaneously influence all three objects, while any object-based SE(3) transformation on reactant, TS, and product should not change a reaction (Fig. 2a). While a vanilla SE(3) GNN fulfills the former requirement, it violates the latter symmetry as it considers all atoms in a system as belonging to the same molecule (Supplementary Table S1). Here, we achieve all required physical symmetries in elementary reactions by developing a general procedure to adapt any SE(3) equivariant GNN as object-aware SE(3) equivariant with minimal effort (Fig. 2b and Supplementary Text S1). In this procedure, we build an object-aware SE(3) interaction layer from a regular SE(3) update layer, a series of

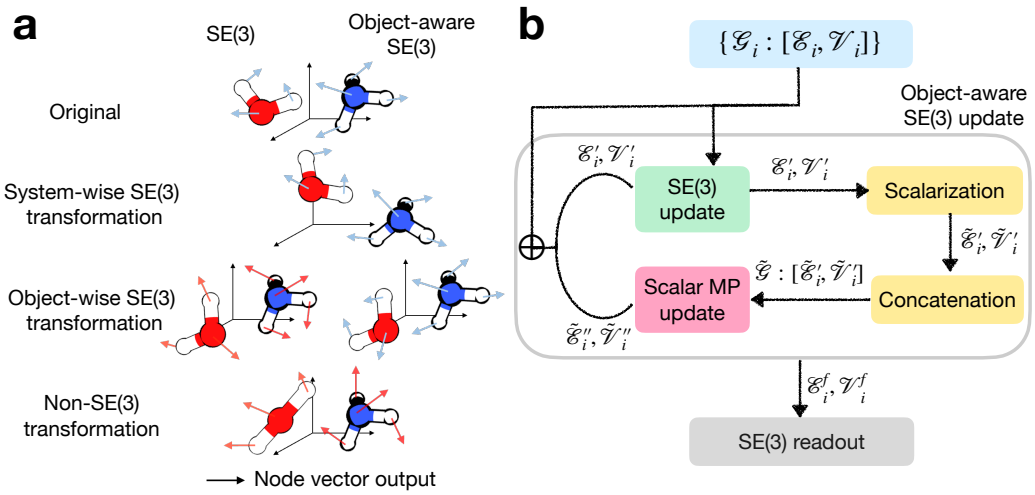

Fig. 2 | Object-aware SE(3) equivariance and its implementation based on SE(3) equivariant graph neural networks. a. Behavior of expected outputs for different transformations (suggested by different rows) on system with SE(3) (left) and object-aware SE(3) equivariance (right). Results that demonstrate the same behavior after transformation are grouped and only shown once for clarity. Node outputs that are preserved by symmetry (i.e., are the same as the original data) are represented by blue arrows, otherwise by red arrows. Atoms are colored as follows: blue for N, red for O, and white for H. **b.** Construction of Object-aware SE(3) equivariant graph neural networks. A system consists of multiple objects (i) are represented by a set of subgraphs $\mathcal{G}_i$ with edges $\mathcal{E}_i$ and nodes $\mathcal{V}_i$, which can be both scalars (e.g., atom types), vectors (e.g., Cartesian coordinates), or higher-order tensors. This set of subgraphs first go through an SE(3) equivariant block for message passing and updating both their scalar and higher-order tensor features (green). The resulting high-order tensors ($\mathcal{E}'_i$ and $\mathcal{V}'_i$) of all objects are scalarized and concatenated as a system-level fully-connected graph with only scalar representation $\tilde{\mathcal{G}}$ (yellow). This graph is then processed by a scalar message-passing (MP) block to include interactions among different objects (i) in the system (pink). The updated nodes and edges are combined with the outputs from equivariant update block as the input for the next object-aware SE(3) interaction block. The process repeats several times until the final object representations are readout (gray).

non-parameterized operations (i.e, scalarization and concatenation), and a scalar-only message-passing update. In essence, the SE(3) update only operates on individual objects, where the relative positioning is not encoded, to learn a comprehensive representation for each molecule, while the scalar-only message passing layer learns interactions among atoms from different molecules (see *Object-aware SE(3) implementation*). Similar to standard SE(3) GNNs, this object-aware SE(3) update repeats several times until a final SE(3) readout layer to pool out the final predicted noise on atom types and Cartesian coordinates. In this work, we choose LEFTNet,[35] our recently-developed SE(3) GNN that reaches comparable state-of-the-art performance on QM9[36] and MD17,[37] as the vanilla SE(3) GNN for OA-ReactDiff (see *LEFTNet*).

OA-ReactDiff training. We trained OA-ReactDiff on Transition1x,[15] a dataset that contains climbing-image NEB[38] calculated reactant, TS structure, and product at the ωB97x/6-31G(d) level of theory[39,40] on 10,073 organic reactions of various types originated from a quite exhaustive enumeration[41,42] of product-reactant pairs based on the GDB7[43] dataset. Each reaction consists of up to seven heavy atoms including C, N, and O, with the largest system consisting of 23 total atoms. The use of climbing-image NEB ensures a relatively accurate TS structure, making

each elementary reaction in Transition1x a unique set of reactant, TS, and product, which guarantees the necessary condition for training OA-ReactDiff. We trained OA-ReactDiff on 9,000 elementary reactions randomly partitioned from Transition1x, leaving 1,073 unseen reactions as the test set. Despite the potential overlap of certain chemical species in the training and test set, there are always at least two species (reactant and TS or TS and product) that are distinct in any test reaction compared to all training data (see *Details for model training*).

In OA-ReactDiff, a molecule is represented by atom types with one-hot encoding and nuclear charges and Cartesian coordinates of its constituent atoms. It is common to consider all components of the atom representation in the diffusion and denoising process as none of them is considered known *a priori*.[26,28] In chemical reactions, however, it is reasonable to assume that we know the atom types due to the conservation of atoms.[27] Therefore, we only diffused and denoised the Cartesian coordinates of the reactant, TS structure, and product in OA-ReactDiff (Fig. 1b). Since OA-ReactDiff satisfies all the symmetries and constraints for describing an elementary reaction, it does not require any pre-processing of reaction data, such as atom order matching for different species and careful alignment of reactants and products, which sometimes can be infeasible to obtain[21] (see *Equivariant diffusion models* and *Object-aware SE(3) implementation*). Due to the use of an object-aware SE(3) GNN, OA-ReactDiff breaks the reflection symmetry and thus can distinguish chiral molecules (see *LEFTNet*). OA-ReactDiff also bypasses the need for data augmentation in the case of reversing reaction direction by enforcing the same graph embedding layer for reactant and product, and thus guarantees the outputs are invariant to the order of reactant and product as inputs.[21] Lastly, there are no post-processing steps (e.g., reconstructing the 3D structure from a distance matrix through optimizations) required as OA-ReactDiff directly yields the Cartesian coordinates of reactant, TS structure, and product.[17] These outstanding features make OA-ReactDiff an end-to-end model for elementary reaction generation and TS search.

Overcoming the stochastic nature of diffusion models with confidence ranking. OA-ReactDiff models the joint distribution of a set of reactant, TS, and product, and thus can generate new elementary reactions without any conditions, including for those which the chemical composition is unseen during the model training (Supplementary Fig. S1). Evaluating the accuracy and building a reaction network from these generated elementary reactions, however, require significant computational resources for running DFT optimizations and may be subject to selection bias on which chemical compositions are included during evaluation. Therefore, we focus on evaluating OA-ReactDiff under the scheme of conditional generation, specifically for TS search where the task is to identify the 3D TS structure provided a pair of reactant and product.

We first consider three example reactions in Transition1x that break and form a varied number of bonds, representing different levels of complexity (Fig. 3a). Due to the stochastic nature of diffusion models, sampled TS structures from OA-ReactDiff will not be unique with a fixed reactant-product pair. For each of the three reactions,

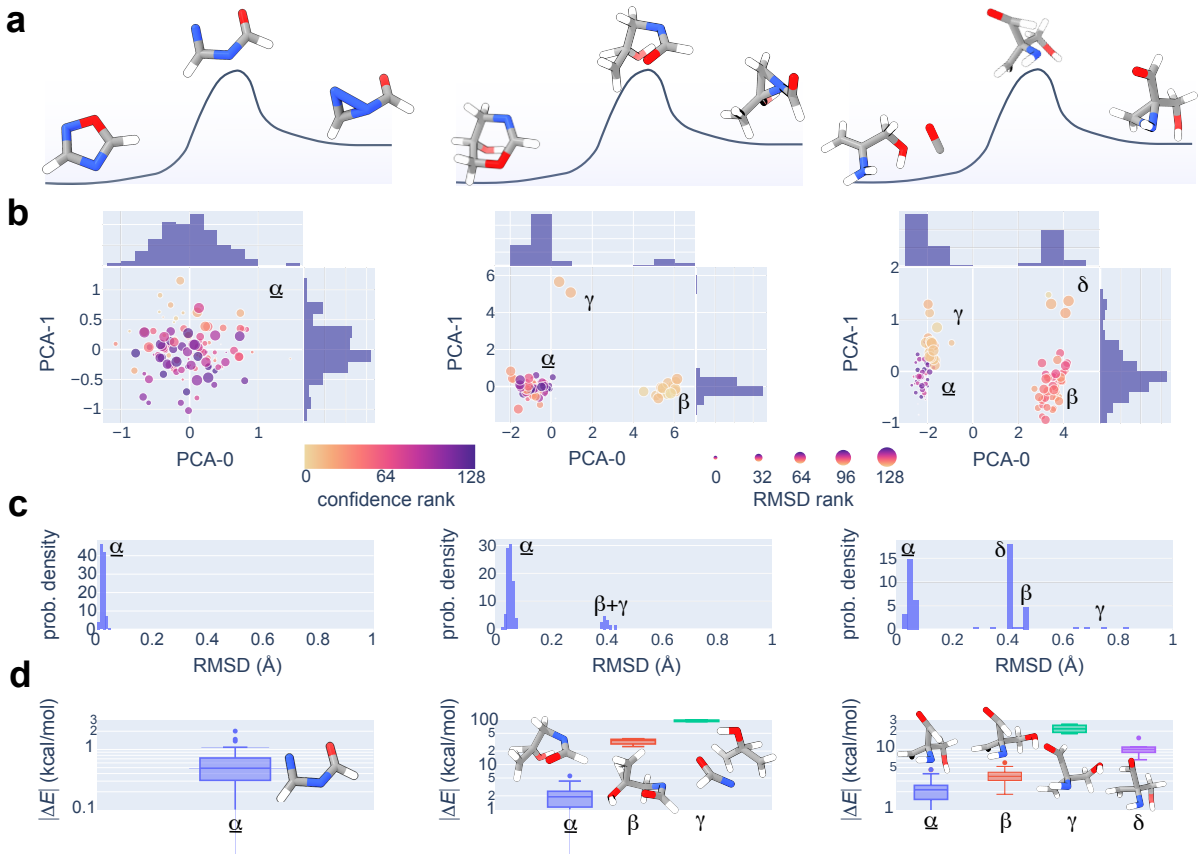

Fig. 3 | Analysis of samples generated by OA-ReactDiff on select elementary reactions. a. Illustration of select reactions: aromatic 5-membered ring to high-valence 3-membered ring rearrangement that breaks and forms one bond (left, b1f1), 6-membered ring to saturated 3-membered ring rearrangement (middle, b2f2), and carbonyl attack reaction (right, b2f3). All three elementary reactions have not been seen during the model training. **b.** Principal component analysis (PCA) on the 128 OA-ReactDiff generated TS structures. Each structure is colored by the rank of confidence score and sized by its rank of RMSD compared to the true TS structure. Structures that form clusters on the PCA space are labeled by Greek letters, following an order of α, β, γ, and δ. The cluster that has the most confident sample is always labeled by $\underline{\alpha}$ for each reaction. **c.** Distribution of RMSD between 128 OA-ReactDiff samples and the true TS structure for each reaction. Each peak in the histogram is labeled by the cluster observed in PCA. **d.** Absolute energy difference ($|\Delta E|$) between 128 OA-ReactDiff samples and the true TS structure for each reaction grouped by the cluster observed in PCA and the corresponding representative structure for each cluster. A log scale on $|\Delta E|$ is presented for better visibility of the difference among multiple clusters. Atoms are colored as follows: gray for C, blue for N, red for O, and white for H.

we ran the OA-ReactDiff under the inpainting scheme 128 times, generating 128 distinct samples. To evaluate the differences among these 128 samples, we featurized them with the Coulomb matrix[44] representation followed by principal component analysis (PCA) to visualize them in a 2D space. Interestingly, the number of distinct clusters observed in PCA increases with the increasing complexity of an elementary reaction (Fig. 3b,c), which can be rationalized by the increasing number of degrees of freedom involved during a reaction. However, there is only one unique TS structure connecting a pair of 3D geometries for reactant and product.

To address this challenge, we further trained an object-aware SE(3) LEFTNet as a confidence model,[27,45] which also satisfies all the symmetries and constraints for elementary reactions. There, provided a set of input reactant, TS,

and product, the confidence model predicts its probability of being a true elementary reaction. During its training, we provided elementary reactions sampled by OA-ReactDiff, which are labeled as good (i.e, 1) if the RMSD between sampled and true TS structure is < 0.2 Å and bad (i.e., 0) otherwise (see *Details for model training*). Once trained, the confidence model successfully distinguishes different 3D TS structures generated by OA-ReactDiff, assigning them a distinct probability score by the identity of clusters in the PCA (Fig. 3b). Moreover, the confidence model always give the highest probability score to generated structures with among the lowest RMSD with respect to the true TS structure (i.e., cluster $\underline{\alpha}$) for all three example reactions. Without the confidence model, random selection from samples generated by OA-ReactDiff may yield a large (> 10 kcal/mol) energy difference for the predicted and true TS structure, which would lead to orders of magnitude differences in predicted reaction rates (Fig. 3d). For example, there is roughly 20% chances that we obtain a TS structure with a > 50 kcal/mol energy difference for the 6-membered-to-3-membered ring conversion and a 25% chance for a TS structure with a > 10 kcal/mol energy difference for the carbonyl attack reaction (Fig. 3 middle and right). Even though it is not guaranteed that the confidence model always selects the sample with the lowest RMSD compared to the true TS structure, the confidence model will likely avoid choosing samples that have incorrect connectivity or geometries with large deviations, especially in reactions that have multiple bonds breaking and forming.

High quality TS structures from OA-ReactDiff. We next systematically evaluated the structural similarity between the OA-ReactDiff and true TS structures for 1,073 set-aside unseen reactions in Transition1x, as judged by RMSD. Notably, in contrast to an average runtime of 12 hours using climbing image NEB[22] with DFT, it only takes 6 seconds on a V100 GPU to generate a TS structure with OA-ReactDiff. Compared to bond lengths, angles, and dihedrals that mostly compare local geometry for a subset of atoms, the RMSD should provide a more accurate assessment on overall structural agreement, which is the ultimate goal of TS search.[1,2] For each reaction, we ran OA-ReactDiff 40 times, generating 40 independent guess TS structures. For a random selection of 40 samples, OA-ReactDiff has already reached an average RMSD of 0.183 Å with a median being 0.076 Å for the 1073 test elementary reactions (Fig. 4a). More than half (two thirds) of the TS structures have an RMSD < 0.1 (0.2) Å compared to their corresponding true TS structures identified by climbing image NEB.

With the confidence model, we can further improve the procedure of sample selection using a recommender approach.[46] Together with the true reactant and product, these guess TS structures are fed into the confidence model to get their probability score. The sample with highest probability score (i.e., top-1 confidence) is chosen as the final predicted TS structure from OA-ReactDiff. With this recommender approach, the quality of selected TS structures is greatly improved, most likely due to the removal of TS structures with incorrect connectivity and geometries with large deviations (Fig. 3). Moreover, the recommended structures mostly reside in the low RMSD and high confidence region, which demonstrates the effectiveness of our combined OA-ReactDiff and confidence recommender

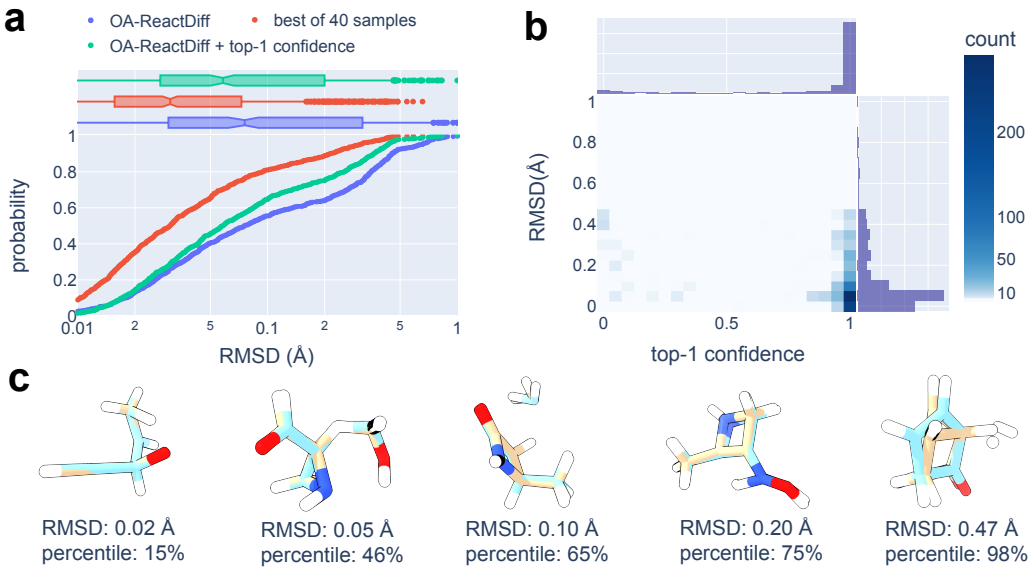

Fig. 4 | Evaluation of structural similarities for TS structures generated by OA-ReactDiff and true TS structures. a. Cumulative probability for RMSD between the true TS structures and OA-ReactDiff samples on 1073 set-aside test reactions. The OA-ReactDiff samples are evaluated under one-shot generation (blue), the top-1 confidence sample (i.e., recommended, green), and the best (yet unachievable) sample out of 40 generated samples for each reaction (red). A log scale of the RMSD is presented for better visibility of the low-RMSD region. **b.** 2D density map for the RMSD vs. top-1 confidence for OA-ReactDiff generated samples. A log-scale color gradient is applied to the color bar to reveal low-density areas, which would otherwise be difficult to distinguish. **c.** Example overlapping true and OA-ReactDiff + recommender TS structures at different RMSD and RMSD percentile rank shown in ascending order from left to right. Atoms are colored as follows: C in the true TS structure are in tan and those in the top-1 confident OA-ReactDiff sample are in skyblue; N for blue, O for red, and H for white.

approach (Fig. 4b). The average and median of the error in RMSD become 0.129 and 0.058 Å, respectively, with approximately two thirds (three fourths) of the recommended TS structures having RMSD < 0.1 (0.2) Å (Fig 4 a). We observed a systematically-improving performance for the OA-ReactDiff + recommender approach as the number of total independent runs increases (Supplementary Fig. S4). Here, we took 40 runs for each sample for a balance between total run time (4 minutes in total) and sampling accuracy. Despite the fact that the recommender is still far from perfect for distinguishing structures with low RMSD (i.e, < 0.2 Å) structures, it helps avoiding TS samples that are very different from the true TS (i.e., > 0.45 Å, Fig. 4c). This feature is particularly useful in end-to-end applications for ML models.

Approaching the energetic accuracy needed in TS search. Besides the exact 3D structure for a TS that provides insights into how an elementary reaction happens microscopically, a TS search algorithm should also evaluate the reaction barrier height, which is crucial for pruning large reaction networks and estimating reaction rates.[2] Here, we evaluate the performance of OA-ReactDiff on predicting the barrier height at the DFT level of theory. Specifically, we compared the electronic energy computed by ωB97x/6-31G(d) for OA-ReactDiff recommended and true TS structures. As one may expect, the absolute energy difference ($|\Delta E_{\mathrm{TS}}|$) between OA-ReactDiff and true TS structures

has positive correlation with their RMSD. We find they follow a power law best across various type of common algebraic fits, giving a Pearson's *r* of 0.56 on a log-log plot (Fig. 5a). This relatively low Pearson's *r* can be explained by the high complexity of potential energy surfaces for molecules, where the direction of displacing an atom has large influences on the energy change. For example, an OA-ReactDiff TS of $C_4H_6O_2$ that has a relatively low RMSD of 0.078 Å compared to the true TS was found to have the highest energy difference (49.3 kcal/mol) among the 1073 test reactions(Fig. 5a). This counter-intuitive result, however, is an artifact due to self-consistent field calculations being converged to different local minima in DFT energy evaluation for OA-ReactDiff and true TS (Supplementary Fig. S2). On the other hand, despite an extremely large RMSD (0.821 Å) between OA-ReactDiff and the true TS structure for $C_6H_{10}O$, the energy difference between the two structures is only 0.4 kcal/mol, due to the fact that this TS consists of two fragments that only weakly interact with each other (Fig. 5a). There, OA-ReactDiff provides quite precise geometry for both fragments, in spite of the incorrect orientation between the two fragments, leading to a small deviation in barrier estimation (Supplementary Fig. S3).

With OA-ReactDiff and the recommender, we reach an average of 4.4 kcal/mol and median of 1.6 kcal/mol for the absolute energy difference between the generated and true TS structure, with 71% of TS barrier errors < 3 kcal/mol (Fig. 5b). OA-ReactDiff + recommender far outperforms semi-empirical methods such as density functional tight binding,[47] which has an MAE of 16.1 kcal/mol and average runtime of 82 seconds.[22] Interestingly, the MAE would only improve marginally to 4.0 kcal/mol if we were able to select the OA-ReactDiff sample with the lowest RMSD compared to the true TS, indicating the power of the recommender for selecting TS structures with low energy deviations (Table 1). The performance of OA-ReactDiff with the recommender on elementary reactions with multiple reactants and/or products is comparable to the performance for rearrangement reactions that only contain one single reactant and product (Supplementary Fig. S5) . The slight deterioration of the performance is likely due to the imbalance of reaction types included in Transition1x, where only one fourth of the elementary reactions contain multiple reactants or products.

We also compare OA-ReactDiff + recommender with two pioneering works where non-diffusion-based approaches were developed for generating 3D TS structures on large diverse organic reaction datasets such as Transition1x[15] or its predecessor.[41] Choi developed a "PSI-based" model[21] combining transformer and bidirectional gated recurrent unit that generates TS structures from refining the linear interpolation of reactant and product, which, however, requires the prior knowledge of atom mapping and careful alignment between reactant and product. Schreiner *et al.*[22] trained a machine learning potential on 10M structures (with both energy and forces) collected during the generation of Transition1x, and, for the first time, applied the trained potential to the TS search problem in place of DFT. There, similar to the problem of DFT-based TS search (e.g., NEB), an attempt may still encounter convergence issues during the saddle point optimization, leading to a null prediction for the final TS structure. We

find OA-ReactDiff + recommender systematically outperforms the prior approaches on both the RMSD and barrier height estimate in terms of both the mean and median of the error distribution (Table 1). This superior performance is attributed the fact that OA-ReactDiff manages to respect all physical symmetries and constraints for describing an elementary reaction, without the need for atom order mapping, reactants or products alignment, reconstruction of 3D geometry from distance matrix, and data augmentation of any kind. In addition, a more gradual increase of absolute energy difference with respect to RMSD was identified in OA-ReactDiff + recommender compared with the two other approaches, suggesting a more accurate barrier estimate can be obtained by OA-ReactDiff at the same level of structural similarity between generated and true TS (Table 1 and Supplementary Fig. S6).

We would ideally aim to select one single TS structure sampled by OA-ReactDiff with the recommender. The recommended sample, however, may not be confident if all 40 samples generated by OA-ReactDiff suffer from a low confidence score (i.e, $p < 0.5$) due to the limited amount of training data. Further removal of these reactions (153, or 14%) from the test set leads to a significantly improved energy difference with a mean of 3.1 kcal/mol and median of 1.4 kcal/mol (Table 1). Moreover, we observe a monotonic behavior between the MAE for barrier height estimates and the confidence threshold imposed for TS structure generation that we consider as valid (Fig. 5c). This desired monotonic behavior suggests that we can use the confidence score for uncertainty quantification to balance the accuracy and number of DFT calculations required in a practical workflow[48] that combines OA-ReactDiff, recommender, and DFT-based TS search. For a set of TS structures generated by OA-ReactDiff and their corresponding confidence score evaluated by the confidence model and recommender, we can decide whether we would accept the recommended TS structure depending on its confidence score or would rather launch a DFT-based NEB. With a confidence threshold of 0.5, we would only perform NEB with DFT on 14% of reactions while directly accepting TS structures from OA-ReactDiff + recommender for the remaining 86% reactions, leading to an overall accuracy of 2.6 kcal/mol. This strategy showcases the power of combining OA-ReactDiff, recommender, and DFT-based NEB for efficient generation of TS structures given a target accuracy level.

Table 1 | Summary of statistics for RMSD and absolute energy differences of TS structures and the linear fitting results obtained by various approaches. The linear fit is obtained between $\log(|\Delta E_{\mathrm{TS}}|)$ vs. $\log(\mathrm{RMSD})$. Some approaches invoke uncertainty quantification (OA-ReactDiff with $p > 0.5$) or subject to convergence issues during optimization (PSI-based model and NeuralNEB), leading to inconclusive predictions some elementary reactions and thus resulting an effective data fraction < 1.

Approach	RMSD (Å)		$\lvert\Delta E_{\mathrm{TS}}\rvert$ (kcal/mol)		fitting coefficeint			data fraction
	mean	median	mean	median	slope	intercept	Pearson's *r*	
OA-ReactDiff	0.183	0.076	6.2	1.7	0.65	0.99	0.59	1.00
OA-ReactDiff + rec.	0.129	0.058	4.4	1.6	0.63	0.99	0.56	1.00
OA-ReactDiff + rec. ($p > 0.5$)	0.106	0.047	3.1	1.4	0.55	0.88	0.53	0.86
OA-ReactDiff + best	0.071	0.031	4.0	1.3	0.72	1.19	0.65	1.00
OA-ReactDiff + best ($p > 0.5$)	0.044	0.021	1.8	1.1	0.56	0.92	0.56	0.74
PSI-based model[21]	0.144	0.122	13.4	8.4	0.96	1.82	0.69	0.96
NeuralNEB[22]	0.136	0.096	6.5	2.1	1.26	1.66	0.77	0.83

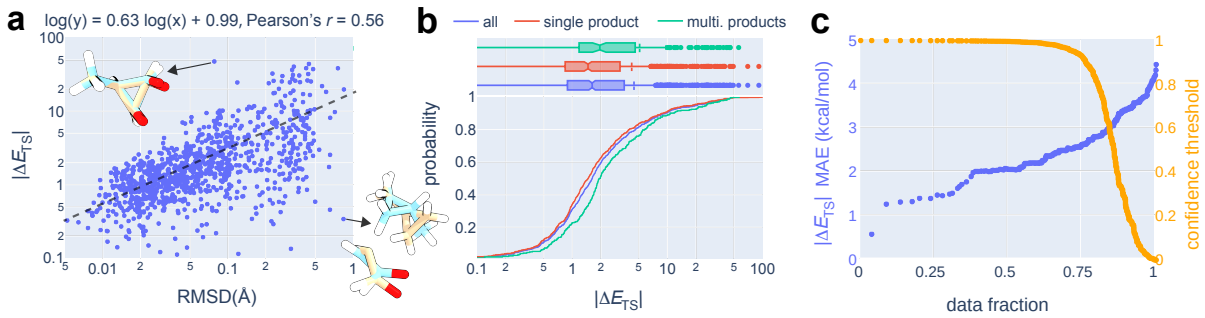

Fig. 5 | Energetic performance for OA-ReactDiff + recommender TS structures. a. Absolute energy error ($|\Delta E_{\text{TS}}|$) vs. RMSD for top-1 confidence sample in log-log scale. A linear regressed black dashed line is shown with the fitted parameters at the top. Two example TS structures that deviate the most from this linear fit are shown. **b.** Cumulative probability for absolute energy error of top-1 confidence sample. The errors are evaluated by 1) all 1073 reactions in the test set (blue), 2) 783 reactions that involves only one single product (red), and 3) 290 reactions that yield multiple products (green). **c.** MAE of $|\Delta E_{\text{TS}}|$ (blue, left y-axis) and the corresponding confidence threshold (orange, right y-axis) as a function of the fraction of data considered in the 1073 test reactions. Atoms are colored as follows: C in the true TS structure are in tan and those in the top-1 confident OA-ReactDiff sample are in skyblue; N for blue, O for red, and H for white.

Discussion

Elucidating TS structures is essential for uncovering the underlying microscopic mechanisms of chemical reactions and estimating reaction barriers for building large reaction networks. In this work, we extended SE(3) equivariant diffusion models to respect the object-wise symmetries, leading to OA-ReactDiff, an object-aware SE(3) equivariant diffusion model that first fulfills all the symmetries and constraints for generating elementary reactions. In addition, we built a confidence-model-enabled recommender to overcome the stochastic nature of diffusion model to select from sampled generated by OA-ReactDiff in multiple runs. OA-ReactDiff + recommender gives an RMSD of 0.129 Å and MAE of 4.4 kcal/mol compared to the true TS structure obtained by computationally demanding climbing image NEB calculations. By further using the confidence score for uncertainty quantification, we can selectively perform climbing image NEB only for 14% of elementary reactions that OA-ReactDiff is most uncertain about, leading to a reduced MAE of 2.6 kcal/mol.

The current OA-ReactDiff approach has two major limitations. First, we describe an elementary reaction as a set of 3D structures (say N atoms for reactant, TS, and product), which leads to a system that is 3x larger (i.e., $3N$ atoms). Although the most expensive equivariant update is still object-wise (i.e., scales with N), the scalar message-passing update requires building a fully-connected graph for the $3N$ atoms, which will be the bottleneck for applying OA-ReactDiff on chemical systems > 100 atoms on a single GPU. Second, despite the workaround of using a confidence model and recommender to select a unique sample generated by OA-ReactDiff, the stochastic nature of diffusion model cannot be avoided. This leads to uncertainty for the sample quality of generated TS structure and accumulated runtime for running OA-ReactDiff repeatedly. These limitations are inherent for diffusion models,

which can potentially be addressed by reformulating elementary reaction generation as a transport problem, where optimal transport *via* flow matching or Schrödinger bridge can be applied. The authors are actively exploring along this direction as a future work.

Together with uncertainty quantification, OA-ReactDiff + recommender reached both the structural and energetic accuracy required in TS search, which can be readily integrated in current high throughput computation workflows for reaction network exploration. In this work, we focus on the relatively well defined TS search problem such that we can evaluate our newly-developed OA-ReactDiff more easily and demonstrate the promise of this new model. OA-ReactDiff, however, models the joint distribution of structures in elementary reactions and thus is not limited to double-ended TS search problem and can be applied in single-ended (i.e, only the reactant is provided) or zero-ended (i.e., only the chemical composition of a system is provided) scenarios. Very recently, a more diverse elementary reaction dataset 17 times larger than Transition1x, named as RGD1, has been established.[49] Provided that the quality of diffusion model is highly dependent on the size of training data, RGD1 has the potential of unleashing the power of OA-ReactDiff for establishing large reaction networks and exploring chemical reactions with unknown mechanisms with a greatly reduced number of DFT calculations. Lastly, despite solely focusing on chemical reactions, the object-aware SE(3) equivariant diffusion model developed in this work can be applied to diverse chemical problems where the system of interest consists of multiple 3D objects, in which their interactions do not depend on their locations in Euclidean space.

Methods

Equivariant diffusion models. *Equivariance.–* A function f is said to be equivariant to a group of actions G if $g \circ f(x) = f(g \circ x)$ for any $g \in G$ acting on x.[50,51] In this paper, we specifically consider the Special Euclidean group in 3D space (SE(3)) which includes permutation, translation and rotation transformations. We intentionally break the reflection symmetry so that our model can describe molecules with chirality.

Diffusion models.– Diffusion models are originally inspired from non-equilibrium thermodynamics.[23–25] A diffusion model has two processes, the forward (diffusing) process and the reverse (denoising) process. The noise process gradually adds noise into the data until it becomes a prior (Gaussian) distribution:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t|\alpha_t x_{t-1}, \sigma_t^2 I),$$

where α_t controls the signal retained and σ_t controls the noise added. A signal-to-noise ratio is defined as $\text{SNR}(t) = \frac{\alpha_t^2}{\sigma_t^2}$. We set $\alpha_t = \sqrt{1-\sigma_t^2}$ following the variance preserving process in.[25]

The *true denoising process* can be written in a closed form due to the property of Gaussian noise:

$$q(x_s|x_0, x_t) = \mathcal{N}(x_s|\mu_{t\to s}(x_0, x_t), \sigma^2_{t\to s} I),$$
$$\mu_{t\to s}(x_0, x_t) = \frac{\alpha_{t|s}\sigma_s^2}{\sigma_t^2} x_t + \frac{\alpha_s \sigma^2_{t|s}}{\sigma_t^2} x \text{ and } \sigma_{t\to s} = \frac{\sigma_{t|s}\sigma_s}{\sigma_t},$$

where $\alpha_{t|s} = \frac{\alpha_t}{\alpha_s}$, $\sigma^2_{t|s} = \sigma^2_t - \alpha^2_{t|s}\sigma^2_s$ and $s < t$. However, this *true denoising process* is dependent on x_0 which is the data distribution and not accessible. Therefore, diffusion learns the denoising process by replacing x_0 with $\hat{x} = \epsilon_\theta(x_t, t)$ predicted by a denoising network ϵ_θ. The training objective is to maximize the variational lower bound (VLB) on the likelihood of the training data:

$$-\log p(x) \leq D_{KL}(q(x_T|x_0)||p_\theta(x_T)) - \log p(x_0|x_1) + \Sigma_{t=2}^T D_{KL}(q(x_{t-1}|x_0, x_t)||p_\theta(x_{t-1}|x_t))$$

Empirically, a simplified objective has been found to be efficient to optimize:[23]

$$\mathcal{L}_{\text{simple}} = \frac{1}{2}||\epsilon - \epsilon_\theta(x_t, t)||^2,$$

Equivariant diffusion models.– To build an SE(3)-equivariant diffusion model, it has been proven that we need an SE(3)-invariant prior and an SE(3)-equivariant transition kernel.[52] To guarantee equivariance on permutation, rotation, and translation, a necessary condition is to use an SE(3)-equivariant transition kernel (i.e. denoising network), as we will explain in details at a later section. (see *LEFTNet*). There are additional requirements for rotation and translation. For rotations, the isotropic Gaussian prior has the nice property to transform equivariantly. For translations, we need to limit the distribution on the linear subspace where the center of mass is the origin.[52]

Inpainting for conditional generation. Inpainting is a flexible technique to formulate the conditional generation problem for diffusion models.[34] Instead of modeling the conditional distribution, inpainting models the joint distribution during training. During inference, inpainting methods combine the conditional input as part of the context through the noising process of the diffusion model before denoising both the conditional input and the inpainting region together. The resampling technique[34] has demonstrated excellent empirical performance in harmonizing the context of the denoising process as there is sometimes mismatch between the noised conditional input and the denoised inpainting region. Specifically, resampling increases the total number of sampling steps in each denoising step by sampling the inpainting region back and forth together with the conditional input. Despite resampling increases the number of total denoising steps, this can be compensated by decreasing the number of total denoising steps accordingly by striding the sampling schedule[53] without significantly sacrificing the model performance.

LEFTNet. We build our denoising network on top of a recently proposed SE(3)-equivariant GNN, LEFTNet.[35] The main idea of LEFTNet relies on building local frames to scalarize the vector (e.g. position, velocity) and higher order tensor (e.g. stress) which becomes invariant to SE(3) transformations. Tensorization can be applied to invert the scalar back to vector and higher order tensor without information loss in each layer to update these quantities. The benefit of scalarization is demonstrated by the flexibility of neural network parameterizations without breaking the symmetry and further proved by the universal approximation theorem[54] such that the resulting neural network has the universality in the space of continuous SE(3) and permutation equivariant functions.

Scalarization and tensorization.– Scalarization and tensorization are two operations in differential geometry to convert geometric quantities. Specifically, scalarization transforms geometric tensors into scalars while tensorization is the inverse of scalarization, transforming scalars back to geometric tensors. In this case, scalarization is used to transform equivariant quantities by three equivariant orthonormal frames:

$$\mathcal{F} := (e_1, e_2, e_3)$$

For simplicity, we use vector as an example. The geometric tensors are scalarized by the inner product between the frames $\mathcal{F}$ and the input vector x as follows:

$$\mathbf{x} := (x \cdot e_1, x \cdot e_2, x \cdot e_3)$$

On the contrary, tensorization reverses the process by:

$$x := \mathbf{x_1} e_1 + \mathbf{x_2} e_2 + \mathbf{x_3} e_3.$$

where $(\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3})$ is the input scalar tuple and x is the converted tensor.

Message passing neural network (MPNN).– A molecular graph can be represented as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V}$ is a set of nodes or atoms and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is a set of edges or bonds connecting pairs of nodes. For each node, we have atom types $h \in \mathbb{R}^{n \times k}$, where n denotes number of nodes and k denotes number of atom types. Edge features are attached to the edges connecting nodes i and j as e_{ij}. Message passing neural network is a common framework to learn embeddings over graphs.[55] MPNNs often have three parts: (1) message, (2) update and (3) readout. The common message between each pair of nodes are:

$$m_i = \sum_{j \in \mathcal{N}(i)} M(h_i, h_j, e_{ij}),$$

where M is the message function, $\mathcal{N}(i)$ denotes neighbors of node i and m_i is the message. Then the message is used to update the node feature as:

$$h_i = U(h_i, m_i),$$

where U is the update function. After a number of layers, the global embedding is calculated by:

$$g = R(h_i | i \in \mathcal{G}),$$

where R is the readout function.

In our case, the molecular graph also has atomic coordinates $x \in \mathbb{R}^{n\times 3}$ and the message function of the MPNN needs to be equivariant to SE(3) transformations. It is achieved by moving the center of mass to the origin (translations) and incorporating the scalarized coordinates s_{ij} in the message function (rotations):

$$m_i = \sum_{j\in\mathcal{N}(i)} M(h_i, h_j, s_{ij}, e_{ij}),$$

where s_{ij} is obtained by scalarizing the input coordinate over the frames $\mathcal{F}_{ij}$ between each pair of nodes i and j:

$$e_1 = \frac{x_i - x_j}{||x_i - x_j||}, e_2 = \frac{x_i \times x_j}{||x_i \times x_j||}, e_3 = e_1 \times e_2,$$
$$s_{ij} = (x_i \cdot e_1, x_i \cdot e_2, x_i \cdot e_3)$$

Building towards LEFTNet.– Motivated by distinguishing local 3D geometric isomorphisms, LEFTNet introduced a local structure encoding module to encode the local atomic environment of each atom in the scalarization operation. In addition, LEFTNet designed another frame transition encoding block to consider the transition between two frames (the atom and its neighbor atom) when calculating the message between them.

Object-aware SE(3) implementation. In general, a system consists of multiple objects (molecules or proteins) that do not have interactions through the 3D Euclidean space can be described by a set of independent graphs, $\{\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i)\}$. In elementary reaction, for example, we have three objects which can index, reactant ($i = 0$), TS ($i = 1$), and product ($i = 2$). One important addition symmetry for these systems is object-wise SE(3) equivariance, meaning any SE(3) transformation on each individual object in a system should not influence its description:

$$\{g_i \circ f(\mathcal{G}_i)\} = \{f(g_i \circ \mathcal{G}_i)\}$$

g_i represents an SE(3) transformation on i^{th} object, which is not necessarily the same for all objects. On the other hand, any non-SE(3) transition on any objects in a system should influence its description:

$$\{g_0 \circ f(\mathcal{G}_0), ..., q_i \circ f(\mathcal{G}_i), ..., g_n \circ f(\mathcal{G}_n)\} \neq \{f(g_0 \circ \mathcal{G}_0), ..., f(q_i \circ \mathcal{G}_i), ..., f(g_n \circ \mathcal{G}_n)\}$$

where q_i represents a non-SE(3) transformation on i^{th} object. An SE(3) GNN would hold for the latter but violate the former symmetry.

To simultaneously fulfill these two requirements (along with other symmetries naturally fulfilled by SE(3) equivariant diffusion models), we developed a generic approach to adapt any SE(3) GNN to its object-aware SE(3) equivalent. The essence of this approach is to still perform equivariant update using an SE(3) GNN for individual object, but only allow scalar-type message passing among different objects to avoid the "leak" of their relative position information. Starting from the original graph representations, $\{\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i)\}$, $\mathcal{G}_i$ first gets updated by an SE(3) equivariant block for message passing. The resulting graphs ($\mathcal{G}_i'$) of all objects are scalarized and concatenated as a system-level fully-connected graph with only scalar representation $\tilde{\mathcal{G}}$. All the scalar node features in $\tilde{\mathcal{G}}$ is then updated by a scalar message-passing block. This way, interactions among different objects (i) in the system are included without introducing the positioning of different objects as all high-order tensors have been scalarized. Finally, the updated nodes scalars are combined with the outputs from equivariant update block. This constitutes an object-aware SE(3) interaction block built on top of a vanilla SE(3) update function. Similar to SE(3) GNNs, this interaction block repeats several times until the final graph representations are readout.

Details for model training. *Dataset and train/test partitioning.*– Built on top a large chemically diverse dataset by Grambow *et al.*,[41] Transition1x[15] dataset consists of 10,073 elementary reactions optimized by climbing image NEB. We partitioned Transition1x randomly, with 9,000 reactions used in training and validation and the remaining 1,073 reactions as set-aside test set. It is not guaranteed that all species in test reactions are unseen by a trained model due to the overlapping structures in different reactions. However, due to the uniqueness of elementary reaction, there is, at most, one chemical species (specifically, reactant or product) that may overlap in multiple reactions. We think this partition is reasonable because all TS structures in the test set are completely new and unseen from model training. In addition, having a certain degree of overlap in reactants/products for the training and set-aside test set is useful to judge whether the diffusion model only memorizes training samples rather than learning to generate new samples.

OA-ReactDiff training.– We trained OA-ReactDiff with LEFTNet as our vanilla SE(3) equivariant GNN. We used a set of hyperparameters similar to that for QM9 dataset in the original paper,[35] with 96 radial basis functions, 196 hidden channels for message passing, and 6 equivariant update blocks. A large neighbor cutoff threshold of 10 Å is used to impose fully connected graphs within each molecule. We mostly adopted hyperparameters of the diffusion process from the EDM paper,[26] where a polynomial noise schedule and L_{simple} loss function is used. We observed a

marginal improvement in model performance as we increase the total diffusion steps and used 5000 steps for our final model. We used a learning rate of 0.0005 and a batch size of 32, which is the largest batch size that we can afford with a V100/16GB GPU. The OA-ReactDiff model was trained for 2,000 epochs. During the training, the 9,000 reactions were further partitioned by a 8:1 ratio as training and validation. In practice, however, we observed that early stopping is not required due to the near monotonic decreasing loss for both the training and validation data during the entire training process.

Confidence model training.– The confidence model shares exactly the same set of hyperparameters as the scoring network, with the only change being the use of *sigmoid* function at the final output layer. To get data for training the confidence model, we ran OA-ReactDiff on the 9,000 training reactions for 40 runs, generating 360,000 synthetic reactions. We labeled a reaction as "good" (i.e., 1) if the generated TS structure has a RMSD < 0.2 Å compared to the true TS, and labeled it as "bad" (i.e., 0) otherwise. Lastly, we train the confidence model as a binary classifier, where the predicted probability is used as the confidence score to estimate the quality of generated TS structure. Note that we used the same partition for both scoring network and confidence model, which ensures the 1,073 reactions in the set-aside test set are unseen to both models during evaluation.

Code and data availability

Source code and data to reproduce results of this work are under review and will be open-sourced upon publication.

References

[1] Dewyer, A. L., Argüelles, A. J. & Zimmerman, P. M. Methods for exploring reaction space in molecular systems. *WIREs Comput. Mol. Sci.* **8**, e1354, DOI: https://doi.org/10.1002/wcms.1354 (2018). https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1354.

[2] Unsleber, J. P. & Reiher, M. The exploration of chemical reaction networks. *Annu. Rev. Phys. Chem.* **71**, 121–142, DOI: 10.1146/annurev-physchem-071119-040123 (2020). PMID: 32105566, https://doi.org/10.1146/annurev-physchem-071119-040123.

[3] Truhlar, D. G., Garrett, B. C. & Klippenstein, S. J. Current status of transition-state theory. *J. Phys. Chem.* **100**, 12771–12800, DOI: 10.1021/jp953748q (1996). https://doi.org/10.1021/jp953748q.

[4] Mardirossian, N. & Head-Gordon, M. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Mol. Phys.* **115**, 2315–2372, DOI: 10.1080/00268976.2017.1333644 (2017). https://doi.org/10.1080/00268976.2017.1333644.

[5] Durant, J. L. Evaluation of transition state properties by density functional theory. *Chem. Phys. Lett.* **256**, 595–602, DOI: 0.1016/0009-2614(96)00478-2 (1996).

[6] Simm, G. N., Vaucher, A. C. & Reiher, M. Exploration of reaction pathways and chemical transformation networks. *J. Phys. Chem. A* **123**, 385–399, DOI: 10.1021/acs.jpca.8b10007 (2019). https://doi.org/10.1021/acs.jpca.8b10007.

[7] Unsleber, J. P. *et al.* High-throughput ab initio reaction mechanism exploration in the cloud with automated multi-reference validation. *J. Phys. Chem.* **158**, 084803, DOI: 10.1063/5.0136526 (2023). https://doi.org/10.1063/5.0136526.

[8] Wang, L.-P. *et al.* Discovering chemistry with an ab initio nanoreactor. *Nat. Chem.* **6**, 1044–1048, DOI: 10.1038/nchem.2099 (2014).

[9] Pieri, E. *et al.* The non-adiabatic nanoreactor: towards the automated discovery of photochemistry. *Chem. Sci.* **12**, 7294–7307, DOI: 10.1039/D1SC00775K (2021).

[10] Zeng, J., Cao, L., Xu, M., Zhu, T. & Zhang, J. Z. H. Complex reaction processes in combustion unraveled by neural network-based molecular dynamics simulation. *Nat. Commun.* **11**, 5713, DOI: 10.1038/s41467-020-19497-z (2020).

[11] Van de Vijver, R. & Zádor, J. Kinbot: Automated stationary point search on potential energy surfaces. *Comput. Phys. Commun.* **248**, 106947, DOI: https://doi.org/10.1016/j.cpc.2019.106947 (2020).

[12] von Lilienfeld, O. A., Müller, K.-R. & Tkatchenko, A. Exploring chemical compound space with quantum-based machine learning. *Nat. Rev. Chem.* **4**, 347–358, DOI: 10.1038/s41570-020-0189-9 (2020).

[13] Margraf, J. T., Jung, H., Scheurer, C. & Reuter, K. Exploring catalytic reaction networks with machine learning. *Nat. Catal.* **6**, 112–121, DOI: 10.1038/s41929-022-00896-y (2023).

[14] Sheppard, D., Terrell, R. & Henkelman, G. Optimization methods for finding minimum energy paths. *J. Chem. Phys.* **128**, 134106, DOI: 10.1063/1.2841941 (2008). https://doi.org/10.1063/1.2841941.

[15] Schreiner, M., Bhowmik, A., Vegge, T., Busk, J. & Winther, O. Transition1x - a dataset for building generalizable reactive machine learning potentials. *Sci. Data* **9**, 779, DOI: 10.1038/s41597-022-01870-w (2022).

[16] Zhao, Q. & Savoie, B. M. Simultaneously improving reaction coverage and computational cost in automated reaction prediction tasks. *Nat. Comput. Sci.* **1**, 479–490, DOI: 10.1038/s43588-021-00101-3 (2021).

[17] Lemm, D., von Rudorff, G. F. & von Lilienfeld, O. A. Machine learning based energy-free structure predictions of molecules, transition states, and solids. *Nat. Commun.* **12**, 4468, DOI: 10.1038/s41467-021-24525-7 (2021).

[18] Zhang, J. *et al.* Deep reinforcement learning of transition states. *Phys. Chem. Chem. Phys.* **23**, 6888–6895, DOI: 10.1039/D0CP06184K (2021).

[19] Pattanaik, L., Ingraham, J. B., Grambow, C. A. & Green, W. H. Generating transition states of isomerization reactions with deep learning. *Phys. Chem. Chem. Phys.* **22**, 23618–23626, DOI: 10.1039/D0CP04670A (2020).

[20] Makoś, M. Z., Verma, N., Larson, E. C., Freindorf, M. & Kraka, E. Generative adversarial networks for transition state geometry prediction. *J. Chem. Phys.* **155**, 024116, DOI: 10.1063/5.0055094 (2021). https://doi.org/10.1063/5.0055094.

[21] Choi, S. Prediction of transition state structures of gas-phase chemical reactions via machine learning. *Nat. Commun.* **14**, 1168, DOI: 10.1038/s41467-023-36823-3 (2023).

[22] Schreiner, M., Bhowmik, A., Vegge, T., Jørgensen, P. B. & Winther, O. Neuralneb—neural networks can find reaction paths fast. *Mach. Learn. Sci. Technol.* **3**, 045022, DOI: 10.1088/2632-2153/aca23e (2022).

[23] Ho, J., Jain, A. & Abbeel, P. Denoising diffusion probabilistic models. In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. & Lin, H. (eds.) *Advances in Neural Information Processing Systems*, vol. 33, 6840–6851 (Curran Associates, Inc., 2020).

[24] Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N. & Ganguli, S. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, 2256–2265 (2015).

[25] Song, Y. *et al.* Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations* (2020).

[26] Hoogeboom, E., Satorras, V. G., Vignac, C. & Welling, M. Equivariant diffusion for molecule generation in 3D. In *Proceedings of the 39th International Conference on Machine Learning*, 8867–8887 (2022).

[27] Corso, G., Stärk, H., Jing, B., Barzilay, R. & Jaakkola, T. DiffDock: Diffusion steps, twists, and turns for molecular docking. *arXiv:2210.01776* (2023).

[28] Schneuing, A. *et al.* Structure-based drug design with equivariant diffusion models. *arXiv:2210.13695* (2022).

[29] Thomas, N. *et al.* Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds. *arXiv:1802.08219* (2018).

[30] Satorras, V. G., Hoogeboom, E. & Welling, M. E(n) equivariant graph neural networks. In *Proceedings of the 38th International Conference on Machine Learning*, 9323–9332 (2021).

[31] Batzner, S. *et al.* E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **13**, 2453, DOI: 10.1038/s41467-022-29939-5 (2022).

[32] Zhou, H.-C., Long, J. R. & Yaghi, O. M. Introduction to metal–organic frameworks. *Chem. Rev.* **112**, 673–674, DOI: 10.1021/cr300014x (2012). https://doi.org/10.1021/cr300014x.

[33] Park, J., Sung, G., Lee, S., Kang, S. & Park, C. ACGCN: Graph convolutional networks for activity cliff prediction between matched molecular pairs. *J. Chem. Inf. Model.* **62**, 2341–2351, DOI: 10.1021/acs.jcim.2c00327 (2022). https://doi.org/10.1021/acs.jcim.2c00327.

[34] Lugmayr, A. *et al.* Repaint: Inpainting using denoising diffusion probabilistic models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, DOI: 10.1109/CVPR52688.2022.01117 (2022).

[35] Du, W. *et al.* A new perspective on building efficient and expressive 3D equivariant graph neural networks. *arXiv:2304.04757* (2023).

[36] Ramakrishnan, R., Dral, P. O., Rupp, M. & von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **1**, 140022, DOI: 10.1038/sdata.2014.22 (2014).

[37] Chmiela, S. *et al.* Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **3**, e1603015, DOI: 10.1126/sciadv.1603015 (2017). https://www.science.org/doi/pdf/10.1126/sciadv.1603015.

[38] Henkelman, G., Uberuaga, B. P. & Jónsson, H. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *J. Phys. Chem.* **113**, 9901–9904, DOI: 10.1063/1.1329672 (2000). https://doi.org/10.1063/1.1329672.

[39] Chai, J.-D. & Head-Gordon, M. Systematic optimization of long-range corrected hybrid density functionals. *J. Phys. Chem.* **128**, 084106, DOI: 10.1063/1.2834918 (2008). https://doi.org/10.1063/1.2834918.

[40] Ditchfield, R., Hehre, W. J. & Pople, J. A. Self-consistent molecular-orbital methods. ix. an extended gaussian-type basis for molecular-orbital studies of organic molecules. *J. Phys. Chem.* **54**, 724–728, DOI: 10.1063/1.1674902 (1971). https://doi.org/10.1063/1.1674902.

[41] Grambow, C. A., Pattanaik, L. & Green, W. H. Reactants, products, and transition states of elementary chemical reactions based on quantum chemistry. *Sci. Data* **7**, 137, DOI: 10.1038/s41597-020-0460-4 (2020).

[42] Grambow, C. A., Pattanaik, L. & Green, W. H. Deep learning of activation energies. *J. Phys. Chem. Lett.* **11**, 2992–2997, DOI: 10.1021/acs.jpclett.0c00500 (2020). https://doi.org/10.1021/acs.jpclett.0c00500.

[43] Ruddigkeit, L., van Deursen, R., Blum, L. C. & Reymond, J.-L. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *J. Chem. Inf. Model.* **52**, 2864–2875, DOI: 10.1021/ci300415d (2012). https://doi.org/10.1021/ci300415d.

[44] Rupp, M., Tkatchenko, A., Müller, K.-R. & von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **108**, 058301, DOI: 10.1103/PhysRevLett.108.058301 (2012).

[45] Jumper, J. *et al.* Highly accurate protein structure prediction with alphafold. *Nature* **596**, 583–589, DOI: 10.1038/s41586-021-03819-2 (2021).

[46] Duan, C., Nandy, A., Meyer, R., Arunachalam, N. & Kulik, H. J. A transferable recommender approach for selecting the best density functional approximations in chemical discovery. *Nat. Comput. Sci.* **3**, 38–47, DOI: 10.1038/s43588-022-00384-0 (2023).

[47] Seifert, G. & Joswig, J.-O. Density-functional tight binding—an approximate density-functional theory method. *WIREs Comput. Mol. Sci.* **2**, 456–465, DOI: 10.1002/wcms.1094 (2012). https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1094.

[48] Duan, C., Chu, D. B. K., Nandy, A. & Kulik, H. J. Detection of multi-reference character imbalances enables a transfer learning approach for virtual high throughput screening with coupled cluster accuracy at dft cost. *Chem. Sci.* **13**, 4962–4971, DOI: 10.1039/D2SC00393G (2022).

[49] Zhao, Q. *et al.* Comprehensive exploration of graphically defined reaction spaces. *Sci. Data* **10**, 145, DOI: 10.1038/s41597-023-02043-z (2023).

[50] Serre, J.-P. *et al. Linear representations of finite groups*, vol. 42 (Springer, 1977).

[51] Bronstein, M. M., Bruna, J., Cohen, T. & Veličković, P. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv:2104.13478* (2021).

[52] Köhler, J., Klein, L. & Noé, F. Equivariant flows: Exact likelihood generative learning for symmetric densities. In *International Conference on Machine Learning*, 5361–5370 (2020).

[53] Nichol, A. Q. & Dhariwal, P. Improved denoising diffusion probabilistic models. In *Proceedings of the 38th International Conference on Machine Learning*, 8162–8171 (2021).

[54] Du, W. *et al.* SE (3) equivariant graph neural networks with complete local frames. In *International Conference on Machine Learning*, 5583–5608 (2022).

[55] Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. & Dahl, G. E. Neural message passing for quantum chemistry. In *International conference on machine learning*, 1263–1272 (2017).

Supplementary Information for "Accurate transition state generation with an object-aware equivariant elementary reaction diffusion model"

Chenru Duan[1, 2, *], Yuanqi Du[3], Haojun Jia[1, 2], and Heather J. Kulik[1, 2]

[1]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, 02139
[2]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, 02139
[3]Department of Computer Science, Cornell University, Ithaca, NY, 14850
[*]Corresponding to: duanchenru@gmail.com

Abbreviation

The following is the list of abbreviation utilized in the main paper.

1. OA-ReactDiff: Object-aware SE(3) GNN for generating sets of 3D molecules in elementary reactions under the diffusion model
2. RMSD: Root mean square deviation.
3. SE(3): Special Euclidean group in 3D space.
4. TS: Transition state.
5. MAE: Mean absolute error.

S1 Physical symmetries and constraints in an elementary reaction.

An elementary reaction that consists of n fragments as reactant and m fragments as product can be described as $\{\mathrm{R}^{(1)}, ..., \mathrm{R}^{(n)}, \mathrm{TS}, \mathrm{P}^{(1)}, ..., \mathrm{P}^{(\mathrm{m})}\}$. This reaction requires the following symmetries:

1. *Permutation symmetry among atoms in a fragment*. For any fragment in $\mathrm{R}^{(i)}, \mathrm{TS}, \mathrm{P}^{(\mathrm{j})}$, change of atom ordering preserves the reaction.
2. *Permutation symmetry among fragments in reactant and product*. The change of ordering in $\{\mathrm{R}^{(1)}, ..., \mathrm{R}^{(n)}\}$ and $\{\mathrm{P}^{(1)}, ..., \mathrm{P}^{(m)}\}$ preserve the reaction.
3. *Rotation and translation symmetry for each fragment*. Rotation and translation operations on any fragment (i.e., $\mathrm{R}^{(i)}, \mathrm{TS}, \mathrm{P}^{(\mathrm{j})}$) preserve the reaction.

Table S1. Ablation studies comparing OA-ReactDiff performance on RMSD evaluation with different models. Vanilla SE(3) LEFTNet[35] is shown to demonstrate the importance of preserve object-wise symmetry in elementary reaction. EGNN[30] is shown to reflect the importance of vanilla SE(3) model.

Approach	RMSD (Å)	
	mean	median
Object-aware SE(3) LEFTNet	0.183	0.076
Vanilla SE(3) LEFTNet	0.638	0.620
Object-aware SE(3) EGNN	0.372	0.360

Preprint. Under review.

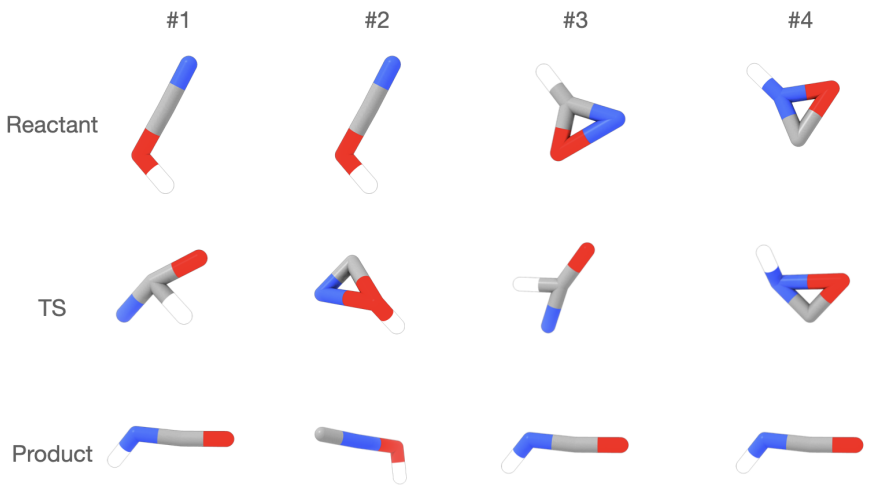

Figure S1. Elementary reactions sampled from OA-ReactDiff by only specifying the chemical composition of interest. Here, we consider a system that contains one C, H, N, and O is chosen. This chemical composition is absent in the Transition1x dataset, and thus is completely new to the trained OA-ReactDiff model. Atoms are colored as follows: gray for C, blue for N, red for O, and white for H.

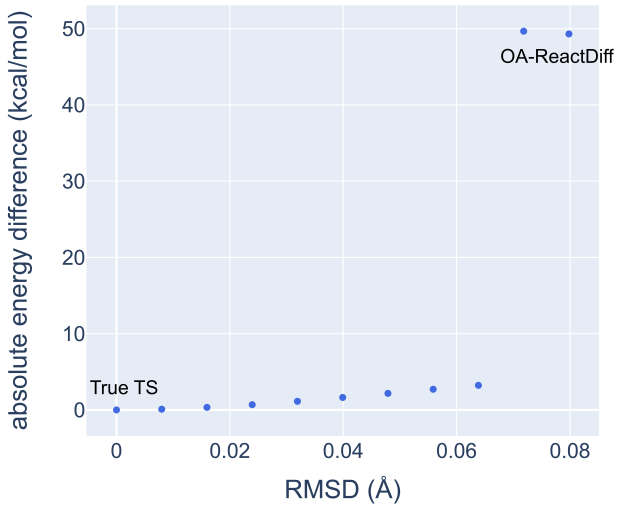

Figure S2. Absolute energy difference vs. RMSD for the ten interpolated structure between true (left) and OA-ReactDiff TS (right) for $C_4H_6O_2$. The abrupt change in energy difference indicates a change in converged electronic state for self-consistent field calculation.

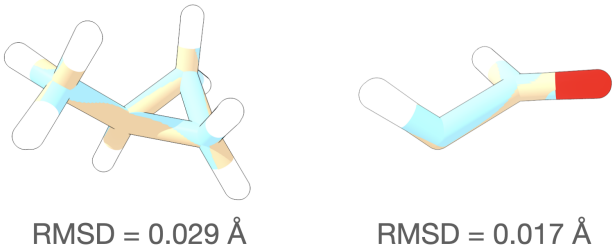

Figure S3. Overlapping OA-ReactDiff and true TS structures of $C_6H_{10}O$ separated as two fragments and their corresponding RMSD. Atoms are colored as follows: C in the true TS structure are in tan and those in the OA-ReactDiff sample are in skyblue; O for red, and H for white.

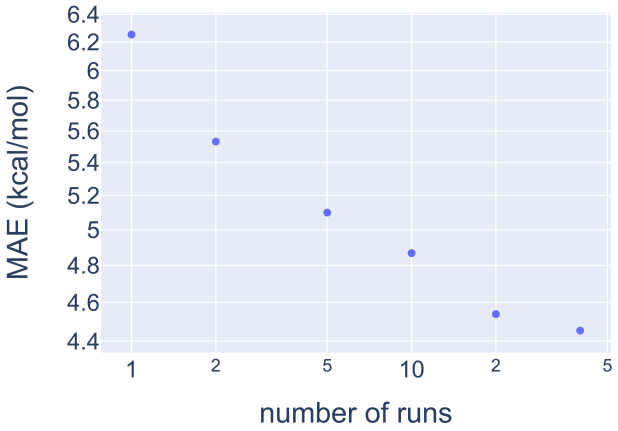

Figure S4. **Mean absolute energy difference vs. number of runs for OA-ReactDiff sampling.**. A log-log axis is used to shown the near power law dependence. The results are shown on 1073 test elementary reactions.

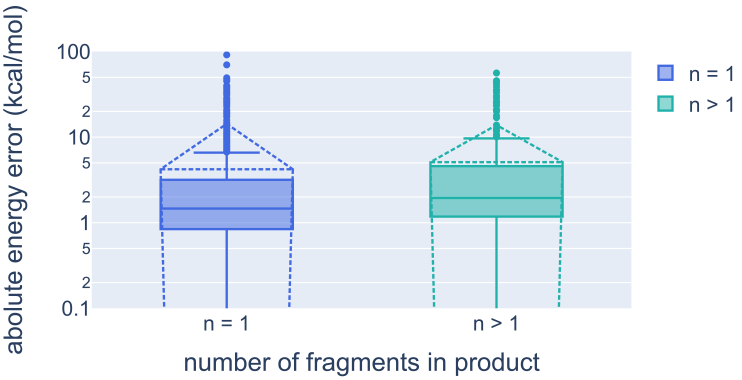

Figure S5. **Box plot for absolute energy difference of OA-ReactDiff + rec. TS structures grouped by single (i.e., n=1) and multi (i.e., n > 1) product cases.**. The solid lines are shown for the quadrants (Q1, median, and Q3) and the dashed lines are shown for the mean and standard deviation. The results are shown on 1073 test elementary reactions.

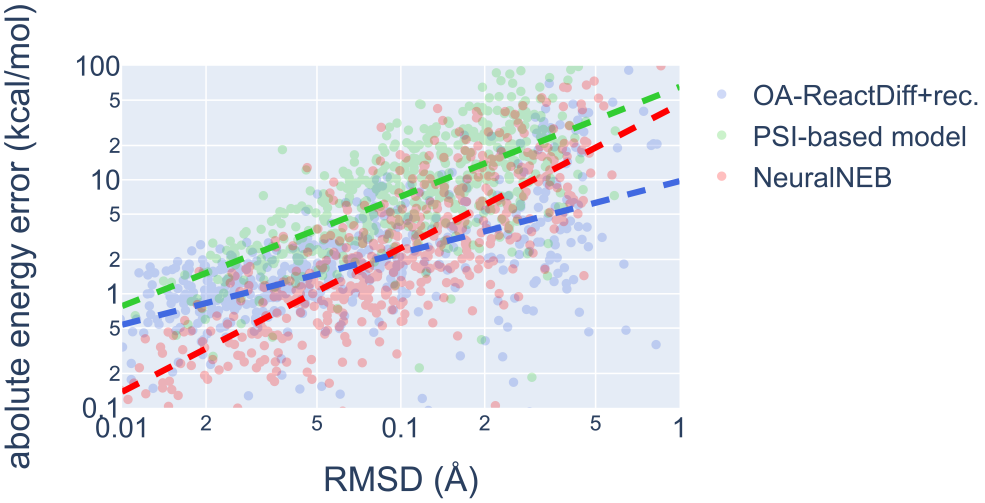

Figure S6. **Absolute energy difference vs. RMSD and the corresponding linear fit in a log-log plot for OA-ReactDiff + rec, PSI-based model,[21] and NeuralNEB[22].**