

React-OT: Optimal Transport for Generating Transition State in Chemical Reactions

Chenru Duan^{1, 2, †, *}, Guan-Horng Liu^{3, †, *}, Yuanqi Du^{4, †, *}, Tianrong Chen⁵, Qiyuan Zhao^{6, 7}, Haojun Jia^{1, 2}, Carla P. Gomes⁴, Evangelos A. Theodorou³, and Heather J. Kulik^{1, 2}

¹Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, 02139

²Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, 02139

³School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, 30332

⁴Department of Computer Science, Cornell University, Ithaca, NY, 14850

⁵School of Electrical and Computer Engineering, Atlanta, GA, 30332

⁶Department of Medicinal Chemistry, College of Pharmacy, University of Michigan, Ann Arbor, MI 48109

⁷Department of Chemistry, University of Michigan, Ann Arbor, MI, 48109

[†]These authors contribute equally

*Corresponding to: duanchenru@gmail.com, guanhong.liu@gmail.com, yuanqidu@cs.cornell.edu

Abstract

Transition states (TSs) are transient structures that are key in understanding reaction mechanisms and designing catalysts but challenging to be captured in experiments. Alternatively, many optimization algorithms have been developed to search for TSs computationally. Yet the cost of these algorithms driven by quantum chemistry methods (usually density functional theory) is still high, posing challenges for their applications in building large reaction networks for reaction exploration. Here we developed React-OT, an optimal transport approach for generating unique TS structures from reactants and products. React-OT generates highly accurate TS structures with a median structural root mean square deviation (RMSD) of 0.053Å and median barrier height error of 1.06 kcal/mol requiring only 0.4 second per reaction. The RMSD and barrier height error is further improved by roughly 25% through pretraining React-OT on a large reaction dataset obtained with a lower level of theory, GFN2-xTB. We envision the great accuracy and fast inference of React-OT useful in targeting TSs when exploring chemical reactions with unknown mechanisms.

Introduction

Transition states (TSs) are central in understanding the kinetics and mechanisms of chemical reactions.^{1,2} Accurate TS structures reveal precise elementary reaction steps on the potential energy surface (PES), enabling the construction of large reaction networks for complex chemical reactions^{3,4} and the design of new catalysts.⁵⁻⁸ However, since they lie at the highest energy point on minimum energy pathways, TSs are transient structures that often live at a time scale of femtoseconds, which makes them difficult to be isolated and characterized experimentally. Very few studies have successfully unraveled elusive TS properties or their structures in experiments. Ultrafast millimeter-wave vibrational spectra, for example, was used to analyze characteristic patterns in isomerizing systems where their TS energies and properties are extracted from frequency-domain data.⁹ More recently, ultrafast electron diffraction was used to get a more direct observation of TS structures for a photochemical electrocyclic ring-opening reaction.¹⁰ These experimental techniques, however, are expensive and cannot be universally applied in all types of reactions.

In parallel, defined as a first-order saddle point on the PES, the TS of any chemical reaction can be searched systemically by a suite of optimization algorithms coupled with quantum chemistry calculations (for example, density functional theory [DFT]¹¹) as the workhorse. These TS search algorithms have been developed and improved over the past twenty years, with salient examples being string methods,¹² growing string methods,¹³ nudged elastic band (NEB) methods,¹⁴ artificial force induced reaction (AFIR),¹⁵ and stochastic surface walking methods (SSWM).¹⁶ Together with these algorithms, comprehensive reaction networks^{4,17} can be constructed by either iteratively enumerating potential elementary reactions on-the-fly,^{18–21} exploration the potential energy surface starting from a local minimum (i.e., given reactant),^{15,16} or propagating biased *ab initio* molecular dynamics for enhanced sampling.^{22–24} For a reasonable-sized reaction network, however, thousands of TS structures need to be optimized, requiring millions of DFT single point calculations.^{25–27} The large number of chemical species involved in reaction networks call reducing the computational cost of searching the TS structure for a single elementary step in chemical reactions.

Recently, machine learning (ML) has demonstrated promise for accelerating the search of TSs. The ideas include developing ML potential as a surrogate for DFT during TS optimizations (for example, in NEB)²⁸ and formulating the transition path sampling problem as a "shooting game" solved by reinforcement learning.²⁹ It has also been formulated as a stochastic optimal control problem to learn a controlled stochastic process from the reaction to the product.³⁰ The TS search problem can also be viewed as a 3D structure generation task, which can then be addressed by generative-adversarial networks,³¹ a combination of gated recurrent neural networks and transformers,³² or denoising diffusion probabilistic models.^{33–35} Our prior work, OA-ReactDiff,³⁵ leverages a diffusion model for elementary chemical reactions, directly generating a set of reactants, TS structure, and products jointly. By preserving all required symmetries of chemical reactions, OA-ReactDiff achieves the state of the art performance on generated structure similarity. Despite its high accuracy in generated TS structures, the stochastic nature of OA-ReactDiff requires multiple runs of sampling for a reaction of interest and the use of a ranking model^{36,37} to recommend one out of those generated TSs. This workflow leads to both an additional cost of sampling and unfavored randomness in a double-ended TS with reactants and products provided, which would be troublesome in its practical usage when replacing or reducing DFT calculations in high throughput TS search.

In this work, we developed React-OT, an optimal transport approach to generate TSs of an elementary reaction in a fully deterministic manner. Compared to OA-ReactDiff, React-OT eliminates the need of training an additional ranking model and reduces the number of inference evaluations of the denoising model from 40,000 to 40, achieving a 1000-fold acceleration. With React-OT, highly accurate TS structures can be deterministically generated in 0.3 seconds. In addition, React-OT outperforms OA-ReactDiff in both structural similarity and barrier height estimation by 30% on Transition1x dataset. When pretrained on RGD1-xTB,³⁸ a dataset containing 760,615 elementary reactions computed with low-cost semi-empirical quantum chemistry methods (i.e., GFN2-xTB³⁹), React-OT improves by 25% in terms

of accuracy, reaching a median RMSD of 0.044 Å and median error of 0.74 kcal/mol in barrier height prediction. To further accelerate the TS search, we replace DFT-optimized reactant and product geometries with those optimized using much cheaper GFN2-xTB. In this scenario, React-OT continues to exhibit comparable performance, achieving a median RMSD of 0.049 Å and median error of 0.79 kcal/mol in barrier height calculations. Furthermore, we integrated React-OT in the high throughput DFT-based TS optimization workflow, where an uncertainty quantification model is used to activate DFT-based TS search only when the generated TS from React-OT is uncertain. With this new workflow, chemical accuracy (mean of 1.58 kcal/mol and median of 0.54 kcal/mol) can be achieved on the generated TS structures using just one seventh of the computational resources required for full reliance on DFT-based TS optimizations. This corresponds to a seven-fold acceleration when building reaction networks while maintaining chemical accuracy. The high quality of generated TS, the extremely low cost in sampling, and the simplicity of its integration to high throughput computational workflow characterize React-OT as a promising model in constructing large reaction networks for studying new chemical reactions with unexplored mechanisms.

Results

Overview of React-OT. A diffusion model learns the underlying distribution of observed samples through training a scoring network by a denoising objective^{33,40,41} (see *Equivariant diffusion models*). Our earlier work, OA-ReactDiff,³⁵ which satisfies all the symmetries in elementary reactions, essentially learns the joint distribution of paired reactant, TSs, and products (Fig. 1a, Supplementary Text 1). With the learned joint distribution, one can either generate a new reaction from scratch or only generate a TS structure conditioned on fixed reactants and products, resembling the setup of the double-ended TS search problem (Fig. 1b, see *Inpainting for conditional generation*). The standard Gaussian distribution, which is far away from a reasonable guess of the TS structure, was used as the starting point for sampling (Supplementary Figure 1). The fact that initial structures at $t = 0$ are randomly sampled from a Gaussian also leads to inevitable stochasticity in the final generated TS structures (Fig. 1b).

To address this challenge, we reformulate the double-ended TS search problem in a dynamic transport setting and utilize an objective resembling flow matching⁴² to achieve the optimal transport in reactions^{43,44} (see *Details about React-OT*). For simplicity, we use the linear interpolation of reactants and products as the guess starting point and keep the reactants and products constant during the transport process (Fig. 1c). This way, we avoid any stochastic processes during the sampling phase. React-OT uses a transition kernel that is object-aware SE(3) equivariant, thus fulfilling all symmetries required in modeling an elementary reaction (see *Equivariance in modeling chemical reactions*). In essence, the key benefit of React-OT are three-fold: First, it simulates the sampling process as an ordinary differential equation instead of as a stochastic differential equation in diffusion models. Therefore, the generated TS with React-OT is deterministic, in line with the fact that there is only one unique TS structure given the

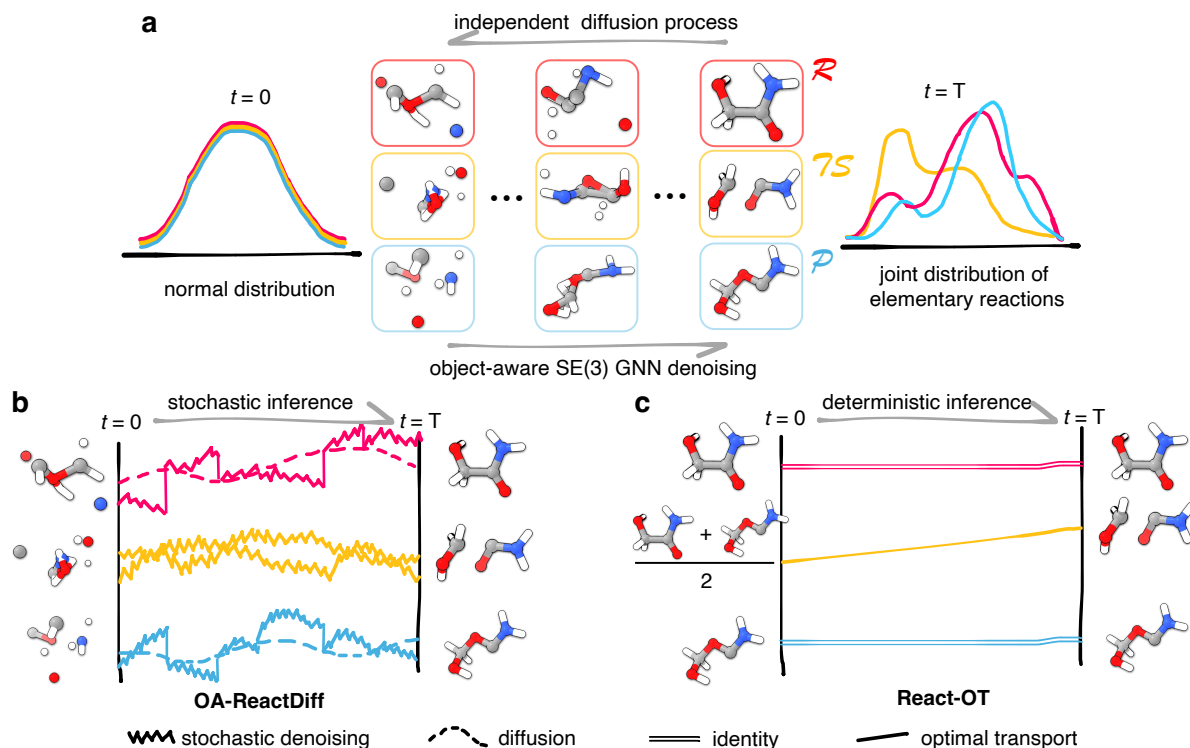


Fig. 1 | Overview of the diffusion model and optimal transport framework for generating TS. **a.** Learning the joint distribution of structures in elementary reactions (reactant in red, TS in yellow, and product in blue). A forward diffusion process brings the joint distribution at $t = T$ to independent normal distribution at $t = 0$. Backward, an object-aware SE(3) GNN is trained with denoising objective to recover the normal distribution to the original joint distribution. **b.** Stochastic inference with inpainting in OA-ReactDiff. Starting with samples drawn from normal distribution, the trained GNN is applied to denoise the reactant, TS, and product. A diffusion process on reactant and product is combined with the denoising process to ensure the end-point reactant and product at $t = T$ are the same as true reactant and product. **c.** Deterministic inference with React-OT. Both the reactant and product are unchanged throughout the entire process from $t = 0$ to $t = T$. The linear interpolation of reactant and product is provided as the initial guess structure at $t = 0$, followed by optimal (i.e., linear) transport to the final TS. Atoms are colored as follows: C for gray; N for blue, O for red, and H for white.

paired reactants and products conformations.^{1,45} Second, utilizing a relatively reasonable initial guess and pushing the sampling path closer to optimal transport, React-OT generates TS structures with higher accuracy and lower number of function evaluation. Third, with fully deterministic inference, React-OT only needs to be run once to generate the final TS structure, greatly simplifying its application workflow and reducing the inference cost. These advantages make React-OT a model that is trustworthy and extremely low-cost, adequately starting to replace actual DFT calculations in high throughput computing.

Generating accurate TS structure in 0.4 seconds. We trained React-OT on Transition1x,⁴⁶ a dataset that contains paired reactants, TSs, and products calculated from climbing-image NEB⁴⁷ obtained with DFT (ω B97x/6-31G(d)).^{48,49} Transition1x contains 10,073 organic reactions with diverse reaction types sampled from exhaustive enumeration^{50,51} of 1,154 reactants in GDB7⁵² dataset, which has up to seven heavy atoms (C, N, and O) and 23 atoms in total. For the

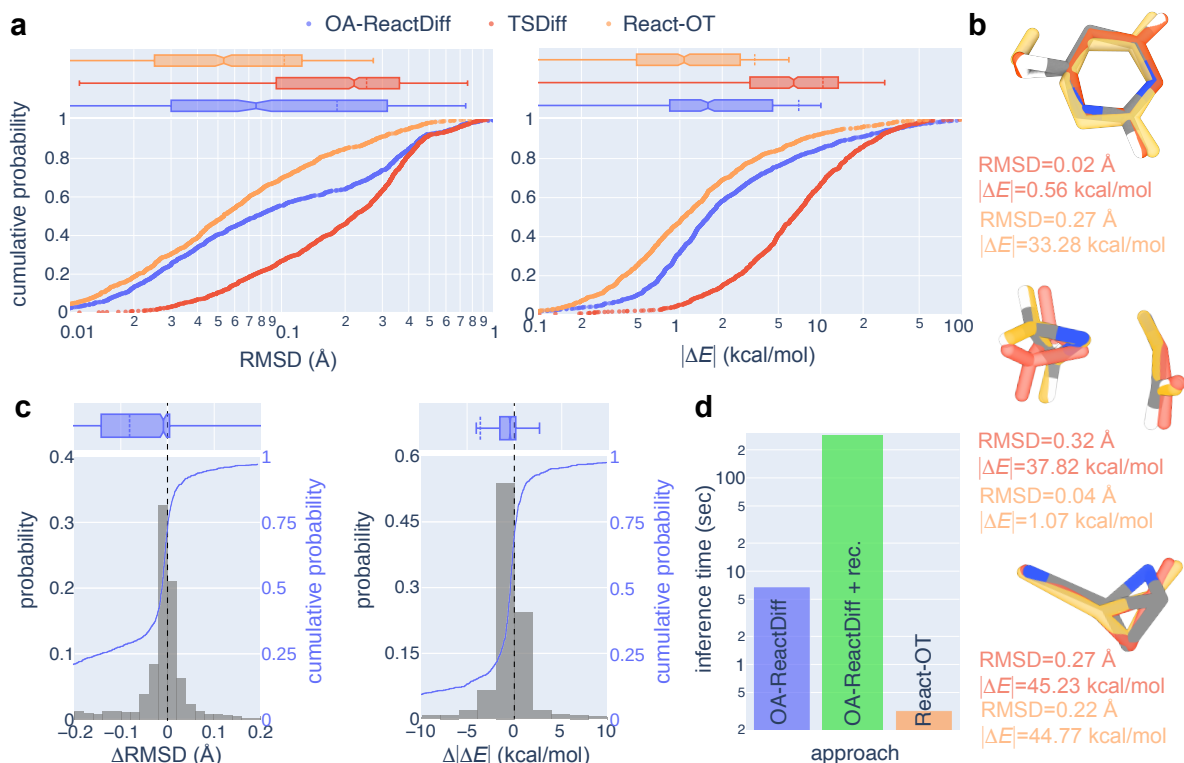


Fig. 2 | Structural and energetic performance of diffusion and optimal transport generated TS structures. a. Cumulative probability for structure root mean square deviation (RMSD) (left) and absolute energy error ($|\Delta E_{TS}|$) (right) between the true and generated TS on 1,073 set-aside test reactions. single-shot OA-ReactDiff³⁵ (blue), single-shot TSDiff³⁴ (red), and React-OT TS (orange) are shown. Both RMSD and $|\Delta E_{TS}|$ are displaced in log scale for visibility of low error regime. **b.** Reference TS structure, OA-ReactDiff TS sample (red), and React-OT structure (orange) for select reactions. RMSD and $|\Delta E_{TS}|$ for OA-ReactDiff and React-OT structures are shown in text with their corresponding color. Atoms in the reference TS are colored as follows: C for gray; N for blue, O for red, and H for white. **c.** Histogram (gray, left y axis) and cumulative probability (blue, right y axis) showing the difference of RMSD (left) and $|\Delta E_{TS}|$ (right) between OA-ReactDiff recommended and React-OT structures compared to reference TS. Negative $\Delta RMSD$ or $\Delta |\Delta E_{TS}|$ suggests React-OT structure is of higher quality. A box plot (blue) for $\Delta RMSD$ and $\Delta |\Delta E_{TS}|$ is shown above the histogram, correspondingly. A dashed vertical line is shown for no deviation between two structures. **d.** Inference time in seconds for single-shot OA-ReactDiff (blue), 40-shot OA-ReactDiff with recommender (green), and React-OT (orange). The y axis is displaced in log scale for visibility of the extremely low inference time for React-OT.

ease of comparison, we use the same train-test partitioning as in OA-ReactDiff, with 9,000 elementary reactions in training and leaving the remaining 1,073 unseen reactions as the test set (see *Details for model training*). In React-OT, the object-aware version of LEFTNet^{35,53} is used as the scoring network to fit the transition kernel (see *LEFTNet*).

React-OT achieves a mean RMSD of 0.103 Å between generated and true TS structures on the set-aside test reactions of Transition1x, significantly improved upon previous state-of-the-art results on ML generation for TS structures.^{34,35} The mean structural RMSD from React-OT is halved compared to a diffusion model that only utilizes 2D graphs of reactants and products (TSDiff,³⁴ 0.252 Å) and that uses 3D structures of reactants and products explicitly (OA-ReactDiff,³⁵ 0.180 Å). With React-OT, the likelihood of finding a TS below a certain RMSD is higher than that of both diffusion models, highlighting the excellent accuracy of React-OT across all reactions in Transition1x

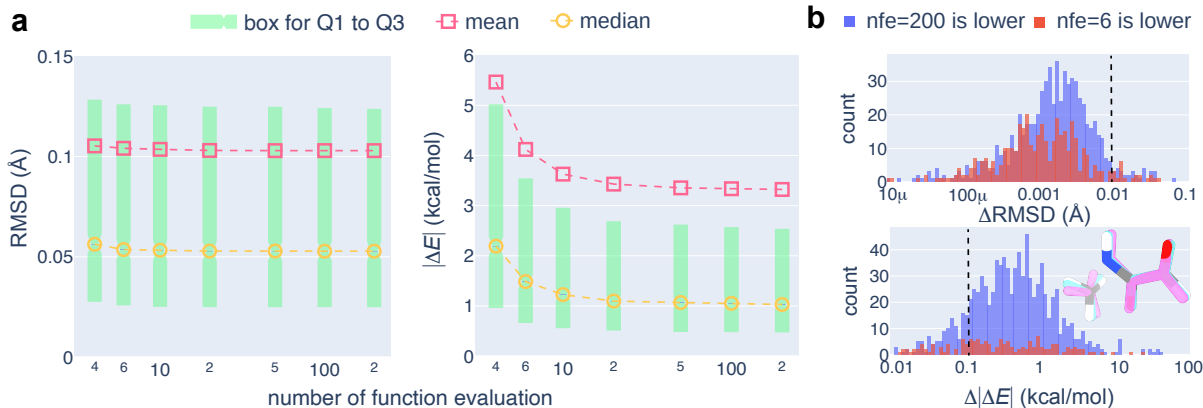


Fig. 3 | Performance of React-OT with respect to the number of function evaluation (nfe). **a.** Distribution of RMSD (left) and $|\Delta E_{\text{TS}}|$ (right) between the true and generated TS on 1,073 set-aside test reactions, where the mean is shown in pink and median is shown in yellow. The first and third quarter is bounded by a green box. **b.** Absolute difference in RMSD (top) and $|\Delta E_{\text{TS}}|$ (bottom) for generated structures at nfe=6 and nfe=200. Structures where nfe=200 gives better quality is shown in blue, otherwise shown in red. A threshold below which the comparison is not chemically meaningful is shown (dashed vertical line). Reference TS structure and React-OT generated structure at nfe=6 (pink) and nfe=200 (skyblue) for a select reaction. Atoms in the reference TS are colored as follows: C for gray; N for blue, O for red, and H for white.

dataset regardless of their level of complexity (Fig. 2a). Similarly, React-OT also at least halves the mean absolute error (MAE) on barrier height estimation from 10.37 kcal/mol of single-shot TSDiff and 6.26 kcal/mol of single-shot OA-ReactDiff to 3.34 kcal/mol. Among 1073 reactions, 64% of the TS structures generated by React-OT gives a more precise TS structure compared to single-shot OA-ReactDiff and 68% of those give a more accurate barrier height (Fig. 2c, Supplementary Figure 2). In chemical reactions, researchers normally use one order of magnitude difference in reaction rate as the characterization of chemical accuracy, which translates to 1.58 kcal/mol in barrier height error assuming a reaction temperature of 70°C. With this definition, over 64% of the generated structures from React-OT falls in chemical accuracy, while the number is only 11% for TSDiff and 50% for OA-ReactDiff.

The superior performance in both the structural similarity and barrier height estimation of React-OT can be ascribed to its more reasonable prior guess at the start of sampling in combination with the optimal transport objective. As a result, React-OT has much higher chance of obtaining the correct inter-molecular arrangement for TS that contains multiple fragments (Fig. 2b, middle) With diffusion models, however, one can take advantage of their stochasticity and keep generating different conformations of TS until getting the most desired one.^{34,35} Still, there are only very few reactions, in 40 runs of sampling, where OA-ReactDiff is able to generate much more accurate TS structures, with H atom transfer in 5H-pyrimidin-5-ide as an example (Fig. 2b, top). Due to the limited size of training data, the parameterization of the scoring network is not perfect, leading to a few reactions where both React-OT and OA-ReactDiff generate the same TS structure that deviates significantly from the true TS (Fig. 2b, bottom). These reactions mostly involve chemical species that are radical or those do not obey octet rule, which are rare in Transition1x and are thus difficult to learn (Supplementary Figure 5). With the ease of starting from a more

reasonable TS guess than Gaussian and trained towards optimal transport paths, React-OT only requires, on average, 0.39 seconds to generate a high-quality TS structure (Fig. 2d). Thanks to its deterministic character, React-OT does not require sampling multiple times and using a recommender to choose one as the final sample as in stochastic diffusion models (for example, OA-ReactDiff). Consequently, the 0.39-second inference time corresponds to a 25-fold acceleration to single-shot OA-ReactDiff and 1,000-fold acceleration to 40-shot OA-ReactDiff with recommender,³⁵ drastically accelerating the TS generation process at the same time as improving the precision.

Given that the initial guess of the sampling path deviates significantly from Gaussian distributions but remains relatively close to the TS, React-OT incorporates a rational implicit bias by assuming that the associations between initial guesses and true TSs are in proximity to the optimal coupling. To further investigate the distance between React-OT and the true optimal transport paths, we gradually reduce the number of function evaluation (nfe) during the sampling process. If the learned transition kernel resembles the true optimal transport, the path would be linear and thus a nfe=2 should suffice. In practice, React-OT converges at nfe=6 for the RMSD of generated TS structures, which corresponds to an inference time of 0.05 seconds, demonstrating its effectiveness in learning the optimal transport path during the structure generation⁴³ (Fig. 3a). However, we observe a much slower convergence at nfe=50 on the barrier height of generated TS structures. This is likely due to the fact that the energy of TS, which usually contains strained bonds, are extremely sensitive to subtle structural changes in some directions compared to others. Since this non-uniform response relationship between structure and energy is not captured by the scoring network, React-OT can hardly optimize the transport path for getting the exact energy, resulting a slower convergence in barrier height. Consequently, React-OT with nfe=6 and nfe=200 generates structurally similar TSs, where most of their differences in RMSD is below a chemically meaningful threshold (that is, 0.01 Å, Fig. 3b). Yet the two sets of generated TSs exhibit a visible difference in their barrier height, with more than 80% showing a deviation above 0.1 kcal/mol. This observation motivates the use of a small nfe (for example, 6) during a fast construction of initial reaction network followed by a large nfe (for example, 50) fine tuning reaction pathways of interest to get a more accurate estimation on their barrier heights and kinetics (Supplementary Figure 3).

Pretraining on additional reactions further improves React-OT. Despite the availability of large computational datasets at DFT accuracy for molecular⁵⁴ and materials⁵⁵ properties, surface-adsorbate structures,⁵⁶ and molecular dynamics trajectories,^{57,58} the size of datasets containing paired reactants, transition states (TSs), and products is often limited.^{38,46,59} This limitation is primarily due to the significantly higher computational cost—two orders of magnitude more expensive—associated with optimizing the minimum energy pathway for chemical reactions.³⁸ In contrast, reaction datasets at more affordable levels of theory, such as semi-empirical quantum chemistry, can be much larger. This disparity highlights the importance of developing efficient schemes for pretraining models with lower-accuracy reaction data to leverage the potential of these larger datasets. For instance, during the generation of

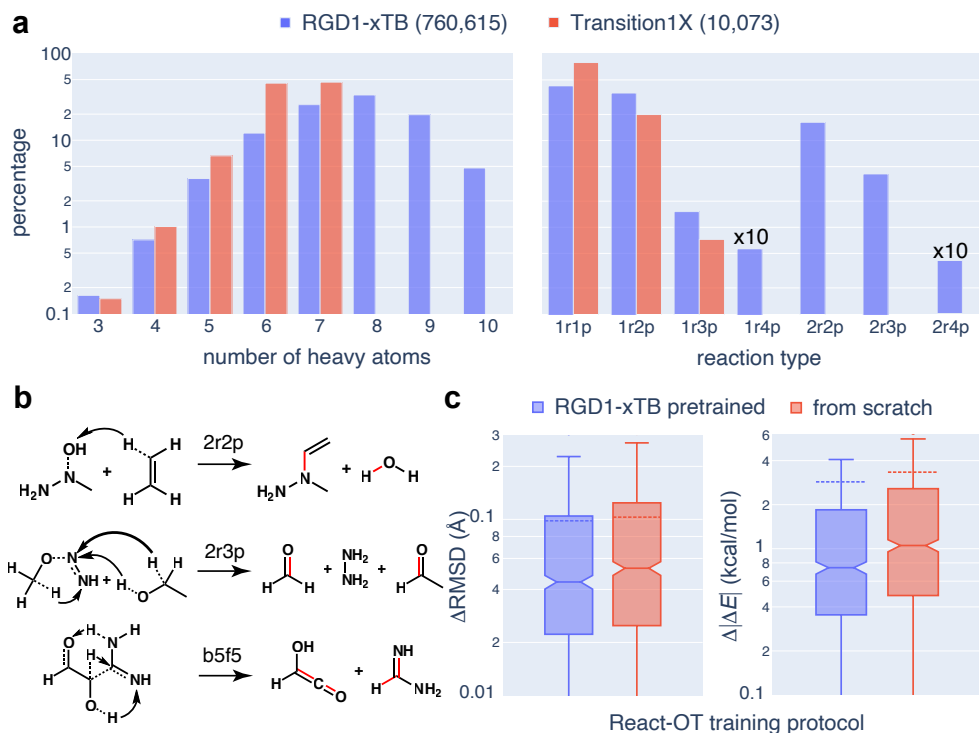


Fig. 4 | React-OT with RGD1-xTB data for pretraining. **a.** Distribution of number of heavy atoms (left) and reaction type (right) for Transition1X (red) and RGD1-xTB (blue). The percentages are shown in log scale for better visibility of rare cases. ArBp refers to a reaction that has A reactants and B products where a rule of $A \leq B$ is set. The percentage of 1r4p and 2r4p is multiplied by 10 to be shown on the plot. **b.** Examples of reaction types that are exclusively presented in RGD1-xTB. **c.** Box plot for RMSD (left) and $|\Delta E_{\text{TS}}|$ (right) in log scale between the true and generated TS on 1,073 set-aside test reactions evaluated by React-OT trained from scratch on Transition1X (red) and React-OT with RGD1-xTB pretraining (blue). Both the mean (dashed horizontal line) and median (notch) are shown in the box plot. The nfe is 100 for this comparison.

DFT-level reactions in RGD1, 760,615 intrinsic reaction coordination calculations were performed at the GFN2-xTB level, resulting in a dataset that can potentially serve as a valuable source of data (named as RGD1-xTB afterwards).³⁸ Compared to Transition1x, RGD1-xTB covers a much larger chemical space, with 57% of reactions both consisting of more than seven heavy atoms (i.e., largest system in Transition1x) as well as more diverse bond rearrangement schemes during chemical reactions (Fig. 4a). Moreover, the dataset's reactants originate from PubChem, and generic enumeration rules were applied to construct the dataset, resulting in a more realistic and diverse set of chemical reactions involving multiple reactants and products and breaking and forming of multiple bonds (Fig. 4b).

With a dataset 75 times larger than Transition1x at hand, we pretrained React-OT on RGD1-xTB, which was then fine-tuned on the training set of Transition1x with a reduced learning rate. RGD1-xTB pretrained React-OT further improves on its quality of generated TS structures, yielding a 0.098Å mean RMSD and 2.86 kcal/mol mean barrier height error (Fig. 4c). The mean error only improves by roughly 10%, mostly due to the persistent high-error TSs generated by React-OT both with and without pretraining (Supplementary Figure 4). In contrast, the median of the structural RMSD and barrier height error is reduced by more than 25%, suggesting the superior performance

Table 1 | Summary of statistics for RMSD, barrier height error, and inference time of TSDiff, OA-ReactDiff and React-OT TS structures. For React-OT, the inference time for nfe=6 and nfe=50 (in parenthesis) are both shown, corresponding to where the RMSD and barrier height error converge, respectively. For barrier height, besides mean and median, the percentage of reactions reaching the 1.58 kcal/mol chemical accuracy (% chem. acc.) is also shown. In case of OA-ReactDiff + recommender, OA-ReactDiff is run 40 times, where a final TS is selected by the recommender according to the ranking score. Note that TSDiff and NeuralNEB use different random seeds for partitioning train/test set compared to OA-ReactDiff and React-OT. Note that the inference time of NeuralNEB was evaluated on CPU.²⁸

Approach	RMSD (Å)		ΔE_{TS} (kcal/mol)			Inference (sec)
	mean	median	mean	median	% chem. acc.	
TSDiff	0.2526	0.2206	10.369	6.428	10.9	–
NeuralNEB	0.1358	0.0959	6.510	2.114	42.6	33.00
OA-ReactDiff	0.1800	0.0752	6.256	1.681	49.7	6.80
OA-ReactDiff + recommender	0.1297	0.0582	4.453	1.617	48.7	272.00
React-OT	0.1029	0.0527	3.337	1.058	63.6	0.05 (0.39)
React-OT + pretraining	0.0981	0.0441	2.864	0.741	71.9	0.05 (0.39)
React-OT + pretraining + xTB RP	0.1056	0.0487	3.283	0.787	68.7	0.05 (0.39)

of React-OT after pretraining on RGD1-xTB (Table 1). This pretraining scheme is helpful, in particular, in cases where the ground-truth reaction data (either DFT or experiment) is limited, which often leads to difficulties in training diffusion models. Considering that GFN2-xTB is roughly 3 orders of magnitude cheaper than DFT, this observation encourages the application of low-level theory on generating reaction datasets coupled with TS search algorithms (such as NEB), especially for large chemical systems where DFT is too computationally demanding.

Utilizing xTB-optimized reactants and products in React-OT. Usually the DFT-optimized reactants and products are assumed to be known explicitly in double-ended TS search problem.¹ This assumption, however, is not practical in ML accelerated workflow such as React-OT since the time spent on optimizing reactants and products with DFT is several orders of magnitude longer than that for generating the TS structure. Therefore, the robustness of being able to utilize approximate reactants and products structures is of great value in ML-based TS search. Here, we re-optimized reactants and products in Transition1x with GFN2-xTB as an approximation to DFT-optimized structures, which on average takes 5.8 seconds per reaction. These xTB-optimized reactants and products still share some degrees of structural similarity to those optimized by DFT (that is, ω B97x/6-31G(d)), especially for unimolecular reactions, with a mean RMSD of 0.07 Å (Fig. 5a). Reactions with multiple products, however, show a much larger RMSD (that is, 0.28 Å) for products optimized by GFN2-xTB and DFT (Supplementary Figure 6 and Supplementary Table 1). This large RMSD is mostly ascribed to the error in the alignment of multiple molecules in the product conformer. Consequently, React-OT with xTB-optimized reactants and products generates TS structures with both higher structural and energetic errors on reactions with multiple products than unimolecular reactions (Fig. 5b). Surprisingly, despite using xTB-optimized reactants and products that demonstrate a 0.07-0.28 Å mean RMSD as inputs, React-OT still generates TS structures at comparable performance, with a 0.106 Å mean RMSD and 0.049 Å median RMSD, only exhibiting less than 10% reduction in structure prediction accuracy compared to that using

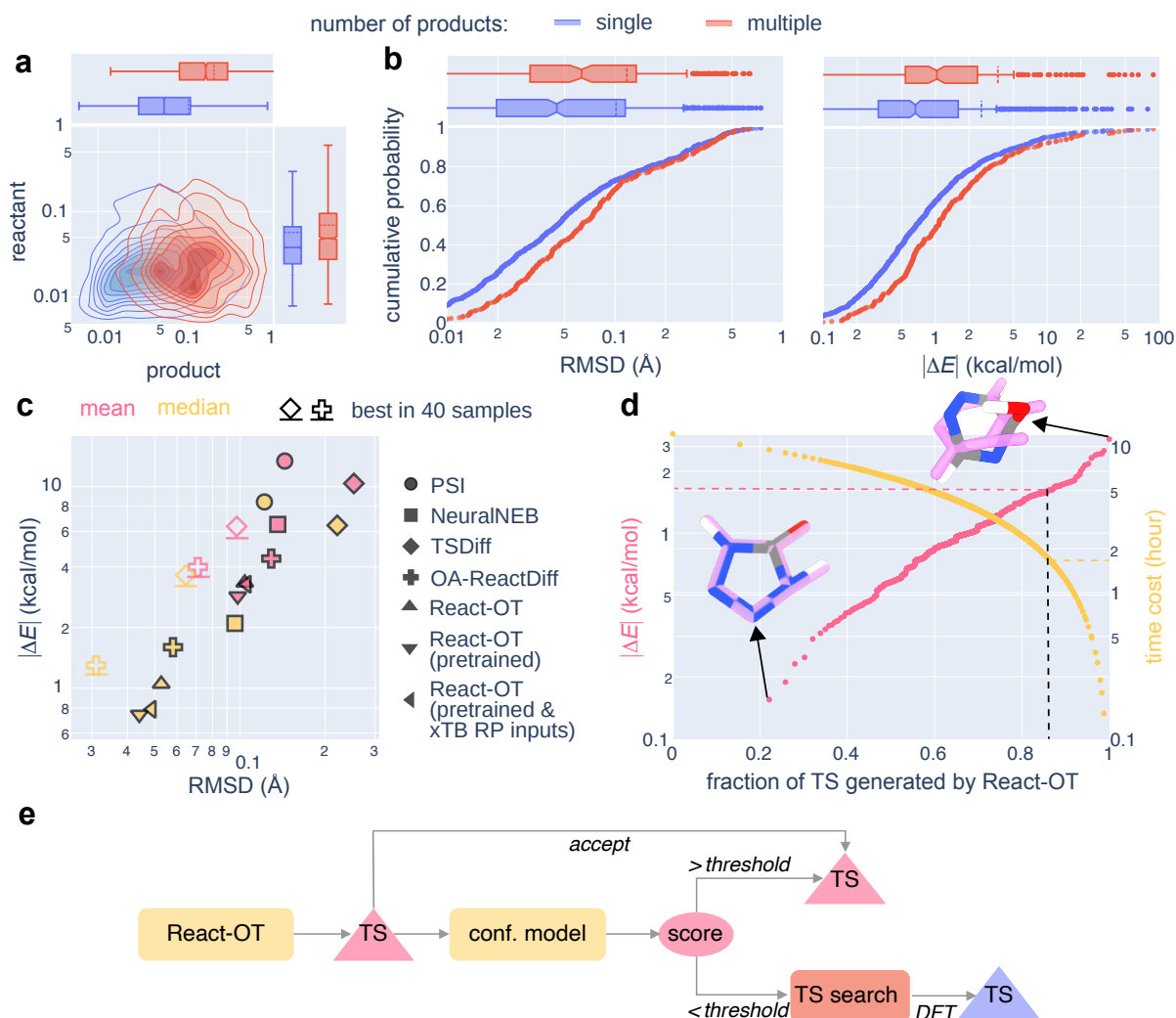


Fig. 5 | React-OT with xTB reactants and products as inputs and its performance in combination with DFT workflow and uncertainty control. **a.** Cumulative probability for structure RMSD (left) and $|\Delta E_{TS}|$ (right) between the true and React-OT generated TS on 1,073 set-aside test reactions. The single- (blue) and multi-product (red) reactions are shown separately. Both RMSD and $|\Delta E_{TS}|$ are displaced in log scale for visibility of low error regime. **b.** 2D contour plot for the RMSD of structures DFT- and xTB optimized reactants and products. The single- (blue) and multi-product (red) reactions are shown separately. **c.** RMSD versus $|\Delta E_{TS}|$ for different TS generation approaches where the mean absolute error is shown in pink and median absolute error is shown in yellow. OA-ReactDiff and TSDiff are also evaluated by the best sample among 40 sampling rounds (symbols without color filling), which, however, is not practical in real application settings. **d.** Average $|\Delta E_{TS}|$ (pink, left y-axis) and time cost per reaction (yellow, right y-axis) with respect to the fraction of TS generated by React-OT under the control by uncertainty quantification. Dashed lines show the statistics at chemical accuracy (1.58 kcal/mol). Reference and generated TS structure (pink) for best and worst test reaction are shown. Atoms in the reference TS are colored as follows: C for gray; N for blue, O for red, and H for white. **e.** Workflow of combining React-OT and conventional DFT-based TS search using a confidence (conf.) model. ML models are shown in orange squares, outputs from ML models are shown in pink triangles (TS) and pink circles (confidence score), DFT-based TS search shown in red square, with its output shown as blue triangle (true TS).

DFT-optimized reactants and products as inputs (Table 1). In addition, React-OT still gives a RMSD as low as 0.12 Å for TSs at multi-molecular reactions, although there are large structural differences between xTB- and DFT-optimized products for multi-molecular reactions (mean RMSD of 0.28 Å, Fig. 5b). The same behavior is observed on

estimating barrier height, where React-OT also shows comparable performance when using xTB-optimized reactants and products. Despite a slight reduction in performance using xTB-optimized reactants and products as inputs, React-OT still maintains state of the art compared to ML approaches with various flavors, such as diffusion models of 3D (OA-ReactDiff³⁵) and 2D reactants and products (TSDiff³⁴), ML potential as surrogate in climbing-image NEB (NeuralNEB²⁸), and transformer-based generative models (PSI³²) (Fig. 5c, Supplementary Figure 7).

There are always desire to integrate ML models in high throughput computational workflows to further accelerate data acquisition and chemical discovery.⁷ With its great accuracy on generating TS structures even using xTB-optimized reactants and products, we investigate the performance of React-OT integrated in typical high throughput computational workflows that mandate TS search. Provided a pair of xTB-optimized reactants and products, we first use React-OT to generate a TS structure (Fig. 5e). This TS structure is then passed through a confidence model to obtain a confidence score on the likelihood of reactants, React-OT predicted TS, and products forming an elementary reaction (see *Details for model training*). One would accept the React-OT structure if the confidence score is higher than a user-defined threshold, otherwise a conventional TS search calculation with DFT is performed. With a confidence threshold of 0, all React-OT generated TS structures are preserved, giving a mean inference time of 0.39 second and barrier height error of 3.28 kcal/mol. With a confidence threshold of 1, none of the React-OT generated TS structures is considered accurate, and thus one would run explicit TS search for all reactions, taking 12.8 hours per reaction on average. By varying the confidence threshold, one can choose between being aggressive (with a small threshold) and conservative (with a large threshold) on activating React-OT in the computational workflow. As we reduce the threshold, a larger fraction of TS structures are generated by React-OT. Meanwhile, the mean barrier height error increases smoothly and monotonically, which indicates the confidence model can successfully assess the quality of generated TS structures (Fig. 5d, Supplementary Figure 8). When the chemical accuracy of barrier height prediction (that is, 1.58 kcal/mol MAE) is targeted, 86% TSs would be generated by React-OT and the remaining 14% searched by climbing-image NEB with DFT, rendering a seven-fold acceleration compared to DFT-only computational workflows. At the same time, an extremely low median structural RMSD of 0.031 Å and median barrier height error of 0.54 kcal/mol would be achieved. These observations showcase the promise of combining React-OT and conventional TS search algorithms in high throughput screening of elementary reactions and building large reaction networks.

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 developed React-OT, an optimal transport approach for deterministically generating TS provided reactants and products. React-OT achieves improved accuracy on generated TS structures compared to OA-ReactDiff, while reducing the cost of inference by 1000 fold.

Pretrained on a large reaction dataset at GFN2-xTB level, React-OT further improves by 25% in TS structural RMSD and barrier height. In a practical setting where xTB-optimized reactants and products are used as inputs, React-OT still shows comparable performance, generating TS at chemical accuracy with only one seventh of the cost for a hybrid ML-DFT high throughput computational workflow.

Despite the use of reactants and products optimized by GFN2-xTB, the preparation of optimized reactants and products is the most time consuming step (10 folds more expensive than others combined) in React-OT. Therefore, ML approaches that accelerates reactant-product conformer optimizations, either by an ML potential or with a 2D graph to 3D conformation sampling model, would be desired to unleash the potential of sub-second inference speed of React-OT in large reaction network exploration. Due to the absence of large reaction datasets with charged species and metals, the demonstration of React-OT is limited to the scope of neutral organic chemical space with CNOH elements. In future, it would be of great interest to test the transferability of React-OT on more diverse chemistry.

With the optimal transport approach, React-OT does not have any stochasticity in the sampling process, making it tailored for double-ended TS search. To construct reaction networks, however, a model is required to generate products from reactants, which does not have a one-to-one mapping. There, diffusion models that carry some degree of stochasticity (for example, OA-ReactDiff) are preferred. In addition, sampling TS from the potential energy surface directly without access to precalculated datasets of TS structures for training is another promising direction to explore. Although we only demonstrate React-OT on the TS search problem, there are many scenarios in biology, chemistry, and materials science, where an optimal transport setting is appropriate, including protein-ligand docking, molecule absorption on metal surfaces, and structural changes of materials in phase transitions. We anticipate this optimal transport approach presented here insightful to applications in other domain of sciences.

Methods

Equivariance in modeling chemical reactions 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 .^{60,61} 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. In diffusion models, SE(3) equivariance is achieved by building an SE(3)-invariant prior and an SE(3)-equivariant transition kernel.⁶² However, in the dynamic optimal transport setting, we can design a more informative prior than Gaussian distribution⁴³ (in this case, we use the linear interpolation between the reactant and product as an initial guess). As the paired data (initial guess and transition state) is available during training, we align the initial and target geometry for permutation, rotation, and reflection, and move the center of mass to the origin to remove the effect of translation. In addition, we have an equivariant transition kernel to transport from the initial to the target distribution.

Equivariant diffusion models Diffusion models are originally inspired from non-equilibrium thermodynamics.^{33,40,41} 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.⁴¹

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

$$\begin{aligned} q(x_s|x_0, x_t) &= \mathcal{N}(x_s|\mu_{t \rightarrow s}(x_0, x_t), \sigma_{t \rightarrow s}^2 I), \\ \mu_{t \rightarrow s}(x_0, x_t) &= \frac{\alpha_{t|s}\sigma_s^2}{\sigma_t^2}x_t + \frac{\alpha_s\sigma_{t|s}^2}{\sigma_t^2}x_0 \text{ and } \sigma_{t \rightarrow s} = \frac{\sigma_{t|s}\sigma_s}{\sigma_t}, \end{aligned}$$

where $s < t$ refer to two different timesteps along the diffusion/denoising process ranging from 0 to T, $\alpha_{t|s} = \frac{\alpha_t}{\alpha_s}$, $\sigma_{t|s}^2 = \sigma_t^2 - \alpha_{t|s}^2\sigma_s^2$. 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:

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

Empirically, a simplified objective has been found to be efficient to optimize:³³

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

Details about React-OT *Optimal transport.*– Optimal transport (OT) problems^{63–65} investigate the most efficient means of transporting samples between two specified distributions. Originating from Gaspard Monge’s work⁶⁶ in the 18th century, the problem has evolved and gained broader significance through Leonid Kantorovich’s relaxation.⁶⁷ Given a pair of boundary distributions $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, where $\mathcal{P}(\mathbb{R}^d)$ denotes the space of probability distributions, the problem involves minimizing the, e.g., squared Euclidean, transport cost:

$$W_2^2(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \int \int \|x - y\|^2 \pi(x, y) dx dy,$$

where $\Pi(\mu, \nu) := \{\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) : \int \pi(x, y) dy = \mu, \int \pi(x, y) dx = \nu\}$ is the set of couplings on $\mathbb{R}^d \times \mathbb{R}^d$ with respective marginals μ, ν . The resulting distance $W_2^2(\mu, \nu)$ between μ and ν , given the optimal transport coupling π^* , is commonly referred to as the squared Wasserstein-2 distance.

Dynamic optimal transport.– Dynamic optimal transport, as introduced by Benamou and Brenier,⁶⁸ presents the squared Wasserstein-2 distance $W_2^2(\mu, \nu)$ in a dynamic formulation. This formulation involves optimizing the squared L^2 norm of a time-varying vector field $u_t(x_t)$, which transports samples from $x_0 \sim \mu(x_0)$ to $x_1 \sim \nu(x_1)$ according to an ODE, $\frac{dx_t}{dt} = u_t(x_t)$. Mathematically, this dynamic formulation is expressed as:

$$\begin{aligned} W_2^2(\mu, \nu) &:= \inf_{q_t, u_t} \int_0^1 \int \frac{1}{2} \|u_t(x_t)\|^2 q_t(x_t) dx_t dt \\ \text{s.t. } \frac{\partial q_t}{\partial t} + \nabla \cdot (u_t q_t) &= 0, \quad q_0 = \mu, \quad q_1 = \nu \end{aligned}$$

subject to the continuity equation and the boundary conditions $q_0 = \mu$ and $q_1 = \nu$. This approach describes a fully deterministic transportation process, distinguishing it from previous methods such as OA-ReactDiff, which rely on stochastic processes. The minimal assumptions placed on the distributional boundaries μ, ν in OT problems, in contrast to the Gaussian priors in standard diffusion models,³³ enable the straightforward incorporation of domain-specific structures, such as the incorporation of pairing information, if available.

Given the OT coupling $\pi^*(x, y)$, the optimal time-marginal q_t^* can be defined using the OT *interpolant* [69, Eq. (7.8)] as

$$q_t^* = ((1-t)x + ty)_\# \pi^*,$$

where $\#$ is the push-forward operator. In essence, the optimal trajectories between samples drawn from $(x_0, x_1) \sim \pi^*$ move along straight lines $x_t = (1-t)x_0 + tx_1$. Consequently, the optimal vector field can be constructed as a constant velocity $u_t^*(x_t) = x_1 - x_0$. In practice, given a tuple of product (P), reactor (R), and transition state (TS), we set the initial and terminal states respectively to $x_0 = \frac{P+R}{2}$ and $x_1 = TS$, then minimize an objective resembling *flow matching*.⁴²

$$\min_{\theta} \|u_{\theta}(x_t, t, z) - (x_1 - x_0)\|^2$$

where additional information (i.e. atom types) are gathered in z . At inference, we can sample $x_1 \sim \nu(x_1|x_0)$ by solving the parametrized ODE $\frac{dx_t}{dt} = u_{\theta}^*(x_t, t, z)$ using off-the-shelf numerical solvers.

It is noteworthy that React-OT adopts an implicit bias where the pairing information is assumed to be close to the optimal coupling, i.e., $(\frac{P+R}{2}, TS) \sim \pi^*$. This assumption generally holds when the pairings are sufficiently close to each other without much overlap with others,^{43,44} which is indeed the case in the application being considered.

Inpainting for conditional generation. Inpainting is a flexible technique to formulate the conditional generation problem for diffusion models.⁷⁰ 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 technique⁷⁰ 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. However, at a given total step T , 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. For example, with a resample size of 10, and total step of 100, the number of function evaluation is 1,000 for a single sampling process, which is significantly larger than that in React-OT.

LEFTNet. We build our equivariant transition kernel on top of a recently proposed SE(3)-equivariant GNN, LEFTNet.⁵³ LEFTNet achieves SE(3)-equivariance based on building local node- and edge-wise equivariant frames that scalarize vector (for example position, velocity) and higher order tensor (for example stress) geometric quantities. The geometric quantities are transformed back from scalars through a tensorization technique without loss of any information. LEFTNet is designed to handle Euclidean group symmetries including rotation, translation and reflection, as well as the permutation symmetry. To tailor the model for chemical reaction, we adopt the object-aware improvement from our previous work, OA-ReactDiff.³⁵ In addition to handling symmetries, LEFTNet has strong geometric and function approximation expressiveness. Specifically, LEFTNet has a local structure encoding module which is proven to distinguish a hierarchy of local 3D isomorphism and a frame transition encoding module which is capable of learning universal equivariant functions. For more detailed descriptions, we refer reading the original LEFTNet⁵³ or OA-ReactDiff³⁵ work.

Details for model training. *React-OT training.*— We directly fine tuned the OA-ReactDiff model checkpoint with the new optimal transport objective and a reduced learning rate of 0.0001 (see *Details about React-OT*). The scoring network, LEFTNet, has 96 radial basis functions, 196 hidden channels for message passing, 6 equivariant update blocks, and an interaction cutoff of 10Å. The React-OT model was trained for an additional 200 epochs.

Confidence model training.— The confidence model is also fine tuned from the OA-ReactDiff model checkpoint, with the change of the final output layer to *asigmoid* function for predicting a probability ranging from 0 to 1 as the confidence score. The OA-ReactDiff were run for 40 rounds of sampling on the 9,000 training reactions, which generated 360,000 synthetic reactions. A reaction was labeled as "good" (that is, 1) if the generated TS structure has a RMSD < 0.2 Å compared to the true TS, otherwise was labeled as "bad" (that is, 0). Here, we directly adopted the confidence model trained in OA-ReactDiff without further changes.

Code and data availability

Code and data are currently under review and will be available as a open source repository on github.

Author contributions

C.D., G.-H.L., and Y.D.: conceptualization, methodology, software, validation, investigation, data curation, writing of original draft, review and editing, and visualization. T.C.: methodology, software, and review and editing. Q.Z.: dataset, writing of original draft, and review and editing. H.J.: data curation, review and editing. C.P.G., E.A.T., and H.J.K.: writing of original draft, review and editing.

Competing interests

The authors declare no competing financial interest at this moment.

References

- ¹ 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>.
- ² E, W. & Vanden-Eijnden, E. Transition-path theory and path-finding algorithms for the study of rare events. *Annu. review physical chemistry* **61**, 391—420, DOI: 10.1146/annurev.physchem.040808.090412 (2010).
- ³ 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>.
- ⁴ 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>.
- ⁵ Klucznik, T. *et al.* Computational prediction of complex cationic rearrangement outcomes. *Nature* **625**, 508–515, DOI: 10.1038/s41586-023-06854-3 (2024).
- ⁶ Back, S. *et al.* Accelerated chemical science with ai. *Digit. Discov.* **3**, 23–33, DOI: 10.1039/D3DD000213F (2024).
- ⁷ Nandy, A. *et al.* Computational discovery of transition-metal complexes: From high-throughput screening to machine learning. *Chem. Rev.* **121**, 9927–10000, DOI: 10.1021/acs.chemrev.1c00347 (2021).
- ⁸ Zhang, S. *et al.* Exploring the frontiers of condensed-phase chemistry with a general reactive machine learning potential. *Nat. Chem.* DOI: 10.1038/s41557-023-01427-3 (2024).
- ⁹ Prozument, K. *et al.* Photodissociation transition states characterized by chirped pulse millimeter wave spectroscopy. *Proc. Natl. Acad. Sci.* **117**, 146–151, DOI: 10.1073/pnas.1911326116 (2020). <https://www.pnas.org/doi/pdf/10.1073/pnas.1911326116>.

- ¹⁰ Liu, Y. *et al.* Rehybridization dynamics into the pericyclic minimum of an electrocyclic reaction imaged in real-time. *Nat. Commun.* **14**, 2795, DOI: 10.1038/s41467-023-38513-6 (2023).
- ¹¹ 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>.
- ¹² Weinan, E., Ren, W. & Vanden-Eijnden, E. String method for the study of rare events. *Phys. Rev. B* **66**, 052301 (2002).
- ¹³ Peters, B., Heyden, A., Bell, A. T. & Chakraborty, A. A growing string method for determining transition states: Comparison to the nudged elastic band and string methods. *The J. chemical physics* **120**, 7877–7886 (2004).
- 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 Maeda, S., Taketsugu, T. & Morokuma, K. Exploring transition state structures for intramolecular pathways by the artificial force induced reaction method. *J. Comput. Chem.* **35**, 166–173 (2014).
- ¹⁶ Shang, C. & Liu, Z. P. Stochastic surface walking method for structure prediction and pathway searching. *J. Chem. Theory Comput.* **9**, 1838–1845 (2013).
- ¹⁷ 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).
- ¹⁸ Zimmerman, P. M. Automated discovery of chemically reasonable elementary reaction steps. *J. Comput. Chem.* **34**, 1385–1392 (2013).
- ¹⁹ 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>.
- ²⁰ 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>.
- ²¹ 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).
- ²² Wang, L.-P. *et al.* Discovering chemistry with an ab initio nanoreactor. *Nat. Chem.* **6**, 1044–1048, DOI: 10.1038/nchem.2099 (2014).
- ²³ Pieri, E. *et al.* The non-adiabatic nanoreactor: towards the automated discovery of photochemistry. *Chem. Sci.* **12**, 7294–7307, DOI: 10.1039/D1SC00775K (2021).

- ²⁴ 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).
- ²⁵ 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).
- ²⁶ 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).
- ²⁷ 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).
- ²⁸ 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).
- ²⁹ Zhang, J. *et al.* Deep reinforcement learning of transition states. *Phys. Chem. Chem. Phys.* **23**, 6888–6895, DOI: 10.1039/D0CP06184K (2021).
- ³⁰ Holdijk, L. *et al.* Stochastic optimal control for collective variable free sampling of molecular transition paths. *Adv. Neural Inf. Process. Syst.* **36** (2024).
- ³¹ 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>.
- ³² 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).
- ³³ 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).
- ³⁴ Kim, S., Woo, J. & Kim, W. Y. Diffusion-based generative ai for exploring transition states from 2d molecular graphs (2023). 2304.12233.
- ³⁵ Duan, C., Du, Y., Jia, H. & Kulik, H. J. Accurate transition state generation with an object-aware equivariant elementary reaction diffusion model. *Nat. Comput. Sci.* **3**, 1045–1055, DOI: 10.1038/s43588-023-00563-7 (2023).
- ³⁶ 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).

- ³⁷ Corso, G., Stärk, H., Jing, B., Barzilay, R. & Jaakkola, T. DiffDock: Diffusion steps, twists, and turns for molecular docking. *arXiv:2210.01776* (2023).
- ³⁸ Zhao, Q. *et al.* Comprehensive exploration of graphically defined reaction spaces. *Sci. Data* **10**, 145, DOI: 10.1038/s41597-023-02043-z (2023).
- ³⁹ Bannwarth, C., Ehlert, S. & Grimme, S. Gfn2-xtb—an accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem. Theory Comput.* **15**, 1652–1671, DOI: 10.1021/acs.jctc.8b01176 (2019).
- ⁴⁰ Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N. & Ganguli, S. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning*, 2256–2265 (2015).
- ⁴¹ Song, Y. *et al.* Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations* (2020).
- ⁴² Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M. & Le, M. Flow matching for generative modeling. In *The Eleventh International Conference on Learning Representations* (2023).
- ⁴³ Liu, G.-H. *et al.* I²SB: Image-to-Image Schrödinger bridge. In *International Conference on Machine Learning* (2023).
- ⁴⁴ Somnath, V. R. *et al.* Aligned diffusion schrödinger bridges. In *Conference on Uncertainty in Artificial Intelligence* (2023).
- ⁴⁵ Zhao, Q., Hsu, H.-H. & Savoie, B. Conformational sampling for transition state searches on a computational budget. *J. Chem. Theory Comput.* **18**, 3006–3016 (2022).
- ⁴⁶ 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).
- ⁴⁷ 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>.
- ⁴⁸ 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>.
- ⁴⁹ 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>.
- ⁵⁰ 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).

- ⁵¹ Grambow, C. A., Pattanaik, L. & Green, W. H. Deep learning of activation energies. *J. Phys. Chem. Lett.* **11**, 2992–2997, DOI: 10.1021/acs.jpcllett.0c00500 (2020). <https://doi.org/10.1021/acs.jpcllett.0c00500>.
- ⁵² 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>.
- ⁵³ Du, W. *et al.* A new perspective on building efficient and expressive 3D equivariant graph neural networks. *arXiv:2304.04757* (2023).
- ⁵⁴ 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).
- ⁵⁵ Jain, A. *et al.* Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002, DOI: 10.1063/1.4812323 (2013). https://pubs.aip.org/aip/apm/article-pdf/doi/10.1063/1.4812323/13163869/011002_1_online.pdf.
- ⁵⁶ Tran*, R. *et al.* The open catalyst 2022 (oc22) dataset and challenges for oxide electrocatalysts. *ACS Catal.* (2023).
- ⁵⁷ 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>.
- ⁵⁸ Chmiela, S. *et al.* Accurate global machine learning force fields for molecules with hundreds of atoms. *Sci. Adv.* **9**, eadf0873, DOI: 10.1126/sciadv.adf0873 (2023). <https://www.science.org/doi/pdf/10.1126/sciadv.adf0873>.
- ⁵⁹ Nandi, S., Vegge, T. & Bhowmik, A. Multixc-qm9: Large dataset of molecular and reaction energies from multi-level quantum chemical methods. *Sci. Data* **10**, 783, DOI: 10.1038/s41597-023-02690-2 (2023).
- ⁶⁰ Serre, J.-P. *et al.* *Linear representations of finite groups*, vol. 42 (Springer, 1977).
- ⁶¹ Bronstein, M. M., Bruna, J., Cohen, T. & Veličković, P. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv:2104.13478* (2021).
- ⁶² Köhler, J., Klein, L. & Noé, F. Equivariant flows: Exact likelihood generative learning for symmetric densities. In *International Conference on Machine Learning*, 5361–5370 (2020).
- ⁶³ Villani, C. *et al.* *Optimal transport: old and new*, vol. 338 (Springer, 2009).
- ⁶⁴ Villani, C. *Topics in optimal transportation*, vol. 58 (American Mathematical Soc., 2021).
- ⁶⁵ Santambrogio, F. Optimal transport for applied mathematicians. *Birkhäuser, NY* **55**, 94 (2015).
- ⁶⁶ Monge, G. Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.* 666–704 (1781).
- ⁶⁷ Kantorovich, L. V. On the translocation of masses. In *Dokl. Akad. Nauk. USSR (NS)*, vol. 37, 199–201 (1942).

- ⁶⁸ Benamou, J.-D. & Brenier, Y. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numer. Math.* **84**, 375–393 (2000).
- ⁶⁹ Peyré, G. & Cuturi, M. Computational optimal transport. *Cent. for Res. Econ. Stat. Work. Pap.* (2017).
- ⁷⁰ 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).

Supplementary Information for "React-OT: Optimal Transport for Generating Transition State in Chemical Reactions"

Chenru Duan^{1, 2, †, *}, Guan-Horng Liu^{3, †, *}, Yuanqi Du^{4, †, *}, Tianrong Chen⁵, Qiyuan Zhao^{6, 7}, Haojun Jia^{1, 2}, Carla P. Gomes⁴, Evangelos A. Theodorou³, and Heather J. Kulik^{1, 2}

¹Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, 02139

²Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, 02139

³School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, 30332

⁴Department of Computer Science, Cornell University, Ithaca, NY, 14850

⁵School of Electrical and Computer Engineering, Atlanta, GA, 30332

⁶Department of Medicinal Chemistry, College of Pharmacy, University of Michigan, Ann Arbor, MI 48109

⁷Department of Chemistry, University of Michigan, Ann Arbor, MI, 48109

[†]These authors contribute equally

*Corresponding to: duanchenru@gmail.com, guanhorng.liu@gmail.com, yuanqidu@cs.cornell.edu

Abbreviation

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

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

1 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 $\{R^{(1)}, \dots, R^{(n)}, TS, P^{(1)}, \dots, P^{(m)}\}$. This reaction requires the following symmetries:

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

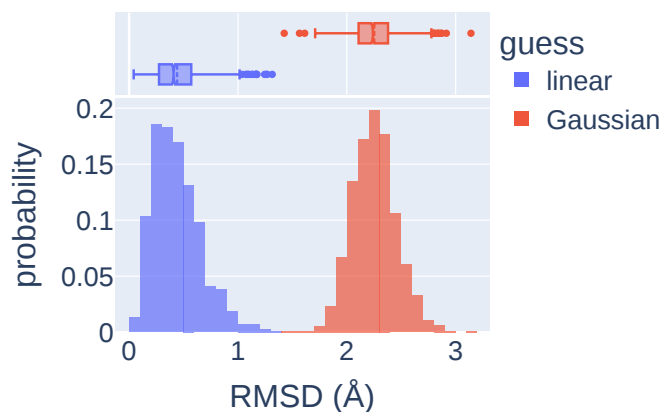


Figure 1. RMSD between the true TS different initial guesses. Linear interpolation between reactants and products (blue) and samples from a Gaussian distribution (red).

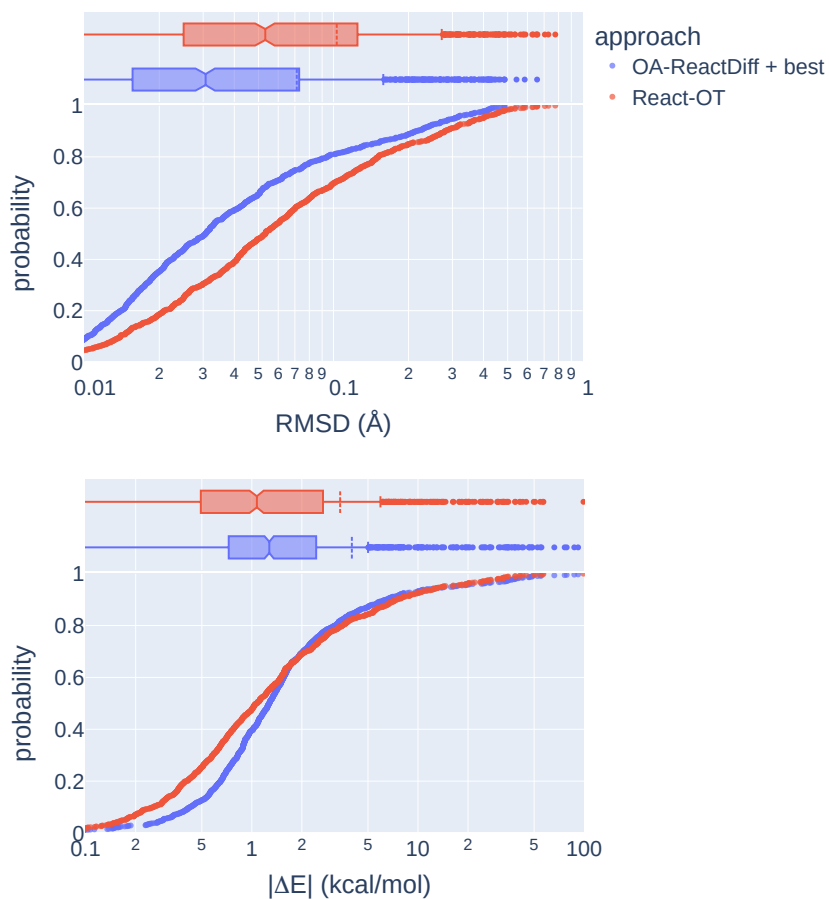


Figure 2. Additional performance comparison between React-OT and OA-ReactDiff. Cumulative probability for RMSD (top) and barrier height (bottom) for React-OT (red) and OA-ReactDiff + best sample in 40 runs (blue) both shown in log scale for visibility. Note that the performance of OA-ReactDiff + best sample in 40 runs is not practically achievable as one does not know the best sample a priori.

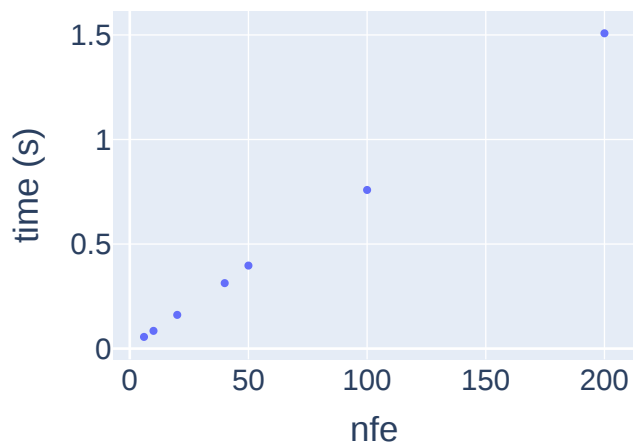


Figure 3. Sampling cost at different number of function evaluation.

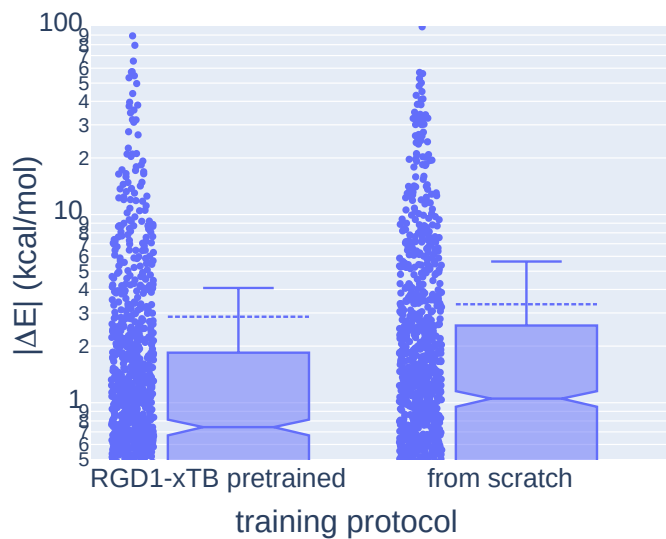


Figure 4. Persistent barrier height errors. Box plots with all points shown for barrier height errors obtained by React-OT with RGD1-xTB pretraining and training from scratch. The barrier height is shown in log scale with a truncation of 0.5 kcal/mol.

Table 1. Statistics of RMSD between xTB and DFT optimized reactants and products..

Species	Reaction type	RMSD (Å)	
		mean	median
reactant	uni-molecular	0.070	0.037
reactant	multi-molecular	0.084	0.048
product	uni-molecular	0.137	0.076
product	multi-molecular	0.289	0.204

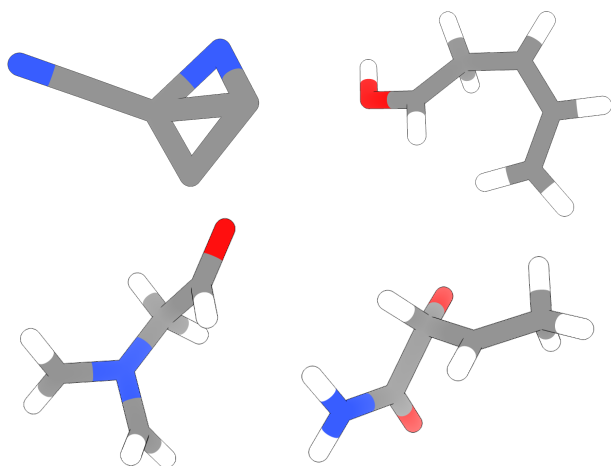


Figure 5. Examples of TS structures breaking the octet rule that React-OT gives large errors. Atoms colored as follows: gray for C, blue for N, red for O, and white for H.

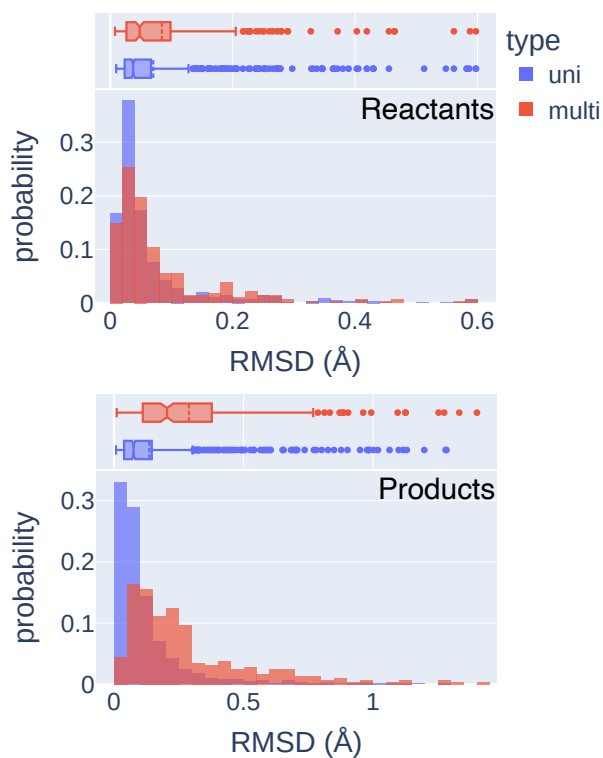


Figure 6. RMSD between xTB and DFT optimized reactants and products. Distribution of RMSD for reactants (top) and products (bottom) grouped by the reaction type: uni-molecular reaction in blue and multi-molecular reaction in red.

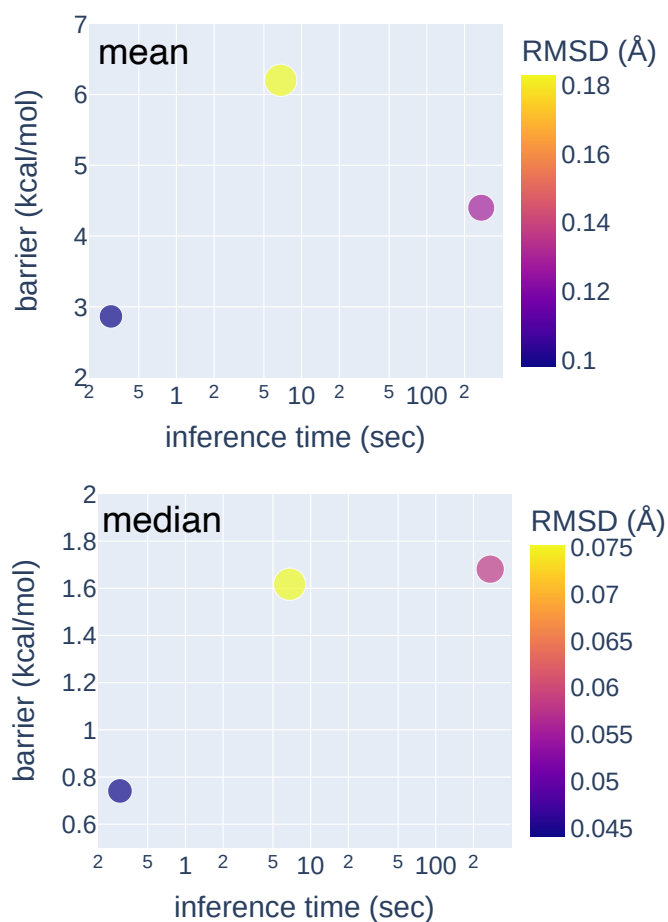


Figure 7. Summary of performance for React-OT and OA-ReactDiff. Barrier height, RMSD (mean as top and median as bottom), and inference time for React-OT (blue), one-shot OA-ReactDiff (yellow), and 40-shot OA-ReactDiff with recommender (purple).

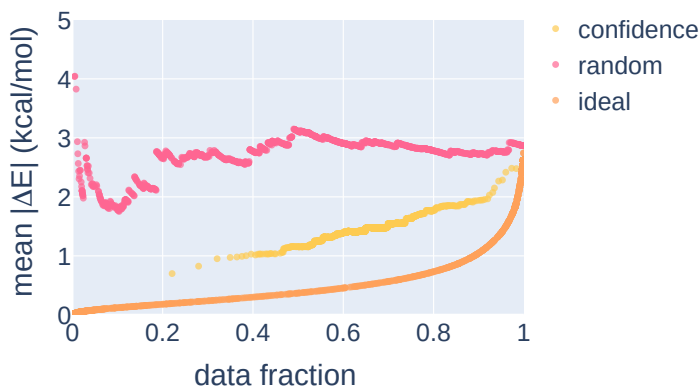


Figure 8. Mean barrier height of React-OT with different methods for adding confidence data. Yellow corresponds to the confidence model used in the main text. Pink corresponds to random selection, giving mean errors oscillating around the overall model performance. Orange corresponds to the ideal case, where the lowest error data point is always selected as the data fraction increases.