nature research

Peer Review Information

Journal: Nature Computational Science
Manuscript Title: Accurate transition state generation with an object-aware equivariant elementary reaction diffusion model
Corresponding author name(s): Dr Chenru Duan

Editorial Notes:

Reviewer Comments & Decisions:

Decision Letter, initial version:

Date: 29th August 23 16:36:54
Last Sent: 29th August 23 16:36:54
Triggered By: Kaitlin McCardle
From: kaitlin.mccardle@us.nature.com
To: crduan@mit.edu
BCC: kaitlin.mccardle@us.nature.com
Subject: Decision on Nature Computational Science manuscript NATCOMPUTSCI-23-0853
Message: ** Please ensure you delete the link to your author homepage in this e-mail if you wish to forward it to your co-authors. **

Dear Dr Duan,

Your manuscript "Accurate transition state generation with an object-aware equivariant elementary reaction diffusion model" has now been seen by 3 referees, whose comments are appended below. You will see that while they find your work of interest, they have raised points that need to be addressed before we can make a decision on publication.

The referees' reports seem to be quite clear. Naturally, we will need you to address *all* of the points raised.

While we ask you to address all of the points raised, the following points need to be substantially worked on:

- Please be sure to provide additional quantitative comparisons to existing methods to verify the level of improvement demonstrated by your approach.

Please use the following link to submit your revised manuscript and a point-by-point response to the referees' comments (which should be in a separate document to any cover letter):

[REDACTED]

** This url links to your confidential homepage and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this e-mail to co-authors, please delete this link to your homepage first. **

To aid in the review process, we would appreciate it if you could also provide a copy of your manuscript files that indicates your revisions by making use of Track Changes or similar mark-up tools. Please also ensure that all correspondence is marked with your Nature Computational Science reference number in the subject line.

In addition, please make sure to upload a Word Document or LaTeX version of your text, to assist us in the editorial stage.

To improve transparency in authorship, we request that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit please visit www.springernature.com/orcid.

We hope to receive your revised paper within three weeks. If you cannot send it within this time, please let us know.

We look forward to hearing from you soon.

Best regards,

Kaitlin McCardle, PhD
Associate Editor
Nature Computational Science

Reviewers comments:

Reviewer #1 (Remarks to the Author):

This manuscript presents a new TS generation model based on the diffusion technique and its validation. The proposed model generates TS structures with high speed and accuracy. This article is highly appealing to many chemists who want TS structures with low computational cost. Although this article includes state-of-the-art

ML models, this manuscript still needs improvement to present ideas clearly to potential readers of Nature Computational Science.

(1) There are some descriptions to be corrected or augmented.

- ‘reaction rate estimation‘ in the abstract should be reaction barrier because reaction rate prediction includes the prediction of pre-Arrhenius factor

- In introduction, the authors emphasize that their model preserve all symmetry that TS structures have but they do not explain what they are. In my perspective, additional paragraph (at least couple of sentences) to explain the symmetries that TS structure have is required somewhere in the manuscript.

- In the method section, ‘t’ and ‘s’ are used without any explanation. Of course, many of authors may infer that it indicates a diffusion step but for the completeness, the explanation should be included in the manuscript.

- In ‘high quality TS structure from OA-reactDiff’ subsection, the author mentioned that generation of TS structure takes 6 second with a single V100 GPU. But it is unclear whether 128 samples are generated during 6 seconds or just a single sample is constructed. (as far as I understood, 128 samples are used for the validation of TS structures in the result section)

(2) Figures are not clear enough.

- For Figure 3 (particularly c and d), labels are rarely readable. Please enlarge the font size.

- Figure 4c shows true and predicted structures. The authors select different colors but the structures are mostly overlapped. I recommend the authors to improve visibility by controlling transparency

(3) Arbitrary numbers (0.2 A) in the ranking model.

The authors construct a ranking model to predict whether generated structures are close to the TS structure in the database. If the ranking model is trained to return 1 if the generated structure has RMSD value smaller than 0.2 otherwise 0. However, to me, 0.2A is arbitrary number. I recommend further experiments using different numbers (e.g. 0.1, 0.15, 0.25A) or using different target. (e.g. sigmoid[1/RMSD])

(4) Chemical meaning and potential usage of alternative structures

As we all know, the TS is not unique. Therefore, the validation of generated structures (beta and gamma structures in Figure 3) may be chemically interesting. Since alpha structures are very close to the true one, they does not required quantum chemical validation (IRC calculation). However, beta, gamma, and delta structures are different structures. Potential readers may be curious whether those results are

just junk structures or one of TS structures. It could be chemically important if those beta and gamma structures are alternative TS of reactions and their energies are lower than that of the true structure. In this perspective, authors need to address the potential usage of beta, gamma, and delta structures in the manuscript..

(5) Size extensivity

- This model can be applied regardless of the size of molecules. However, the accuracy of the model is not discussed in the manuscript. Unlike the model that requires atom-mapping, the diffusion model may dramatically fail in predicting TS structures whose size is out of training dataset. I think additional experiments to validate the model accuracy for the large-scale reactions may extend the applicability of the proposed model. If appropriate reaction database is not available, authors can measure the performance of the model trained with reactions having a small number of atoms.

(6) Additional explanations are required to reference model.

- In the comparison of accuracy with respect to other models, PSI-model is trained with Grambow's set, which is different from what the author used. I know that transition1x is constructed based on Grambow's set but it would impact the result. So I think it should be noticed somewhere in the manuscript.

- On the 10th page, the authors used "(PSI-model) requires the prior knowledge of atom mapping and careful alignment between reactant and product." It is true that PSI-model requires atomic mapping but as far as I understood, the alignment between reactant and product is not required because test time augmentation is used in prediction, which means the model runs two times and the final structures are derived from both inference results.

Reviewer #2 (Remarks to the Author):

The paper deals with an important point in computational chemistry. In particular, the findings herein reported are important and may have a relevant impact on the study of reaction mechanisms. The paper is well-written and presented. There are, however, some important points to be taken into consideration and addressed in more detail:
1/Several times in the paper, the authors refer to DFT calculations. Most of the time, they also report quantitative comparisons and errors with respect to them. I found this aspect very elusive since without the specification of the functional and the computational set-up, it is impossible to assess/evaluate the numbers they reported.
2/The same considerations apply to the training set they have used: Was a specific choice made (e.g., functional)? If yes/no, why? any foreseen impact?
3/It is known that the geometry of the TS is "less" affected by the chemical accuracy in terms of geometry (see e.g., W. G. Liu and W. A. Goddard, J. Am. Chem. Soc., 2012, 134, 12970 —12978). How did the authors deal with geometry and energy in the selection of the training set?

4/ Would it be possible to exploit specific physical insights in the algorithm? E.g., nature of the TS (late, early, ...).

Reviewer #3 (Remarks to the Author):

The authors have developed a state-of-the-art method which can predict transition state (TS) of molecular reaction using diffusion model which exhibits outstanding performance in terms of RMSD of TS structure with respect to reference TS. At the same time, computational cost involved in TS optimization has been significantly reduced without potential convergence issue. What makes this model even more remarkable is SE(3)-equivariance which makes it free from a priori dependence toward atom-mapping between reactant and products and initial 3D arrangement of reactant molecules which are often non-trivial. I believe this manuscript has its sufficient novelty to be published in this journal nevertheless I'd like to suggest a few points for authors to consider prior to the publication.

1. In page 2, "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.", I think authors meant translation but transition could be misleading to readers. Also, in the same page, please provide unabbreviated form for MAE as you did for RMSD for consistency.

2. In page 4, "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", there are typo in this sentence.

3. I think it's very impressive that this model doesn't require atom-mapping and careful alignment of reactant molecules. However, the way author tested the novelty of the model does not sufficiently demostrate this feature because it was tested within Transition 1x dataset which already has atom-mapping and aligned reactant molecules. I think author can generate several reactions which are not included in Transition 1x dataset. For these reactions, there are no atom-mapping and one has no prior knowledge how multiple reactants should be initially aligned in order reaction to be proceeded and the result of model prediction can be compared with DFT optimized TS structure. The value of this model would be better highlighted by its performance toward unknown reactions we encounter during reaction network exploration where we have to generate our own dataset of TS structures.

4. Along with previous point, it would be very helpful to see the trend of error decline with the size of the training set. (Like a learning curve) At least for Transition 1x dataset, one can estimate how much error one could expect when there are different sizes of training set. I hope this could be used in the reaction network exploration in a similar fashion to active learning of machine-learning interatomic potential which starts from small dataset.

5. In Fig 3b, I wonder how authors have classified the clusters. Especially in the third right-most column, cluster alpha and gamma are neighboring and their distinction might need some clarification. There should be some consistent way that works for all

the reactions included in the work. Also author may want to elaborate what is the difference between "confidence rank" and "RMSD rank" which corresponds to the color and the size of the points. I understood that confidence rank is related to the predicted accuracy of TS structure generated by the model but the meaning of RMSD rank a bit unclear.

6. I think how TS energy has been trained in this work is not sufficiently explained in the manuscript. If I understood correctly, the diffusion based model is solely predicting the TS structure, not the energy. The details on how energy is trained and predicted should be elaborated.

7. In page 10, authors compare their model performance with density functional tight binding (DFTB) but cited paper is general review on DFTB. There could be various flavors of DFTB which have different parameterization. Therefore, it would be better and fair if authors can specify which particular DFTB model has been used for their comparison.

8. In Table 1, the Pearson's r value of the developed model is less than the other models used for comparison. Authors may want to discuss further how this can be interpreted.

9. Which value of alpha t or sigma t is used for noising in this study? Is it adaptive? I think its value has not been mentioned in the manuscript.

Author Rebuttal to Initial comments

nature research

Dr. Chenru Duan
Department of Chemistry
Department of Chemical Engineering

Massachusetts Institute of Technology
77 Massachusetts Avenue, Cambridge
Massachusetts 02139

Phone (857) 928-4855
Email crduan@mit.edu

September 14, 2023

Dr. Kaitlin McCardle, Editor
Nature Computational Science

Dear Dr. McCardle and Reviewers,

Please find enclosed a submission our manuscript entitled "Accurate transition state generation with an object-aware equivariant elementary reaction diffusion model" by Chenru Duan, Yuanqi Du, Haojun Jia, and Heather J. Kulik for consideration of publication in *Nature Computational Science*. The corresponding author is myself – contact information is given in the letterhead.

We have made changes according to concerns raised by the reviewers, as indicated below. Reviewers' comments are provided in full on pages 2-16 of this letter with response to each comment labeled "Response" and reproduction of corresponding changes to the text labeled "Changes to text" with associated page numbers.

For the ease of the review of this reply letter, we provide a "general response" section, aiming to address two common questions from the reviewers, which motivated us to perform new model training and analysis during this revision. All the original comments from the reviewers are in black, our responses in blue, and text update outside of this response letter (main text, supplementary information, etc.) in red. We also provide a copy of the original manuscript with all changes marked in red and uploaded it as supplementary information for review. If you have any questions, please do not hesitate to contact me as indicated above.

Sincerely,

Chenru Duan

General response to Reviewers #1-3:

We thank you for reading the manuscripts thoroughly and providing us positive feedback. Many comments are inspiring and helped us improved the manuscript significantly. There are two common questions from multiple reviewers that are particularly interesting, which we would like to address upfront. Note that these statements are short summaries for the finding of these two common questions after we performed new model training and analysis during this revision. For more detailed answer for each question, we refer the reviewers to read the response at the relevant question in the later part of this response letter.

1. R#1Q#5 and R#3Q#4 both ask for the model performance with respect to larger molecules and smaller training set sizes.

The OA-ReactDiff model was originally trained on 9000 reactions randomly sampled from the Transition1x dataset. In this revision, we further trained three models: one on 1000 reactions randomly sampled from the original 9000 training data, one on 5000 reactions randomly sampled from the original 9000 training data, and one on all reactions with system size < 15 atoms in the original 9000 training data (5734 in total). We find that

1) models trained with 1000, 5000, and 9000 reactions give near power-law decay RMSD with respect to the training data, which is not depleted with 9000 training reactions. This behavior suggests the room of improvement for OA-ReactDiff on larger reaction datasets.
2) the model trained with 5734 reactions with small systems lies on the same trend as those of the three models that do not have constraints on system size. This behavior actually highlights the capability of OA-ReactDiff on generalizing to larger systems not including in the training data.

2. R#1Q#4, R#2Q#4, and R#3Q#5 for whether the generated TS structures from OA-ReactDiff were chemically meaningful and how the 128 structures are clustered.

We really appreciate this common question from all reviewers. Originally, we did emphasize the chemical relevance of all sampled structures due to the stochastic nature of OA-ReactDiff. The clustering results were simply obtained by PCA and k-medoids sampling where the number of medoids was determined by eye. However, this comment from the reviewers raised our attention on this ambiguity and motivated us to perform a more thorough study on these three example reactions in Figure 3. During this revision, we computed the Hessian of all 128 generated TS structures in each reaction and found many of them only have one imaginary frequency and thus are good candidates for TS optimization (Fig. 3b). For those with only one imaginary frequency, we performed single-ended TS optimization with DFT. We then analyzed these optimized TS structures by their pairwise RMSD to identify unique TS structures by clustering structures where the RMSD is smaller than 0.1 Å (Fig. 3c). Lastly, we calculated the RMSD and absolute energy difference between the OA-ReactDiff generated and DFT optimized TS structures (Fig. 3d). In addition, we performed internal reaction coordinate calculation on the unique TS structures to get the reactant and product (Supplementary Data). As one may expect, these different 3D TS structure lead to different reactant and product conformation or connectivity. These results suggest that

1) Despite the stochasticity, many TS structures generated by OA-ReactDiff are chemically meaningful, which can lead to true saddle point of different conformation or connectivity. These different TS structures can be exploited during the reaction network exploration for reactions that would otherwise be neglected.
2) Sometimes, OA-ReactDiff can generate TS structures that are close to saddle point, even when the reactant and product provided are in different (i.e., incorrect) conformation or connectivity.
3) Our recommender can reliably identify the TS structure of the correct conformation given the 3D geometry of reactant and product

These new observations and analysis lead to a major update to the Fig. 3.

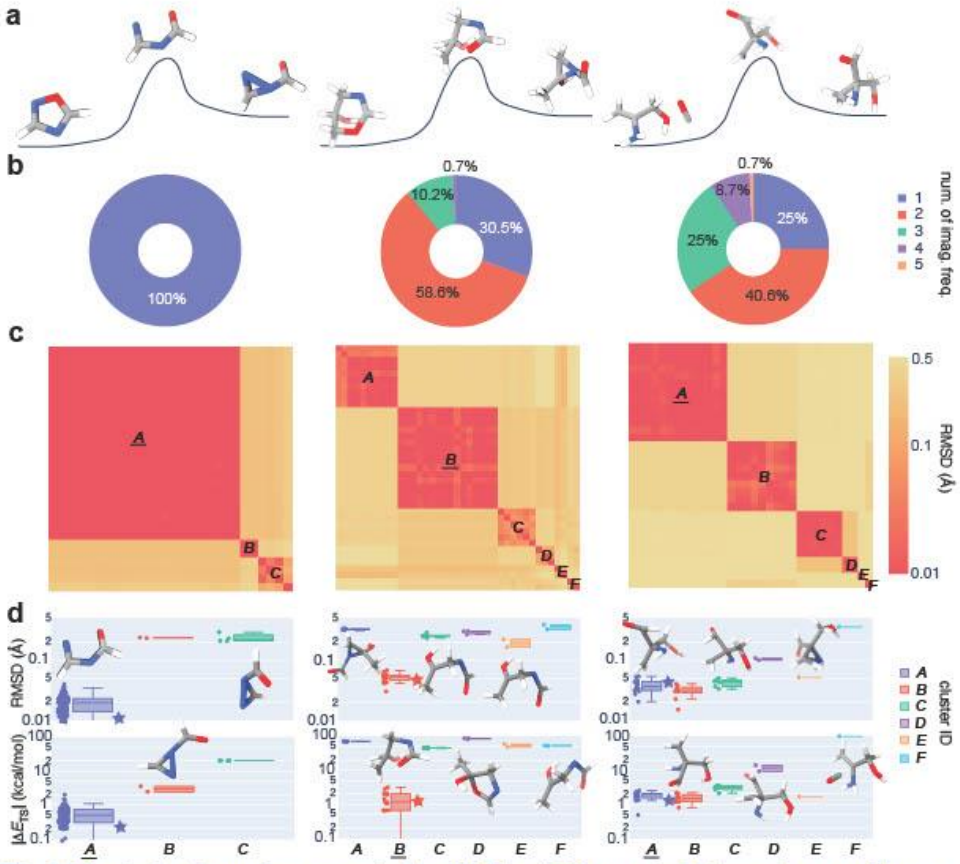


Fig. 3 | Analysis of samples generated by OA-ReactDiff on select elementary reactions. a. Illustration of select reactions: aromatic 5-membered ring to high-valence 3-membered ring rearrangement that breaks and forms one bond (left, b1f1), 6-membered ring to saturated 3-membered ring rearrangement (middle, b2f2), and carbonyl attack reaction (right, b2f3). All three elementary reactions have not been seen during the model training. **b.** Pie chart showing the percentage of OA-ReactDiff generated TS structure with different numbers of imaginary frequencies (imag. freq.) evaluated by a Hessian calculation by ωB97x/6-31G(d). **c.** Pairwise RMSD matrix for OA-ReactDiff generated and ωB97x/6-31G(d)-optimized TS structures. Only the generated structures with one imag. freq. are optimized by DFT. The structures are sorted to form clusters (where the RMSDs are smaller than 0.1 Å) on the diagonal line of the matrix. Each cluster is labeled by a capital bold italic letter. The cluster corresponding to the true TS is underlined. RMSD is colored through a red-to-yellow color bar in the log scale to show the clustering effect more clearly. **d.** RMSD (top) and absolute energy difference ($|\Delta E_{TS}|$, bottom) between OA-ReactDiff generated and their corresponding ωB97x/6-31G(d)-optimized TS structures for each reaction. These structures are grouped by the cluster observed in c.. A representative structure for each cluster is shown. **The RMSD and $|\Delta E_{TS}|$ for top-1 confidence TS structure picked by the recommender are shown with the star symbol.** Atoms are colored as follows: gray for C, blue for N, red for O, and white for H.

During the revision of this manuscript, we found a concurrent work from Kim et al. named TSDiff (https://arxiv.org/abs/2304.12233), where the 3D geometry of the TS is generated from a 2D graph composed of molecular connectivity of reactant and product via a diffusion model. They circumvent the need to handle object-wise SE(3) symmetry by removing 3D conformations of reactant and product as inputs. This simplification comes with the price of not being able to select "the best" TS structure out of all generated samples. However, it would be of interest to compare OA-ReactDiff plus 2D-to-3D conformation sampling and TSDiff in reaction exploration in future work. We added this as a comment in the Discussion section.

Changes to text: At the second last paragraph of the Discussion section, added in red, "During the revision of this manuscript, we found a concurrent work from Kim et al. named TSDiff[49], which showcases the use of diffusion model for TS generation in another perspective. There, the TS structure is generated from a 2D graph composed of molecular connectivity of reactant and product via a diffusion model. They circumvent the need to handle object-wise SE(3) symmetry by removing 3D conformations of reactant and product as inputs. This simplification comes with the price of not being able to select "the best" TS structure out of all generated samples[50]. However, it would be of interest to compare OA-ReactDiff plus 2D-to-3D conformation sampling and TSDiff in reaction exploration in future work."

Reviewer #1 (Remarks to the Author):

This manuscript presents a new TS generation model based on the diffusion technique and its validation. The proposed model generates TS structures with high speed and accuracy. This article is highly appealing to many chemists who want TS structures with low computational cost. Although this article includes state-of-the-art ML models, this manuscript still needs improvement to present ideas clearly to potential readers of Nature Computational Science.
Response: We thank the reviewer for their time and careful consideration of our work.
Changes to text: None.

(1) There are some descriptions to be corrected or augmented.

- 'reaction rate estimation' in the abstract should be reaction barrier because reaction rate prediction includes the prediction of pre-Arrhenius factor
Response: We agree with the reviewer that we do not predict the pre-Arrhenius factor and thus "barrier" is a more appropriate word than "rate".
Changes to text: Change "reaction rate" to "reaction barrier" throughout.

- In introduction, the authors emphasize that their model preserve all symmetry that TS structures have but they do not explain what they are. In my perspective, additional paragraph (at least couple of sentences) to explain the symmetries that TS structure have is required somewhere in the manuscript.
Response: We thank the reviewer for this suggestion. We have description for the symmetry at "Overview of OA-ReactDiff", "Object-aware SE(3) implementation", and SI "Physical symmetries and constraints in an elementary reaction". But we agree elaborating more in the introduction would improve the readability for non-ML experts.
Changes to text: Added at the last paragraph of introduction after the first sentence, colored in red, "These symmetries include permutation among atoms in a fragment of reactant or product, permutation among fragments in reactant or product, and rotation and translation symmetry for each fragment in reactant and product (Supplementary Text S1). With all these symmetries satisfied, we bypassed the need of atom order mapping and fragment alignment in TS search, generating 3D TS structure with only the 3D geometry of fragments in reactant and product."

- In the method section, 't' and 's' are used without any explanation. Of course, many of authors may infer that it indicates a diffusion step but for the completeness, the explanation should be included in the manuscript.
Response: We thank the reviewer for pointing out this issue. We added the explanation at their first occurrence in the Method section.
Changes to text: At "Equivariant diffusion models", added in red, "where $s < t$ refer to two different timesteps along the diffusion/denoising process ranging from 0 to T,".

- In 'high quality TS structure from OA-reactDiff' subsection, the author mentioned that generation of TS structure takes 6 second with a single V100 GPU. But it is unclear whether 128 samples are generated during 6 seconds or just a single sample is constructed. (as far as I understood, 128 samples are used for the validation of TS structures in the result section)
Response: We thank the reviewer for pointing out this potential ambiguity in the phrasing. By within 6 second on a V100 GPU, we refer to the runtime per sample with batching. For example, the walltime is 10 minutes with a batch size of 128, and the average runtime per sample is 4.6 second. This is a reasonable metric as a batch size of 128 requires 11.2 GB GPU memory, which is standard for modern GPUs. If we restrict ourselves to no batching, which is an uncommon use case, the runtime would be 17.1 seconds. We revised the text to make this point more clear and added this Table in the SI.

Table S2. Runtime and GPU memory consumption with different batch sizes.

Batch size	1	2	4	8	16	32	64	128	256
Walltime (sec)	17.1	18.9	28.0	44.7	87.1	171.8	322.9	582.1	1106.1
Runtime per sample (sec)	17.1	9.5	7.0	5.6	5.4	5.4	5.1	4.6	4.3
GPU Memory (GB)	1.1	1.3	1.4	1.6	2.2	3.6	5.7	11.2	21.8

Changes to text: At the first paragraph at the "High quality TS structures from OA-ReactDiff" section, changes in red, "Notably, in contrast to an average runtime of 12 hours using climbing image NEB[22] with DFT, it only takes on average within 6 seconds runtime to generate a TS structure with OA-ReactDiff with proper batching on a V100 GPU (single sample generation without batching takes 17 seconds, Supplementary Table S2)."

(2) Figures are not clear enough.

- For Figure 3 (particularly c and d), labels are rarely readable. Please enlarge the font size.
- Figure 4c shows true and predicted structures. The authors select different colors but the structures are mostly overlapped. I recommend the authors to improve visibility by controlling transparency

Response: We thank the reviewer for the suggestions. We enlarged the label for Figure 3. For Figure 4, we altered the transparency but found this made the figure less clear, potentially due to that some structures overlap a lot. We provided the raw xyz files in the SI data in case the reader found the structures are unclear in the Figure and readers would like to examine the structures further.
Changes to text: Updated Figure 3.

(3) Arbitrary numbers (0.2 A) in the ranking model.

The authors construct a ranking model to predict whether generated structures are close to the TS structure in the database. If the ranking model is trained to return 1 if the generated structure has RMSD value smaller than 0.2 otherwise 0. However, to me, 0.2A is arbitrary number. I recommend further experiments using different numbers (e.g. 0.1, 0.15, 0.25A) or using different target. (e.g. sigmoid[1/RMSD])

Response: We agree with the reviewer that the choice of 0.2 A is a bit arbitrary. The fact is that we choose 0.2 A because we went through some examples in the dataset (similar to Fig. 4c) and found 0.2 A usually give quite indistinguishable pairs of structures. To attempt a more rigorous process, we trained another regression model predicting the RMSD between generated and true TS given {Reactant, TS_gen, Product}. We observe the same behavior of using this regressor as recommender, where the TS with the lost predicted RMSD is selected (see the Figure below). We would like to keep the original classifier-based recommender in the main chuck of the text due to the simplicity and more intuitive confidence score. We mentioned this regressor in the main text and added this Figure in the SI.

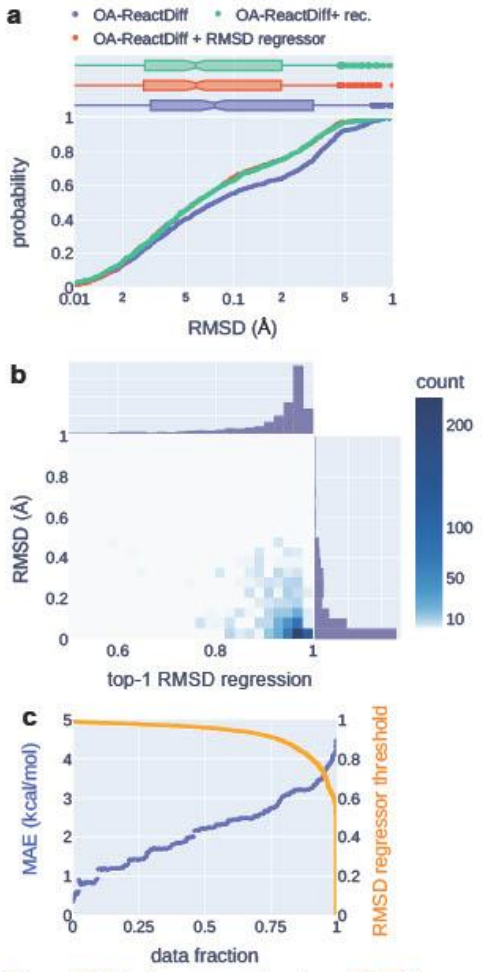


Figure S6. Performance of using a RMSD regressor as the confidence model. a. Cumulative probability for RMSD between the true TS structures and OA-ReactDiff samples on 1073 set-aside test reactions. The OA-ReactDiff samples are evaluated under one-shot generation (blue), the top-1 confidence sample via classifier-based recommender (green), and the top-1 confidence sample via RMSD regressor 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 RMSD regressor 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. MAE of $|\Delta E_{TS}|$ (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 TS structures selected via the RMSD regressor.

Changes to text: At the last paragraph of section "High quality TS structures from OA-ReactDiff", added in red, "We also trained a regressor that predicts the RMSD between generated and true TS given a set of reactant, product, and generated TS structure as the confidence model, where the same quantitative behavior is observed (Supplementary Fig. S6)."

(4) Chemical meaning and potential usage of alternative structures

As we all know, the TS is not unique. Therefore, the validation of generated structures (beta and gamma structures in Figure 3) may be chemically interesting. Since alpha structures are very close to the true one, they does not required quantum chemical validation (IRC calculation). However, beta, gamma, and delta structures are different structures. Potential readers may be curious whether those results are just junk structures or one of TS structures. It could be chemically important if those beta and gamma structures are alternative TS of reactions and their energies are lower than that of the true structure. In this perspective, authors need to address the potential usage of beta, gamma, and delta structures in the manuscript..

Response: We thank the reviewer for this inspiring question! As we mentioned in the point #2 of the "General response", we computed the Hessian of all 128 generated TS. For those with only one imaginary frequency, we performed single-ended TS optimization with DFT. There, we indeed identified multiple saddle points that correspond to different TS structures. By clustering these TSes and performing IRC calculations on the unique TS structures, we confirmed that these TS lead to different reactant and product conformations or connectivity, and thus form distinct elementary reactions compared to the intended one. These different TS structures can still be exploited during the reaction network exploration for reactions that would otherwise be neglected.

Changes to text: Updated Fig. 3. At the second paragraph of "Overcoming the stochastic nature of diffusion models with confidence ranking.", changes in red, "We then computed the Hessian for these 128 generated TS structures with ωB97x/6-31G(d) and found many only contained one imaginary frequency and thus are good candidates for single-ended TS optimizations (Fig. 3b). To evaluate the differences among the structures with one imaginary frequency, we performed TS optimization on these structures with ωB97x/6-31G(d). We then computed the pairwise RMSD and identified multiple distinct TS structures discovered by OA-ReactDiff (Fig. 3c). Further internal reaction coordinate calculations with ωB97x/6-31G(d) confirmed that these TS structures lead to different reactant and product conformation or connectivity, and thus form distinct elementary reactions compared to the intended one (Supplementary Data). These generated TS structures sometimes show relatively large structural deviations (i.e., > 0.2 Å) compared to the optimized TS structures, which is expected since the input reactant and product do not actually correspond to the generated TS structure (Fig. 3d). However, these different TS structures can still be exploited during the reaction network exploration for elementary reactions that would otherwise be neglected."

(5) Size extensivity

- This model can be applied regardless of the size of molecules. However, the accuracy of the model is not discussed in the manuscript. Unlike the model that requires atom-mapping, the diffusion model may dramatically fail in predicting TS structures whose size is out of training dataset. I think additional experiments to validate the model accuracy for the large-scale reactions may extend the applicability of the proposed model. If appropriate reaction database is not available, authors can measure the performance of the model trained with reactions having a small number of atoms.

Response: This is a great question! We bin the performance of OA-ReactDiff + recommender and do not see a very obvious trend of performance deterioration with increasing system size (see the Figure below). To further investigate on the system size extensivity, we trained models with 1000 and 5000 data points randomly sampled from the original 9000 training data, as well as another model with only system size < 15 reactant/product (5734 out of 9000 training data). These models are then evaluated on the original 1073 test reactions. We find all model performances roughly lies on the trend as the system size increases (see the Table below). Particularly, the model trained with 5734 small systems

(i.e., < 15 atoms) show comparable RMSD statistics compared to that trained with 5000 randomly drawn samples that contain larger systems. This observation actually highlights the extensiveness of OA-ReactDiff. We added these results and commented on this point in the main text. Reviewer #3 asked a relevant question about the error scaling with respect to the number of training data at Question #4. We refer the reviewer to read more about this topic if you are interested at the response there.

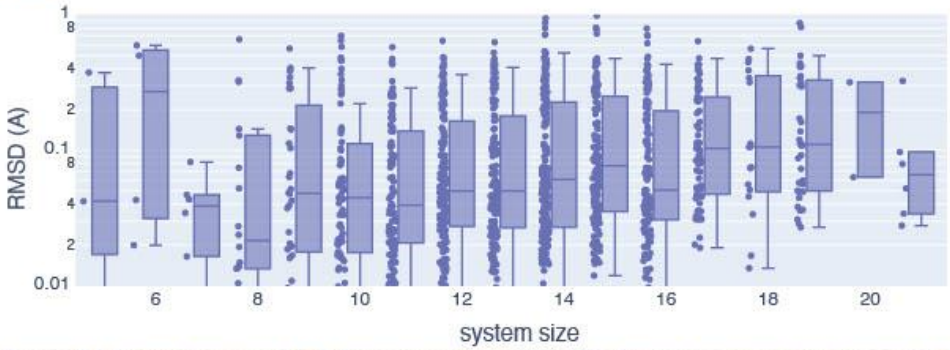


Figure S4. Distributions of RMSD binned by system size. The RMSD is computed between top-1 confidence TS structure generated by OA-ReactDiff and the true TS structure for the 1073 test reactions. A standard box plot with median as the horizontal bar is shown with all RMSDs at that system size displayed on its left hand side.

Table S3. Performance of OA-ReactDiff at different number of training data and constraints.

Number of sample	1000	5000	5734	9000
Mean RMSD (Å)	0.453	0.252	0.240	0.183
Median RMSD (Å)	0.424	0.174	0.180	0.076
Constraints in training data	--	--	system size < 15 atoms	--

Changes to text: At the first paragraph of section "High quality TS structures from OA-ReactDiff", added in red, "We observed a near power-law dependence of RMSD for OA-ReactDiff generated samples with the number of training data (Supplementary Table S3 and Fig. S3). The performance of OA-ReactDiff shows no sign of depletion when we reach the maximum of our training data available, suggesting room for further improvement through training OA-ReactDiff on larger datasets. More interestingly, we find that given a similar amount of training data, an OA-ReactDiff model trained only on system within 15 atoms performs similarly well to that trained on randomly sampled reactions (Supplementary Table S3). Meanwhile, the OA-ReactDiff model yields similar RMSD distribution regardless of system size (Supplementary Fig. S4). This observation showcases great size extensiveness of OA-ReactDiff, which is encouraging for its application in practical reaction exploration for large molecules."

(6) Additional explanations are required to reference model.

- In the comparison of accuracy with respect to other models, PSI-model is trained with Grambow's set, which is different from what the author used. I know that transition1x is constructed based on Grambow's set but it would impact the result. So I think it should be noticed somewhere in the manuscript.

Response: We thank the reviewer pointing this out and we fully agree with the reviewer that the two datasets are slightly different. In addition, we do not use exactly the same train/test partitioning despite

the train/test ratio is the same, which may also impact the result. We added these notes in the Table caption.
Changes to text: At the caption of Table 1, added in red, "Note that PSI-based model[21] used dataset from Grambow et al.[41], which is the predecessor of Transition1x. Although all three approaches used a 8:2 training/test ratio, the random seeds for the partition are not the same, which may have influence on the comparison".

- On the 10th page, the authors used "(PSI-model) requires the prior knowledge of atom mapping and careful alignment between reactant and product." It is true that PSI-model requires atomic mapping but as far as I understood, the alignment between reactant and product is not required because test time augmentation is used in prediction, which means the model runs two times and the final structures are derived from both inference results.
Response: We thank the reviewer for the comment. We read the PSI-model paper and its peer review file again. To the best of our understanding, PSI-model still requires fragment alignment of reactant and product for multi-molecular reactions. The test time augmentation that the reviewer mentioned refers to a process that the run the model twice with inputs of [reactant, product] and [product, reactant] to cancel asymmetry of the model predictions.

In their peer review file (https://static-content.springer.com/esm/art%3A10.1038%2Fs41467-023-36823-3/MediaObjects/41467_2023_36823_MOESM2_ESM.pdf), reviewer #3 asked about the alignment issue for their PSI-model. From us reading the author's response, PSI-model still requires fragment alignment as a pre-processing step. Here is a quote from their response: "Although off-equilibrium conformations are utilized to predict TS structures, they are still sampled from well-optimized and aligned structures, which are obtained from quantum chemical calculations. This process is not trivial, particularly for bimolecular and trimolecular reactions, because the initial relative pose of molecules is a critical factor in determining reaction profiles." To avoid confusion and potential ambiguity, we slightly changed our wording in this sentence.

Changes to text: Updated the sentence with changes in red, "requires the prior knowledge of atom mapping and careful alignment among fragments in reactant and product for multi-molecular reactions."

Reviewer #2 (Remarks to the Author):

The paper deals with an important point in computational chemistry. In particular, the findings herein reported are important and may have a relevant impact on the study of reaction mechanisms. The paper is well-written and presented. There are, however, some important points to be taken into consideration and addressed in more detail:
Response: We thank the reviewer for their time and careful consideration of our work.
Changes to text: None.

1/Several times in the paper, the authors refer to DFT calculations. Most of the time, they also report quantitative comparisons and errors with respect to them. I found this aspect very elusive since without the specification of the functional and the computational set-up, it is impossible to assess/evaluate the numbers they reported.
Response: We thank the reviewer for this comment. We use the Transition1x dataset throughout the entire manuscript, which was generated under ωB97x/6-31G(d). During the evaluation, whenever we mentioned DFT results, they are computed with the same functional and basis set ωB97x/6-31G(d). We added more description about this choice in the DFT evaluation part to make this point more clear.

Changes to text: At the first paragraph of "Approaching the energetic accuracy needed in TS search", changes in red, "During the DFT evaluation, we used ωB97x/6-31G(d) throughout this work, consistent with the method and basis set choice with Transition1x. Specifically, we compared the electronic energy ~~computed by ωB97x/6-31G(d)~~ between OA-ReactDiff recommended and true TS structures for evaluating the barrier height."

2/The same considerations apply to the training set they have used: Was a specific choice made (e.g., functional)? If yes/no, why? any foreseen impact?
Response: We thank the reviewer for this comment. We did not generate any new reaction data through transition state optimizations. All reaction data was adopted from Transition1x, which used ωB97x/6-31G(d). We emphasize this point in the manuscript now to avoid potential confusion.
Changes to text: At the first paragraph of "Details for model training", added in red, "Throughout this manuscript, we do not generate new reaction data through DFT-based TS optimizations, except for ones used as examples in Fig. 3. Therefore, all DFT results discussed here are computed with ωB97x/6-31G(d)."

3/It is known that the geometry of the TS is "less" affected by the chemical accuracy in terms of geometry (see e.g., W. G. Liu and W. A. Goddard, J. Am. Chem. Soc., 2012, 134, 12970 — 12978). How did the authors deal with geometry and energy in the selection of the training set?
Response: We thank the reviewer for this great question! We agree with the reviewer that the choice of functional should not impact the geometry of TS structure that much, especially for small organics with CNOH. We only adopted data from Tranisition1x, which was computed with ωB97x/6-31G(d). From all 10073 reactions, we randomly partitioned them to 9000 training reactions and 1073 test reactions. Since the reviewer raised this point about functional dependency, we re-evaluated the barrier height error from OA-ReactDiff + recommender under two additional methods, B3LYP/6-31G(d) and PBE/6-31G(d). We confirmed a similar quantitative behavior across these three functionals, demonstrating weak functional dependence of performance reported in this work. We added a comment in the main text and added this Table in the SI.

Table S4. Barrier height error of OA-ReactDiff evaluated by different methods. All energies are reported in the unit of kcal/mol.

	ωB97x		B3LYP		PBE		Data fraction
	mean	median	mean	median	mean	median	
OA-ReactDiff + rec.	4.4	1.6	4.3	1.6	4.9	1.7	1.0
OA-ReactDiff + rec. (p>0.5)	3.1	1.4	3.0	1.4	3.7	1.6	0.86

Changes to text: At the second paragraph of the section "Approaching the energetic accuracy needed in TS search", added in red, "It is known that the TS structure is less sensitive to the choice of DFT functional,[48] especially or small organic molecules with only CNOH. Here, we also evaluated the error for barrier height estimation of OA-ReactDiff with two other functionals and found quantitatively similar performance (Supplementary Table S4)."

4/ Would it be possible to exploit specific physical insights in the algorithm? E.g., nature of the TS (late, early, ...).
Response: We thank the reviewer for this inspiring question! As we mentioned in the point #2 of the "General response", we performed a series of analysis on the chemical meanings of the generated TS structures from OA-ReactDiff. There, we found that many OA-ReactDiff generated structures that are not picked by the recommender are not "garbage" – They can lead to non-intended elementary reactions that would otherwise be neglected during the reaction exploration. For the particular question about whether OA-ReactDiff can classify whether a TS is early (i.e., closer to reactant) or late (i.e.,

closer to product), we do not see an easy way to tease this information out, mostly because OA-ReactDiff generates a TS structure from noises sampled from normal distributions (which does not have bias to either reactant or product).

Changes to text: At the third paragraph of "Overcoming the stochastic nature of diffusion models with confidence ranking.", changes in red, "Without the confidence model, random selection from samples generated by OA-ReactDiff may result in a TS structure of completely different conformation or connectivity compared to the reactant and product. It would in turn 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."

Reviewer #3 (Remarks to the Author):

The authors have developed a state-of-the-art method which can predict transition state (TS) of molecular reaction using diffusion model which exhibits outstanding performance in terms of RMSD of TS structure with respect to reference TS. At the same time, computational cost involved in TS optimization has been significantly reduced without potential convergence issue. What makes this model even more remarkable is SE(3)-equivariance which makes it free from a priori dependence toward atom-mapping between reactant and products and initial 3D arrangement of reactant molecules which are often non-trivial. I believe this manuscript has its sufficient novelty to be published in this journal nevertheless I'd like to suggest a few points for authors to consider prior to the publication.
Response: We thank the reviewer for their time and careful consideration of our work.
Changes to text: None.

1. In page 2, "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.", I think authors meant translation but transition could be misleading to readers. Also, in the same page, please provide unabbreviated form for MAE as you did for RMSD for consistency.
Response: We thank the reviewer for their careful reading and catching these two small issues. We revised them accordingly.
Changes to text: Introduction, paragraph 3, "transition" to "translation". Introduction, last paragraph, changes in red, "..., we obtain a mean absolute error (MAE) 2.6 kcal/mol on barrier height ..."

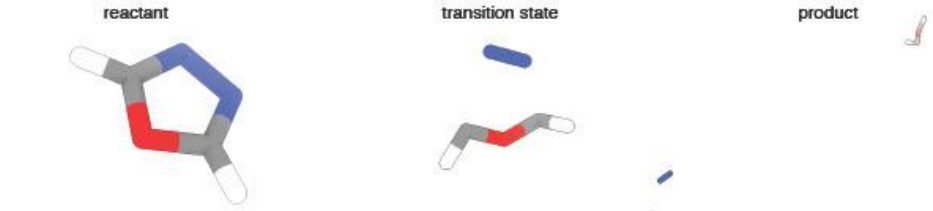
2. In page 4, "In TS search, specifically, it combines distributions from the diffused reactant an product (i.e., known parts) and denoised TS structure (i.e, unknown parts) at each step before proceeding to the following denoising step", there are typo in this sentence.
Response: We thank the reviewer for their careful reading and catching these two small issues. We revised them accordingly.
Changes to text: This sentence, changes in red, "In TS search, specifically, it combines distributions from the diffused reactant and product (i.e., known parts) and denoised TS structure (i.e., unknown parts) at each step before proceeding to the following denoising step", there are typo in this sentence."

3. I think it's very impressive that this model doesn't require atom-mapping and careful alignment of reactant molecules. However, the way author tested the novelty of the model does not sufficiently demostrate this feature because it was tested within Transition 1x dataset which already has atom-mapping and aligned reactant molecules. I think author can generate

several reactions which are not included in Transition 1x dataset. For these reactions, there are no atom-mapping and one has no prior knowledge how multiple reactants should be initially aligned in order reaction to be proceeded and the result of model prediction can be compared with DFT optimized TS structure. The value of this model would be better highlighted by its performance toward unknown reactions we encounter during reaction network exploration where we have to generate our own dataset of TS structures.

Response: We thank the reviewer for pointing out this great point on potential improvement on our demonstration and writing. We would like to note that although Transition1x dataset already has atom-mapping and aligned molecules as the reviewer mentioned, we have swapped the atom ordering and broken the alignment as a preprocessing step before the training and validation of OA-ReactDiff. Therefore, all results we reported in this work do not take advantage of the atom mapping and fragment alignment of Transition1x. We added a comment in the main text to make this point clear.

We agree with the reviewer that generating new reactions would be the ultimate use case of this method, which is why we have an example in the SI where we used OA-ReactDiff sampled all possible elementary reactions that may happen in a simple system of 4 atoms (C, N, O, H each, Supplementary Fig. S2). Along with the reviewer's comment, we added a new example where we generated the TS structure with reactant and products as inputs for a multi-molecular reaction. As can be seen in the raw xyz file, both the atom mapping and alignment are quite random. Still, we generated a reasonable TS structure which only has a RMSD of 0.03 A compared to the true TS after a DFT (ωB97x/6-31G(d)) optimization.




```
7

O 0.89812744 0.3577404 -0.068295956
C 0.35732 -0.4092097 0.90621525
N -0.87802404 -0.7156354 0.6834425
N -1.2095765 -0.11866519 -0.5262311
C -0.14500502 0.48708305 -0.9293382
H 0.9695818 -0.66973865 1.7556692
H 0.007576242 1.0684255 -1.8214616
```

```
7

C -0.9807652 -0.41762903 -0.67331475
N -1.1892109 0.17415379 1.1667829
C 1.3269237 -0.016659755 0.06412991
N -0.32823536 0.20446682 1.8621631
H 2.2551227 -0.18935578 -0.5450948
O 0.34054843 -0.22559318 -0.7199818
H -1.4243833 0.47061718 -1.1546844
```

```
7

N 15.999035 13.324711 13.365774
C -5.2679567 -4.3791122 -5.5255594
N 15.338476 12.598929 12.876601
C -7.0493574 -6.172478 -5.5539536
O -6.238264 -5.208907 -5.358236
H -8.039697 -5.784987 -5.2473087
H -4.7422347 -4.378155 -4.557316
```

Figure S1. Mutil-molecular elementary reactions sampled from OA-ReactDiff by specifying reactant and product. Here, the atom mapping and fragment alignment are randomized to demonstrate the capability of OA-ReactDiff not relying on these factors. The generated TS structure only has a RMSD of 0.03 Å compared to the DFT (ωB97x/6-31G(d)) optimized true TS. Atoms are colored as follows: gray for C, blue for N, red for O, and white for H.

Changes to text: At the second paragraph of the section "OA-ReactDiff training", added in red, "Transition1x dataset has atom mapping and aligned product molecules as constructed. To showcase of the capability of OA-ReactDiff not replying on the atom mapping and fragment alignment, we intentionally swapped the atom ordering and broke the alignment in Transition1x beforehand and verified that OA-ReactDiff functions well without these pre-processing requirement (Supplementary Fig. S1)."

4. Along with previous point, it would be very helpful to see the trend of error decline with the size of the training set. (Like a learning curve) At least for Transition 1x dataset, one can estimate how much error one could expect when there are different sizes of training set. I hope

this could be used in the reaction network exploration in a similar fashion to active learning of machine-learning interatomic potential which starts from small dataset.

Response: This is an interesting question! We further trained models with 1k and 5k training data randomly sampled from the original 9k training data and found a power law decay of RMSD with respect to the number of training data (see the Figure below). Interestingly, the performance improvement shows no sign of saturation with 9k training data, suggesting the room for improvement with increasing training data. For applying OA-ReactDiff at reaction network exploration, we would recommend one to take advantage of pretrained checkpoint of OA-ReactDiff model, which is usually more data efficient. Reviewer #1 asked a relevant question about the model size extensiveness at Question #5. We refer the reviewer to read more about this topic if you are interested at the response there.

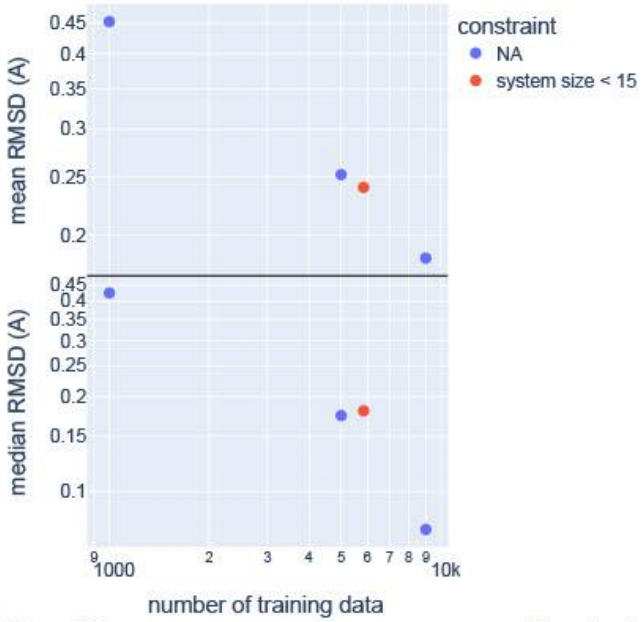


Figure S3. **RMSD vs. the number of training data.** Mean (top) and median (bottom) RMSD for OA-ReactDiff models trained on different number of training data that are either randomly sampled (blue) or sampled under the constraint that only systems within 15 atoms in size (red) from the original 9000 training reactions.

Changes to text: At the At the first paragraph of section "High quality TS structures from OA-ReactDiff", added in red, "We observed a near power-law dependence of RMSD for OA-ReactDiff generated samples with the number of training data (Supplementary Table S3 and Fig. S3). The performance of OA-ReactDiff shows no sign of depletion when we reach the maximum of our training data available, suggesting room for further improvement through training OA-ReactDiff on larger datasets. More interestingly, we find that , given a similar amount of training data, an OA-ReactDiff model trained only on system within 15 atoms performs similarly well to that trained on randomly sampled reactions (Supplementary Table S3). Meanwhile, the OA-ReactDiff model yields similar RMSD distribution regardless of system size (Supplementary Fig. S4). This observation showcases great size extensiveness of OA-ReactDiff, which is encouraging for its application in practical reaction exploration for large molecules."

5. In Fig 3b, I wonder how authors have classified the clusters. Especially in the third right-most column, cluster alpha and gamma are neighboring and their distinction might need some clarification. There should be some consistent way that works for all the reactions included in the work. Also author may want to elaborate what is the difference between "confidence rank" and "RMSD rank" which corresponds to the color and the size of the points. I understood that confidence rank is related to the predicted accuracy of TS structure generated by the model but the meaning of RMSD rank a bit unclear.

Response: We thank the reviewer for this inspiring question, which pushes us to think deeper on the potential ambiguity of the clustering results! As we mentioned in the point #2 of the "General response", we performed a series of analysis on the chemical meanings of the generated TS structures from OA-ReactDiff. Now we completely remove the clustering and replace it with more rigorous analysis. We thank the reviewer again for pointing out this issue!

Changes to text: Updated Fig. 3 and its caption. Updated the discussion around this figure. See "General response" and response in Review #1 Question #4.

6. I think how TS energy has been trained in this work is not sufficiently explained in the manuscript. If I understood correctly, the diffusion based model is solely predicting the TS structure, not the energy. The details on how energy is trained and predicted should be elaborated.

Response: We thank the reviewer for the comment. We do not train a model predicting the TS energy or barrier in this work. All the energy and barrier related evaluations were performed with DFT, using the same functional and basis set combination (i.e., ωB97x/6-31G(d)) as Tranistion1x dataset. As the reviewer pointed out, we can readily extend our method on predicting barrier height by swapping the output of the current model. However, we would like to focus on structure prediction in this manuscript to avoid further complication as the direction of making diffusion model works for TS structure generation is already quite new and method intensive. We added a comment along this direction in the Discussion.

Changes to text: At the last paragraph of the "Discussion" section, added in red, "Although we perform DFT calculations for evaluating the barrier height throughout this work, OA-ReactDiff framework can be readily adopted to predict the reaction barrier by swapping the model output layer."

7. In page 10, authors compare their model performance with density functional tight binding (DFTB) but cited paper is general review on DFTB. There could be various flavors of DFTB which have different parameterization. Therefore, it would be better and fair if authors can specify which particular DFTB model has been used for their comparison.

Response: We thank the reviewer for this question. We extracted the DFTB results directly from the NeruaINEB paper (i.e., the ML potential paper comes with Transition1x dataset, https://iopscience.iop.org/article/10.1088/2632-2153/aca23e/meta). We followed the citation in the NeuralNEB paper on DFTB. We also double checked in the paper and do not find a more detailed description about which flavor of DFTB was applied there. We modified our description in the manuscript to reflect more clearly about that fact that the DFTB comparison is extracted from the NeuralNEB paper.

Changes to text: At the second paragraph of section "Approaching the energetic accuracy needed in TS search", changes in red, "OA-ReactDiff + recommender far outperforms semi-empirical methods such as density functional tight binding,[47] which was reported to have an MAE of 16.1 kcal/mol and average runtime of 82 seconds on Transition1x dataset.[22]"

8. In Table 1, the Pearson's r value of the developed model is less than the other models used for comparison. Authors may want to discuss further how this can be interpreted.

Response: We thank the reviewer for noticing this detail! By analyzing SI Figure. S10 (attached below), we ascribe the slightly lower Pearson's *r* to the presence of many high RMSD but low barrier height samples (i.e., those in the bottom right corner). Those correspond to cases where OA-ReactDiff can generate good structures for individual fragments but not their alignment (thus high RMSD). Yet these fragments are pretty far away and barely interact with each other (thus low energy error). We have detailed one example like this at Fig. 5a. We added a comment for this point in the main text.

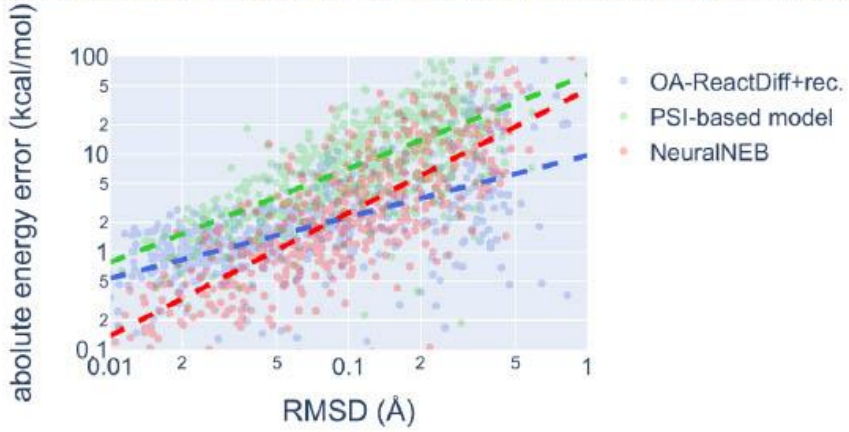


Figure S10. Absolute energy difference vs. RMSD and the corresponding linear fit in a log-log plot for OA-ReactDiff + rec, PSI-based model,[21] and NeuralNEB[22].

Changes to text: At the second last paragraph of section "Approaching the energetic accuracy needed in TS search", added in red, "We ascribe the slightly lower Pearson's r between RMSD and barrier height error of OA-ReactDiff to the presence of many high RMSD but low barrier height TS samples, which correspond to cases where OA-ReactDiff can generate good 3D structures for nearly non-interacting individual fragments but not their alignment (Fig. 5a and Supplementary Fig. S10)."

9. Which value of alpha t or sigma t is used for noising in this study? Is it adaptive? I think its value has not been mentioned in the manuscript.

Response: We thank the reviewer for this comment. We mostly adopted hyperparameters of the diffusion process from the EDM paper (https://arxiv.org/abs/2203.17003), where a second order polynomial schedule is applied. To be more precise, alpha_t = 1 – (t / T)², sigma_t = sqrt(1 – alpha_t²). We made this more clear in the Method section.

Changes to text: At the second paragraph of section "Approaching the energetic accuracy needed in TS search", changes in red, "We mostly adopted hyperparameters of the diffusion process from the EDM paper,[26] where second-order polynomial noise schedule (i.e., $\alpha_t = 1 - (t / T)^2$)} and L_{simple} loss function is used."

Decision Letter, first revision:

Date: 17th October 23 16:56:48
Last Sent: 17th October 23 16:56:48
Triggered By: Kaitlin McCardle
From: kaitlin.mccardle@us.nature.com
To: crduan@mit.edu
CC: computationalscience@nature.com
BCC: kaitlin.mccardle@us.nature.com
Subject: AIP Decision on Manuscript NATCOMPUTSCI-23-0853A
Message: Our ref: NATCOMPUTSCI-23-0853A

17th October 2023

Dear Dr. Duan,

Thank you for submitting your revised manuscript "Accurate transition state generation with an object-aware equivariant elementary reaction diffusion model" (NATCOMPUTSCI-23-0853A). It has now been seen by the original referees and their comments are below. The reviewers find that the paper has improved in revision, and therefore we'll be happy in principle to publish it in Nature Computational Science, pending minor revisions to comply with our editorial and formatting guidelines.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements in about a week. Please do not upload the final materials and make any revisions until you receive this additional information from us.

TRANSPARENT PEER REVIEW
Nature Computational Science offers a transparent peer review option for original research manuscripts. We encourage increased transparency in peer review by publishing the reviewer comments, author rebuttal letters and editorial decision letters if the authors agree. Such peer review material is made available as a supplementary peer review file. **Please remember to choose, using the manuscript system, whether or not you want to participate in transparent peer review.**
Please note: we allow redactions to authors' rebuttal and reviewer comments in the interest of confidentiality. If you are concerned about the release of confidential data, please let us know specifically what information you would like to have removed. Please note that we cannot incorporate redactions for any other reasons. Reviewer names will be published in the peer review files if the reviewer signed the comments to authors, or if reviewers explicitly agree to release their name. For more information, please refer to our FAQ page.

Thank you again for your interest in Nature Computational Science. Please do not hesitate to contact me if you have any questions.

Sincerely,

Kaitlin McCardle, PhD
Senior Editor
Nature Computational Science

ORCID
IMPORTANT: Non-corresponding authors do not have to link their ORCIDs but are encouraged to do so. Please note that it will not be possible to add/modify ORCIDs at proof. Thus, please let your co-authors know that if they wish to have their ORCID added to the paper they must follow the procedure described in the following link prior to acceptance: https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research

Reviewer #1 (Remarks to the Author):

The authors completely resolved my comments. I do not have any additional comments.

Reviewer #2 (Remarks to the Author):

On the basis of the rebuttal and of the revised manuscript, the clarity of the paper has improved. The points that I have raised in the assessment are now well addressed in the paper. This makes it possible to follow in detail the discussion and understand the potential impact.

Reviewer #3 (Remarks to the Author):

The authors have presented all the points raised by the reviewer. I suggest publishing this manuscript as modified by the authors.

Final Decision Letter:

Date: 3rd November 23 15:12:07
Last Sent: 3rd November 23 15:12:07
Triggered By: Kaitlin McCardle
From: kaitlin.mccardle@us.nature.com
To: crduan@mit.edu

nature research

BCC: fernando.chirigati@us.nature.com,kaitlin.mccardle@us.nature.com,rjsart@springernature.com,computationalscience@nature.com,rjsproduction@springernature.com

Subject: Decision on Nature Computational Science manuscript NATCOMPUTSCI-23-0853B

Message: Dear Dr Duan,

We are pleased to inform you that your Article "Accurate transition state generation with an object-aware equivariant elementary reaction diffusion model" has now been accepted for publication in Nature Computational Science.

Once your manuscript is typeset, you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required.

Please note that Nature Computational Science is a Transformative Journal (TJ). Authors may publish their research with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. Find out more about Transformative Journals

Authors may need to take specific actions to achieve compliance with funder and institutional open access mandates. If your research is supported by a funder that requires immediate open access (e.g. according to Plan S principles) then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route, the journal's standard licensing terms will need to be accepted, including self-archiving policies. Those licensing terms will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

Acceptance of your manuscript is conditional on all authors' agreement with our publication policies (see https://www.nature.com/natcomputsci/for-authors). In particular your manuscript must not be published elsewhere and there must be no announcement of the work to any media outlet until the publication date (the day on which it is uploaded onto our web site).

Before your manuscript is typeset, we will edit the text to ensure it is intelligible to our wide readership and conforms to house style. We look particularly carefully at the titles of all papers to ensure that they are relatively brief and understandable.

Once your manuscript is typeset, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately.

If you have queries at any point during the production process then please contact the production team at rjsproduction@springernature.com. Once your paper has been scheduled for online publication, the Nature press office will be in touch to confirm the details.

Content is published online weekly on Mondays and Thursdays, and the embargo is set at 16:00 London time (GMT)/11:00 am US Eastern time (EST) on the day of publication. If you need to know the exact publication date or when the news embargo will be lifted, please contact our press office after you have submitted your proof corrections. Now is the time to inform your Public Relations or Press Office about your paper, as they might be interested in promoting its publication. This will allow them time to prepare an accurate and satisfactory press release. Include your manuscript tracking number NATCOMPUTSCI-23-0853B and the name of the journal, which they will need when they contact our office.

About one week before your paper is published online, we shall be distributing a press release to news organizations worldwide, which may include details of your work. We are happy for your institution or funding agency to prepare its own press release, but it must mention the embargo date and Nature Computational Science. Our Press Office will contact you closer to the time of publication, but if you or your Press Office have any inquiries in the meantime, please contact press@nature.com.

An online order form for reprints of your paper is available at https://www.nature.com/reprints/author-reprints.html. All co-authors, authors' institutions and authors' funding agencies can order reprints using the form appropriate to their geographical region.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

We look forward to publishing your paper.

Best regards,

Kaitlin McCardle, PhD
Senior Editor
Nature Computational Science

P.S. Click on the following link if you would like to recommend Nature Computational Science to your librarian: https://www.springernature.com/gp/librarians/recommend-to-your-library

** Visit the Springer Nature Editorial and Publishing website at www.springernature.com/editorial-and-publishing-jobs for more information about our career opportunities. If you have any questions please click here.**