Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Predicting reaction barriers of hydrogen atom transfer in proteins†

Kai Riedmiller[a], Patrick Reiser[b,c], Elizaveta Bobkova[a], Kiril Maltsev[a], Ganna Gryn'ova[a], Pascal Friederich[b,c,*], Frauke Gräter[a,d,*]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Hydrogen atom transfer (HAT) reactions are important in many biological systems. As these reactions are hard to observe experimentally, it is of high interest to shed light on them using simulations. Here, we present a machine learning model based on graph neural networks for the prediction of activation energies of HAT reactions in proteins. It is trained on more than 17,000 energy barriers calculated using hybrid density functional theory. We built and evaluated the model in the context of HAT in collagen, but the same workflow can easily be applied to HAT reactions in other biological or synthetic polymers. We obtain for relevant reactions (small reaction distances) a model with good predictive power ($R^2 \sim 0.9$ and mean absolute error of < 3 kcal/mol). As the inference speed is high, this model enables evaluations of many chemical situations in rapid succession. When combined with molecular dynamics in a kinetic Monte-Carlo scheme, the model paves the way toward reactive simulations.

1 Introduction

Free radicals critically impact and can be deleterious for biological systems.[1,2] They are highly reactive and lead to unspecific damage of proteins, DNA, and lipids, causing various diseases and aging.[3] Radical formation is followed by a plethora of subsequent reactions, most importantly radical propagation through hydrogen atom transfer (HAT).[4] Radical formation and propagation are not only at play in biomolecules but very analogously occur in synthetic polymers, and similarly lead to damage and material aging.[5,6] Due to the high reactivity of radicals, intermediate products of radical reactions can be very short-lived and therefore hard to capture experimentally. Predicting the fate of radicals in proteins or other (bio)polymers is thus of utmost relevance to better understand and combat radical-induced damage.

A major challenge in predicting chemical reactivity in proteins, such as unspecific radical transfer reactions, is the molecular environment of the reaction: it decides on the reactivity but is both chemically very diverse and highly dynamic. This leads to a virtually infinite number of possible reaction scenarios, in which reactants represent instances within a vast chemical and conformational space. As a consequence, directly computing this amount of radical reactions by *ab initio* calculations is computationally not feasible. Instead, machine learning can leverage quantum chemical calculations by predicting reactivity based on an initial quantum chemical data set. We here set out to predict the energy barriers of hydrogen atom transfer reactions in proteins using Graph Neural Networks that are trained on computed energy barriers.

Machine learning has come of age to predict structures, energies, and properties of molecules, while predicting kinetics by machine learning is a relatively new field. In a recent review by Lewis-Atwell, Townsend, and Grayson, various methods for the prediction of activation energies are summarized.[7] The used techniques range from kernel ridge regression over gradient-boosted decision trees to neural networks and Gaussian processes. No single method performs best on all kinds of data sets, data set sizes, and reaction types. Reaction representations are key to faithful predictions, and can be built from established molecular ones, as shown by van Gerwen et al.[8] They also discovered that representations for reactions should incorporate information from the educt and product, something we also found important in this work.

Also for the prediction of HAT energy barriers in this work, we evaluated different regression techniques. Based on performance measures, we settled on graph neural networks, which is representative of the overall development in representations

[a] *Heidelberg Institute for Theoretical Studies, Heidelberg, Germany. *E-mail: frauke.graeter@h-its.org*
[b] *Institute of Theoretical Informatics, Karlsruhe Institute of Technology, Engler-Bunte-Ring 8, 76131 Karlsruhe, Germany. *E-mail: pascal.friederich@kit.edu*
[c] *Institute of Nanotechnology, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1: 76344 Eggenstein-Leopoldshafen, Germany.*
[d] *Interdisciplinary Center for Scientific Computing, Heidelberg University, Heidelberg, Germany*
† Electronic Supplementary Information (ESI) available: Method details, additional figures. See DOI: 00.0000/00000000.

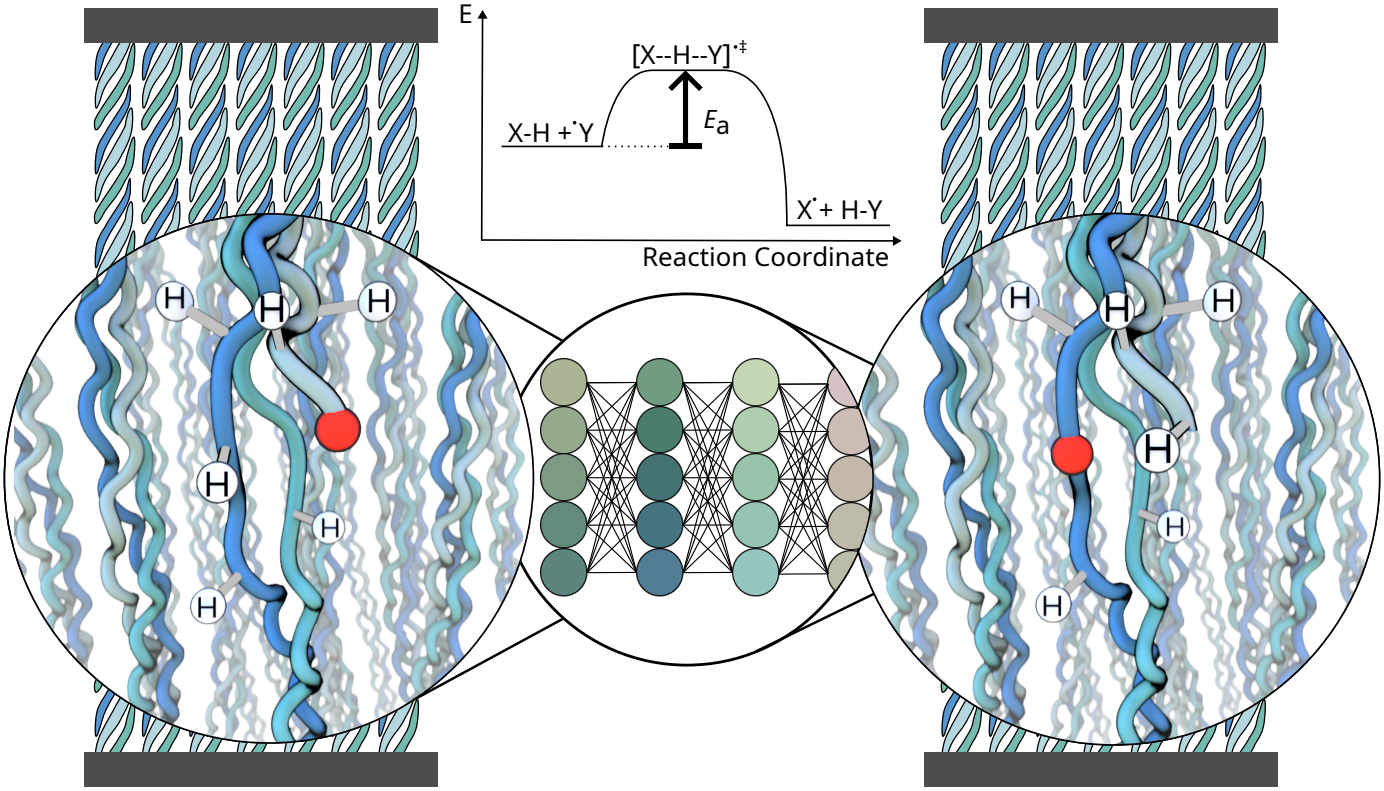

Fig. 1 Schematic of the workflow. The system of interest here is a collagen fibril under tension, containing radicals (red). A cutout of the fibril is presented to the neural network, which predicts the energy of activation E_a for every possible reaction, one at a time. This information can be used to decide which reaction is likely to occur and continue the simulation after the reaction.

of chemical systems for machine learning. Representations such as the Coulomb Matrix[9] perform well on simpler tasks, but are typically outperformed by more expressive representations, e.g., atom-centered symmetry functions,[10] smooth overlap of atomic orbitals (SOAP),[11] and, in the last years, graph neural networks (GNNs).[12,13] The latest accuracy increases can be partially attributed to moving from invariant to equivariant models, where we arrive at the GNN used in this work, namely PaiNN.[14,15]

In this work, we focus on predicting HAT reaction barriers within one particular protein system, collagen (Figure 1). The model is also applicable to proteins of similar composition, and the developed workflow can be used for other, not necessarily biological, polymers. As shown earlier by some authors of this article, stretching collagen generates mechanoradicals within the protein.[16] These radicals rapidly localize on specific protein residues, dihydroxyphenylalanine (DOPA), plausibly through a sequence of HAT reactions. However, intermediates of the migration and stabilization process are challenging to identify. This renders collagen an ideal model system, yet allows transferability of our approach to any other protein or chemically similar polymer.

We built thousands of molecular fragments as they occur in collagen and calculated HAT energy barriers on the level of hybrid density functional theory (DFT). The computed reaction barriers range between 0 kcal/mol and 175 kcal/mol and are highly dependent on the local environment, rationalizing the machine learning approach. We used our quantum chemical data to train the GNN, which is able to predict barriers approaching chemical accuracy.

Our machine learning model predicts the energy of activation for one selected reaction at a time while taking the chemical environment around the radical as input.(Figure 1) It thus can be used as a surrogate model of hydrogen atom transfer within classical molecular dynamics (MD) simulations to model radical propagation within collagen or other (bio)materials on the fly, e.g., by using hybrid kinetic Monte-Carlo and MD simulations.[17] Our GNN-based approach tackles the challenge of predicting reaction barriers in a heterogeneous and dynamic chemical setting, and will likely prove useful for other complex soft matter systems.

2 Methods

2.1 Data generation

The geometries to learn HAT activation energies were generated in two ways: in a bottom-up approach from single amino acids, and in a top-down approach by extracting reactive systems from a larger atomistic model. In the following, the structures from the bottom-up and top-down approaches will be called synthetic systems and trajectory systems, respectively. Synthetic systems are pairs of amino acids arranged in a way that two hydrogen atoms are in a defined position to one another. As shown in Figure 2 A, the translation distance between the hydrogens, the rotation, and the tilt angle are varied. The positions of these two central

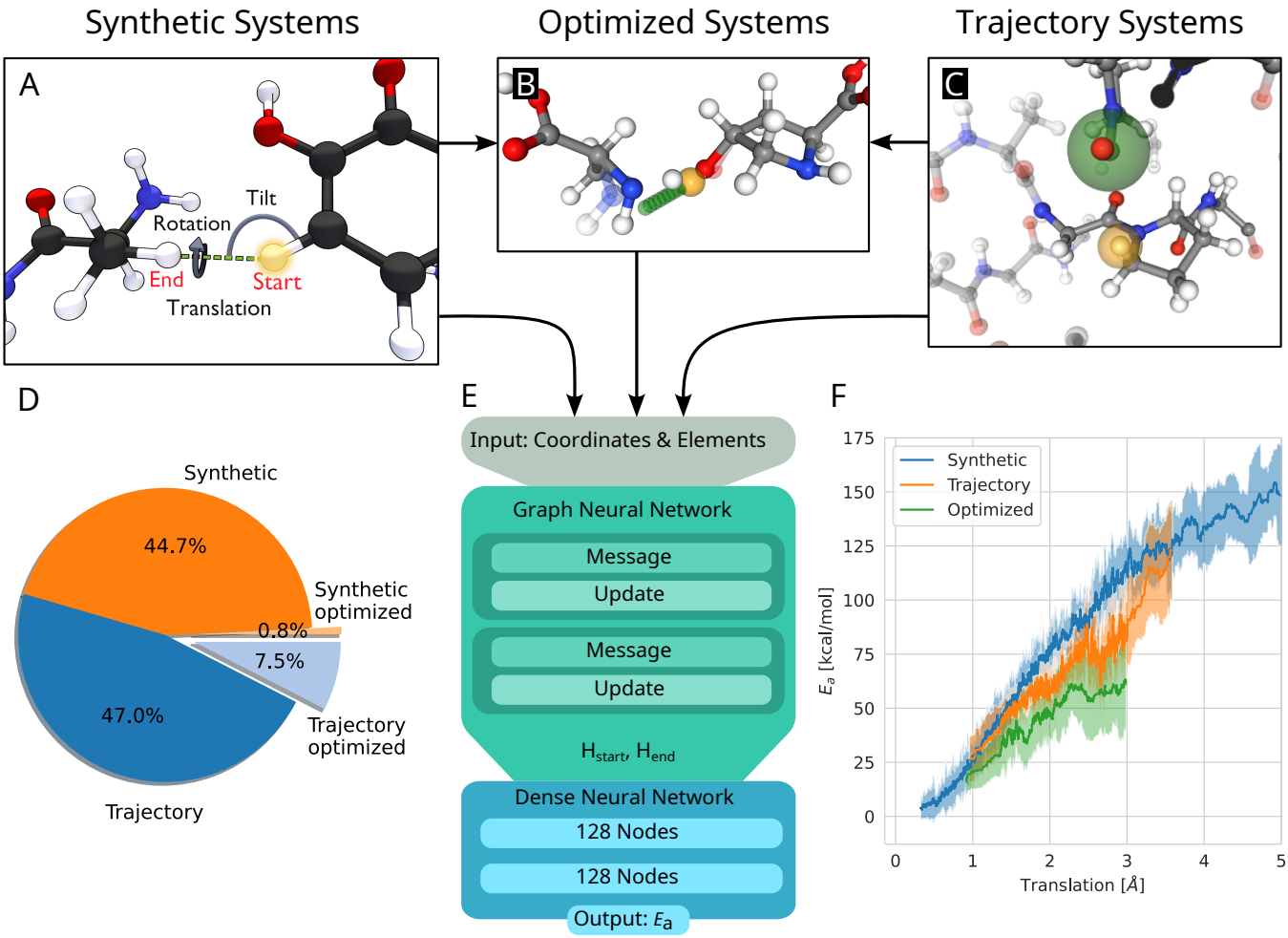

Fig. 2 (A) Build process of reactive HAT systems for the synthetic data set. The HAT reaction between two example molecules is shown. The distance between the start and end position of the transferring hydrogen (translation), the angle formed by the transferring hydrogen with the donating and accepting heavy atoms (tilt), and the dihedral angle around the hydrogen atom transfer axis (rotation) were varied to construct the synthetic systems. (B) Data set of optimized structures, built from synthetic and trajectory systems. The optimized transition state is shown, alongside the interpolated reaction path of the hydrogen in green and the start position in orange. The non-optimized structure is shown translucently. (C) A trajectory system with its environment shown translucently. The radical heavy atom is highlighted in green and the reacting hydrogen in yellow. The solid-drawn atoms at the border to the translucent environment are used in the construction of the capping groups, the translucent atoms are discarded. (D) Data distribution of the synthetic and trajectory data sets. (E) Architecture of the used graph neural network, based on the PaiNN architecture.[15] (F) The calculated energy barriers of HAT reactions in the data set vs the distance the hydrogen has to move during the reaction.

hydrogen atoms represent the start and end positions of a single hydrogen atom undergoing the HAT reaction. Furthermore, intramolecular reactions are generated from within single amino acids. Combinations of hydrogen atoms with less than 4 Å distance are considered. Systems with atoms close to or in the way of the transition path are removed.

The generation of reactive systems from molecular dynamics (MD) trajectories starts from a collagen model obtained from Colbuilder.[18] The model is simulated using GROMACS 2020. In the resulting trajectory, possible reaction sites are identified by monitoring H-H distances. As activation energy, E_a, heavily depends on the translation distance, an emphasis is put on smaller translations when sampling. The HAT candidates are cut out together with their close surrounding from the bigger system. To generate chemically meaningful systems and to allow reference computations, the cut-out sections of the protein are capped using N-methyl and acetyl groups. In Figure 2 C, the capping procedure is visualized.(Also See Figure SI.2 B) For a given set of selected atoms in a trajectory, only the system with the smallest translation distance is kept, as otherwise a large amount of highly correlated systems would be generated.

At this point, the reactive systems have been defined. Further preparation steps are applied to synthetic as well as trajectory systems. One of the two central hydrogen atoms is removed and therefore acts as the starting location of the radical. Then, the reaction path is estimated by interpolating the position of the remaining hydrogen atom from its starting position to the position of the removed hydrogen atom. Along the reaction path, the energy of the system is calculated using the hybrid functional BMK[19] together with the 6–31+G(2df,p) basis set in seven in-

crements, using Gaussian 09 (rev D.01).[20] We originally used 11 equally spaced increments, but in the first data set of 1600 HAT reactions, all barriers were found to be located within steps 4-8, 1 or 11, so steps 2, 3, 9, and 10 are omitted in the data used here. More details on the building of the systems and the calculations are given in the ESI†.

In many cases, the guess of the transition path by interpolating between the start and end position of the hydrogen whilst keeping the rest of the system frozen is reasonable, however, it introduces a significant error in conformationally flexible systems. (Figure SI.3) To address this, a subset of the data set described above is optimized at the same QM level as used for the energy calculations. During optimization, we constrain all atom positions except the donor and acceptor atoms and directly bonded hydrogens, to reflect the embedding of the reactants into the structure of the material, here the protein backbone, and to prevent contributions to the calculated energy barrier from rearrangements in the reactants unrelated to the HAT.

In this work, 4393 structures are created using the bottom-up approach, and 5261 from trajectories. From these structures, 10% are set aside randomly for testing. 803 structures are optimized, most of them (725) originate from the trajectory data set, 78 are synthetic systems. Figure 2 D shows the distribution between the data sets. Note, that each structure has two associated energy barriers, resulting in twice the amount of data points for training/testing.

2.2 Graph Neural Network

To predict the activation energy for a given structure, the graph neural network PaiNN is used.[15] Figure 2 E shows our workflow: the inputs for the model are the atom positions and elements in the educt configuration. The start and end positions of the transferred hydrogen atom are encoded as two unique elements to define the reaction direction. After two message passing iterations, the invariant node features of the hydrogen and the pseudo-atom are concatenated and fed into a dense neural network, consisting of two layers with 128 nodes each, using the swish activation function,[21] followed by one output node with linear activation. The GNN is trained to minimize the mean absolute error using the Adam optimizer[22] with learning rate decay and early stopping. Hyperparameters are optimized using a Bayesian optimizer as implemented in the Keras Tuner package.[23] To increase accuracy and obtain a measure of uncertainty, an ensemble of ten models with random initializations is trained. The models are validated using 10% of the training data, each model using a random training/validation split.

2.3 Dense Neural Network

As an alternative to the graph neural network, a simple feed-forward dense neural network is tested together with the local many-body tensor representation (L-MBTR).[24] This descriptor is based on histograms of distances and angles between one central atom and its surrounding. It is calculated on the position of the missing hydrogen atom, next to the radical-carrying atom. This position is encoded as a special element 'X', and the reacting hydrogen as 'Y' to present a well-defined task to the network. The L-MBTR descriptor is generated using the DScribe library.[25]. Parts of the descriptor, which correspond to interactions between multiple elements 'X' or 'Y', are always zero and therefore removed to improve efficiency. Three hidden layers of shrinking size (1000, 500, and 100 neurons) are used in the network, utilizing the ReLU activation function, followed by a single output node with linear activation. The hyperparameters were determined by a non-exhaustive manual grid search. Similar to the graph neural network, an ensemble of ten models is trained.

3 Results and Discussion

As a starting point of the training, we generated a data set of structures where HAT reactions can occur, along with the associated energy barriers. The dataset spans the relevant conformational and chemical space and provides valuable insight into the behavior of HAT reactions in collagen, on top of enabling the creation of predictive models.

Unsurprisingly, the calculated barriers show a strong dependence on the distance the hydrogen atom has to travel during the reaction, as can be seen in Figure 2 F. However, variations in barriers for a given translation are significant, substantiating the need to use more complex models for predicting barriers across different amino acids, orientations, and local environments. The synthetic data, by construction, includes data at translations down to 0.3 Å and with barriers smaller than 20 kcal/mol, cases not covered in the trajectories, where small interatomic distances are disfavored and thus rare. Translations found in trajectories start from 0.7 Å, but the vast majority is larger than 1.2 Å. (Figure SI.1 B)

The correlation of barriers with translation allows the introduction of a cutoff after a certain translation, rather than one dependent on the energy barrier, which would be unknown beforehand. Here, we chose a translation cutoff of 2 Å or 3 Å to focus on thermochemically probable HAT reactions, i.e., those with barriers mostly lower than 100 kcal/mol and to thereby consider HAT reactions most relevant in an actual protein material. In the following, performance metrics for both cutoffs are presented.

To assess the quality of the method, an ensemble of models was trained on all available training data, and evaluated on the whole test data or only on the trajectory systems of the test data. We chose to evaluate on the trajectory data as a major application of the energy prediction is the prediction of HAT for conformations of proteins (here collagen) encountered during MD simulations.

Performance metrics are summarized in Figure 3. Panels A and B show the performance for trajectory systems with translations below 2 Å, i.e., focus on the most feasible HAT reactions, while panels D to F show measurements using all available trajectory data. As can be seen from Figure 3 A and B, we achieve MAEs of 2.4 kcal/mol for the model ensemble and 2.7 kcal/mol for individual models on the trajectory data with the translation cutoff in place. The accuracy improved significantly when including more quantum chemical data into the training, all the way up to ~ 90 % of the training data, which corresponds to 12662 individual barriers (Figure 3 C). The amount of training data generated at the BMK/6–31+G(2df,p) level thus is approximately required

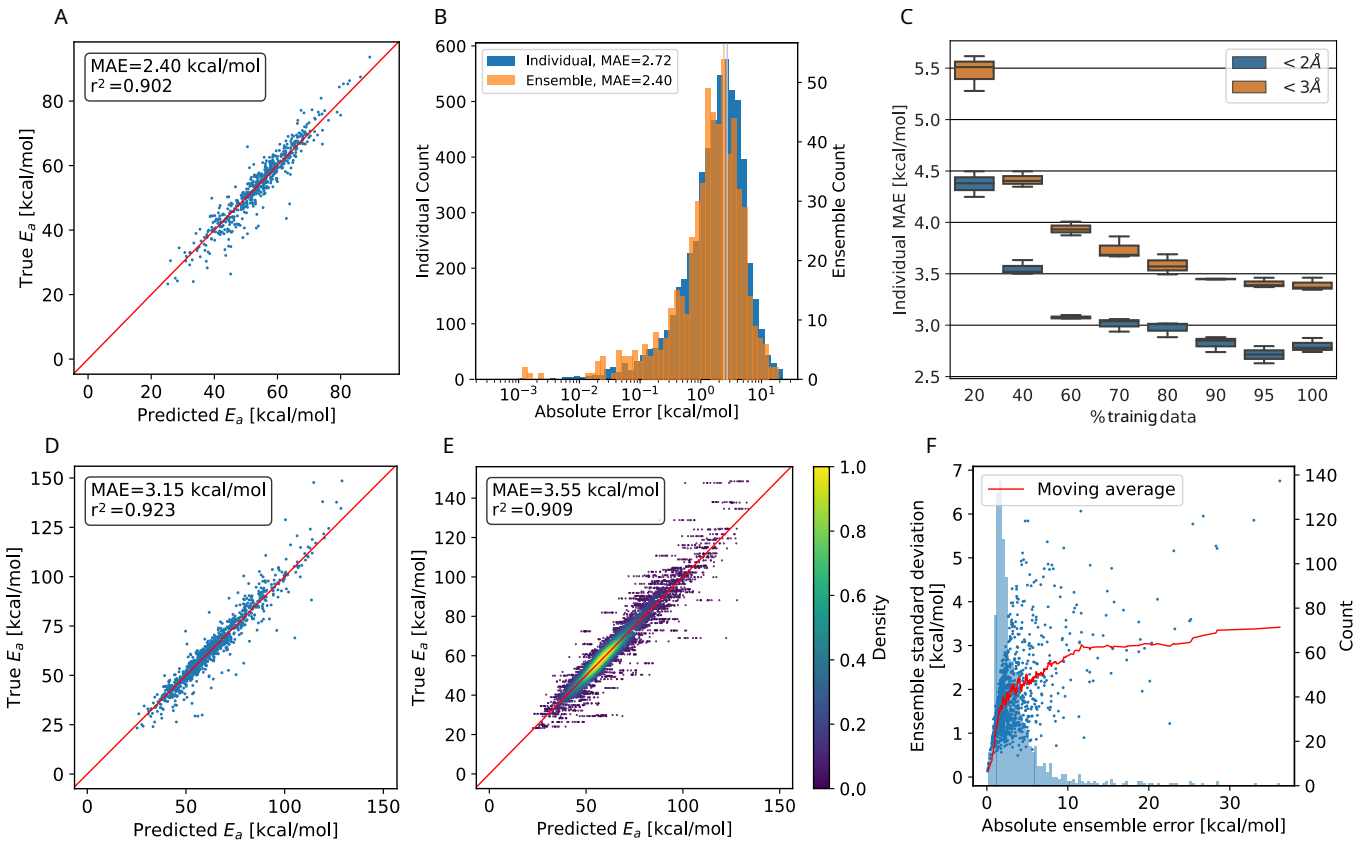

Fig. 3 Model performance predicting HAT barriers on test data. (A) Predicted energy barriers vs. ground truth using the PaiNN ensemble model on trajectory test data with translation < 2 Å. (B) Histogram of the prediction errors of individual PaiNN models and the ensemble model. The mean of both distributions is shown as a vertical line. (C) Performance of three individual models trained on fractions of the complete training set. 100% corresponds to 15633 training points. (D) Predicted energy barriers vs. ground truth using the PaiNN ensemble model on all trajectory test data. (E) Predicted energy barriers vs. ground truth using ten PaiNN models on all trajectory test data. (F) The absolute error of the PaiNN ensemble model on all trajectory data versus the standard deviation of the predictions of individual models within the ensemble. In red, the mean cumulative sum is plotted, and in the background a frequency plot of occurring errors.

to reach this accuracy, but also appears to suffice, as the learning curve flattens towards the end. The accuracy decreases for systems with bigger translations and energy barriers, as shown in Figure 3 D and E: The removal of the translation cutoff introduces more high-barrier systems to the test set, which seem to be harder to predict exactly. For the prediction of the propagation pathway of a radical in a complex environment, this might be acceptable though, as reactions with high energy barriers are unlikely to occur under ambient conditions. As mentioned, the use of an ensemble model also brings the advantage of an uncertainty measure: the standard deviation between the models. In Figure 3 F, the absolute ensemble error is plotted against the ensemble standard deviation together with a rolling average. For a low standard deviation (smaller than 1.7 kcal/mol), one can assume a low prediction error (< 3 kcal/mol) quite confidently. On the other hand, higher standard deviations no longer scale reliably with the error.

To understand to what degree a given part of the data improves the model, multiple models were trained on different parts of the training data and evaluated on a subset of the test data. The models were trained on trajectory and synthetic data, or on trajectory data only. Additionally, several translation cutoffs were used, either at 2 or 3 Å or without a cutoff. In all cases, the models were evaluated on trajectory data below 2 or 3 Å. This setup shows that the model benefits from being trained on synthetic systems alongside trajectory systems, even if it is evaluated only on trajectory systems. (Figure 4 A) Similarly, the model performance for systems with translations smaller than 2 Å improves when it is trained on larger translations. Overall, data that is rather distant from the target prediction with regard to its chemistry and geometry still significantly adds to the predictive power of the model.

So far, we showed that the model can well reproduce activation energies close to DFT accuracy on structures from MD trajectories and synthetic ones. However, as mentioned previously, realistic energy barriers are expected to be closer to those computed for at least partially optimized systems. But, since optimizations at the DFT level of theory for the entire data set are prohibitively expensive, only a subset of reaction paths were optimized BMK/6–31+G(2df,p) method. The mean absolute deviation between the barriers computed for frozen geometries and barriers of optimized reaction paths is 13.6 kcal/mol (Figure SI.3). Notably, the activation energies of synthetic systems

change less when optimized compared to barriers from trajectory systems. This is likely due to higher atom density in trajectory systems: it is more likely in trajectory systems that atoms interfere with the transition path. In other words, the lowest energy reaction path changes more in trajectory systems relative to the interpolated path. This analysis also serves as a validation for the bottom-up structure building process. If the built structures were unreasonable, they would change more drastically during optimization compared to the structures produced by MD simulation. At a given translation, the optimized reactions generally show lower barriers (see also Figure 2 F).

To correct for the deviation between the barriers computed for non-optimized and optimized reaction paths, models already trained on non-optimized systems were retrained in a transfer learning scheme to be as data-efficient as possible. When using transfer learning, one often freezes most of the network and retrains only parts of it. Here, however, we found the best results when not freezing any part of the model. Still, models trained with transfer learning substantially outperform models trained on optimized data alone. (Figure 4 B) Note that only the training target, i.e., the barrier, was changed for transfer learning, and not the input to the model. The same non-optimized structures are fed into the model, as it is intended to be used on non-optimized structures from MD simulations. In other words, the model learns a mapping between non-optimized MD structures and DFT barriers of the optimized reaction paths.

The ensemble model for predicting barriers after transition state optimization achieves an MAE of 3.6 kcal/mol when trained on trajectory data with translations of less than 2 Å and 4.9 kcal/mol when trained on all trajectory data (Figure 5 A and B, respectively). The learning curve of the transfer learning procedure in Figure 5 C suggests that the model is data limited, as the accuracy increases in particular from 90% over 95% to 100% of the optimized test data.

While the model does not reproduce the DFT results perfectly, its error is approaching the accuracy of the underlying target method. The authors of the functional of our choice, BMK, targeted an accuracy of 2 kcal/mol on energy barriers,[19] and, depending on the benchmarks, BMK achieves an MAE of 0.8 kcal/mol to 5 kcal/mol relative to CCSD(T) calculations.[26,27]

To justify the use of an arguably complex graph neural network, we compared it to two simpler methods, a densely connected feed-forward neural network and a random forest model (as implemented in scikit-learn[28]). For both methods, we used the L-MBTR[24] descriptor as input and the barriers computed for non-optimized reactions as targets. Details on the network architecture and input generation are given in the Methods section. Using an ensemble of ten models, the feed-forward neural network accomplishes an MAE of 4.8 kcal/mol, and single models achieve 5.1 kcal/mol on average over all trajectory data (see Figure SI.4). The random forest model achieved an MAE of 6.6 kcal/mol (Figure SI.5). Taken together, these results highlight the need for more sophisticated representations and models to capture subtle structural differences.

Our ultimate aim is to model the chemistry of radical-induced damage to collagen whilst simultaneously capturing the dynamic

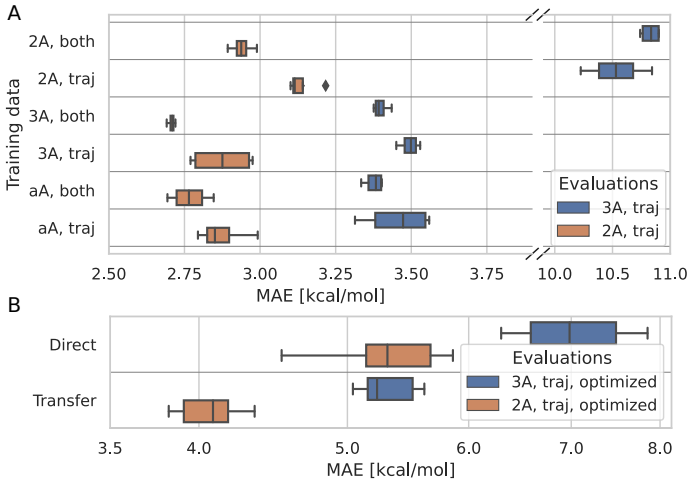

Fig. 4 Differently trained models evaluated on comparable subsets of the test data. (A) Comparison between models trained on different data sets as indicated on the y-axis. 2A, 3A, and aA correspond to data with translations below 2A, 3A, or all translations, respectively. 'traj' refers to data only from trajectories, 'both' includes in addition the synthetic data. Four models were trained per data set.
(B) Comparison between transfer learning and training directly on the optimized data only. Again, both models were evaluated on trajectory data < 3 Å and < 2 Å. All ten original models are used in transfer learning, ten new models were trained for the direct learning approach.

nature of this system. Kinetic Monte-Carlo (KMC) method enables incorporating reactions into MD on timescales beyond those covered by conventional MD simulations. A hybrid KMC-MD approach models reactions in a Markov-process, allowing arbitrarily big time jumps between reaction steps.[29] Previously, we coupled KMC with MD to simulate homolytic bond rupture in stretched collagen fibrils in a method called KIMMDY.[17] Our GNN-based model for predicting reaction barriers allows applying the KIMMDY approach to radical transfer reactions. Inferring an energy barrier of a reaction from a trained neural network within a reactive MD simulation substitutes the otherwise computationally costly quantum chemical calculation, and only marginally compromises the efficiency of standard MD simulations.

In the past, several methodologies were developed to achieve reactive MD, including reactive force fields, such as ReaxFF[30] and AIREBO[31], hybrid quantum mechanical / molecular mechanical (QM/MM) simulations, and, more recently, molecular dynamics simulations paired with machine-learned force fields (MLFF).[32,33] However, all these methods are slower compared to regular MD[34] and are by default restricted to reactions on the timescale of the simulation. KIMMDY overcomes these drawbacks but relies on the availability of reaction rates, which can now be provided with the model introduced here.

While the model extends the applicability of KIMMDY, one should keep the underlying limitations in mind. The type of reactions one can predict is predefined. The model predicts hydrogen atom transfer only, while we can not rule out proton coupled electron transfer to play a role in this system. Further, tunneling effects are ignored, and all QM calculations are performed in the gas phase. The advantage of our approach is that it can be im-

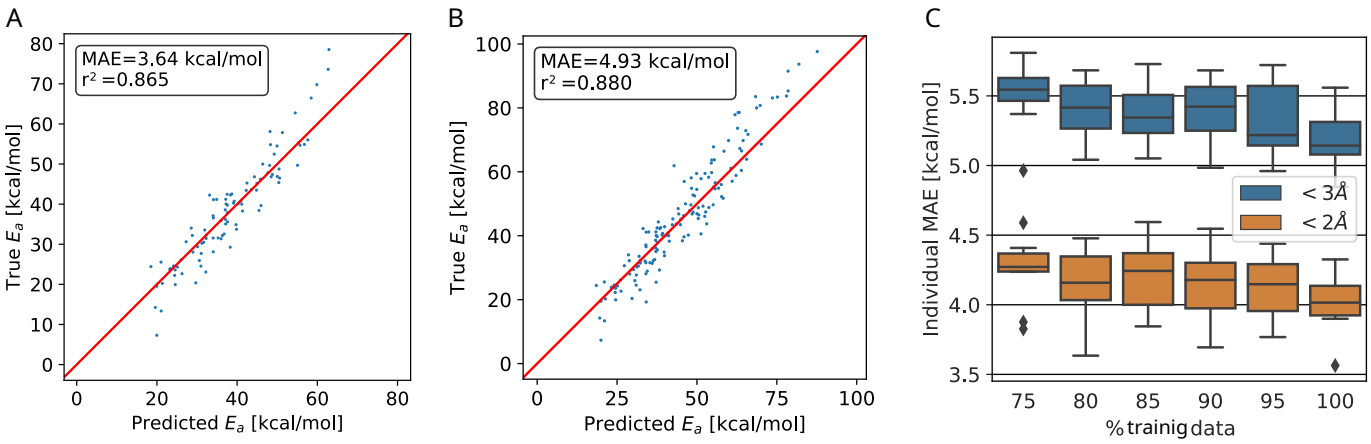

Fig. 5 Performance of the transfer-learned ensemble model on (A) trajectory test data < 2 Å and on (B) all trajectory test data. (C) Learning curve of the transfer learning process. The test MAE of the individual models is shown.

proved and adopted straightforwardly, e.g. by transfer learning.

4 Conclusion

In this work, we introduced a workflow to train machine learning models for fast predictions of activation energies of hydrogen atom transfer reactions. Our model was trained and evaluated in the context of radical migration in collagen fibrils, but can be transferred to other chemical systems subject to HAT reactions. Since the predicted reaction barriers are based on 3D structures of molecules, the model can be used in direct conjunction with MD simulations. For example, utilizing predicted barriers in a kinetic Monte-Carlo scheme, one can extend MD simulation to allow HAT reactions to take place in a dynamically evolving molecular system. Our study emphasizes the strength of graph neural networks for predicting chemical reactivity – even in such challenging cases as dynamic biopolymers.

Data availability

Structures and energies are available at `https://doi.org/10.11588/data/TGDD4Y`. Trained models and example code available on GitHub: `https://github.com/HITS-MBM/HAT_prediction_GNN`

Author Contributions

Kai Riedmiller: Methodology, Software, Formal analysis, Investigation, Data Curation, Writing – Original Draft, Visualization **Patrick Reiser:** Software, Methodology **Elizaveta Bobkova:** Software, Investigation **Kiril Maltsev:** Conceptualization, Methodology **Ganna Gryn'ova:** Conceptualization, Methodology, Writing – Review & Editing **Pascal Friederich:** Conceptualization, Methodology, Writing – Review & Editing **Frauke Gräter:** Conceptualization, Resources, Writing – Review & Editing, Supervision, Project administration, Funding acquisition.

Conflicts of interest

There are no conflicts to declare.

Acknowledgment

We acknowledge generous financial support from the Klaus Tschira Foundation through the HITS Lab and SIMPLAIX. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 101002812 RADICOL).

References

1 C. C. Winterbourn, *Nat. Chem. Biol.*, 2008, **4**, 278–286.
2 M. Gutowski and S. Kowalczyk, *Acta Biochim. Pol.*, 2013, **60**, 1–16.
3 K. J. Davies, *Journal of Biological Chemistry*, 1987, **262**, 9895–9901.
4 C. L. Hawkins and M. J. Davies, *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 2001, **1504**, 196–219.
5 G. Gryn'ova, J. L. Hodgson and M. L. Coote, *Org. Biomol. Chem.*, 2011, **9**, 480–490.
6 D. L. Allara, *Environmental health perspectives*, 1975, **11**, 29–33.
7 T. Lewis-Atwell, P. A. Townsend and M. N. Grayson, *WIREs Computational Molecular Science*, 2022, **12**, e1593.
8 P. van Gerwen, A. Fabrizio, M. D. Wodrich and C. Corminboeuf, *Machine Learning: Science and Technology*, 2022, **3**, 045005.
9 M. Rupp, A. Tkatchenko, K.-R. Müller and O. A. von Lilienfeld, *Phys. Rev. Lett.*, 2012, **108**, 058301.
10 J. Behler, *The Journal of Chemical Physics*, 2011, **134**, 074106.
11 A. P. Bartók, R. Kondor and G. Csányi, *Phys. Rev. B*, 2013, **87**, 184115.
12 K. Schütt, P.-J. Kindermans, H. E. Sauceda Felix, S. Chmiela, A. Tkatchenko and K.-R. Müller, *Advances in Neural Information Processing Systems*, 2017, **30**, 992–1002.
13 P. Reiser, M. Neubert, A. Eberhard, L. Torresi, C. Zhou, C. Shao, H. Metni, C. van Hoesel, H. Schopmans, T. Sommer *et al.*, *Communications Materials*, 2022, **3**, 1–18.
14 S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt and B. Kozinsky, *Nature Communications*, 2022, **13**, 2453.
15 K. T. Schütt, O. T. Unke and M. Gastegger, *arxiv*, 2021, preprint, DOI: 10.48550/arXiv.2102.03150.
16 C. Zapp, A. Obarska-Kosinska, B. Rennekamp, M. Kurth, D. M. Hudson, D. Mercadante, U. Barayeu, T. P. Dick, V. Denysenkov, T. Prisner, M. Bennati, C. Daday, R. Kappl and F. Gräter, *Nature Communications*, 2020, **11**, 2315.
17 B. Rennekamp, F. Kutzki, A. Obarska-Kosinska, C. Zapp and F. Gräter, *Journal of Chemical Theory and Computation*, 2019, **16**, 553–563.
18 A. Obarska-Kosinska, B. Rennekamp, A. Ünal and F. Gräter, *Biophys. J.*, 2021, **120**, 3544–3549.
19 A. D. Boese and J. M. L. Martin, *The Journal of chemical physics*, 2004, **121**, 3405–3416.
20 M. J. F. G. W. T. H. B. S. G. E. S. M. A. R. J. R. C. G. S. V. B. B. M. G. A. P. H. N. M. C. X. L. H. P. H. A. F. I. J. B. G. Z. J. L. S. M. H. M. E. K. T. R. F. J. H. M. I. T. N. Y. H. O. K. H. N. T. V. J. A. M. J. J. E. P. F. O. M. B. J. J. H. E. B. K. N. K. V. N. S. T. K. R. K. J. N. K. R. A. R. J. C. B. S. S. I. J. T. M. C. N. R. J. M. M. M. K. J. E. K. J. B. C. V. B. C. A. J. J. R. G. R. E. S. O. Y. A. J. A. R. C. C. P. J. W. O. R. L. M. K. M. V. G. Z. G. A. V. P. S. J. J. D. S. D. A. D. D. O. F. J. B. F. J. V. O. J. C. D. J. Fox, *Gaussian 09, revision D. 01*, 2013.

21 P. Ramachandran, B. Zoph and Q. V. Le, *arxiv*, 2017, preprint, DOI: 10.48550/arXiv.1710.05941v2.

22 D. P. Kingma and J. Ba, *arxiv*, 2014, preprint, DOI: 10.48550/arXiv.1412.6980v9.

23 T. O'Malley, E. Bursztein, J. Long, F. Chollet, H. Jin, L. Invernizzi *et al.*, *KerasTuner*, `https://github.com/keras-team/keras-tuner`, 2019.

24 H. Huo and M. Rupp, *Mach. Learn.: Sci. Technol.*, 2022, **3**, 045017.

25 L. Himanen, M. O. J. Jäger, E. V. Morooka, F. Federici Canova, Y. S. Ranawat, D. Z. Gao, P. Rinke and A. S. Foster, *Comput. Phys. Commun.*, 2020, **247**, 106949.

26 M. Korth and S. Grimme, *Journal of chemical theory and computation*, 2009, **5**, 993–1003.

27 G. F. Mangiatordi, E. Brémond and C. Adamo, *J. Chem. Theory Comput.*, 2012, **8**, 3082–3088.

28 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.

29 D. T. Gillespie, *Journal of computational physics*, 1976, **22**, 403–434.

30 A. C. T. van Duin, S. Dasgupta, F. Lorant and W. A. Goddard, *J. Phys. Chem. A*, 2001, **105**, 9396–9409.

31 T. C. O'Connor, J. Andzelm and M. O. Robbins, *J. Chem. Phys.*, 2015, **142**, 024903.

32 P. Friederich, F. Häse, J. Proppe and A. Aspuru-Guzik, *Nature Materials*, 2021, **20**, 750–761.

33 O. T. Unke, M. Stöhr, S. Ganscha, T. Unterthiner, H. Maennel, S. Kashubin, D. Ahlin, M. Gastegger, L. M. Sandonas, A. Tkatchenko and K.-R. Müller, *arxiv*, 2022, preprint, DOI: 10.48550/arXiv.2205.08306v1.

34 M. J. Buehler and S. Keten, *Rev. Mod. Phys.*, 2010, **82**, 1459.