

Manuscript version: Author's Accepted Manuscript

The version presented in WRAP is the author's accepted manuscript and may differ from the published version or Version of Record.

Persistent WRAP URL:

<http://wrap.warwick.ac.uk/166422>

How to cite:

Please refer to published version for the most recent bibliographic citation information. If a published version is known of, the repository item page linked to above, will contain details on accessing it.

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions.

Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:

Please refer to the repository item page, publisher's statement section, for further information.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk.

Deep learning study of tyrosine reveals that roaming can lead to photodamage

Julia Westermayr,¹ Michael Gastegger,² Dóra Vörös,³ Lisa Panzenboeck,^{3,4} Florian Joerg,^{3,5} Leticia González,^{3,6} and Philipp Marquetand^{3,6,7, a)}

¹⁾*University of Warwick, Department of Chemistry, Gibbet Hill Rd, Coventry, CV4 7AL, UK*

²⁾*Technical University of Berlin, Machine Learning Group, 10587 Berlin, Germany*

³⁾*University of Vienna, Faculty of Chemistry, Institute of Theoretical Chemistry, Währinger Str. 17, 1090 Vienna, Austria.*

⁴⁾*Present address: University of Vienna, Faculty of Chemistry, Department of Analytical Chemistry, Währinger Str. 38, 1090 Vienna, Austria.*

⁵⁾*Present address: University of Vienna, Faculty of Chemistry, Institute of Computational Biological Chemistry, Währinger Str. 17, 1090 Vienna, Austria.*

⁶⁾*University of Vienna, Vienna Research Platform on Accelerating Photoreaction Discovery, Währinger Str. 17, 1090 Vienna, Austria.*

⁷⁾*University of Vienna, Research Network Data Science @ Uni Vienna, Währinger Str. 29, 1090 Vienna, Austria.*

Amino acids are among the building blocks of life forming peptides and proteins and have been carefully ‘selected’ to prevent harmful reactions caused by light. To prevent photodamage, molecules relax from electronic excited states to the ground state faster than the harmful reactions can occur, however, such photochemistry is not fully understood, in part because theoretical simulations of such systems are extremely expensive — with only smaller chromophores accessible. Here, we study the excited-state dynamics of tyrosine using a method based on deep neural networks that leverages the physics underlying quantum chemical data and combines different levels of theory. We reveal unconventional and dynamically-controlled ‘roaming’ dynamics in excited tyrosine that are beyond chemical intuition and compete with other ultrafast deactivation mechanisms. Our findings suggest that the roaming atoms are radicals that can lead to photodamage, offering a new perspective on the photostability and photodamage of biological systems.

INTRODUCTION

Amino acids form functional peptides and proteins that enable human life on earth and fundamental reactions in nature such as photosynthesis. [1, 2] These systems have been carefully selected to prevent harmful reactions caused by external stimuli, such as UV/visible light. The uptake of UV light leaves the molecule in a highly electronic excited state – potentially driving toward harmful reactions such as photodegradation, aggregation or bond cleavage. [3–6] To prevent photodamage, molecules undergo ultrafast, nonradiative transitions from electronic excited states to the ground state on a time scale much faster than irreversible, harmful reactions. Yet, the mechanisms underlying photo-induced reactions in amino acids remain elusive. Thus, knowledge about these mechanisms can substantially contribute to a better understanding of the photostability of peptides and proteins and can further help the design of novel drugs in phototherapy [7] as well as functional systems with special excited-state properties. [8–10]

Mainly three amino acids are prone to photoexcitation by sunlight: phenylalanine, tyrosine, and tryptophan. Their photodynamics can be studied experimentally, for instance via pump-probe [11] or high-harmonic spectroscopy, [12] but also theoretically via excited-state dynamics. [13] Photodynamics simulations are very powerful to decipher mecha-

nisms underlying photo-excitation and to provide explanations to experimental observables. While tryptophan, tyrosine, and phenylalanine were studied experimentally, [14–17] theoretical simulations of their excited states are extremely expensive, such that only the smaller chromophores of these molecules (such as benzene, phenol, and indole) are often the focus of theoretical investigations. [15, 18–20] However, size-dependent deactivation pathways suggested by experiments question the use of chromophores as model systems to study the photochemistry of the respective amino acids. [14, 16, 17] In particular, the photochemistry of tyrosine leaves many questions unanswered.

Photodissociation of the O–H bond located on the phenol ring (abbreviated as PhO–H in the following) is found to be a major deactivation pathway. Two main dissociation channels, which operate on a slow and a fast time scale, have been proposed for tyrosine and its chromophores after photo-excitation using 200 nm laser pulses. [16, 21] However, a significantly lower signal-to-noise ratio was found in tyrosine, in contrast to its chromophores, phenol and tyramine. Existing theoretical studies confirmed a repulsive ${}^1\pi\sigma^*$ state that can lead to photodissociation. [22] More elaborate theoretical simulations are needed to unravel the excited-state dynamics of tyrosine, but remain computationally infeasible. Studies have been either limited to static calculations or to low accuracy. [22, 23] Neither experiments nor theoretical simulations could suggest time constants or a comprehensive picture of the processes that take place in photoexcited tyrosine.

In this work, we present a computational method that can predict the excited-state dynamics of tyrosine with high accu-

^{a)}Electronic mail: philipp.marquetand@univie.ac.at

racy and on experimentally relevant time scales – that is on the order of picoseconds with respect to recent experiments. [21] This is achieved by extending our previously proposed [24, 25] photodynamics approach based on deep neural networks. We combine different levels of theory and exploit underlying physics when training neural network models. In this way, an unexpected reaction pathway, namely roaming, [26–29] for highly excited tyrosine has been discovered putting photochemical processes in biology into a new perspective.

ROAMING DISCOVERED IN HIGHLY EXCITED TYROSINE

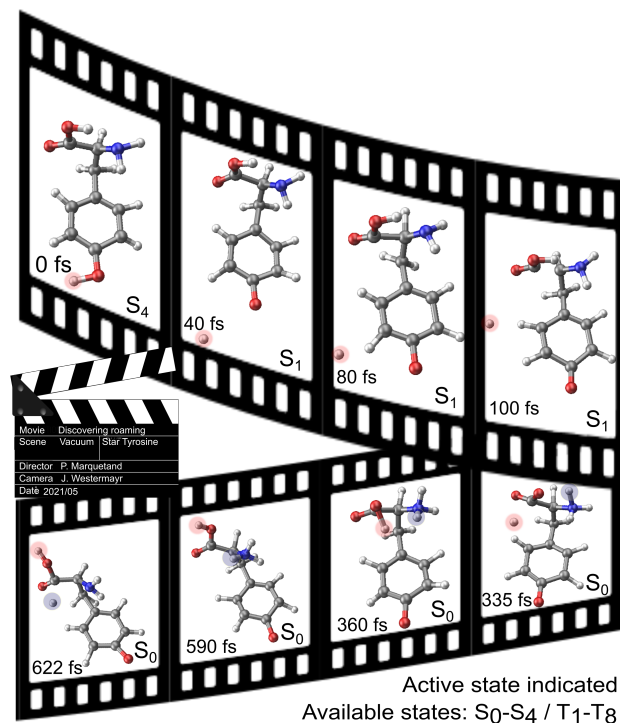


Figure 1. Roaming hydrogen atoms in tyrosine discovered during a photodynamics simulation. Selected frames of the trajectory in which roaming atoms in tyrosine were first observed. Each frame shows the time step and active state at which the geometry was visited. As can be seen, the hydrogen atom attached at the PhO–H group dissociates, but changes direction in between 40 and 80 fs. After this time, it roams around the molecule and eventually, attaches the carboxy group at the peptide backbone, after which hydrogen transfer from the carboxy group to the amino group is observed. One hydrogen atom dissociates from the amino group and starts roaming around the molecule again. A movie that shows this trajectory is attached as a supplementary file.

Roaming, originally explored in formaldehyde in the pioneering work of Bowman and coworkers from 2004 [29] and actively investigated since then, [30–35] describes an atom or fragment that moves along a dissociative potential and suddenly changes direction toward an unconventional, dynamically controlled path. The path of roaming fragments deviates from and competes with known deformation pathways and is not what one would expect based on chemical intuition.

Characteristics derived from known examples are a large interatomic distance up to 2–3 Å between the dissociated fragment and the remaining molecule [26–28, 36] and time scales that can range from hundreds of femtoseconds up to nanoseconds. [34] Yet, we are just at the beginning of understanding this special reaction mechanism and its role in nature. It was only in 2020 that roaming fragments have been observed experimentally in real-time. [34]

Experimental studies conducted on tyrosine [16, 21] are blind to the mechanisms underlying photoexcitation and cannot trace roaming atoms. Another fact that complicates the investigation of the reaction is that, at least in principle, roaming can lead to the same photoproducts. Whether this reaction prevents or promotes photodissociation or leads to long-lived excited states prone to ionization is not known.

Here, we shed light on the dynamical processes that take place after photoexcitation of tyrosine by carrying out dynamics simulations with our SchNarc approach that combines trajectory surface hopping with deep neural networks for excited-state properties. [25] The training set for the neural networks is to a large part based on the algebraic diagrammatic construction to second order perturbation theory (PT2) method, ADC(2). [37] To describe reactions that involve the breaking and formation of bonds, data points based on the complete active space perturbation theory of second order method, CASPT2, [38, 39] are added and amended to match the ADC(2) potential energy surfaces (see Supplementary section 3 for details). Corresponding simulations of only 1 picosecond directly based on the CASPT2 approach would have taken about 8 years on a high performance computer but likely would have encountered problematic geometries for the rather small, but just still computationally feasible active space (see Methods and Extended Data Fig. 1 a) and crashed. In this sense, our machine learning (ML) approach offers the unique possibility to carry out such dynamics simulations, while, to the best of our knowledge, neither multi-reference methods nor single reference methods can directly be used in practice to simulate the photodynamics of tyrosine. A more detailed description on the reference methods and training set generation can be found in the Methods section and SI, i.e., in supplementary sections 1–3, Supplementary Figs 1–4, and Supplementary Tables 1–4.

A total number of 29 spin-mixed states, i.e., 5 singlet and 8 triplet states, are learned including the forces as derivatives of the fitted potential energy surfaces and the spin-orbit couplings (SOCs) between states of different multiplicity. To the best of our knowledge, previous studies on medium-sized molecular systems like tyrosine only described a few states, while we treat more than dozens of states with ML including SOCs. One data point comprises 29 energy values, 29 force vectors, and 812 SOC values. The large number of values that have to be trained per data point adds additional complexity to the model training, which led us to introduce different aspects of the underlying physics into the neural network model. We developed a network architecture that uses the Hamiltonian as directly obtained from quantum chemistry, which consists of spin-diabatic states as diagonal elements and SOCs as off-diagonal elements. In addition, informa-

tion of the diagonalized Hamiltonian is added. The diagonalized Hamiltonian contains spin-mixed energies as diagonal elements with off-diagonal elements being zero. The model is trained on energies and SOC by using information of both the spin-diabatic (adiabatic) and diagonal representations. It predicts a spin-diabatic Hamiltonian and diagonalizes it to get the diagonal Hamiltonian. Both representations are included in the loss function, which leads to improved training of SOC and ensures that these SOC in combination with the spin-diabatic potentials yield accurate diagonal potentials that are afterwards used in the dynamics. Supplementary section 7.B provides a detailed explanation of the architecture (Supplementary Fig. 15) and accuracy (Supplementary Fig. 16) of the neural network models.

We simulated over 1,000 trajectories based on the neural network potentials to obtain statistically significant results and to discover possible reactions that take place after light excitation. We get a first impression of the photodynamics of tyrosine from a representative trajectory shown in Fig. 1, which is also attached as a movie in the supplementary information. The active state is indicated at each frame; corresponding excited state potential energy curves are given in Supplementary Fig. 8. As can be seen, the hydrogen atom that is located at the PhO–H group of the molecule follows a dissociative path and suddenly changes direction at about 80 femtoseconds to roam around the molecule. In this case, hydrogen abstraction is in competition with the hydrogen transfer from the carboxy group to the amino group of the peptide chain, forming a zwitter-ionic species that is known from recent studies and reflects much better our chemical intuition. [22] Interestingly, after hydrogen transfer, the roaming atom attaches to the carboxy group of the peptide chain and the hydrogen atom of the NH_3^+ group is transferred toward the phenyl-ring to roam again around the molecule. Roaming is accompanied by internal conversion from the first excited singlet state, S_1 , to the ground state, S_0 , which takes place between 300 and 400 femtoseconds. In contrast, the deactivation from the fourth, bright excited singlet state, S_4 , to the first excited singlet state is much faster. To make sure that roaming is not an artifact introduced by our neural network potentials, we verified this reaction in Supplementary section 5 using CASPT2 reference calculations (see Supplementary Fig. 7).

A. Fragmentation analysis

To investigate whether roaming atoms are radicals or protons, we characterize partial charges of 10,767 molecular geometries with roaming atoms that were obtained from different time steps of all trajectories (>3 million data points). We find that roaming atoms are present as radicals. The analysis was carried out with deep neural networks trained on dipole moment vectors, which internally form latent partial charges by exploiting the underlying physics of dipole moment vectors. [40] This workflow needs to be followed as partial charges derived from conventional population analysis, especially those for excited states, would necessitate a tedious recomputation of all conformations sampled during the

neural network dynamics with quantum chemistry. This process is not only computationally costly, but also tedious, and existing partitioning schemes are often unreliable or counter-intuitive. [41, 42] The model for dipole moment vectors is discussed in the Methods section and Supplementary section 6.B and its performance and architecture is evaluated in supplementary sections S7.B and S7.C, respectively, with scatter plots for dipole moments shown in Supplementary Fig. 17.

The impact of roaming on the photochemistry of tyrosine was obtained from over 1,000 dynamics simulations that would have been computationally infeasible without the help of deep learning. As already mentioned, simulations were set up according to earlier pump-probe experiments conducted by Iqbal *et al.* [16] (with further details in Supplementary section 6.D/Supplementary Fig. 5). Every dynamics trajectory was simulated at least up to 1 picosecond or until the photoproducts were formed. This picosecond time scale was suggested by experimental studies to be sufficiently long to capture all relevant reactions taking place. Nevertheless, a few hundred trajectories were additionally simulated up to 2 and 10 picoseconds. Fig. 2 (a) shows the distribution of the products that we split into those that were obtained from non-roaming and roaming trajectories. As can be seen, we found roaming hydrogen atoms that were originally located at the phenol ring in about 17% of all 1,022 trajectories characterized. Due to the large amount of data to analyze, i.e., over 3 million data points, we used k-means clustering [43] (Supplementary section 6.A and Supplementary Fig. 10) to identify different groups of products. For this analysis, we separated all trajectories into roaming and non-roaming trajectories and analysed them separately:

a. Non-roaming trajectories. The results suggest that most molecules that do not show roaming atoms remain stable during the photodynamics and can prevent most of the harmful reactions that can take place. Only about 9% of the non-roaming trajectories show direct dissociation of a hydrogen atom. The hydrogen atoms that dissociated were originally located at the oxygen of the phenol-group. Remarkably, in about 15% of the cases of hydrogen dissociation, the formation of a zwitter-ionic species can be found (see Fig. 2 (b) upper right). An additional 3.5% of non-roaming trajectories show other fragmentation paths within the conducted simulation time: p-methylphenol and peptide chain fragments (see Fig. 2 (b) upper left) are likely to be generated. Decarboxylation is another possible route that can happen before C–C bond breaking of the peptide chain or after (see lower left example).

b. Roaming trajectories. In contrast to non-roaming trajectories, only 36% of all excited molecules stay stable in roaming trajectories during the simulation time of 2 ps, while 64% eventually undergo dissociation or fragmentation of some sort, see second purple/yellow pie chart in Fig. 2 (a). All pathways start with roaming of the hydrogen atom from the phenol group. Later on, most of these hydrogen atoms leave the vicinity of the parent molecule and are classified as dissociated. Hence, roaming increases the likelihood of subsequent dissociation. Other roaming pathways lead the hydrogen atom to different parts of the tyrosine molecule and induce fragmentation into smaller pieces, as indicated in the lower

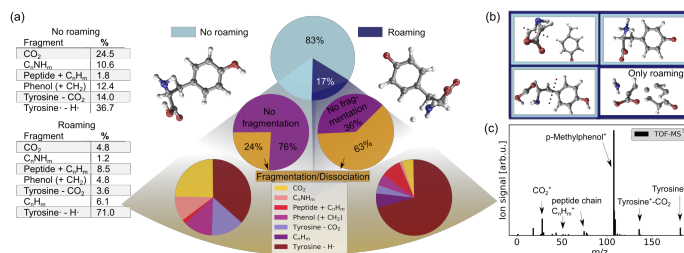


Figure 2. Photoproducts of tyrosine observed in "non-roaming" and "roaming" trajectories. (a) Products discovered by the dynamics simulations split into "no roaming" trajectories and "roaming" trajectories. Clustering techniques are used to identify the different types of products in each set of trajectories. Both types of mechanisms can lead to dissociation and fragmentation, but with different probability. (b) Representative structures found in "non-roaming" trajectories marked by light-blue rectangulars and "roaming" trajectories marked by dark-blue rectangulars. Importantly, only "roaming" trajectories lead to much smaller fragments. Other pronounced signatures that are found during the dynamics are carboxylation, H-dissociation, and dissociation of the peptide chain to form p-methylphenol and phenol. (c) Time-of-flight (TOF) mass spectrum with data extracted from ref. [21] that confirms the existence of the discovered reaction outcomes and fragments observed. The fragments that correspond to a specific, pronounced m/z ratio are indicated in the plot.

right pie chart of Fig. 2 (a) and also in panel (b). The fragments that are found in roaming trajectories are much more diverse than those in non-roaming trajectories, see Fig. 2 (b), and some smaller fragments are only found in roaming trajectories. Signatures corresponding to these smaller fragments are in agreement to those detected in experiments using multi-mass ion imaging [16, 21]. The time-of-flight mass spectrum shown in Fig. 2 (c) extracted from ref. [21] further confirms the dynamics. Large peaks in the spectrum are related to p-methylphenol, decarboxylation, hydrogen dissociation and hydrocarbons. The findings are in agreement with our current understanding of the impact of roaming on photo-excited molecules.[36]

B. Kinetics

Besides geometrical characterization, the dynamics simulations can be used to obtain population kinetics of the reaction. The average populations, after excitation to the S_4 state, in the different excited states are plotted in Fig. 3 (c). After 1 picosecond the populations change only slightly, so they are shown up to 10 picoseconds in Supplementary Fig. 6 in Supplementary section 4.

In contrast, the population transfer to the S_1 state is extremely fast and happens on an average of 66 femtoseconds. After these ultrafast transitions the molecule is found to be stuck in the S_1 state. About 40% of the trajectories show population transfer back to the ground state within 1 picosecond.

Intersystem crossing between singlet and triplet states can play an important role in biomolecules [44, 45] but here it is found to be negligible and does not have an impact on the dynamics as suggested by the population curves shown in Fig. 3 (c). The latter are similar to the ones by Mitrić and co-workers,[22] who simulated about 50 trajectories for up to 200 fs using TDDFT and did not include any triplets in the dynamics. There are some trajectories in triplets by the end of the simulation, e.g. dissociated or roaming structures, but in such conformational regions of the molecule singlets and triplets

are degenerate, hence they are treated on the same footing.

The pump-probe experiments by Stavros and co-workers [16] suggest ultrafast photodissociation and two different reaction channels – one operating on a slow and one on a fast time scale after excitation with 200 nm (pump) and probe pulses of 243.1 nm. The kinetic energy spectrum of tyrosine, extracted from ref. [16], with the two peaks corresponding to the different kinetic components can be seen in Fig. 3 (a). To get an insight into the impact of roaming on the time scales of photodissociation, the PhO–H bond distance of each trajectory is plotted along the simulation time in panel (b). Trajectories are split into those that show roaming and those that do not show roaming – the latter are additionally split into dissociative and non-dissociative trajectories. The roaming fragments are characterized by bond distances that are mainly within 2-5 Å to the parent oxygen atom initially. The relatively small distances in the range of 2-3 Å refer to recoiling hydrogen atoms that bounce against the oxygen atom several times before they either recombine, dissociate or roam around the molecule. An example trajectory of this event is attached as a supplementary movie. Noticeably, these large amplitude vibrations were also found in formaldehyde. [29]

Unfortunately, due to a low signal-to-noise ratio, no experimental time constants are reported for tyrosine. Experiments on smaller chromophores of tyrosine and other amino acids [14, 16, 17] suggest that the fast component with time constant, k_1 , results from dissociation in the ground state, whereas the slow component with time constant, k_2 , is attributed to take place in an excited state. To investigate this assumption, we carry out a two-fold analysis. On one side, the experimentally found time constants for p-ethylphenol and tyramine are used to fit an exponential function, $f(t)$, of the form: $f(t) = a_1 \cdot \exp(-\frac{t}{k_1}) + a_2 \cdot \exp(-\frac{t}{k_2})$ with a_1 and a_2 being constants of 0.15 and 0.1. The function is subtracted from its maximum value and plotted in panel (d). As it is visible, the dissociation is slower in tyramine. In addition, the amount of dissociation found in the dynamics are plotted up to 1 picosecond (solid dark orange curve), which fits well to

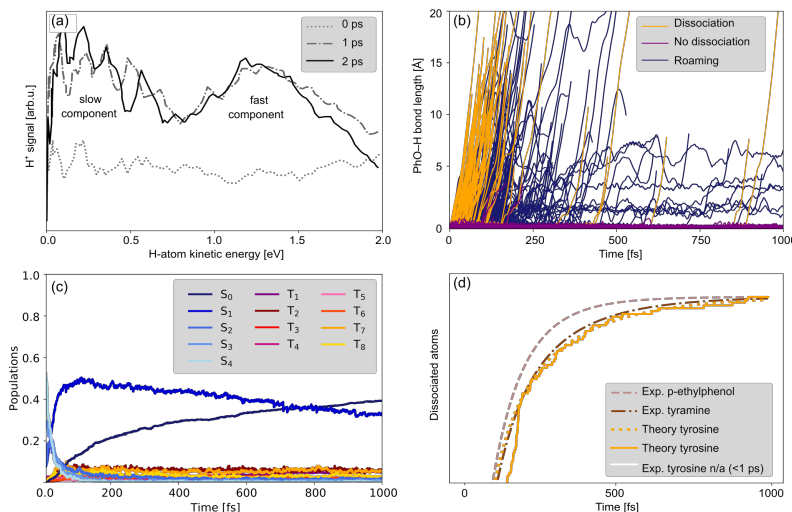


Figure 3. Kinetic analysis of photodissociation in tyrosine. (a) Kinetic energy release spectrum for tyrosine with data extracted from ref. [16], in which it was obtained via velocity map ion imaging techniques and which shows that the dynamics operate on two different time scales and reaction mechanisms. (b) The bond distance of the hydrogen atom located at the hydroxy group at the phenyl ring (PhO–H) is plotted against simulation time of 1 picosecond and separated into "roaming" and "non-roaming" (dissociative and non-dissociative) trajectories. As can be seen, roaming atoms can be present for several hundred femtoseconds, but can also be short-lived, resulting in re-attachment to the PhO–H group or dissociation. (c) Population plots averaged from 1022 trajectories simulated up to 1 picosecond. Population transfer from the S_4 to the S_1 state is ultrafast, but slow from the S_1 state to the S_0 state. Intersystem crossing is negligible and populations on short time scales are similar to those obtained with time-dependent density functional theory and without consideration of triplet states.[22] (d) Amount of trajectories that show dissociation (solid orange) against time. Experimentally found time constants for p-ethylphenol (dashed orange line) and tyramine (dashed-dotted purple line) as well as the theoretically fit time constants for tyrosine (dotted orange line) used to fit an exponential function. The determined time constant of tyrosine fits well to the experimental assumption that the dynamics are slower in tyrosine compared to those of its smaller chromophores.

the experimental observables. To verify that the slow and fast components are due to dissociation in an excited and ground state, respectively, we split the trajectories into these two categories. The population curves from the S_1 to the S_0 state are fitted (see Supplementary Fig. 12 in Supplementary section 6.B). The fast component has a reaction constant of $k_1 = 66 \pm 9$ femtoseconds and the slow component has a time constant of $k_2 = 237 \pm 77$ femtoseconds. These time constants are used to fit the previously defined function $f(t)$ and are shown by dotted lines in panel (d). The results confirm that the two time scales are due to dissociation on different potential energy surfaces. In addition, the calculated constants agree very well with the size-dependent kinetics suggested by Iqbal *et al.* [16] and Tseng *et al.*[17] The reaction rates of p-ethylphenol are reported to be 80 ± 28 femtoseconds (k_1) and 140 ± 22 femtoseconds (k_2) and those of tyramine are in the range of 80 ± 40 femtoseconds (k_1) and 210 ± 24 femtoseconds (k_2) for the fast and slow components, respectively.

CONCLUSION

In summary, we unravelled the photodynamics of tyrosine by using a combination of different high-level *ab initio* data that are learned with deep neural networks that exploit underlying physics of potential energy surfaces and couplings

in the spin-diabatic and spin-mixed representations. The distinct characteristics of the photochemistry of this amino acid could be explained and theoretically observed photoproducts and computed reaction kinetics are in agreement with experimental findings. [16, 21]

Besides the expected photodissociation, we discovered roaming atoms that are beyond chemical intuition and compete with other ultrafast deactivation mechanisms. Roaming atoms are characterized by large interatomic distances and deviate from the minimum energy paths. Analysis of latent partial charges obtained from another physically-inspired neural network [40] at the accuracy of the reference method can provide information on the type of the roaming atom. Analysis revealed that roaming atoms are present as radicals. While they are found in both, dissociative and non-dissociative trajectories, analysis with ML clustering models suggests that roaming leads to higher yields of dissociated structures and smaller fragments. Two time components of the dissociation pathway can be distinguished, in line with lifetimes proposed experimentally. [16, 21] The simulations confirm that the slow and fast time scales originate from dissociation in the ground and excited states, respectively.

The dynamics simulations as well as the analysis could only be achieved with the help of different types of ML methods. The dynamics were conducted with deep neural network potentials that were fitted by using combined data of multi-reference methods to capture the bond-breaking and -

formation in tyrosine and a single-reference method to provide smooth potential energy surfaces in conformational regions that do not describe any bond-breaking and -formation. Due to the complexity of the system and the many states, the underlying physics of the system have been considered, both in the curation of the quantum chemical data as well as in the neural network models, to enable the fitting of 29 electronic states and over 800 coupling values. Due to the computational efficiency of the prediction of potential energies, derivatives, and couplings provided by the neural networks, over thousand trajectories could be simulated on time scales comparable to experiment.

Discovering theoretical evidence for roaming in highly excited tyrosine, one of the main building blocks of life, brings our knowledge one step further toward a better understanding of the photostability and -damage of biological systems. Our results suggest that roaming might be a competing relaxation pathway in peptides and proteins, especially in phenomena like hydrogen transfer reactions that are fundamental to nature. [46, 47] Yet we are still at the beginning of understanding the secrets behind these mechanisms and their role in nature.

ACKNOWLEDGMENTS

This work was financially supported by the Austrian Science Fund, W 1232 (MolTag) and [J- 4522-N] (J.W.) the uni:docs program of the University of Vienna (J.W.). The computational results presented have been achieved in part using the Vienna Scientific Cluster. P. M. and L. G. thank the University of Vienna for continuous support, also in the frame of the research platform ViRAPID. J.W. and P. M. are grateful for an NVIDIA Hardware Grant. M. G. works at the BASLEARN Joint Lab for Machine Learning, co-financed by TU Berlin and BASF SE.

AUTHOR CONTRIBUTIONS

P.M. and L.G. proposed the project and supervised it. J.W. and P.M. implemented and designed the methods. M.G. and D.V. helped fitting neural networks. D.V., L.P., and F.J. contributed to reference calculations and the training set generation, *i.e.*, the generation of adjusted data points. J.W. performed the model training, data acquisition and model analysis. J.W. and P.M. interpreted the data, designed the analysis and wrote the initial manuscript. L.G., P.M., J.W., and M.G. revised the manuscript. All authors proofread the final manuscript and supplementary information.

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

METHODS

Choice of Reference Method

The simulation of the photodynamics of tyrosine is enabled with deep neural network potentials that are trained by combining two different levels of theory. To the best of our knowledge, ML remains one of few, if not the only, viable method to tackle the excited state dynamics of tyrosine on a quantitatively and qualitatively accurate picture. This is, because ML models can outsource the problem of costly quantum chemistry calculations from the dynamics simulations by pre-fitting the potential energy surfaces from *ab initio* data. [48, 49] Unfortunately, this is only true if a reference method capable of describing the investigated phenomena is available, which is not the case for tyrosine. The question that thus had to be addressed and lies at the heart of the fitting is how to combine different methods in the best possible way to enable the photodynamics simulations of tyrosine?

Available methods that could, at least in principle, describe the photochemistry of tyrosine qualitatively correctly are multi-reference methods such as CASPT2/SA-CASSCF (complete active space perturbation theory of second order / state-average CAS self-consistent field). [38, 39] However, they are far too expensive to study the dynamics of this system. Simulations on a picosecond time scale would take up to 8 years on a high performance computer, let alone the requirement for statistical averaging by sampling of hundreds of individual trajectories. Even if one decided to invest such computational resources, the simulations would likely crash before reaching the desired time scale due to inconsistencies in the active space along different trajectories. In addition, the fitting of potential energy surfaces obtained from CASSCF or CASPT2 fails with ML methods, due to the many excited states that are energetically close to each other in tyrosine. As can be seen in the dissociation curves in Extended Data Fig. 1 (a) in the left panel, states are likely to switch their character along a reaction path and lead to inconsistent potential energy curves. Significant jumps in the potential energy prohibit a meaningful fitting with ML. [49]

Single-reference methods, such as the algebraic diagrammatic construction to second order (ADC(2)) [37] (Extended Data Fig. 1 (a) right panel) provide smooth potential energy curves in non-dissociative regions. However, single-reference methods are usually less suitable to describe strongly correlated systems, such as tyrosine, [49, 50], as well as bond-breaking and -formation. [51] Consequently, the potential energy curves at large interatomic distances are qualitatively wrong and show a splitting of singlet and triplet states, which should be degenerate in the dissociation limit.

We thus conclude that neither multi-reference methods, nor single reference methods can be used in practice to simulate the photodynamics of tyrosine. However, ADC(2) gives a reasonably accurate description of tyrosine in non-dissociative regions, which can also be seen from the absorption spectra obtained with ADC(2) in Extended Data Fig. 1 (b). The spectrum of ADC(2) is shown with contributions from different excited states and redshifted by 0.3 eV in panel (c). It

can be seen that ADC(2) can reproduce the experimentally measured spectrum of tyrosine, while other methods such as CASSCF(12,11) or time-dependent density functional theory (TD-DFT) as used in literature [22] exhibit pronounced differences (see panel b). The spectra of CASPT2, CASSCF, and TDDFT are shown separately in Supplementary Fig. 3 with contributions from each state.

Still, the advantage of a multi-reference method like CASPT2 is that it can describe the correct asymptotic behaviour (see panel a). Thus, we resorted to exploit the expressive power of deep neural networks to combine the best of both worlds: 1) the smooth potential energy surfaces for singlet and triplet states including their derivatives and SOC's are provided from the ADC(2) method in regions that are not characterized by large interatomic distances, [37] and 2) the correct description of bond breaking and formation from CASPT2 reference data. A similar approach was applied by Kidwell *et al.*, [52] in which the ground-state potential energy surface for describing syn-CH₃CHO dissociation was obtained by combining CCSD(T) and CASPT2 energies.

Training Data Generation and *Ad Hoc* Data

A detailed description of the training set generation and the training of models is provided in supplementary sections 2-3 and 7, respectively. Here, we only briefly describe the concepts and process used for the computation of the training set. A total number of 29 spin-mixed states, i.e., 5 singlet and 8 triplet states, were learned. Chemical accuracy could be achieved by introducing underlying physics into the NN model. The numbers and types of states were chosen to match the experimental conditions, where molecules are excited with 200 nm (≈ 6.2 eV) laser pulses. [16, 21] With the shift of ADC(2) energies (Extended Data Fig. 1(b)), this excitation energy corresponds to 6.5 eV. The full excitation window was set to 6.5-7.0 eV. Using this excitation window, stochastic selection of the states populated via excitation was carried out based on oscillator strengths and excitation energies. This wavelength corresponds to an excitation to the fourth excited singlet state as can be seen from Extended Data Fig. 1(c). All triplet states that cover the same energy range as the considered singlet states are included to allow for assessing the importance of intersystem crossing between singlet and triplet states in the dynamics as phosphorescence in tyrosine cannot be excluded. [53]

The final training set comprised 17,265 data points, of which 1,967 data points were obtained from initial sampling with ADC(2) along different normal modes and combinations thereof using different conformers. Every data point was phase corrected [24] to remove the influence of the arbitrary sign of the wave function on the learning of excited-state properties like SOC's and transition dipole moments that arbitrarily switch their signs. Six simple multi-layer feed-forward neural networks as described in ref. [24] and discussed in Supplementary section 7.A (Supplementary Table 5 and Supplementary Fig.s 13-14), which train faster than SchNarc models, were then trained on the initial data points and adaptive

sampling was carried out with ADC(2) starting from different conformers. Adaptive sampling uses two or more ML models during dynamics simulations. At every time step, the predictions of the different models are compared and whenever the variance of the predictions exceeds a pre-specified threshold, simulations terminate and reference calculations are carried out to expand the training set. We started this iterative process with six neural networks as their mean allows for more robust simulations with little amount of data. [41] We ended this scheme with two neural networks and 16,738 data points. 105 *ad hoc* data points were added subsequently that were amended based on CASPT2 to reflect the correct physics in dissociative regions as will be described below (see Supplementary section 3 for additional information).

105 data points were added from 5 scans along different dissociation coordinates using the energetically lowest lying conformer of tyrosine. The corresponding hydrogen atoms are indicated in Extended Data Fig. 1(d) and were initially attached at the hydroxy group located at the phenyl-ring (1), at two C-atoms located at the phenyl ring (2 and 3), at the amino group (4), and at the carboxyl group (5). The hydrogens were detached along the vectors in which their bonds originally pointed. Unrelaxed scans were carried out, i.e., the rest of the atoms did not change position. The reaction coordinates were chosen with the aim of including a dissociation from every heavy atom type (C, N, O) and based on literature, where for example a hydrogen transfer was reported from the carboxyl group or from the amino group and to the carbon atom of the phenyl-ring. [22] To check the maximum bond length up to which ADC(2) provided valid reference data, we recalculated several points along the dissociation scans using CASPT2(12,11) and MP2. Comparison of the energy values from these different theory levels led us to define different maximum bond lengths up to which the ADC(2) data were usable along the different X-H coordinates. These maximum distances for the use of the ADC(2) data, $r_{max}^{ADC(2)}$, are given in Supplementary Table 3 along with the respective X-H equilibrium distances.

As an example for the artificial data generation, the potential scans along the PhO-H bond (bond 1) and the N-H bond (bond 4) of tyrosine are analysed in Extended Data Fig.s 1(e) and (f), respectively. In the leftmost panels, ADC(2) scans are illustrated using solid lines up to the molecular geometry that was included in the training set. Noticeably, at that point, also the phase correction algorithm [24] that tracks the phase back to a reference geometry could not be used anymore, because states that were very high in energy at the equilibrium structure (which is the reference point used in this work) entered the range of considered states here, so that their phases could not be tracked reliably. The dashed lines indicate ADC(2) results that were not included in the training set. CASPT2 energies are shown by dots and unrestricted Møller-Plesset perturbation theory (MP2) energies by blue and red crosses. The absolute energy at the respective equilibrium distance was set to zero for all methods.

As can be seen in the leftmost plots of panels (e) and (f), CASPT2(12,11) describes the dissociation event accurately and shows that at large distances, the S_0/T_1 , S_1/T_2 ,

S_2/T_3 , S_3/T_4 , and S_4/T_5 states become degenerate. In contrast, ADC(2) incorrectly describes this event and shows a splitting of singlet and triplet energies. To include dissociated geometries in the training set, the relative energies of the *ad hoc* data points were adjusted from CASPT2(12,11) calculations to the MP2 energy of the ground state (since the ground state in ADC(2) was taken from MP2 calculations) and scaled to the energy range obtained in the ADC(2) calculations at $r_{max}^{ADC(2)}$. The bond distances that were added manually to the training set are given in Supplementary Table 4. In more details the data points were manipulated as follows: In a first attempt, the degeneracy of the respective states (S_0/T_1 , S_1/T_2 , S_2/T_3 , S_3/T_4 , and S_4/T_5) at large bond distances was confirmed with CASPT2(12,11) calculations for each scan separately. Afterwards, several unrestricted MP2 calculations were carried out. At large bond distances, unrestricted MP2 also shows that the S_0 and T_1 states are degenerate. The energies at this point were taken for the artificial data points for the S_0 and T_1 states. The energies of T_5 , T_6 , and T_7 were left unchanged from ADC(2) calculations at the point $r_{max}^{ADC(2)}$. The remaining energies were obtained by taking the energy difference from CASPT2(12,11) calculations of each pair of degenerated states, *i.e.* S_1/T_2 , S_2/T_3 , S_3/T_4 , and S_4/T_5 , to the S_0/T_1 pair. The SOC values were kept constant from $r_{max}^{ADC(2)}$. The forces were first taken from the equilibrium geometry since an unrelaxed scan was carried out. Only the forces along the dissociation coordinate are then set to zero and the remaining ones adapted from the original forces by applying Gram-Schmidt orthogonalization. The ADC(2) training data and the *ad hoc* generated training data are illustrated with solid lines in the second left-most plots of panels (e) and (f).

The third and fourth plots of panels (e) and (f) show the neural network potentials that were obtained from the previously generated training set of 16,654 data points without and with artificial data points, respectively. As can be seen, models without manipulated data that reflect physically correct behaviour fail to reproduce dissociation accurately. In contrast, the neural network potentials that were additionally trained on dissociated geometries can reproduce this reaction, as expected. Close inspection of the potentials generated with a simple feed-forward neural network based on the matrix of inverse distances as input (see supplementary section 7.A) shows that the singlet and triplet states are not fully degenerate. In fact, degeneracies in energies are known to pose a problem for ML algorithms, [54, 55] which can also be seen here. Nevertheless, the potentials are sufficiently close to each other and the underlying dataset is deemed sufficient to proceed with training SchNarc models, which yield higher accuracy predictions (see Supplementary Fig. 4), and conduct photodynamics simulations. Potential energy curves along the Ph-OH minimum path are shown in Supplementary Fig. 4.

After we trained two SchNarc models on energies and forces and one SchNarc model on SOCs, accurate photodynamics were ensured by using two neural network models for energies. Their energy predictions were compared every time step and two additional adaptive sampling runs were needed until the neural network potentials converged for the targeted simulation time. The mean absolute errors (root-

mean squared errors) for energies, forces, and SOCs on a hold-out test set were 9.6 meV, 163 meV/Å, and 0.137 cm^{-1} , respectively, with scatter plots shown in Supplementary Fig. 16. Note that these values are averaged from all electronic states used. State-specific errors are reported in Supplementary Table 6.

Photodynamics simulations

All excited-state dynamics simulations were carried out with the SchNarc approach for deep-learning enhanced nonadiabatic dynamics [25] with Tully’s fewest switching surface hopping method [56, 57] as implemented in the SHARC program. [58, 59] The photodynamics approach SchNarc was validated with small model systems in pilot projects. [24, 25, 60] The accuracy and usability of this method was validated for long time scale simulations, [24] ultrafast and slow nonadiabatic transitions, [25] and the description of different spin multiplicities. [25, 61]

The neural network that lies at the heart of SchNarc is SchNet, [62, 63] adapted by us for excited states. [25] A special feature of SchNarc is that it can compute the nonadiabatic couplings between electrons and nuclei via Hessians of the squared energy-difference potentials provided by the neural network model. Such approximated nonadiabatic couplings were used to determine the hopping probability from one electronic state to another during the dynamics simulations. As nonadiabatic couplings are not available in the open quantum chemistry code that support ADC(2) and were thus not available to us, the coupling approximation was crucial to our study although they come with considerable computational costs that are outlined in Supplementary section 4.

Neural Network Models for Atomic Partial Charges

To find out whether the bond between the roaming atom and the parent oxygen atom is homo- or heterolytically broken, leading to radicals or cations and anions, respectively, the partial atomic charges were analyzed. The partial atomic charges were predicted by another deep neural network model trained on dipole moment vectors. We used the extended SchNarc approach of ref. [40] to fit dipole moment vectors, $\vec{\mu}$, via latent partial atomic charges for a given state i , $q_{i,a}$. This model was used here to analyze the atomic partial charges of the roaming atoms and the parent oxygen atoms. It is worth mentioning that partial atomic charges cannot be obtained by solving the electronic Schrödinger equation and that their computation requires a post-processing step. Different charge models exist, such as the Hirshfeld [64] or Mulliken partitioning scheme. [65] However, especially the latter scheme, is often considered as unreliable and less accurate than the former. Further, these schemes are often not implemented for the excited states. [40–42] Nevertheless, the models can be validated with dipole moment vectors for the ground and excited states. The dipole moment vectors are obtained as the sum of partial atomic charges multiplied by the vector that points from an

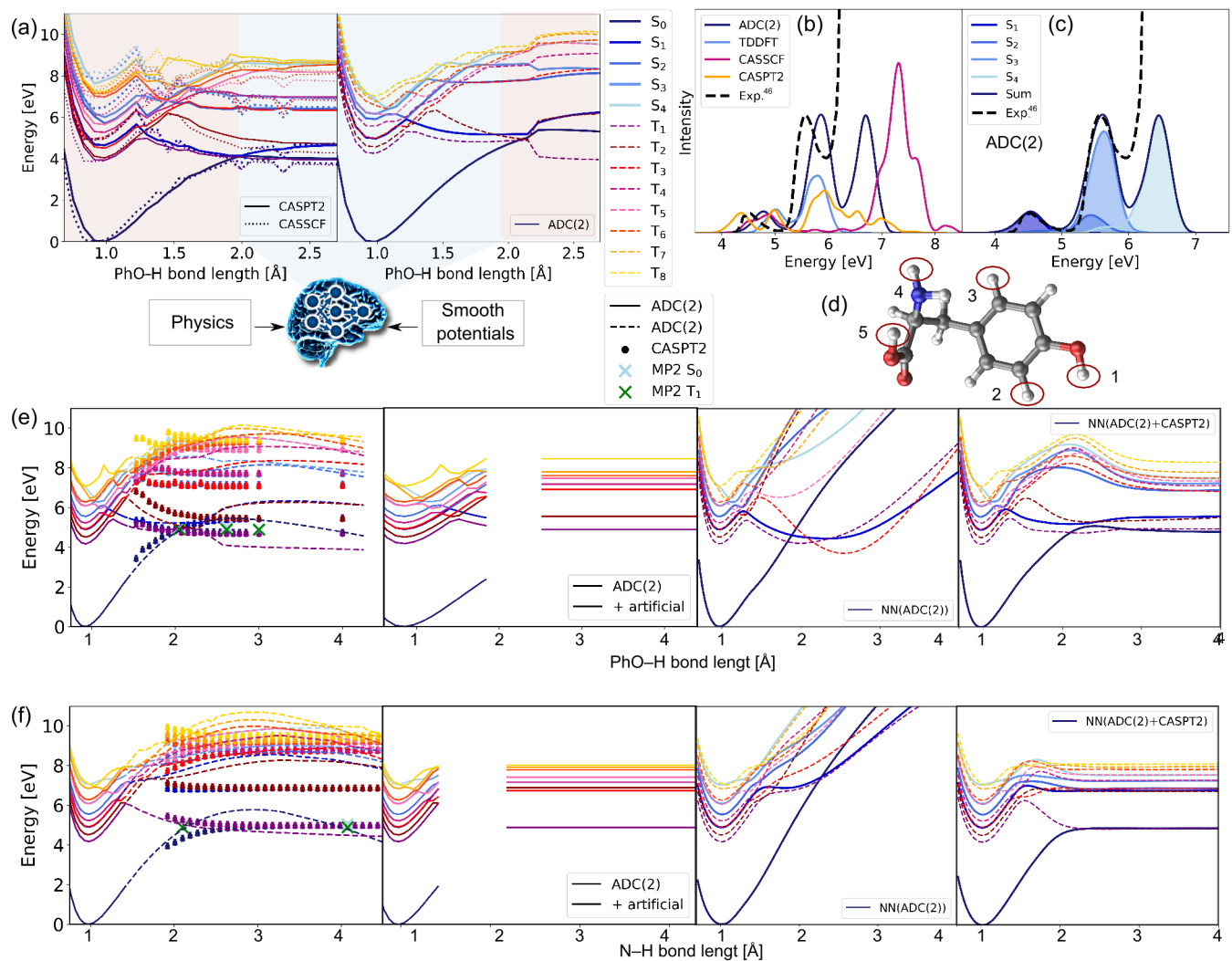


Figure 4. Validation of our method to combine ADC(2) and CASPT2, the generation of ad hoc data, and their influence on neural network training. (a) Potential energy curves along the PhO-H bond using CASSCF(12,11) (left panel, solid lines), CASPT2(12,11) (left panel, dotted lines), and ADC(2) (right panel). (b) Absorption spectra of tyrosine computed with ADC(2), TD-DFT/PBE0/SV(P), CASSCF(12,11), and CASPT(12,11) spectra using 1000 Wigner-sampled conformations. For CASPT2(12,11) and CASSCF(12,11) 758 calculations converged. The full width at half maximum for the Gaussian convolution was 0.2 eV in ADC(2) and TD-DFT spectra and 0.5 eV in CASPT2(12,11) and CASSCF(12,11) spectra and absorption peaks are scaled such that the lowest-energy peak has the same height as the energetically lowest-lying experimental absorption peak that was extracted from ref. 46. (c) Spectrum obtained from ADC(2) redshifted by 0.3 eV with contributions from different excited states. (d) Tyrosine shown with circled hydrogen atoms that were detached to generate ad hoc data (Supplementary Table 3 and 4). (e) The Ph-OH and (f) the N-H bond as example reaction coordinates to show the differences of ADC(2) (solid and dashed lines), CASPT2(12,11) (circles), and MP2 calculations (crosses) that manifest the degeneracy of singlet and triplet states. Second plots illustrates the data used for training and the third and fourth plots show neural network predictions using a simple multi-layer feed-forward neural network (Supplementary section 7.A) faster trainable than SchNarc models that were trained without ad hoc data points and with ad hoc data points, respectively. S refers to singlet and T to triplet states.

atom to the center of mass, r_a^{CM} :

$$\vec{\mu}_i = \sum_a^{N_a} q_{i,a} r_a^{CM} \quad (1)$$

The model fits the permanent and transition dipole moments. The latter properties are fitted in the same way as permanent dipole moment vectors to preserve rotational covariance. The

MAE (RMSE) on a hold-out test is 0.14 (0.32) Debye and models are comparable in accuracy with models from our previous study,[40] which could accurately represent the charge distribution in the methylenimmonium cation and ethylene. The scatter plots that show the predicted dipole moment values against the reference values are shown in Supplementary Fig. 17 with additional information in Supplementary section 7.C.

Data availability

The supporting information is freely available at doi:10.1038/s41557-022-00950-z. The molecular coordinates of the used conformers in this study that are shown in Supplementary Fig. 1 and Supplementary Fig. 9 are available as a supplementary file. Additionally, the data set is made available at doi:10.6084/m9.figshare.15132081[66] in the Atomic Simulation Environment (ase) [67] data base format including the initial conditions (geometries and velocities) to set up the dynamics. A detailed description of the machine learning models, the training set generation, training procedure, dynamics simulations and analysis is provided in the supplementary material. For reproduction of this work replace the example data set provided in the tutorial with the provided data set and set the necessary parameters indicated in the supplementary information and listed in the tutorial-instructions. For dynamics, set up the trajectories with SHARC [68] using the initial conditions file available at doi:10.6084/m9.figshare.15132081[66] and follow the tutorial on how to run dynamics with SchNarc. Source data for figures shown in the main text and the supplementary information are included.

Code availability

The ML code used in this work is available at <https://github.com/schnarc/schnarc> [25] and included as a Supplementary Code. The development branch includes a tutorial for training SchNarc models and running dynamics with it.

REFERENCES

- Collini, E. *et al.* Coherently wired light-harvesting in photosynthetic marine algae at ambient temperature. *Nature* **463**, 644–647 (2010).
- Cerullo, G. *et al.* Photosynthetic Light Harvesting by Carotenoids: Detection of an Intermediate Excited State. *Science* **298**, 2395–2398 (2002).
- Domcke, W. & Sobolewski, A. L. Peptide deactivation: Spectroscopy meets theory. *Nat. Chem.* **5**, 257–258 (2013).
- Ashfold, M. N. R., Cronin, B., Devine, A. L., Dixon, R. N. & Nix, M. G. D. The Role of $\pi\sigma^*$ Excited States in the Photodissociation of Heteroaromatic Molecules. *Science* **312**, 1637–1640. ISSN: 0036-8075 (2006).
- Schreier, W. J. *et al.* Thymine Dimerization in DNA Is an Ultrafast Photoreaction. *Science* **315**, 625–629. ISSN: 0036-8075 (2007).
- Rauer, C., Nogueira, J. J., Marquetand, P. & González, L. Cyclobutane Thymine Photodimerization Mechanism Revealed by Nonadiabatic Molecular Dynamics. *J. Am. Chem. Soc.* **138**, 15911–15916 (2016).
- Wang, Y. *et al.* Intravenous treatment of choroidal neovascularization by photo-targeted nanoparticles. *Nat. Commun.* **10**. <<https://doi.org/10.1038/s41467-019-08690-4>> (2019).
- Marder, S. R., Kippelen, B. & Jen Alex K.-Y. and Peyghambarian, N. Design and synthesis of chromophores and polymers for electro-optic and photorefractive applications. *Nature* **388**, 845–851 (1997).
- Sanchez-Lengeling, B. & Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **361**, 360–365 (2018).
- Chen, J. *et al.* Optical nano-imaging of gate-tunable graphene plasmons. *Nature* **487**, 77–81 (2012).
- Zewail, A. H. in *Femtochemistry: Ultrafast Dynamics of the Chemical Bond* 3–22 (World Scientific, 1994). doi:10.1142/9789814287609_0001. <https://www.worldscientific.com/doi/abs/10.1142/9789814287609_0001>.
- Wörner, H. J., Bertrand, J. B., Kartashov, D. V., Corkum, P. B. & Villeneuve, D. M. Following a chemical reaction using high-harmonic interferometry. *Nature* **466**, 604–607 (2010).
- Mai, S. & González, L. Molecular Photochemistry: Recent Developments in Theory. *Angew. Chem. Int. Ed.* **n/a**. doi:10.1002/anie.201916381. <<https://onlinelibrary.wiley.com/doi/pdf/10.1002/anie.201916381>> (2020).
- Tseng, C.-M. *et al.* Photostability of amino acids: photodissociation dynamics of phenylalanine chromophores. *Phys. Chem. Chem. Phys.* **12**, 4989–4995 (19 2010).
- Roberts, G. M. & Stavros, V. G. The role of $\pi\sigma^*$ states in the photochemistry of heteroaromatic biomolecules and their subunits: insights from gas-phase femtosecond spectroscopy. *Chem. Sci.* **5**, 1698–1722 (5 2014).
- Iqbal, A. & Stavros, V. G. Active Participation of $^1\pi\sigma^*$ States in the Photodissociation of Tyrosine and its Subunits. *J. Phys. Chem. Lett.* **1**, 2274–2278 (2010).
- Tseng, C.-M., Lee, Y. T., Ni, C.-K. & Chang, J.-L. Photodissociation Dynamics of the Chromophores of the Amino Acid Tyrosine: p-Methylphenol, p-Ethylphenol, and p-(2-Aminoethyl)phenol. *J. Phys. Chem. A* **111**, 6674–6678 (2007).
- Sobolewski, A. L. & Domcke, W. Ab initio investigations on the photo-physics of indole. *Chemical Physics Letters* **315**, 293–298. ISSN: 0009-2614 (1999).
- Oliver, T. A. A., Zhang, Y., Roy, A., Ashfold, M. N. R. & Bradforth, S. E. Exploring Autoionization and Photoinduced Proton-coupled Electron Transfer Pathways of Phenol in Aqueous Solution. *J. Phys. Chem. Lett.* **6**, 4159–4164 (2015).
- Xie, C. *et al.* Nonadiabatic Tunneling in Photodissociation of Phenol. *J. Am. Chem. Soc.* **138**, 7828–7831 (2016).
- Iqbal, A. *Towards Understanding the Photochemistry of tyrosine* PhD thesis (University of Warwick, 2010).
- Tomasello, G., Wohlgemuth, M., Petersen, J. & Mitrić, R. Photodynamics of Free and Solvated Tyrosine. *J. Phys. Chem. B* **116**, 8762–8770 (2012).
- Sobolewski, A. L., Shemesh, D. & Domcke, W. Computational Studies of the Photophysics of Neutral and Zwitterionic Amino Acids in an Aqueous Environment: Tyrosine-(H₂O)₂ and Tryptophan-(H₂O)₂ Clusters. *J. Phys. Chem. A* **113**, 542–550 (2009).
- Westermayr, J. *et al.* Machine Learning Enables Long Time Scale Molecular Photodynamics Simulations. *Chem. Sci.* **10**, 8100–8107 (35 2019).
- Westermayr, J., Gastegger, M. & Marquetand, P. Combining SchNet and SHARC: The SchNarc Machine Learning Approach for Excited-State Dynamics. *J. Phys. Chem. Lett.* **11**, 3828–3834 (2020).
- Bowman, J. M. & Suits, A. G. Roaming Reactions: The Third Way. *Phys. Today* **64**, 33 (2011).
- Bowman, J. M. & Shepler, B. C. Roaming Radicals. *Ann. Rev. Phys. Chem.* **62**, 531–553 (2011).
- Herath, N. & Suits, A. G. Roaming Radical Reactions. *J. Phys. Chem. Lett.* **2**, 642–647 (2011).
- Townsend, D. *et al.* The Roaming Atom: Straying from the Reaction Path in Formaldehyde Decomposition. *Science* **306**, 1158–1161 (2004).

30. Ekanayake, N. *et al.* Mechanisms and time-resolved dynamics for tri-hydrogen cation (H_3^+) formation from organic molecules in strong laser fields. *Sci. Rep.* **7**, 1–12 (2017).
31. Lu, Z., Chang, Y. C., Yin, Q.-Z., Ng, C. Y. & Jackson, W. M. Evidence for direct molecular oxygen production in CO_2 photodissociation. *Science* **346**, 61–64 (2014).
32. Mereshchenko, A. S., Butaeva, E. V., Borin, V. A., Eyzips, A. & Tarnovsky, A. N. Roaming-mediated ultrafast isomerization of geminal tri-bromides in the gas and liquid phases. *Nat. Chem.* **7**, 562–568 (2015).
33. Tso, C.-J., Kasai, T. & Lin, K.-C. Roaming Dynamics and Conformational Memory in Photolysis of Formic Acid at 193 nm using Time-Resolved Fourier-Transform Infrared Emission Spectroscopy. *Sci. Rep.* **10**, 1–10 (2020).
34. Endo, T. *et al.* Capturing Roaming Molecular Fragments in Real Time. *Science* **370**, 1072–1077 (2020).
35. Nandi, A., Zhang, P., Chen, J., Guo, H. & Bowman, J. M. Quasiclassical simulations based on cluster models reveal vibration-facilitated roaming in the isomerization of CO adsorbed on NaCl. *Nat. Chem.* **13**, 249–254 (2021).
36. Suits, A. G. Roaming Reactions and Dynamics in the van der Waals Region. *Annu. Rev. Phys. Chem.* **71**, 77–100 (2020).
37. Dreuw, A. & Wormit, M. The algebraic diagrammatic construction scheme for the polarization propagator for the calculation of excited states. *WIREs comput. Mol. Sci.* **5**, 82–95 (2015).
38. Roos, B. O., Taylor, P. R. & Siegbahn, P. E. A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach. *Chem. Phys.* **48**, 157–173 (1980).
39. Finley, J., Malmqvist, P.-A., Roos, B. O. & Serrano-Andrés, L. The multi-state {CASPT2} method. *Chem. Phys. Lett.* **288**, 299–306 (1998).
40. Westermayr, J. & Marquetand, P. Deep Learning for UV Absorption Spectra with SchNarc: First Steps Toward Transferability in Chemical Compound Space. *J. Chem. Phys.* **153**, 154112 (2020).
41. Gastegger, M., Behler, J. & Marquetand, P. Machine learning molecular dynamics for the simulation of infrared spectra. *Chem. Sci.* **8**, 6924–6935 (10 2017).
42. Gastegger, M. & Marquetand, P. in *Machine Learning Meets Quantum Physics* (eds Schütt, K. T. *et al.*) 233–252 (Springer International Publishing, Cham, 2020). ISBN: 978-3-030-40245-7. doi:10.1007/978-3-030-40245-7_12. <https://doi.org/10.1007/978-3-030-40245-7_12>.
43. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).
44. Richter, M., Marquetand, P., González-Vázquez, J., Sola, I. & González, L. Femtosecond Intersystem Crossing in the DNA Nucleobase Cytosine. *J. Phys. Chem. Lett.* **3**, 3090–3095. ISSN: 1948-7185 (Oct. 2012).
45. Richter, M., Mai, S., Marquetand, P. & González, L. Ultrafast Intersystem Crossing Dynamics in Uracil Unravelling by Ab Initio Molecular Dynamics. *Phys. Chem. Chem. Phys.* **16**, 24423–24436 (44 2014).
46. Marazzi, M., Sancho, U., Castano, O., Domcke, W. & Frutos, L. M. Photoinduced proton transfer as a possible mechanism for highly efficient excited-state deactivation in proteins. *J. Phys. Chem. Lett.* **1**, 425–428 (2009).
47. Shemesh, D., Sobolewski, A. L. & Domcke, W. Efficient Excited-State Deactivation of the Gly-Phe-Ala Tripeptide via an Electron-Driven Proton-Transfer Process. *J. Am. Chem. Soc.* **131**, 1374–1375 (2009).
48. Behler, J. Four Generations of High-Dimensional Neural Network Potentials. *Chem. Rev.* **121**, 10037–10072 (2021).
49. Westermayr, J. & Marquetand, P. Machine Learning for Electronically Excited States of Molecules. *Chem. Rev.* **121**, 9873–9926 (2021).
50. Crespo-Otero, R. & Barbatti, M. Recent Advances and Perspectives on Nonadiabatic Mixed Quantum–Classical Dynamics. *Chem. Rev.* **118**, 7026–7068 (2018).
51. Giesbertz, K. & Baerends, E. Failure of time-dependent density functional theory for excited state surfaces in case of homolytic bond dissociation. *Chem. Phys. Lett.* **461**, 338–342 (2008).
52. Kidwell, N., Li, H., Wang, X., Bowman, J. M. & Lester, M. I. Unimolecular dissociation dynamics of vibrationally activated CH_3CHOO Criegee intermediates to OH radical products. *Nature Chem.* **8**, 509–514 (2016).
53. Truong, T., Behrsohn, R., Brumer, P., Luk, C. K. & Tao, T. Effect of pH on the Phosphorescence of Tryptophan, Tyrosine, and Proteins. *J. Biol. Chem.* **242**, 2979–2985 (1967).
54. Schütt, K. T., Gastegger, M., Tkatchenko, A., Müller, K.-R. & Maurer, R. J. Unifying Machine Learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nat. Commun.* **10**, 5024 (2019).
55. Westermayr, J. & Maurer, R. J. Physically inspired deep learning of molecular excitations and photoemission spectra. *Chem. Sci.* **12**, 10755–10764 (2021).
56. Tully, J. C. Molecular Dynamics with Electronic Transitions. *J. Chem. Phys.* **93**, 1061–1071 (1990).
57. Tully, J. C. Nonadiabatic molecular dynamics. *Int. J. Quantum Chem.* **40**, 299–309. ISSN: 1097-461X (1991).
58. Richter, M., Marquetand, P., González-Vázquez, J., Sola, I. & González, L. SHARC: Ab Initio Molecular Dynamics with Surface Hopping in the Adiabatic Representation Including Arbitrary couplings. *J. Chem. Theory comput.* **7**, 1253–1258. ISSN: 1549-9618 (2011).
59. Mai, S., Marquetand, P. & González, L. Nonadiabatic Dynamics: The SHARC Approach. *WIREs comput. Mol. Sci.* **8**, e1370 (2018).
60. Westermayr, J., Faber, F. A., Christensen, A. S., von Lilienfeld, O. A. & Marquetand, P. Neural Networks and Kernel Ridge Regression for Excited States Dynamics of CH_2NH_2^+ : From Single-State to Multi-State Representations and Multi-Property Machine Learning Models. *Mach. Learn.: Sci. Technol.* **1**, 025009 (2020).
61. Westermayr, J. M. *Machine Learning for Excited-State Molecular Dynamics Simulations* PhD thesis (University of Vienna, 2020).
62. Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. SchNet – A Deep Learning Architecture for Molecules and Materials. *J. Chem. Phys.* **148**, 241722 (2018).
63. Schütt, K. T. *et al.* SchNetPack: A Deep Learning Toolbox For Atomistic Systems. *J. Chem. Theory Comput.* **15**, 448–455 (2019).
64. Hirshfeld, F. Bonded-atom fragments for describing molecular charge densities. *Theoret. Chim. Acta* **44**, 129–138 (1977).
65. Mulliken, R. S. Electronic Population Analysis on LCAO–MO Molecular Wave Functions. I. *J. Chem. Phys.* **23**, 1833–1840 (1955).
66. Westermayr, J. *Tyrosine_ExcitedStates* https://figshare.com/articles/dataset/Tyrosine_ExcitedStates/15132081. Sept. 2021. doi:{10.6084/m9.figshare.15132081.v4}.
67. Larsen, A. H. *et al.* The atomic simulation environment—a Python library for working with atoms. *J. Phys. Condens. Matter* **29**, 273002 (2017).
68. Mai, S. *et al.* SHARC2.0: Surface Hopping Including Arbitrary couplings – Program Package for Non-Adiabatic Dynamics sharc-md.org. 2018.