


# Excited state non-adiabatic dynamics of large photoswitchable molecules using a chemically transferable machine learning potential

Simon Axelrod<sup>1,2</sup>, Eugene Shakhnovich<sup>1</sup> & Rafael Gómez-Bombarelli<sup>2</sup>  

Light-induced chemical processes are ubiquitous in nature and have widespread technological applications. For example, photoisomerization can allow a drug with a photo-switchable scaffold such as azobenzene to be activated with light. In principle, photoswitches with desired photophysical properties like high isomerization quantum yields can be identified through virtual screening with reactive simulations. In practice, these simulations are rarely used for screening, since they require hundreds of trajectories and expensive quantum chemical methods to account for non-adiabatic excited state effects. Here we introduce a *diabatic artificial neural network* (DANN), based on diabatic states, to accelerate such simulations for azobenzene derivatives. The network is six orders of magnitude faster than the quantum chemistry method used for training. DANN is transferable to azobenzene molecules outside the training set, predicting quantum yields for unseen species that are correlated with experiment. We use the model to virtually screen 3100 hypothetical molecules, and identify novel species with high predicted quantum yields. The model predictions are confirmed using high-accuracy non-adiabatic dynamics. Our results pave the way for fast and accurate virtual screening of photoactive compounds.

<sup>1</sup>Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA. <sup>2</sup>Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. email: [rafagb@mit.edu](mailto:rafagb@mit.edu)

Light is a powerful tool for manipulating molecular systems. It can be controlled with high spatial, spectral and temporal precision to facilitate a variety of processes, including energy transfer, intermolecular reactions, and photoisomerization<sup>1</sup>. These processes are used in areas as diverse as synthesis, energy storage, display technology, biological imaging, diagnostics and medicine<sup>1–3</sup>. Photoactive drugs, for instance, are photoswitchable compounds whose bioactivity can be toggled through light-induced isomerization. Precise spatiotemporal control of bioactivity allows photoactive drugs to be delivered in high doses with minimal off-target activity and side effects. Such therapeutics are a promising path for the treatment of cancer, neurodegenerative diseases, bacterial infections, diabetes, and blindness<sup>4,5</sup>.

Theory plays a key role in explaining and predicting photochemistry because empirical heuristics learned from thermally activated ground state processes typically do not apply to excited states<sup>3</sup>. Computer simulations based on quantum mechanics can achieve impressive accuracy in the prediction of experimental observables. These include the isomerization efficiency and absorption spectrum of photoswitchable compounds<sup>6</sup>, which are key quantities in the design of photoactive drugs.

However, *ab initio* methods in photochemistry are severely limited by their computational cost<sup>7</sup>. In order to gather meaningful statistics for one molecule, hundreds of replicate simulations are needed, each of which involves thousands of electronic structure calculations performed in series with sub-femtosecond timesteps. The individual quantum chemical calculations are particularly demanding, requiring excited state gradients and some treatment of multireference effects. In some cases, both the ground and excited state gradients are required at each time step<sup>8,9</sup>. Using *ab initio* methods to compute photochemical properties of tens or hundreds of molecules is impractical, and photodynamic simulations have not yet been used for large-scale virtual screening.

Among the most accurate and expensive electronic structure methods are multireference perturbation techniques<sup>10–15</sup>, but their cost and requirement for manual active space selection limit their use in virtual screening. The photochemistry community has made exciting developments over several years to overcome both of these hurdles. For example, reduced scaling techniques<sup>16,17</sup> and graphics processing units<sup>18</sup> can significantly accelerate multi-reference calculations. The density matrix renormalization group (DMRG)<sup>19,20</sup> and multi-reference density functional theory (DFT) methods<sup>21–23</sup> have expanded the size of systems that can be treated with high accuracy. DMRG has also been used to automate the selection of active spaces for multi-reference methods<sup>24,25</sup>. Less accurate but more affordable black-box methods include spin-flip time-dependent DFT (SF-TDDFT)<sup>26</sup> and hole-hole Tamm-Dancoff DFT<sup>27</sup>, among others<sup>28–33</sup>. Despite these developments, the cost of non-adiabatic simulations remains high. As discussed below, even SF-TDDFT is prohibitively expensive for virtual screening. Semi-empirical methods<sup>34–36</sup> are currently the only affordable approach for large-scale screening. They provide qualitatively correct results across many systems, but are ultimately bounded by their approximations, with average energy errors of 15 kcal/mol<sup>35</sup>.

A different approach is to use data-driven models in place of quantum chemistry (QC) calculations. Machine learning (ML) models trained on quantum chemical data can now routinely predict ground state energies and forces with sub-chemical accuracy<sup>37–39</sup>, and take only milliseconds to make predictions. These models have been successfully used in a variety of ground state simulations<sup>38,40,41</sup>. They have also been used to accelerate non-adiabatic simulations in a number of model systems<sup>42–48</sup>. However, excited state ML has not yet offered affordable photodynamics for hundreds of molecules of realistic size, which is

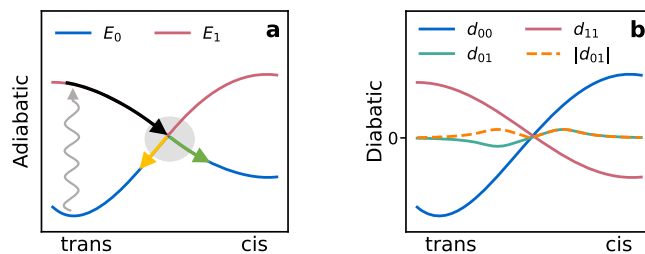
the ultimate goal for predictive simulation in photopharmacology. Further, no excited state interatomic potentials have been developed that are transferable to different compounds. They therefore require thousands of QC calculations for every new species to serve as training data.

Here we make significant progress toward affordable, large-scale photochemical simulations and virtual screening with ML. To develop a transferable potential we focus on molecules from the same chemical family, studying derivatives of azobenzene, a prototypical photoswitch. The derivatives studied here contain up to 100 atoms, making them the largest systems fit with excited state ML potentials to date. Combining an equivariant neural network<sup>38</sup> and a physics-informed diabatic model, together with data generated by combinatorial exploration of chemical space, and configurational sampling through active learning, we produce a model that is transferable to large, unseen derivatives of azobenzene. This yields computational savings in excess of six orders of magnitude. Predicted isomerization quantum yields of unseen species are correlated with experimental values. The model is used to predict the quantum yield for over 3100 hypothetical species, revealing rare molecules with high *cis*-to-*trans* and *trans*-to-*cis* quantum yields.

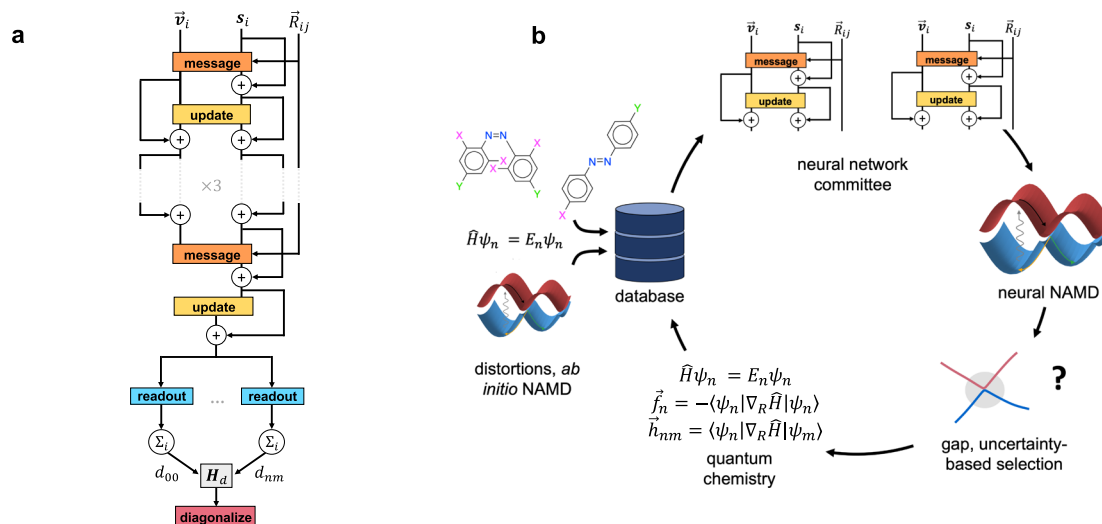
## Results

**Azobenzene photoswitches.** This work focuses on the photo-switching of azobenzene derivatives, but the methods are general and can be applied to other chemistries and other excited state processes. Azobenzene derivatives can exist as *cis* or *trans* isomers. The conformations are local minima in the ground state, but not in the excited state. Photoexcitation of either can therefore induce isomerization into the other (see the potential energy schematics in Figs. 1(a) and 2(b)). A key experimental observable is the quantum yield, defined as the probability that excitation leads to isomerization. The yield depends critically on the dynamics near conical intersections (CIs), configurations in which the excitation energy is zero. In these regions the electrons can return to the ground state with non-zero probability.

Many approaches have been developed over several decades to model such non-adiabatic transitions. These include *ab initio* multiple spawning<sup>49</sup> and cloning<sup>50</sup>; Ehrenfest dynamics<sup>8,9</sup>; coherent switching with decay of mixing<sup>51</sup>; the variational multi-configurational Gaussian method<sup>52</sup>; exact factorization<sup>53–57</sup>; the multi-configuration time-dependent Hartree (MCTDH) method<sup>58,59</sup>; Gaussian MCTDH<sup>60</sup>; and trajectory surface hopping<sup>61</sup>. A recent review of these methods can be found in ref. 3. Surface hopping is a popular approach because of its simplicity and efficiency. In this method, independent trajectories are simulated with stochastic hops



**Fig. 1** Depiction of the potential energy surfaces in azobenzene derivatives. **a**  $S_0$  and  $S_1$  adiabatic energies, with the CI region shaded in gray. Initial excitation is shown with a vertical zigzag line. Trajectories prior to hopping are shown in black. Reactive and unreactive trajectories after hopping are shown in green and yellow, respectively. **b** Diabatic energies  $d_{nm} \equiv (H_e)_{nm}$ . The diagonal diabatic elements cross and become re-ordered along the isomerization coordinate. A CI occurs when the diagonal diabatic elements cross and the off-diagonal element becomes zero.



**Fig. 2** Neural network architecture and active learning loop. **a** Schematic of the DANN architecture, which is based on the PaiNN model. Scalar atomic features  $\mathbf{s}_i$  and vectorial atomic features  $\vec{\mathbf{v}}_i$  are updated through messages from neighboring atoms. The  $\mathbf{s}_i$  are then mapped to atomic energies, which are summed to produce the diabatic Hamiltonian  $\mathbf{H}_d$ . The diabatic matrix is diagonalized to produce adiabatic quantities. **b** Schematic of the active learning loop. Geometries and QC data are first generated through ab initio NAMD, normal mode sampling, and inversion/rotation about the central N=N double bond. Two neural networks are then trained on the data and used to perform DANN-NAMD. Newly generated geometries with high committee variance and/or low predicted gaps receive QC calculations. The new calculations are added to the training data, the networks are retrained, and the cycle is repeated until convergence.

between potential energy surfaces (PESs). Depending on the curvature of the PESs and the location of the hop, a trajectory can end in the original isomer or in a new isomer (Figs. 1(a) and 2(b)). The quantum yield is the proportion of trajectories that end in a new isomer. Our goal is to predict the quantum yield of azobenzene derivatives after excitation from the singlet ground state ( $S_0$ ) to the first singlet excited state ( $S_1$ ). This can be accomplished with the surface hopping approach described above, using a fast surrogate ML model to generate the PESs. The impact of considering only the first excited state is discussed in Supplementary Note 3.

**ML architecture and training.** Our model is based on the PaiNN neural network<sup>38</sup>, which uses equivariant message-passing to predict molecular properties. In this approach, an initial feature vector is generated for each atom using its atomic number. The vector is then updated through a set of neural network operations involving “messages”, which incorporate the distance, orientation, and features of atoms within a cutoff distance. A series of updates leads to information being aggregated from increasingly distant atoms. Once the updates are complete, the atomic features are mapped to molecular energies using a neural network.

This architecture can be used to predict energies and, through automatic differentiation, the forces for each state. However, models that predict adiabatic energies have a basic shortcoming for non-adiabatic molecular dynamics (NAMD). Since surface hopping is largely controlled by the energy gap when it is close to zero, small errors in the energies can lead to exponentially large errors in the hopping probability<sup>62,63</sup>. This in turn can cause large errors in observable quantities like the quantum yield. This point is discussed in further detail in Supplementary Methods A. Further, since CIs are non-differentiable cusps in the energy gap, they are difficult to fit with neural networks. For  $N$  atoms in a molecule, the network must predict two different energies that are exactly equal in  $3N - 8$  dimensions. We found this to be particularly challenging for *trans* species that are outside the training set. As shown in Supplementary Note 6, small errors in the gap lead to the incorrect prediction that many species never hop to the ground state.

To remedy this issue we introduce a model based on diabatic states, which we call DANN (diabatic artificial neural network; Fig. 2(a)). The approach builds on previous work using neural networks for diabaticization<sup>64–66</sup>. Much of the previous work could only be used for specific system types, such as semi-rigid molecules<sup>65</sup> and coupled monomers, and is thus not applicable to azobenzene. None of the methods have been used for large systems with significant conformational changes<sup>64,66</sup>, such as azobenzene derivatives. Further, our work uses diabaticization to ease the fitting of adiabatic states across chemical space. In particular, it addresses the issue of gap overestimation near conical intersections of unseen species, as described in Supplementary Notes 1 and 6. Our work uses diabaticization to address this problem, whereas previous work only used diabaticization in single, model species. We also note that gap overestimation in unseen species is both a newly-identified and newly-addressed problem, as previous work in ML-NAMD focused on single species only<sup>42–48</sup>.

The diabatic energies form a non-diagonal Hamiltonian matrix,  $\mathbf{H}_d$ , which is diagonalized to yield adiabatic energies. When a  $2 \times 2$  sub-block of  $\mathbf{H}_d$  has diagonal elements that cross, and off-diagonal elements that pass through zero, a CI cusp is generated (Fig. 1). The diabatic energies that generate the cusp are smooth, which makes them easier to fit with an interpolating function than the adiabatic energies. In the DANN architecture, smoothness is imposed through a loss function related to the non-adiabatic coupling vector (NACV). The loss minimizes the value that the NACV takes when it is rotated from the adiabatic basis (Eq. (3)) into the diabatic basis. The NACV measures the change in overlap between two wavefunctions after a small nuclear displacement. If the NACV between two states is zero, then their wavefunctions must change slowly in response to a nuclear perturbation. Therefore, their energies cannot form the cusp in Fig. 1(a), and must instead resemble the smooth energies in Fig. 1(b).

The DANN model was trained on SF-TDDFT<sup>26</sup> calculations for 567,037 geometries, using the 6-31G\* basis<sup>67</sup> and BHHLYP<sup>68</sup> exchange-correlation functional. Unlike traditional TDDFT<sup>69</sup>,

SF-TDDFT provides an accurate description of the CI region<sup>70</sup>, and, unlike multi-reference methods, is fairly fast and requires no manual parameter selection. The configurations were sampled from 8,269 azobenzene derivatives, of which 164 were taken from the experimental literature. The remaining molecules were generated from combinatorial substitution using common literature patterns (Supplementary Tables X and XI).

The data generation process is shown in Fig. 2. Initial data was generated through ab initio NAMD with 164 species from the literature, together with normal-mode sampling and distortions of the combinatorial species to near-CI regions. The remaining data was generated through active learning. In each cycle we trained a committee of models, used one model to perform NN-NAMD, and used the committee variance and energy gap to choose NAMD geometries for new quantum chemistry calculations. The cycle was repeated five times in total; further details can be found in the Methods section.

**Validation.** To test whether the model could reproduce experimental results for unseen molecules, we evaluated it on species that were outside the training set. The test set contained 40 species (20 *cis/trans* pairs), including 33 with experimental  $S_1$  quantum yields in non-polar solution. Non-polar solution was chosen because it is the closest to the gas-phase conditions simulated here. Solvent effects can be easily incorporated into the model through transfer learning to implicit solvent calculations. This was previously shown to require new calculations for only a small proportion of the training set<sup>40</sup>.

The performance of the model is summarized in Table 1. Statistics are shown for both seen and unseen species. The former contains species that are in the training set, but geometries that are outside of it. The geometries were selected with the balanced sampling criteria described in Supplementary Note 9. Geometries from unseen species were generated with DANN-NAMD using the final trained model. Half of the DANN-NAMD geometries were selected randomly from the full trajectory and half by proximity to a CI (Supplementary Eq. (13)). 100 configurations were chosen for each molecule.

For species in the training set, all quantities are accurate to within approximately 1 kcal/mol(Å). Apart from the NACV, all quantities have  $R^2$  correlation coefficients close to 1. The  $R^2$  of the NACV is 0.84. This may be somewhat low because diabaticization cannot remove the curl component of the NACV in the diabatic basis<sup>71,72</sup>. This would also explain the low  $R^2$  value for the NACV in ref. 45, which computed it as the gradient of a scalar. For molecules outside the training set, all quantities apart from the energies have an error below 3 kcal/mol(Å). The energy gaps and ground state forces have  $R^2$  correlation coefficients near 1. The gap error of 1.89 kcal/mol should be contrasted with the error of 15 kcal/mol for the semi-empirical method in ref. 35. The errors in the excited state forces are slightly larger, but still quite low.

The correlation coefficient for the force NACV  $\vec{h}_{01}$  is rather

poor. As described in Supplementary Note 6, the yields of *trans* derivatives are better correlated with experiment when using Zhu-Nakamura surface hopping than Tully's method. The latter uses the NACV and the former does not, so part of the difference may be explained by the high error in the force NACV. Nevertheless, there is still reasonable agreement between Tully's method and experiment, suggesting that errors in the force NACV do not spoil the dynamics.

Figure 3(a) shows snapshots from an example DANN-NAMD trajectory, and panel (b) shows random samples of the hopping geometries. Reactive hopping geometries are shown on top, and non-reactive ones are shown below. The molecule is the (aminomethyl)pyridine derivative 26, with the species numbering given in Supplementary Data 2 and 3. The overlays show *cis-trans* isomerization proceeding through inversion-assisted rotation, consistent with previous work<sup>73</sup>. The dominant motion is rotation, with the CNNC dihedral angle increasing in magnitude from  $-10^\circ$  at equilibrium to  $-86^\circ$  at the hopping points. Significant changes also occur in the CNN and NNC angles, with each transitioning from  $123^\circ$  to either  $113^\circ$  or  $135^\circ$ .

The predicted PES in the branching space  $(\vec{g}, \vec{h})$  is shown beside the geometries.  $\vec{h}$  is the direction of the force NACV and  $\vec{g} \propto \nabla_R(\Delta E_{01})$  is the direction of the gap gradient. Each vector was computed with automatic differentiation using Eq. (1). The diabatic energies, adiabatic energies, and gap are shown from top to bottom. We see that the model generates a true CI, in which the  $S_0$  and  $S_1$  energies are exactly equal. Further, the degeneracy is lifted in both the  $\vec{g}$ - and  $\vec{h}$ -directions, so that the  $S_1$  energy and gap each form a characteristic cone. These hallmarks of CIs are built into the model because the adiabatic energies are eigenvalues of a diabatic matrix. For example, the cone emerges from the fact that  $d_{11} - d_{00}$  and  $d_{01}$  each pass linearly through zero in different directions<sup>74</sup>.

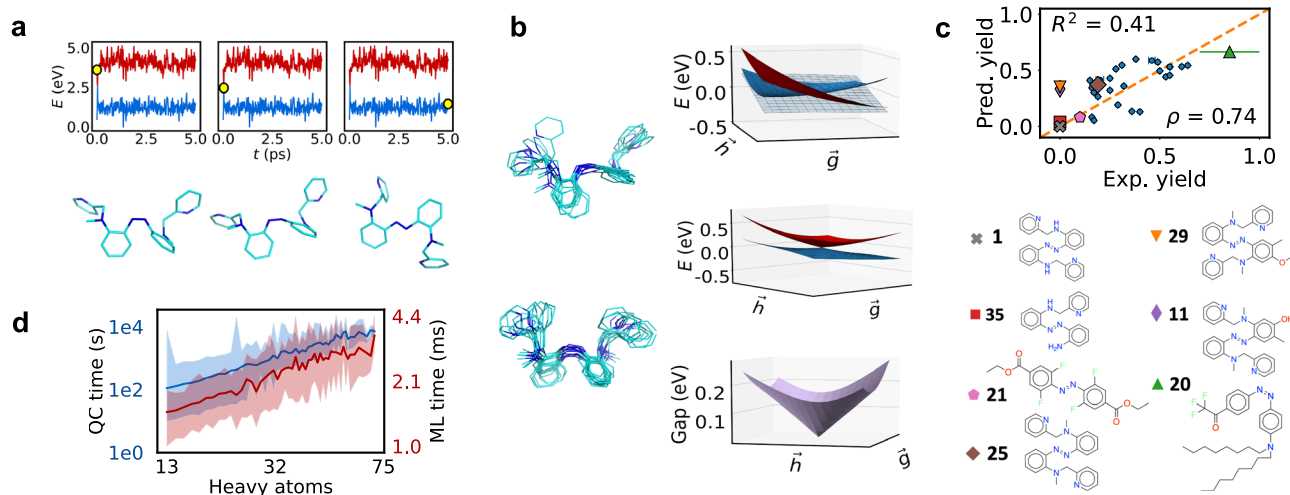
Figure 3(c) indicates that the predicted and experimental quantum yields of unseen species are correlated. The yields are for the 33 *cis* and *trans* species with experimental data in Supplementary Data 1. The  $R^2$  value is 0.42, and the Spearman rank correlation coefficient  $\rho$  is 0.74. While the  $R^2$  value is somewhat low, the Spearman rank correlation is high. The Spearman coefficient measures the accuracy with which the model ranks species by quantum yield.  $\rho$  only compares orderings, while  $R^2$  compares the model error to the error of a mean predictor. This means that  $\rho$  is a more forgiving metric, and also a more relevant metric for virtual screening. Since *cis* isomers have yields two to three times higher than *trans* isomers, the high value of  $\rho$  means that the model properly separates the isomers into low- and high-yield groups.

Further, as shown in Supplementary Figs. 5 and 7, the model produces meaningful rankings among *trans* species. The correlation coefficients are  $\rho = 0.32$  using Tully's method<sup>61</sup> and  $\rho = 0.57$  using the Zhu-Nakamura approach<sup>75</sup>. The model is largely able to differentiate between high- and low-yield *trans* derivatives.

**Table 1 MAE and coefficient of determination ( $R^2$ ) of the DANN model for various quantities.**

		$E_0$	$E_1$	$\Delta E_{01}$	$(\Delta E_{01})_{\text{small}}^a$	$\vec{f}_0$	$\vec{f}_1$	$\vec{h}_{01}$
Seen species	MAE (↓)	0.86	1.01	0.75	0.47	1.00	1.17	0.87
	$R^2$ (↑)	1.00	1.00	1.00	0.97	0.99	0.99	0.84
Unseen species	MAE (↓)	3.06	3.77	1.89	0.97	1.72	2.31	1.36
	$R^2$ (↑)	0.99	0.98	0.98	0.95	0.97	0.86	0.50

<sup>a</sup>For these  $R^2$  calculations, we computed the total sum of squares using  $\text{mean}\{\Delta E_{01}\}$  instead of  $\text{mean}\{(\Delta E_{01})_{\text{small}}\}$ . The mean predictor should not know a priori which gaps are small, and hence should predict the mean of all gaps. Units are kcal/mol for energies and kcal/mol/Å for forces and force couplings.  $E_i$  are energies,  $\vec{f}_i$  are forces,  $\Delta E_{01}$  is the energy gap, and  $\vec{h}_{01}$  is the force NACV.  $(\Delta E_{01})_{\text{small}}$  denotes the energy gap when it is under 4.6 kcal/mol (0.2 eV).



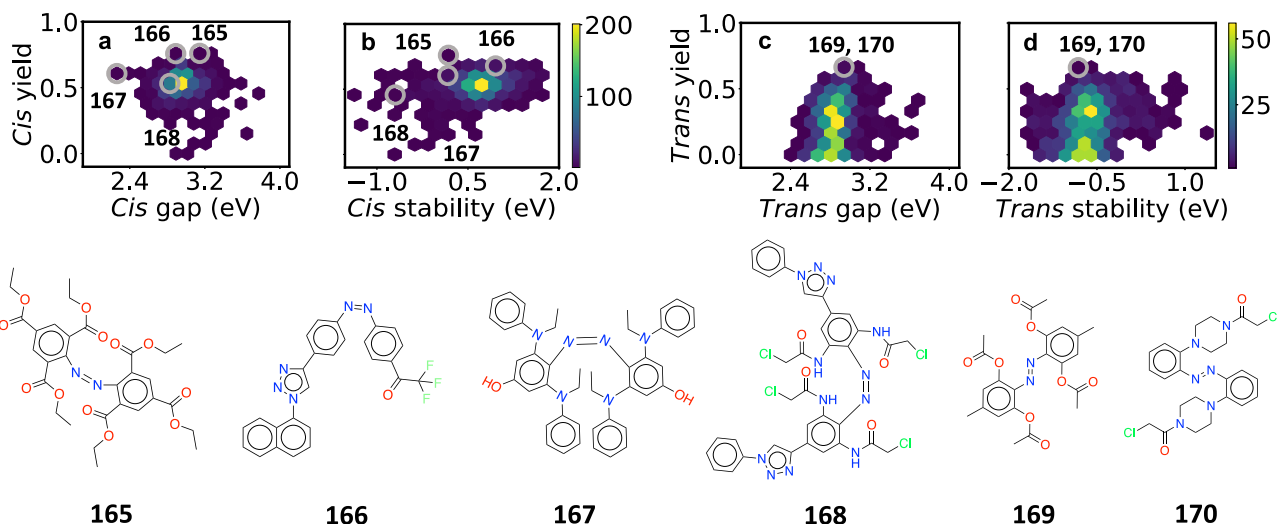
**Fig. 3** Speed and accuracy of DANN-NAMD. **a** Selected trajectory frames for a molecule outside the training set. The top panels show the  $S_0$  and  $S_1$  energy as a function of time. A yellow dot indicates the time at which the snapshot below was taken. **b** Left: Overlay of selected hopping geometries from reactive (top) and unreactive (bottom) trajectories. Right: PES as a function of branching plane coordinates at one of the reactive hopping geometries. Diabatic energies, adiabatic energies, and adiabatic gaps are shown from top to bottom. The diabatic coupling is shown in gray. **c** Predicted vs. experimental quantum yield for 33 species outside the training set. The  $R^2$  value and Spearman rank correlation  $\rho$  are both shown. Color-coded data points are defined below. Prediction error bars represent one standard deviation of 1000 bootstrapped samples. Experimental error bars represent reported uncertainties or standard deviations from multiple observations, when applicable. **d** Node time for QC and ML calculations.

Several such molecules are of interest. They are color-coded in the plots, with the legend given below. A full list of predictions is given in Supplementary Data 1. We see, for example, that the (aminomethyl)pyridine derivatives **1** and **35** are both predicted to have near-zero yields. These species do not isomerize from *trans* to *cis*, because strong N-H hydrogen bonds lock the planar *trans* conformation in place<sup>76</sup>. Replacing the NH group in **1** with N-CH<sub>3</sub> gives species **25**. This molecule isomerizes because there is no hydrogen bonding. This, too, is predicted by the model. Further, the hepta-tert-butyl derivative **17** has an experimental and predicted yield of zero. This is likely because of steric interactions among the bulky tert-butyl groups. While able to account for these two different mechanisms, the model fails to predict the subtle electronic effects in species **11** and **29**. Resonance interactions between oxygen lone pairs and the azo group modify the PES, such that there is no rotational CI<sup>77</sup>. There is instead a concerted inversion CI, which occurs too early along the path between *trans* and *cis* to allow for isomerization. The changes in the PES may either be too small or too specific to the substituents for the model to predict without fine tuning. Finally, derivatives with high yields are partly distinguished from those with low but non-zero yields. An example is **21**, whose experimental yield of 10% is half that of *trans*-azobenzene. The model properly identifies this molecule as having a low yield, but also mistakenly does the same for several high-yield species. The accuracy for unseen species could always be improved with transfer learning, in which the model is fine-tuned with a small number of calculations from a single molecule (discussed below). This would increase the computational cost, but would still be orders of magnitude less expensive than ab initio NAMD.

While meaningful correlations are produced for *trans* species, the same is not true of *cis* molecules ( $\rho = 0.02$ ). This may be because there are no *cis* derivatives with zero yield. Nevertheless, the model properly identifies **20** as having the highest yield. Further, it does not mistakenly assign a zero yield to any derivative. This is noteworthy because, as shown in Fig. 4(a) and (b), some hypothetical *cis* species are predicted to have zero yield. Synthesis of non-switching *cis* derivatives and comparison to predictions could therefore be of interest in the future.

Overall, we observe moderate correlation between predicted and experimental yields. The Spearman correlation is high when including both isomers, moderate for *trans* isomers, and low for *cis* isomers. The  $R^2$  value, a measure of numerical error compared to that of a mean predictor, is moderate when including both isomers and near-zero when separating them. Indeed, the MAEs of the mean predictor are 9.5%, 10.3%, and 17.7% for *trans*, *cis*, and all species, respectively. The model MAEs before (after) subtracting the mean signed error are 14.4% (13.5%), 11.5% (11.2%) and 13.2% (13.0%). In addition to model error, sources of error include inaccuracies in SF-TDDFT, approximations in surface hopping, solvent effects, and experimental uncertainty. These are discussed in depth in Supplementary Note 3. Each source of error affects both  $R^2$  and  $\rho$ , but is expected to have a larger effect on  $R^2$ . The rank correlation with experiment is encouraging given the difficulty of the task, as captured by the sensitivity of the yield to model errors in the PES<sup>75</sup>, and given the sources of error outside the model. Further, as discussed below, DANN provides an excellent starting point for fine-tuned, molecule-specific models that can be used for high-accuracy simulations of single species.

Figure 3(d) shows that DANN-NAMD is extremely fast. The plot shows the node time, defined as  $t_{\text{calc}}/n_{\text{calc}}$ , where  $t_{\text{calc}}$  is the calculation time per geometry, and  $n_{\text{calc}}$  is the number of parallel calculations that can be performed on a single node. We see that ML speeds up calculations by five to six orders of magnitude. The direct comparison of the pre-trained model node times and QC node times is appropriate because the model generalizes to unseen species. This means that it incurs no extra QC cost for any future simulations. The minimum speedup corresponds to the smallest molecules (14 heavy atoms or 24 total atoms), and the maximum to the largest molecules (70 heavy atoms or 99 total atoms). This reflects the different scaling of the QC and ML calculations. Empirically we see that DANN scales as  $N^{0.49}$  for  $N$  heavy atoms, while SF-TDDFT scales as  $N^{2.8}$ . These values come from fitting the timings to  $t = A \cdot N^x$ , where  $t$  is the computational time,  $A$  and  $x$  are fitted constants, and  $N$  is the number of heavy atoms. DANN's apparent sub-linear scaling is an artifact of diagonalizing  $\mathbf{H}_i$ ; when the diagonalization is removed, the



**Fig. 4 Results of virtual screening.** Species of interest are circled in gray and shown below the plots. **a** Predicted yield vs. excitation energy for *cis* derivatives. **b** Predicted yield vs. stability for *cis* derivatives. **c-d** As in (**a-b**), but for *trans* derivatives.

scaling becomes linear. This is the expected scaling for a message-passing neural network with a fixed cutoff radius. Evidently diagonalizing  $H_d$  introduces a large overhead with weak dependence on system size. Nevertheless, we see that DANN is still quite fast.

**Virtual screening.** Having shown that the model is fast and generalizes in the chemical and configurational space of azobenzenes, we next used it for virtual screening of hypothetical compounds. We first retrained the network on all available data, including species that were originally held out, for a total of 631,367 geometries in the training set. We then predicted the quantum yields of 3100 combinatorial species generated through literature-informed substitution patterns, as in ref. <sup>78</sup>. This screen served two purposes. The first was to gather statistics about the distribution of photo-physical properties of azobenzenes at a scale not accessible to experiments or traditional simulations. The second was to identify molecules with rare desirable properties. In particular, we sought to find molecules with high  $c \rightarrow t$  or  $t \rightarrow c$  quantum yields and redshifted absorption spectra. The former is important because increasing the ratio  $QY_{a \rightarrow b}/QY_{b \rightarrow a}$  where QY is the quantum yield, can lead to more complete  $a \rightarrow b$  transformation under steady state illumination. This is critical for precise spatial control of drug activity when the two isomers have different biological effects<sup>79</sup>. Redshifting is a crucial requirement for photo-active drugs, since human tissue is transparent only in the near-IR<sup>79</sup>.

The results are shown in Fig. 4. Panels (a) and (c) show the predicted yield vs. mean gap. For each species we averaged the gap over the configurations sampled during neural network ground state MD. The thermal averaging led to a typical blueshift of 0.2–0.3 eV relative to the gaps of single equilibrium geometries. The mean excitation energies are 2.95 eV for *cis* derivatives and 2.84 eV for *trans* species; the gaps are 2.98 eV and 2.97 eV for the respective unsubstituted compounds. The average gaps and their differences are similar to experimental measurements for azobenzene<sup>80</sup>. The average  $c \rightarrow t$  and  $t \rightarrow c$  yields are 54% and 24%, respectively, while those of the unsubstituted species are 59% and 37%. These are consistent with experimental results in non-polar solution, for which the base compound has yields of 44–55% and 23–28%<sup>80</sup>; the former is closer to 55% on average. However, the yield of the base *trans* compound is overestimated with respect to both theory and

experiment<sup>6,75,80</sup>. The mean (median) proportion of trajectories ending in the ground state after 2 ps are 92% (100%) for *cis* species and 31% (17%) for *trans* species. The standard deviations are 25% and 30%, respectively.

Panels (b) and (d) show the yield plotted against the isomeric stability, defined as  $E_{trans} - E_{cis}$  for *trans* isomers and  $E_{cis} - E_{trans}$  for *cis* isomers. The energy  $E$  is the median value of the configurations sampled in the ground state; we used the median to reduce the effect of outlier geometries. On average the *trans* isomers are more stable than the *cis* isomers by 0.66 eV (15.3 kcal/mol), which is similar to experimental values over 10 kcal/mol for azobenzene<sup>81</sup>. The stability is of interest for three reasons. First, a large absolute value indicates that one isomer is dominant at room temperature. This is essential for photoactive drugs, and is the case for regular azobenzene. Second, an inverted stability, in which *cis* is more stable than *trans*, allows for stronger absorption at longer wavelengths. This is because the dipole-forbidden  $n - \pi^*$  ( $S_1$ ) transition is significantly stronger for *cis* than for *trans*<sup>80</sup>. Third, in photopharmacology, one often wants to deliver a drug in inactive form, and activate it with light in a localized region. If *trans* happens to be active and *cis* inactive, then localized activation is only possible if *cis* is more stable.

Several species of interest are shown in Fig. 4. The molecules **165** and **166** have predicted  $c \rightarrow t$  yields of  $75 \pm 6\%$  and  $72 \pm 6\%$ , respectively, which are well above the *cis* average of 55%. The species **169** and **170** have predicted  $t \rightarrow c$  yields of  $66 \pm 7\%$  and  $63 \pm 10\%$ , respectively, which are three times the average *trans* yield. Molecule **167** is highly redshifted, with a mean predicted gap of 2.26 eV (548 nm), and a standard deviation of 0.87 eV. QC calculations on the geometries sampled with DANN gave a gap of  $2.26 \pm 0.61$  eV, in good agreement with predictions. The mean gap is lower than the median of 2.52 eV, which reflects the presence of several ultra-low gap structures. The low gap and large variance mean that **167** may be able to absorb in the near IR. The redshifting is likely because of the six electron donating groups, which increase the HOMO energy, together with the crowding of the four *ortho* substituents. The latter distorts the molecule, leading to twisted configurations with smaller gaps. Finally, species **168** is more stable in *cis* form than *trans* form. The predicted *cis* stability is  $-0.79$  eV ( $-18$  kcal/mol), in good agreement with the QC prediction of  $-0.92$  eV ( $-21$  kcal/mol). As mentioned above, this inverted stability can be a desirable property for photopharmacology.

To validate the yield results, we performed DANN-NAMD using highly accurate species-specific models. As described in Supplementary Note 13 B, we generated a model for each species by refining the base network with data from that species alone. The data was generated through several active learning cycles, resulting in 1200–2500 training geometries for each compound. We used this approach in place of ab initio NAMD because of the latter's prohibitive cost for large molecules. The QC computational cost for fine-tuning was at most 3% of that of an ab initio simulation, and hence far less demanding. The average gradient calculation for a molecule with 50 atoms took 58 min for two surfaces using 8 cores, and the average NACV calculation took 55 min. Fine-tuning with 2000 geometries for a medium-sized molecule would thus take 30,000 core hours. For ab initio NAMD, a conservative estimate of 100 trajectories run for 1 ps each, with only one gradient computed per frame, would take 780,000 core hours.

We also computed the yields of *cis* and *trans* azobenzene for comparison. For these species we used full ab initio simulations, because of the central role of the unsubstituted compound as a reference point and because simulations were fairly affordable for such small molecules. Issues with spin contamination also hampered the fine-tuning process for these compounds (see Supplementary Note 13 B).

Initially we generated refined models for species **165**, **166**, **169** and **170**. It became clear early on that only **165** and **169** had high yields, and so we focused on those molecules thereafter. Using molecule-specific models, we computed the quantum yields of **165** and **169** to be  $66 \pm 1\%$  and  $37 \pm 1\%$ , respectively. The computed yields for *cis* and *trans* azobenzene are  $60 \pm 4\%$  and  $26 \pm 3\%$ , respectively, which are in excellent agreement with experiment<sup>80</sup>. Both of the new molecules have higher quantum yields than the associated base compounds. The improvement is particularly large for species **169**: its yield is 11 points higher than *trans* azobenzene, which is a relative enhancement of 42 percent. We show below that this significant increase has an intuitive physical explanation.

The dynamics of the four molecules are shown in Fig. 5. Panels (a) and (b) show the CNNC dihedral angle vs. time for azobenzene, and panels (d) and (e) show the same for the derivatives. Both the substituted and unsubstituted *cis* isomers rapidly proceed through a rotational CI, but the derivative rotates much more quickly. Indeed, we see that the isomerization of the derivative is complete within 75 fs, while the base compound takes nearly 130 fs. The excited state lifetimes are  $34.2 \pm 0.3$  fs and  $63 \pm 3$  fs for the derivative and base compound, respectively, indicating that the former reaches the CI earlier than the latter. These observations may explain the enhanced yield, since a higher rotational velocity leads to more efficient isomerization<sup>82</sup>. We also note that the derivative rotates in only the counter-clockwise direction, while *cis* azobenzene rotates in both directions, but this is not expected to affect the yield.

The two *trans* molecules behave in qualitatively different ways. In *trans* azobenzene, the distribution of dihedral angles slowly widens with time (Fig. 5(b)). This is consistent with a rotational barrier<sup>6,75</sup>, as different trajectories overcome the barrier at different times, and so the torsion angle becomes uniformly distributed. Additionally, as seen in the marginal dihedral distribution of Fig. 5(c), many of the geometries hop near  $180^\circ$ . This agrees with ref. 6, which identified a non-reactive planar CI and a reactive twisted CI as the main hopping points for *trans* azobenzene. The non-reactive CI leads exclusively back to *trans*, while the reactive CI leads to *cis* and *trans* in different proportions. Using the method described in Supplementary Note 13 C, we found that 26% of the trajectories proceed through the planar CI and 74% through the rotational CI. This is similar

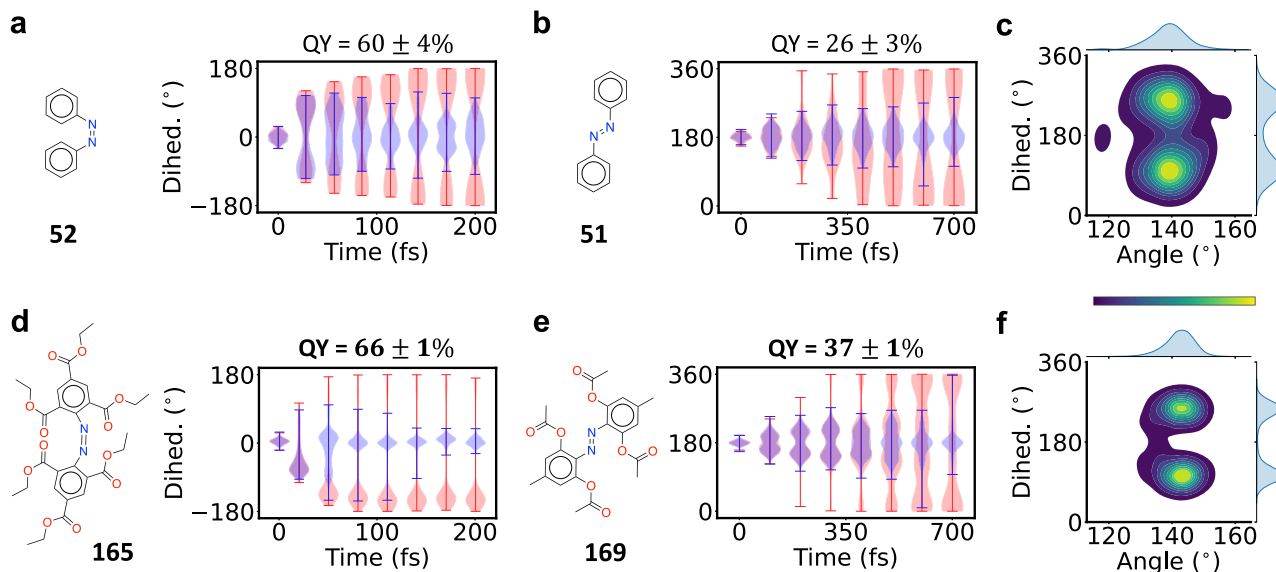
to the distribution reported in ref. 75. Approximately 36% of the rotational trajectories generate *cis* azobenzene, giving an overall yield of 26%. This is in good agreement with previous computational and experimental values<sup>6</sup>.

By contrast, nearly all trajectories of **169**, including non-reactive trajectories, rotate significantly. This can be seen in the marginal dihedral distribution in Fig. 5(f), in which the hops are tightly localized around  $180 \pm 77^\circ$ . Only 5% of the trajectories hop at the planar CI, which is five times lower than the base compound. Additionally, the yield of the rotational trajectories increases from 36 to 40%. The inhibition of the planar CI pathway, together with the enhancement of the rotational yield, leads to an overall yield increase from 26 to 37%. While the enhanced reactive yield does not have a simple explanation, the reason for the planar pathway inhibition can be clearly seen in Fig. 5(e). Whereas the rotation of **51** is stochastic, leading to a uniform distribution of angles, the rotation of **169** is initially concerted. Nearly all trajectories rotate in unison to a dihedral angle of  $180 \pm 45^\circ$  at 300 fs. Past 300 fs, hopping begins and the trajectories separate from each other. Hence they proceed through the rotational reactive CI, and become distributed between  $0^\circ$  and  $360^\circ$  after hopping. The planar non-reactive CI is avoided because of the molecule's initial rotation. This explanation is consistent with the presence of bulky *ortho* groups, which twist the equilibrium structure and hence weaken the N=N double bond. This lowers the excited state barrier to rotation, which leads to an initial torsion and hence increases the yield.

## Discussion

The DANN model shows high accuracy and transferability among azobenzene derivatives. One limitation is that the unseen species contained functional groups that were present to some degree in the training set. Model performance was generally higher for more highly represented functional groups, though some groups were highly represented yet difficult to fit, while others were weakly represented and well-fit (Supplementary Note 4). Moreover, the model cannot be applied to other chemical families without additional training data. Further, as shown in Supplementary Note 6, it substantially overestimates the excited state lifetime for a number of *trans* derivatives. On the other hand, semi-empirical methods provide qualitatively correct predictions across a variety of chemistries, but cannot match DANN's in-domain accuracy, and cannot be improved with more reference data. Adding features from semi-empirical calculations, as done in the OrbNet model<sup>83</sup>, may therefore prove useful in the future. Recent developments accounting for non-local effects and spin states have improved neural network transferability<sup>39</sup>, and could also be beneficial for excited states. The model could be further improved with high-accuracy multi-reference calculations, solvent effects, and the inclusion of the bright  $S_2$  state. The use of spin-complete methods in particular is of crucial importance, since spin contamination prevented fine-tuning the model for the base compounds. It may also have affected the accuracy of the DANN model in general. Thus spin-complete, affordable alternatives are of particular interest<sup>27</sup>. Active learning could be accelerated through differentiable sampling with adversarial uncertainty attacks<sup>84</sup>, which would improve the excited state lifetimes. Transfer learning could also be used to improve performance for specific molecules. Only a small number of ab initio calculations would be required to fine-tune the model for an individual species.

Diabatization may also prove to be useful for reactive ground states. Reaction barriers can often be understood as transitions from one diabatic state to another<sup>85</sup>. The diabatic basis may make reactive surfaces easier to fit with neural networks.



**Fig. 5 Visualization of high-accuracy non-adiabatic dynamics for several compounds of interest.** **a–b, d–e** Violin plots showing the CNNC dihedral angle vs. time. Reactive and non-reactive NAMD trajectories are shown in red and blue, respectively. The violin width at a given dihedral angle indicates the density of trajectories with that angle. The yield of each compound is shown above the plots. For ease of visualization we have used the range  $[-180, 180]$  for *cis* dihedral angles and  $[0, 360]$  for *trans* dihedral angles. **c** Distribution of hopping geometries for *trans* azobenzene. **f** As in (**c**), but for the derivative **169**. The density is visualized with kernel density estimation as a function of the CNNC dihedral and  $\max(\alpha_{\text{CNN}}, \alpha_{\text{NCC}})$ , where  $\alpha$  is an angle. Yellow corresponds to the highest density and blue to the lowest. The marginal distributions over single coordinates are shown above and to the right of each plot.

In conclusion, we have introduced a diabatic multi-state neural network potential trained on over 630,000 geometries at the SF-TDDFT BHHLYP/6-31G\* level of theory, covering over 8000 unique azobenzene molecules. We used DANN-NAMD to predict the isomerization quantum yields of derivatives outside the training set, and the results were correlated with experiment. We also identified several hypothetical compounds with high quantum yields, redshifted excitation energies, and inverted stabilities. The network architecture, diabaticization approach, and chemical and configurational diversity of the training data allowed us to produce a robust and transferable potential. The model can be applied off-the-shelf to new molecules, producing results that approximate those of SF-TDDFT at orders of magnitude lower computational cost.

## Methods

**Network and training.** As explained in Supplementary Methods A, a unique challenge for non-adiabatic simulations is their sensitivity to the energy difference between states. Using a typical neural network to predict energies and forces for NAMD leads to some molecules becoming incorrectly trapped in the excited state. This is partly caused by overestimation of the gap and/or an incorrectly shaped PES in the vicinity of the CI. To address this issue we introduce an architecture based on diabatic states, whose smooth variation leads to more accurate neural network fitting (Fig. 1(b)).

In general diabatic states must satisfy<sup>86</sup>

$$(\mathbf{U}^\dagger [\nabla_{\mathbf{R}} \mathbf{H}_d] \mathbf{U})_{nm} = \begin{cases} -\vec{f}_n, & \text{if } n = m \\ \vec{h}_{nm}, & \text{if } n \neq m. \end{cases} \quad (1)$$

where  $\nabla_{\mathbf{R}}$  is the gradient with respect to  $\vec{\mathbf{R}}$ ,  $\mathbf{U}$  diagonalizes the diabatic Hamiltonian through

$$(\mathbf{U}^\dagger \mathbf{H}_d \mathbf{U})_{nm} = E_n \delta_{nm}, \quad (2)$$

and  $\vec{f}_n = -\nabla_{\mathbf{R}} E_n$  is the adiabatic force for the  $n^{\text{th}}$  state. The dependence on  $\vec{\mathbf{R}}$  has been suppressed for ease of notation.  $\vec{h}_{nm}$  is the force NACV,

$$\begin{aligned} \vec{h}_{nm}(\vec{\mathbf{R}}) &= \langle \psi_n(\vec{\mathbf{r}}; \vec{\mathbf{R}}) | \nabla_{\mathbf{R}} \hat{H}(\vec{\mathbf{r}}; \vec{\mathbf{R}}) | \psi_m(\vec{\mathbf{r}}; \vec{\mathbf{R}}) \rangle \\ &= (E_m(\vec{\mathbf{R}}) - E_n(\vec{\mathbf{R}})) \vec{k}_{nm}(\vec{\mathbf{R}}), \end{aligned} \quad (3)$$

where  $\hat{H}(\vec{\mathbf{r}}, \vec{\mathbf{R}})$  is the clamped nucleus Hamiltonian,  $\psi_n(\vec{\mathbf{r}}; \vec{\mathbf{R}})$  is the  $n^{\text{th}}$  adiabatic wavefunction, and the matrix element is an integral over the electronic degrees of freedom  $\vec{\mathbf{r}}$ . The vector  $\vec{k}_{nm}(\vec{\mathbf{R}})$  is the derivative coupling:

$$\vec{k}_{nm}(\vec{\mathbf{R}}) = \langle \psi_n(\vec{\mathbf{r}}; \vec{\mathbf{R}}) | \nabla_{\mathbf{R}} \psi_m(\vec{\mathbf{r}}; \vec{\mathbf{R}}) \rangle \quad (4)$$

Combined with the following reference geometry conditions (Supplementary Methods C),

$$(E_0, E_1) = \begin{cases} (d_{00}, d_{11}), & \text{if } \vec{\mathbf{R}} \in \text{trans} \\ (d_{22}, d_{00}), & \text{if } \vec{\mathbf{R}} \in \text{cis}, \end{cases} \quad (5)$$

we arrive at three sets of constraints, Eqs. (1), (2), and (5). In principle only Eqs. (1) and (2) are required for the states to be diabatic. However, we found the reference loss to provide a minor improvement in the gap near CIs (Supplementary Table I).

We use a neural network to map the nuclear positions  $\vec{\mathbf{R}}_i$  and charges  $Z_i$  to the diabatic matrix elements  $d_{nm}$ , and a loss function to impose Eqs. (1), (2) and (5). The adiabatic energies  $E_n$  are generated by diagonalizing  $\mathbf{H}_d$ , and the forces and couplings by applying Eq. (1) and using automatic differentiation. The design of the network is shown schematically in Fig. 2(a). The general form of the diabatic loss function is

$$\mathcal{L} = \mathcal{L}_{\text{core}} + \mathcal{L}_{\text{ref}} + \mathcal{L}_{\text{nacv}}. \quad (6)$$

Here  $\mathcal{L}_{\text{core}}$  penalizes errors in the adiabatic energies, forces, and gaps,  $\mathcal{L}_{\text{ref}}$  imposes Eq. (5) and  $\mathcal{L}_{\text{nacv}}$  imposes Eq. (1) for  $n \neq m$ . The terms are defined explicitly in Supplementary Eqs. (1)–(3).

For the network itself we adopt the PaiNN equivariant architecture<sup>38</sup>. In this approach a set of scalar and vector features for each atom are iteratively updated through a series of convolutions (Fig. 2(a)). In the message block, the features of each atom gather information from atoms within a cutoff distance, using the interatomic displacements. The scalar and vector features for each atom are then mixed in the update phase. Hyperparameters can be found in Supplementary Table IV. Most were taken from ref. <sup>38</sup>, but some were modified based on experiments with azobenzene geometries. Further details of the PaiNN model can be found in ref. <sup>38</sup>. Once the elements of  $\mathbf{H}_d$  are generated, the diabatic matrix is diagonalized to yield the transformation matrix  $\mathbf{U}$  and the adiabatic energies  $E_n$ .

The vector quantities  $\vec{f}_n$  and  $\vec{h}_{nm}$  are given by Eq. (1). When non-adiabatic couplings are not required, the  $\vec{f}_n$  can be calculated by directly differentiating the  $E_n$ . This is more efficient than Eq. (1), since it requires only  $M_{\text{ad}} = 2 \times M_d(M_d + 1)/2 = 6$  gradient calculations. This approach was used for NAMD runs, which required only diabatic energies, adiabatic energies, and adiabatic forces, while Eq. (1) was used for training.



**Data generation and active learning.** Data was generated in two different ways. First, we searched the literature for azobenzene derivatives that had been synthesized and tested experimentally. This yielded 164 species (82 *cis* and 82 *trans*). For these species we performed ab initio NAMD, yielding geometries that densely sampled configurational space. Second, to enhance chemical diversity, we generated nearly 10,000 species through combinatorial azobenzene substitution. This was done using 48 common literature substituents and four common substitution patterns (Supplementary Tables X and XI). We then performed geometry optimizations, normal mode sampling, and inversion/rotation about the central N=N bond to generate configurations. QC calculations were performed on 25,212 combinatorial geometries. All calculations were performed with Q-Chem 5.3<sup>87</sup>, using SF-TDDFT<sup>26</sup> with the BHHLYP functional<sup>68</sup> and 6-31G\* basis<sup>67</sup>.

Two neural networks were trained on the initial data and used to perform DANN-NAMD. Initial positions and velocities for DANN-NAMD were generated from classical MD with the Nosé-Hoover thermostat<sup>88,89</sup>. The initial trajectories were unstable because the networks had not been trained on high-energy configurations. To address this issue we used active learning<sup>40,41</sup> to iteratively improve the network predictions (Fig. 2(b)). After each trajectory we performed new QC calculations on a sample of the generated geometries. For all but the last two rounds of active learning, geometries were selected according to the variance in predictions of two different networks, where the networks were initialized with different parameters and trained with different random batches. In the last two rounds, half the geometries were selected by network variance, and half by proximity to a CI. Further details are given in Supplementary Note 12. The new data was then added to the training set and used to retrain the networks. The cycle was repeated three times with all species and another two times with azobenzene alone. In all, we computed ground state gradients, excited state gradients, and NACVs with SF-TDDFT for 641,367 geometries. 96% of the geometries were from the 164 literature species. In total, 88% were generated through ab initio NAMD and 8% through active learning. The remaining 4% were from the combinatorial species. 1.5% were generated through geometry optimizations, 1.5% through inversion/rotation, and 1% through normal-mode sampling.

We initially set out to train a model using energies and forces alone. Since analytic NACVs are unavailable for many ab initio methods, an adiabatic architecture could have been used with a wider variety of methods. NACVs also add computational overhead, and so generating training data for an adiabatic model would have taken less time. To this end we initially used the Zhu-Nakamura (ZN) surface hopping method<sup>82</sup>, which only requires adiabatic energies and forces. However, the issues with adiabatic models described in Supplementary Note 6 led us to develop the diabatic approach. Since diabatic states can be used with any surface hopping method, we used the diabatic model to perform Tully's fewest switches (FS) surface hopping<sup>61</sup> after the last round of active learning. All results in the main text use the FS method. A comparison of FS and ZN results is given in Supplementary Note 6.

## Data availability

The quantum chemistry data generated in this study has been deposited in the Materials Data Facility database at <https://doi.org/10.18126/unc8-336t>. A detailed description of how to load and interpret the data is given in the README file. Source data of experimental and predicted quantum yields are provided in the Supplementary Information/Source Data file.

## Code availability

Trained models and computer code are available in the Neural Force Field repository at <https://github.com/learningmatter-mit/NeuralForceField>. Notebook tutorials explain how to train the models and perform DANN-NAMD.

Received: 27 October 2021; Accepted: 23 May 2022;

Published online: 15 June 2022

## References

- Evans, R. C., Douglas, P. & Burrow, H. D. *Applied photochemistry* (Springer, 2013).
- Kolpak, A. M. & Grossman, J. C. Azobenzene-functionalized carbon nanotubes as high-energy density solar thermal fuels. *Nano Lett.* **11**, 3156–3162 (2011).
- Mai, S. & González, L. Molecular photochemistry: Recent developments in theory. *Angewandte Chemie Int. Ed.* **59**, 16832–16846 (2020).
- Broichhagen, J., Frank, J. A. & Trauner, D. A roadmap to success in photopharmacology. *Acc. Chem. Res.* **48**, 1947–1960 (2015).
- Lerch, M. M., Hansen, M. J., van Dam, G. M., Szymanski, W. & Feringa, B. L. Emerging targets in photopharmacology. *Angewandte Chemie Int. Edition* **55**, 10978–10999 (2016).
- Yu, J. K., Bannwarth, C., Liang, R., Hohenstein, E. G. & Martínez, T. J. Nonadiabatic dynamics simulation of the wavelength-dependent photochemistry of azobenzene excited to the  $n\pi^*$  and  $\pi\pi^*$  excited states. *J. Am. Chem. Soc.* **142**, 20680–20690 (2020).
- Bannwarth, C., Yu, J. K., Hohenstein, E. G. & Martínez, T. J. Hole-hole Tamm–Dancoff-approximated density functional theory: A highly efficient electronic structure method incorporating dynamic and static correlation. *J. Chem. Phys.* **153**, 024110 (2020).
- Tully, J. C. Mixed quantum–classical dynamics. *Faraday Discussions* **110**, 407–419 (1998).
- Shalashilin, D. V. Quantum mechanics with the basis set guided by Ehrenfest trajectories: Theory and application to spin-boson model. *J. Chem. Phys.* **130**, 244101 (2009).
- Nakano, H. Quasidegenerate perturbation theory with multiconfigurational self-consistent-field reference functions. *J. Chem. Phys.* **99**, 7983–7992 (1993).
- Finley, J., Malmqvist, P., Roos, B. O. & Serrano-Andrés, L. The multi-state CASPT2 method. *Chem. Phys. Lett.* **288**, 299–306 (1998).
- Angeli, C., Cimiraglia, R., Evangelisti, S., Leininger, T. & Malrieu, J.-P. Introduction of n-electron valence states for multireference perturbation theory. *J. Chem. Phys.* **114**, 10252–10264 (2001).
- Malmqvist, P. A., Pierloot, K., Shahi, A. R. M., Cramer, C. J. & Gagliardi, L. The restricted active space followed by second-order perturbation theory method: Theory and application to the study of CuO<sub>2</sub> and Cu<sub>2</sub>O<sub>2</sub> systems. *J. Chem. Phys.* **128**, 204109 (2008).
- Shiozaki, T., Györfy, W., Celani, P. & Werner, H.-J. Communication: Extended multi-state complete active space second-order perturbation theory: Energy and nuclear gradients. *J. Chem. Phys.* **135**, 081106–081106 (2011).
- Ma, D., Manni, G. L., Olsen, J. & Gagliardi, L. Second-order perturbation theory for generalized active space self-consistent-field wave functions. *J. Chem. Theory Comput.* **12**, 3208–3213 (2016).
- Song, C. & Martínez, T. J. Reduced scaling extended multi-state CASPT2 (XMS-CASPT2) using supporting subspaces and tensor hyper-contraction. *J. Chem. Phys.* **152**, 234113 (2020).
- Song, C., Neaton, J. B. & Martínez, T. J. Reduced scaling formulation of CASPT2 analytical gradients using the supporting subspace method. *J. Chem. Phys.* **154**, 014103 (2021).
- Seritan, S. et al. TeraChem: Accelerating electronic structure and ab initio molecular dynamics with graphical processing units. *J. Chem. Phys.* **152**, 224110 (2020).
- Marti, K. H. & Reiher, M. New electron correlation theories for transition metal chemistry. *Phys. Chem. Chem. Phys.* **13**, 6750–6759 (2011).
- Sharma, S. & Chan, G. K.-L. Spin-adapted density matrix renormalization group algorithms for quantum chemistry. *J. Chem. Phys.* **136**, 124121 (2012).
- Marian, C. M., Heil, A. & Kleinschmidt, M. The DFT/MRCI method. *Wiley Interdisciplinary Rev.: Comput. Mol. Sci.* **9**, e1394 (2019).
- Manni, G. L. et al. Multiconfiguration pair-density functional theory. *J. Chem. Theory Comput.* **10**, 3669–3680, 2014.
- Gagliardi, L. et al. Multiconfiguration pair-density functional theory: A new way to treat strongly correlated systems. *Acc. Chem. Res.* **50**, 66–73 (2017).
- Stein, C. J. & Reiher, M. Automated selection of active orbital spaces. *J. Chem. Theory Comput.* **12**, 1760 (2016).
- Stein, C. J. & Reiher, M. autoCAS: A program for fully automated multiconfigurational calculations. *J. Comput. Chem.* **40**, 2216 (2019).
- Shao, Y., Head-Gordon, M. & Krylov, A. I. The spin-flip approach within time-dependent density functional theory: Theory and applications to diradicals. *J. Chem. Phys.* **118**, 4807–4818 (2003).
- Yu, J. K., Bannwarth, C., Hohenstein, E. G. & Martínez, T. J. Ab initio nonadiabatic molecular dynamics with hole-hole Tamm–Dancoff approximated density functional theory. *J. Chem. Theory Comput.* **16**, 5499–5511 (2020).
- Li, S. L., Marenich, A. V., Xu, X. & Truhlar, D. G. Configuration interaction-corrected Tamm–Dancoff approximation: A time-dependent density functional method with the correct dimensionality of conical intersections. *J. Phys. Chem. Lett.* **5**, 322–328 (2014).
- Filatov, M. Spin-restricted ensemble-referenced Kohn–Sham method: basic principles and application to strongly correlated ground and excited states of molecules. *Wiley Interdisciplinary Rev.: Comput. Mol. Sci.* **5**, 146–167 (2015).
- Yang, Y., Shen, L., Zhang, D. & Yang, W. Conical intersections from particle–particle random phase and Tamm–Dancoff approximations. *J. Phys. Chem. Lett.* **7**, 2407–2411 (2016).
- Shu, Y., Parker, K. A. & Truhlar, D. G. Dual-functional Tamm–Dancoff approximation: a convenient density functional method that correctly describes S<sub>1</sub>/S<sub>0</sub> conical intersections. *J. Phys. Chem. Lett.* **8**, 2107–2112 (2017).
- Lee, S., Filatov, M., Lee, S. & Choi, C. H. Eliminating spin-contamination of spin-flip time dependent density functional theory within linear response formalism by the use of zeroth-order mixed-reference (MR) reduced density matrix. *J. Chem. Phys.* **149**, 104101 (2018).

33. Teh, H.-H. & Subotnik, J. E. The simplest possible approach for simulating  $S_0$ - $S_1$  conical intersections with DFT/TDDFT: Adding one doubly excited configuration. *J. Phys. Chem. Lett.* **10**, 3426–3432 (2019).
34. Cusati, T. et al. Semiempirical Hamiltonian for simulation of azobenzene photochemistry. *J. Phys. Chem. A* **116**, 98–110 (2012).
35. Inamori, M., Yoshikawa, T., Ikabata, Y., Nishimura, Y. & Nakai, H. Spin-flip approach within time-dependent density functional tight-binding method: Theory and applications. *J. Comput. Chem.* **41**, 1538–1548 (2020).
36. de Wergifosse, M., Bannwarth, C. & Grimme, S. A simplified spin-flip time-dependent density functional theory approach for the electronic excitation spectra of very large diradicals. *J. Phys. Chem. A* **123**, 5815–5825 (2019).
37. Qiao, Z., Welborn, M., Anandkumar, A., Manby, F. R. & Miller III, T. F. OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features. *J. Chem. Phys.* **153**, 124111 (2020).
38. Schütt, K., Unke, O. & Gastegger, M. Equivariant message passing for the prediction of tensorial properties and molecular spectra. In *International Conference on Machine Learning*, 9377–9388. PMLR, (2021).
39. Unke, O. T. et al. Spookynet: Learning force fields with electronic degrees of freedom and nonlocal effects. *Nat. Commun.* **12**, 1–14 (2021).
40. Ang, S. J., Wang, W., Schwalbe-Koda, D., Axelrod, S. & Gómez-Bombarelli, R. Active learning accelerates ab initio molecular dynamics on reactive energy surfaces. *Chem* **7**, 738 (2021).
41. Wang, W., Yang, T., Harris, W. H. & Gómez-Bombarelli, R. Active learning and neural network potentials accelerate molecular screening of ether-based solvate ionic liquids. *Chem. Commun.* **56**, 8920 (2020).
42. Chen, W.-K., Liu, X.-Y., Fang, W.-H., Dral, P. O. & Cui, G. Deep learning for nonadiabatic excited-state dynamics. *J. Phys. Chem. Lett.* **9**, 6702–6708 (2018).
43. Dral, P. O., Barbatti, M. & Thiel, W. Nonadiabatic excited-state dynamics with machine learning. *J. Phys. Chem. Lett.* **9**, 5660–5663 (2018).
44. Hu, D., Xie, Y., Li, X., Li, L. & Lan, Z. Inclusion of machine learning kernel ridge regression potential energy surfaces in on-the-fly nonadiabatic molecular dynamics simulation. *J. Phys. Chem. Lett.* **9**, 2725–2732 (2018).
45. Li, J. et al. Automatic discovery of photoisomerization mechanisms with nanosecond machine learning photodynamics simulations. *Chem. Sci.* **12**, 5302–5314 (2021).
46. Westermayr, J. Machine learning enables long time scale molecular photodynamics simulations. *Chem. Sci.* **10**, 8100–8107 (2019).
47. 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).
48. Westermayr, J. & Marquetand, P. Machine learning for electronically excited states of molecules. *Chem. Rev.* **121**, 9873–9926 (2020).
49. Ben-Nun, M., Quenneville, J. & Martínez, T. J. Ab initio multiple spawning: Photochemistry from first principles quantum molecular dynamics. *J. Phys. Chem. A* **104**, 5161–5175 (2000).
50. Makhov, D. V., Glover, W. J., Martínez, T. J. & Shalashilin, D. V. Ab initio multiple cloning algorithm for quantum nonadiabatic molecular dynamics. *J. Chem. Phys.* **141**, 054110 (2014).
51. Zhu, C., Nangia, S., Jasper, A. W. & Truhlar, D. G. Coherent switching with decay of mixing: an improved treatment of electronic coherence for non-Born-Oppenheimer trajectories. *J. Chem. Phys.* **121**, 7658–7670 (2004).
52. Richings, G. W. et al. Quantum dynamics simulations using Gaussian wavepackets: the vMCG method. *Int. Rev. Phys. Chem.* **34**, 269–308 (2015).
53. Abedi, A., Maitra, N. T. & Gross, E. K. U. Exact factorization of the time-dependent electron-nuclear wave function. *Phys. Rev. Lett.* **105**, 123002 (2010).
54. Abedi, A., Agostini, F. & Gross, E. K. U. Mixed quantum-classical dynamics from the exact decomposition of electron-nuclear motion. *EPL Europhys. Lett.* **106**, 33001 (2014).
55. Min, S. K., Agostini, F., Tavernelli, I. & Gross, E. K. U. Ab initio nonadiabatic dynamics with coupled trajectories: A rigorous approach to quantum (de) coherence. *J. Phys. Chem. Lett.* **8**, 3048–3055 (2017).
56. Curchod, B. F. E. & Agostini, F. On the dynamics through a conical intersection. *J. Phys. Chem. Lett.* **8**, 831–837 (2017).
57. Ha, J.-K., Lee, I. S. & Min, S. K. Surface hopping dynamics beyond nonadiabatic couplings for quantum coherence. *J. Phys. Chem. Lett.* **9**, 1097–1104 (2018).
58. Beck, M. H., Jäckle, A., Worth, G. A. & Meyer, H.-D. The multiconfiguration time-dependent Hartree (MCTDH) method: a highly efficient algorithm for propagating wavepackets. *Phys. Rep.* **324**, 1–105 (2000).
59. Wang, H. & Thoss, M. Multilayer formulation of the multiconfiguration time-dependent Hartree theory. *J. Chem. Phys.* **119**, 1289–1299 (2003).
60. Burghardt, I., Meyer, H.-D. & Cederbaum, L. S. Approaches to the approximate treatment of complex molecular systems by the multiconfiguration time-dependent Hartree method. *J. Chem. Phys.* **111**, 2927–2939 (1999).
61. Tully, J. C. Molecular dynamics with electronic transitions. *J. Chem. Phys.* **93**, 1061–1071 (1990).
62. Zhu, C. & Nakamura, H. The two-state linear curve crossing problems revisited. II. Analytical approximations for the Stokes constant and scattering matrix: The Landau-Zener case. *J. Chem. Phys.* **97**, 8497–8514 (1992).
63. Zhu, C. & Nakamura, H. The two-state linear curve crossing problems revisited. III. Analytical approximations for Stokes constant and scattering matrix: Nonadiabatic tunneling case. *J. Chem. Phys.* **98**, 6208–6222 (1993).
64. Shu, Y. & Truhlar, D. G. Diabatization by machine intelligence. *J. Chem. Theory Comput.* **16**, 6456–6464 (2020).
65. Williams, D. M. G. & Eisfeld, W. Neural network diabatization: A new ansatz for accurate high-dimensional coupled potential energy surfaces. *J. Chem. Phys.* **149**, 204106 (2018).
66. Guan, Y., Zhang, D. H., Guo, H. & Yarkony, D. R. Representation of coupled adiabatic potential energy surfaces using neural network based quasi-diabatic Hamiltonians:  $1,2^2A'$  states of LiFH. *Phys. Chem. Chem. Phys.* **21**, 14205–14213 (2019).
67. Francl, M. M. et al. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **77**, 3654–3665 (1982).
68. Becke, A. D. A new mixing of Hartree-Fock and local density-functional theories. *J. Chem. Phys.* **98**, 1372–1377 (1993).
69. Levine, B. G., Ko, C., Quenneville, J. & Martínez, T. J. Conical intersections and double excitations in time-dependent density functional theory. *Mol. Phys.* **104**, 1039–1051 (2006).
70. Lee, S., Shostak, S., Filatov, M. & Choi, C. H. Conical intersections in organic molecules: Benchmarking mixed-reference spin-flip time-dependent DFT (MRSF-TD-DFT) vs spin-flip TD-DFT. *J. Phys. Chem. A* **123**, 6455–6462 (2019).
71. Mead, C. A. & Truhlar, D. G. Conditions for the definition of a strictly diabatic electronic basis for molecular systems. *J. Chem. Phys.* **77**, 6090–6098 (1982).
72. Baer, M. & Englman, R. A study of the diabatic electronic representation within the Born-Oppenheimer approximation. *Mol. Phys.* **75**, 293–303 (1992).
73. Toniolo, A., Ciminelli, C., Persico, M. & Martínez, T. J. Simulation of the photodynamics of azobenzene on its first excited state: Comparison of full multiple spawning and surface hopping treatments. *J. Chem. Phys.* **123**, 234308 (2005).
74. Köppel, H., Gronki, J. & Mahapatra, S. Construction scheme for regularized diabatic states. *J. Chem. Phys.* **115**, 2377 (2001).
75. Yue, L., Liu, Y. & Zhu, C. Performance of TDDFT with and without spin-flip in trajectory surface hopping dynamics: cis=trans azobenzene photoisomerization. *Phys. Chem. Chem. Phys.* **20**, 24123–24139 (2018).
76. Bandara, H. M. D. et al. Proof for the concerted inversion mechanism in the trans → cis isomerization of azobenzene using hydrogen bonding to induce isomer locking. *J. Organic Chem.* **75**, 4817–4827 (2010).
77. Bandara, H. M. D., Cawley S., Gascón A. & Burdette S. C. Short-circuiting azobenzene photoisomerization with electron-donating substituents and reactivating the photochemistry with chemical modification. *Eur. J. Org. Chem.* **2011**, 2916–2919 (2011).
78. Gómez-Bombarelli, R. et al. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nat. Mater.* **15**, 1120–1127 (2016).
79. Velema, W. A., Szymanski, W. & Feringa, B. L. Photopharmacology: beyond proof of principle. *J. Am. Chem. Soc.* **136**, 2178–2191 (2014).
80. Bandara, H. M. D. & Burdette, S. C. Photoisomerization in different classes of azobenzene. *Chem. Soc. Rev.* **41**, 1809–1825 (2012).
81. Dias, A. R. et al. Enthalpies of formation of cis-azobenzene and trans-azobenzene. *J. Chem. Thermodyn.* **24**, 439–447 (1992).
82. Yu, L., Xu, C., Lei, Y., Zhu, C. & Wen, Z. Trajectory-based nonadiabatic molecular dynamics without calculating nonadiabatic coupling in the avoided crossing case: Trans=cis photoisomerization in azobenzene. *Phys. Chem. Chem. Phys.* **16**, 25883–25895 (2014).
83. Qiao, Z. et al. Multi-task learning for electronic structure to predict and explore molecular potential energy surfaces. arXiv preprint arXiv:2011.02680 (2020).
84. Schwalbe-Koda, D., Tan, A. R. & Gómez-Bombarelli, R. Differentiable sampling of molecular geometries with uncertainty-based adversarial attacks. *Nat. Commun.* **12**, 1–12 (2021).
85. Van Voorhis, T. et al. The diabatic picture of electron transfer, reaction barriers, and molecular dynamics. *Ann. Rev. Phys. Chem.* **61**, 149–170 (2010).
86. Schuurman, M. S. & Yarkony, D. R. On the vibronic coupling approximation: A generally applicable approach for determining fully quadratic quasiadiabatic coupled electronic state Hamiltonians. *J. Chem. Phys.* **127**, 094104 (2007).
87. Shao, Y. et al. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **113**, 184–215 (2015).
88. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.* **81**, 511–519 (1984).
89. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695 (1985).

## Acknowledgements

We thank Wujie Wang, Daniel Schwalbe-Koda, Shi Jun Ang (MIT), Kristof Schütt, and Oliver Unke (Technische Universität Berlin) for scientific discussions and access to computer code. Harvard Cannon cluster, MIT Engaging cluster, and MIT Lincoln Lab Supercloud cluster at MGHPCC are gratefully acknowledged for computational resources and support. Financial support from DARPA (Award HR00111920025) and MIT-IBM Watson AI Lab is acknowledged.

## Author contributions

S.A. conceived the project, and developed the methodology with R.G.-B. and E.S. S.A. performed the calculations under the guidance of R.G.-B. S.A. wrote the first draft of the manuscript, and all authors contributed to the final version.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-30999-w>.

**Correspondence** and requests for materials should be addressed to Rafael Gómez-Bombarelli.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022