Electronic Supplementary Material (ESI) for Chemical Science.
This journal is © The Royal Society of Chemistry 2021

Supporting Information

Automatic Discovery of Photoisomerization Mechanisms with Nanosecond Machine Learning Photodynamics Simulations

Jingbai Li[1], Patrick Reiser[2], Benjamin R. Boswell[4], André Eberhard[3], Noah Z. Burns[4*], Pascal Friederich[2,3]*, and Steven A. Lopez[1]*
[1] Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA 02115, USA
[2] Institute of Nanotechnology, Karlsruhe Institute of Technology, Karlsruhe, Germany
[3] Institute of Theoretical Informatics, Karlsruhe Institute of Technology, Karlsruhe, Germany
[4] Department of Chemistry, Stanford University, Stanford, CA, USA.
* Correspondence to nburns@stanford.edu, pascal.friederich@kit.edu, s.lopez@northeastern.edu

Table of Contents

S1. Machine learning model S2
S2. Initial training set generation S4
S3. Forces and non-adiabatic couplings S5
S4. NNs training S8
S5. Adaptive sampling S10
S6. The computational cost of ML-NAMD S13
S7. Data and code availability S13
S8. Quantum chemical calculations of *trans*-hexafluoro-2-butene S14
S9. Quantum chemical calculations of norbornyl cyclohexadiene S18
S10. Experimental details for the 4π-electrocyclic ring-closing of norbornyl cyclohexadiene. S21
S11. Cartesian coordinates of optimized geometries S26
S12. Reference S30

S1. Machine learning model

In this work, we investigate Gaussian Processes (GPs) and Deep Neural Networks (NNs) as machine learning models to fit *ab initio* quantum calculations of target properties like energies, forces, and non-adiabatic couplings (NACs). GPs can predict energies but require further modifications to predict vector quantities like forces or NACs from rotational covariant features. We initially investigated GPs but found differentiable NNs to perform better and to be the method of choice for our purpose. Nonetheless, we intend to conduct further studies on GPs in the future. We constructed multilayer feedforward NNs using TensorFlow/Keras (v2.3) API for Python.

S1.1 Gaussian Process.

Gaussian Processes (GPs) have an inherent uncertainty prediction and are therefore well suited for adaptive sampling. We trained a GP using the GPyTorch library (v1.2) and PyTorch (v1.6) for the energies using inverse distances as input. In contrast to NNs, a prediction of vector quantities from rotational invariant input features is difficult with GPs. To obtain force information, we calculated the derivative of the GP mean prediction using PyTorch's autograd module. However, we expect forces to be notably worse, so we did not modify the GP's hyperparameter optimization to incorporate forces error. For the final dataset, we obtain a mean absolute error (MAE) of 0.1 eV for energies and about 0.6 eV A^{-1} for the forces. Since the validation error for energies and forces could not outperform NNs, and GPs suffer from longer prediction times (1 s on CPU compared to 0.5–1 ms on GPU for NN), we chose NNs (see below) over GPs. We do not yet have a proper way to learn NACs with GPs in this way.

S1.2 Neural Network.

We developed a TensorFlow/Keras implementation to meet the unique requirements of non-adiabatic molecular dynamics (NAMD) simulation. As described in the main article, the machine learning kernel in PyRAI^{2}MD provides tools for active learning, forces, or NACs prediction as a derivative from learned potentials, parallel training, and further optimizations like feature standardization. The implementation details are given below.

In ***Adaptive Sampling***, we trained multiple NNs with separate weight initialization and different train-test splits, and a different set of hyperparameters for the same target property. We used the standard deviation (SD) between an ensemble of NNs (N = 2) to estimate the prediction error indicating the uncertain conformational region. The complementary NNs will generally disagree in the out-of-training areas (*i.e.*, completely unknown data categories). The SD of predicted value Y among N NNs is defined by Eq S1.

$$SD(Y) = \sqrt{\frac{1}{N-1}\sum_{j}^{N}\left(Y_j - \bar{Y}\right)^2} \#(Eq\ S1)$$

All NNs for adaptive sampling were trained in **parallel** on a CPU cluster. Within the framework, it is possible to train both on separate GPUs or in parallel on a single GPU with sufficient memory. This accelerates the training phase between trajectory explorations. Each model runs a separate training process, which is evenly distributed between cluster nodes.

Most importantly***, differentiable NNs*** are needed to predict forces for the MD simulation by taking the first-order derivative of the NN energy prediction with respect to the input coordinates. We chose a smooth activation function for hidden layers (*i.e.*, a shifted or leaky

soft plus activation). Forces and energy are jointly trained with combined mean squared error loss, proposed by Schütt *et al.*[1] The loss function is balanced by weighting the respective loss parts with coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ as Eq S2. We used $\boldsymbol{\alpha}$ = 1 and $\boldsymbol{\beta}$ = 1 in this work. The summation index *i* runs over *N* atoms and 3*N* coordinates. The ground truth $\hat{E}$ is obtained from quantum chemical calculation.

$$L = \alpha \|\hat{E} - E\|^2 + \beta \sum_{i}^{3N} \left\| \frac{\partial \hat{E}}{\partial x_i} - \frac{\partial E}{\partial x_i} \right\|^2 \#(Eq\ S2)$$

A prediction of vector quantities from rotational- and translational invariant input features is generally considered possible by modeling the force as a gradient of the NN energy prediction as Eq S3. For the NACs, we choose a similar approach motivated by Zhang *et al.*[2] Virtual atom-wise potentials are differentiated to arrive at vector quantities that are then trained without any loss restriction on the direct NN output. The TensorFlow *batch_jacobian()* method offers a fast implementation to calculate the Jacobian matrix of the model output. As a result, a single NN can predict energies for the ground state plus multiple excited states simultaneously and have each state differentiated accordingly. We speculate that this may improve model performance since the NNs can learn a more versatile underlying representation of the molecule that maps the different energy states. We expect the same improvement in the training of NACs.

$$\vec{F}_{i,j} = \frac{\partial E_i}{\partial \vec{x}_j}, i \in (States), j \in (Atoms)\ \#(Eq\ S3)$$

Furthermore, we adopt the approach of a **phase-independent (phase-less) loss** for training the NACs suggested by Westermayr *et al.*[3] We find that a preliminary training period without a phase-independent loss improves convergence. Simultaneously, the final error slightly benefits from a phase-independent loss L also for phase corrected data. With T_{ij}representing the NACs between state *i* and *j*, the ground truth $\hat{T}$ and a phase factor $\epsilon_i \in \{-1, 1\}$, the loss for *S* states is defined as Eq S4:

$$L = min_{\varepsilon_{ij}} \sum_{i,j}^{\frac{S(S-1)}{2}} || \hat{T}_{ij} - \epsilon_i \epsilon_j T_{ij} ||^2 \#(Eq\ S4)$$

In this loss definition, the relative phase factor ϵ_{ij}requires self-consistency, *i.e.*, if the relative phase between states 1, 2, and 2, 3 has been defined, the phase between 1,3 is also fixed.

The **feature descriptors** must be integrated into the model for differentiation. We use inverse distances, bond and dihedral angles as possible geometric descriptors computed from nuclear coordinates. To reduce training time, the gradients of the geometric feature representation with respect to atomic coordinates were pre-computed beforehand. This was previously exploited by Zhang *et a*l. Given potential or energy output $E = NN(f(x))$, the relation to feature description $f(x)$, and coordinates $x$is defined as Eq S5.

$$\frac{\partial E}{\partial x} = \frac{\partial NN(f)}{\partial f_i} \frac{\partial f_i}{\partial x} \#(Eq\ S5)$$

The derivative of the NN output is multiplied with the pre-computed feature gradient matrix of the form $(\partial f_i / \partial x_j)_{j=1...N}$ so that only $\partial NN / \partial f_i$ has to be computed during training. The training of NACs uses a similar way.

Weight optimization and **training** are carried out using Adam optimizer with a stepwise decrease of the initial learning rate of $1 \cdot 10^{-3}$ on validation error plateaus if no improvement

was found. For the initial exploration of MD trajectories, we prefer small NNs, which can be retrained quickly and switched to a larger model with a large number of hidden layers for the final MD population analysis.

The feedforward NN is further **regularized** by dropout and weight regularization directly inherited by the Keras API. We also find that the standardization of individual features or a general scaling of features within the differentiable model improves training.

S2. Initial training set generation

S2.1 *cis-trans* isomerization model for *trans*-1

The initial training set for the *cis-trans* isomerization model for ***trans*-1** was generated with a composite approach, including *Wigner sampling*, *Geometrical Interpolation,* and *Trajectories*. The idea is to construct a chemically intuitive sampling in the conformational space. Figure S1b illustrates the data distribution from 129 CASSCF(2,2)/cc-pVDZ trajectories of ***trans*-1**. It represents an "ideal set", including all necessary data. Our proposed composite approach allows us to efficiently generate data to resemble the "ideal set", shown in Figure S1a

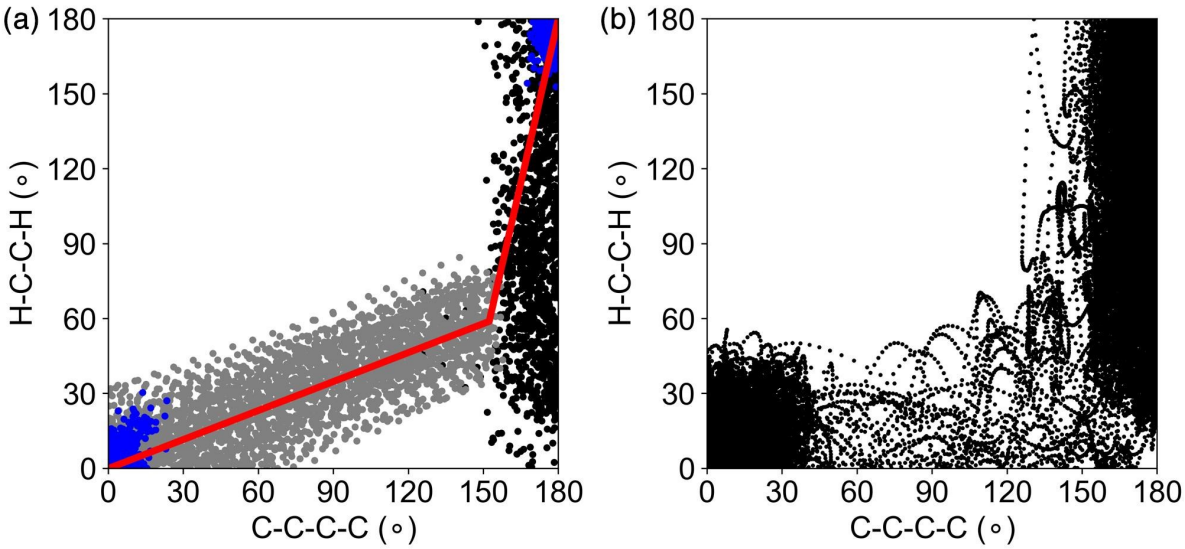


Figure S1. The spatial distribution of (a) the initial training set and (b) the 129 reference NAMD trajectories with CASSCF(2,2)/cc-pVDZ in the conformation space of **1**, defined by the dihedral angle ∠H-C-C-H and ∠C-C-C-C. In the initial training set, the Winger sampled geometries of reactant and products are in blue; the interpolated geometries and the reaction coordinate are in red; the interpolated geometries with Wigner sampled distortions are in grey; the geometries from CASSCF(2,2)/cc-pVDZ NAMD trajectories are in black.

We generated 300 geometries for both ***trans*-1** and ***cis*-1** using Wigner sampling at 300K (600 in total, blue points in Figure S1a). We interpolated 401 geometries from the optimized ***trans*-1** to ***cis*-1** via the **MECP-*trans*-1** as the middle point (red points in Figure S1a). We chose 132 out of 750 CASSCF(2,2)/cc-pVDZ NAMD trajectories, which successfully propagated more than 50 fs (~100 steps). We evenly sampled 10 geometries in the first 50 fs (1320 in total) in each of the selected trajectories. As shown in Figure S1a (black points), these data points have reached the crossing region at ∠H-C-C-H=0–60° and ∠C-C-C-C=150–180°. To enumerate the possible pathways after the crossing region, we added the Wigner sampled

geometrical distortions to the interpolated reaction coordinate diagram from the **MECP-*trans*-1** to ***cis*-1** (grey points in Figure S1a). The Wigner sampled geometrical distortions are the differences between the equilibrium geometry of ***trans*-1,** and the Wigner sampled initial conditions. We uniformly took 20 of 200 geometries in each perturbed reaction coordinate (2640 in total). Overall, the initial data set for ***trans*-1** has 4961 data points.

S2.2 4π-electrocyclic ring-closing model for 3

We used the same approach to generate the initial data for the 4π-electrocyclic ring-closing model for **3**. The underlying reactions have two distinct pathways, one bends the cyclohexadiene plane upward (***syn*-4**: θ = 120°) the other goes downward (***anti*-4**: θ = 240°). Figure S2 shows the data distribution of the initial set of **3**.

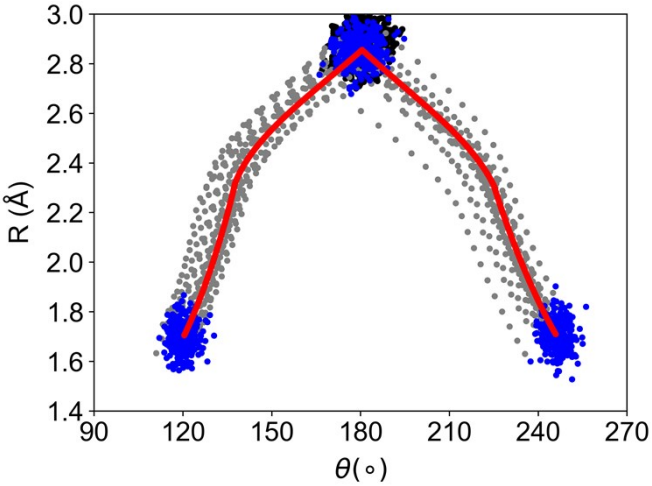


Figure S2. The spatial distribution of the initial training set of **3**, defined by the 1,4-carbon distance R and the inversion angle θ. The Winger sampled geometries of reactant and products are in blue; the interpolated geometries and the reaction coordinate are in red; the interpolated geometries with Wigner sampled distortions are in grey; the geometries from CASSCF(4,3)/ANO-S-VDZP NAMD trajectories are in black.

Wigner sampling at 300K generated 300 geometries for **3**, ***syn*-4,** and ***anti*-4** (900 in total, blue points in Figure S2). We interpolated 200 geometries from the optimized **3** to ***syn*-4** and ***anti*-4** via the **MECP-*syn*-4** and **MECP-*anti*-4** as the middle point (400 in total, red points in Figure S2). We evenly sampled 5 geometries from each of the 250 CASSCF(4,3)/ANO-S-VDZP NAMD trajectories in 50 fs simulations (1250 in total). To expand the conformational space around the interpolated reaction pathways, we added the Wigner sampled geometrical distortions (20) to the interpolated geometries (grey points in Figure S2). We uniformly took 20 of 200 geometries in each perturbed reaction pathway (800 in total). The total number of the initial data for **3** is 3349.

S3. Forces and non-adiabatic couplings

Westermayr *et al.* have recently described the particular challenge of predicting forces and NACs because the vector components are rotationally covariant (*i.e.*, they depend on the molecule's orientation).[3] Figure S3 shows a rotation of geometry associated with the covariantly revolved force vectors.

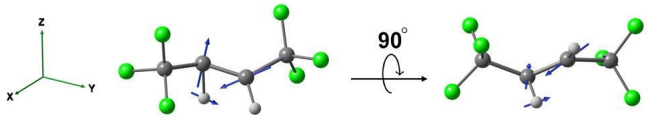


Figure S3. 3D representation of the rotational covariance of forces and NACs. The blue arrows represent the same force vectors. The geometry is rotated by 90°, leading to different x,y, and z components of force vectors in each orientation.

However, typical ML representations (*e.g.*, ML predictions are scalar) are rotationally invariant and have no information about the global orientation of the molecule. Thus, it is necessary to implement efficient and differentiable NNs that predict forces according to the first-order derivatives of the energy with respect to nuclear coordinates. For predicting NACs, Westermayr *et al.* introduced a non-physical, anti-derivative of NACs, and used its first-order derivatives to predict NACs.[3] We further tested different variants of this idea and decided to implement atom-wise virtual potentials for predicting NACs.

NACs depend on two electronic state wavefunctions (*i.e.*, state *i* and *j* in Eq S6), where the phases of CASSCF wavefunctions do not necessarily cancel out. OpenMolcas computes wavefunctions with an arbitrary phase (*i.e.*, sign). The phase change can give the opposite NACs values per Eq S7. Westermayr *et al.* have shown the uncorrected sign of NACs is undesired when training ML models.[4] Figure S4a demonstrates the frequent and random alternations of the NACs along the isomerization reaction coordinate of **1**. Figure S4b shows the phase-correction NACs.

$$d_{i,j} = \frac{\left\langle \Psi_i \middle| \frac{\partial H^{el}}{\partial R} \middle| \Psi_j \right\rangle}{E_j - E_i} \#(Eq\ S6)$$

$$\left\langle -\Psi_i \middle| \frac{\partial H^{el}}{\partial R} \middle| \Psi_i \right\rangle = -\left\langle \Psi_i \middle| \frac{\partial H^{el}}{\partial R} \middle| \Psi_i \right\rangle \#(Eq\ S7)$$

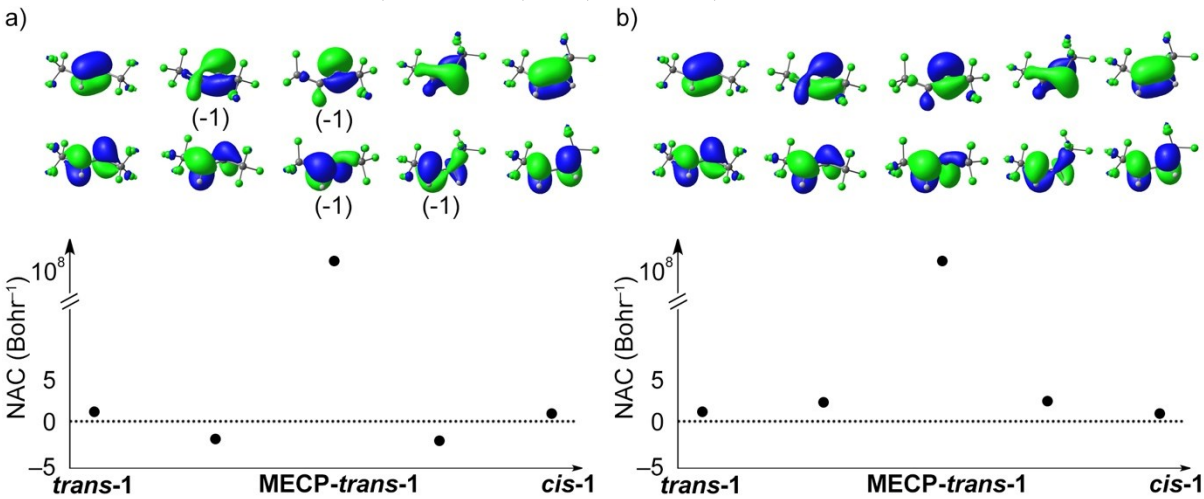


Figure S4. NAC representation using π- and π*-orbitals as a function of interpolated geometries from ***trans*-1** to **MECP-*trans*-1** to ***cis*-1**. (a) The phases are randomly assigned, resulting in an unsmooth NACs function. The –1 indicates the undesired phase changes if we do not apply corrections. (b) The phases of electronic state wavefunctions are corrected, leading to a smooth NACs function. The NACs in the equilibrium geometry of ***trans*-1** are the reference.

We employed a phase correction scheme[4-6] based on wavefunction overlap between two adjacent geometries along with the interpolated reaction path. The phase correction to the i-th state at t+Δt uses a factor $p_i(t+\Delta t)$ computed by the electronic state wavefunction overlap between two adjacent geometries, $S_i(t+\Delta t) = \langle\Phi_i(t)|\Phi_i(t+\Delta t)\rangle$. The factor $p_i(t+\Delta t) = 1$ when $S_i(t+\Delta t)$ is close to 1 and $p_i(t+\Delta t) = -1$ when $S_i(t+\Delta t)$ is close to –1. In CASSCF formalism, the electronic state wavefunction is a linear combination of all considered electronic configurations, $\Phi_k = \sum C_k\Psi_k$, where C_k is the configuration interaction (CI) coefficient. Due to the electronic configurations being orthonormal, the factor becomes the sum of multiplication of CI coefficients, $p_i(t+\Delta t) = \sum C_i(t)C_i(t+\Delta t)$. The magnitude of $p_i(t+\Delta t)$ is usually close to 1 when two states are largely separate. At the crossing region, the interstate wavefunction overlap $S_{i,j}(t+\Delta t)$ might be greater than the intrastate overlap $S_i(t+\Delta t)$ as the diabatic contributions switching. In this case, the maximum $S_{i,j}(t+\Delta t)$ will be use to track the phase.

The active orbitals are degenerate, and the order of orbitals does not affect the CASSCF calculation results. But it does change the CI coefficients. For instance, in CASSCF(2,2), switching the order of orbital φ_1 and φ_2 in the singly-excited configuration $\Psi = \varphi_{1,\alpha}\varphi_{2,\beta} - \varphi_{2,\alpha}\varphi_{1,\beta}$ results in an opposite configuration $\Psi' = \varphi_{2,\alpha}\varphi_{1,\beta} - \varphi_{1,\alpha}\varphi_{2,\beta} = -\Psi$. Therefore, phase correction based on the CI overlap requires the orbital order unchanged in the active space of adjacent geometries. To check if the orbital order flipped, we evaluated the orbital overlap $s_j(t+\Delta t) = \langle\Psi_j(t)|\Psi_j(t+\Delta t)\rangle = \sum c_j(t)c_j(t+\Delta t)$ using the natural orbital coefficients, c_j. The orbitals remaining in the same order give $s_j(t+\Delta t) \approx 1$. Otherwise, the CI coefficients must first multiply by –1 to correct the CI overlap.

The phase-corrected NACs become the reference for subsequent trajectory points. For example, we chose the converged wavefunction of the optimized ***trans*-1** as the reference geometry because it is the reactant. Figure S5 illustrates the phase correction scheme.

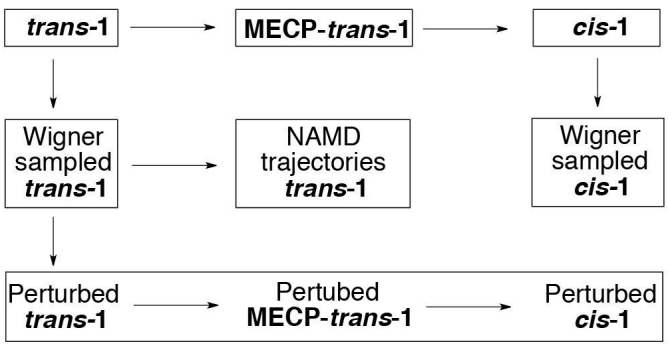


Figure S5. The phase correction scheme. The arrow represents the tracking and correcting procedures.

We first tracked the phase from ***trans*-1** to **MECP-*trans*-1,** then to ***cis*-1** along with the interpolated reaction coordinate diagram. Then, we used ***trans*-1** and ***cis*-1** to correct the phase in their Wigner sampled initial geometries, respectively, assuming the generated geometries are adjacent to the equilibrium. The corrected initial geometries become the reference for the subsequent points in the trajectories and the perturbed reaction coordinate.

The discontinuous nature of NACs near the surface crossings make NNs learning extremely difficult; instead of training with the full expression of NACs, we trained NNs with only the

numerator, known as interstate coupling.[7-10] The ML-NAMD trajectory evaluates NACs on-the-fly based on the predicted state energy differences and interstate coupling using Eq S6.

In adaptive sampling, our code automatically corrects NACs of the new geometries using the most similar geometries in the initial set as the reference by computing their root-mean-square difference RMSD of Cartesian coordinates. The RMSD depends on the atom order and the molecular alignment. Thus, we implemented the Hungarian algorithm[11] to sort and align the molecule for more accurate RMSD calculation.

In the 4π-electrocyclic ring-closing of **3**, the phase correction with (4,3) space is more challenging to manage because of the increasing combination of the orbital orders and configuration state functions. An alternative phase correction scheme directly compute the sign of the NACs using the sign of the overlap between the interstate couplings at t and t+Δt as $S_{i,j}(t+\Delta t) = \sum h_{i,j}(t)h_{i,j}(t+\Delta t)$, where $h_{i,j}$ is the normalized interstate coupling vectors between state i and j. A better choice is to do an internal phase correction in the NNs with a phase-less loss function proposed by Westermayr *et al.*[3] We only used the phase-corrected NACs in developing PyRAI^{2}MD. All results discussed in the main text are obtained with the phase-less loss in NNs.

S4. NNs training

We trained NNs on Intel(R) Xeon(R) CPU E5-2680v4@2.40GHz. We shuffled the initial training set and split them into training and validation sets in a 9:1 ratio. We performed grid-search to optimize the NNs hyperparameters. We used a learning rate scheduler to monitor the learning efficiency. The learning rate for energies and forces reduces in 2700 epochs as: 1–1500: $1\cdot10^{-3}$, 1501–2500: $1\cdot10^{-4}$, 2501–2700: $1\cdot10^{-5}$; the learning rate for NACs reduces in 1,600 epochs as : 1–900: $1\cdot10^{-3}$, 901–1400: $1\cdot10^{-4}$, 1401–1600: $1\cdot10^{-5}$. Table S1 lists the explored hyperparameters in the grid-search for the *cis-trans* isomerization of ***trans*-1** and 4π-electrocyclic ring-closing of **3**.

Table S1. The NNs hyperparameters in grid-search using the initial training set.

Hyperparameters	Values
Hidden layers	3, 4, 5, 6, 7, 8
Neurons/layer	300, 400, 500, 600, 700, 800
Batch size	64, 128
Regularization method	dropout ($5\cdot10^{-3}$), L2 ($1\cdot10^{-9}$)

The grid-search included 864 NNs and many different combinations of hyperparameters have achieved a similar accuracy. We select the optimal NNs hyperparameters based on the computation time. Table S2 and S3 collect all hyperparameters of the chosen model for the *cis-trans* isomerization of ***trans*-1** and 4π-electrocyclic ring-closing of **3**, respectively.

Table S2. The NNs hyperparameters and the mean absolute errors with R^2 of predicted energies (eV), forces (eV Å^{-1}), and interstate coupling term of NACs (eV Å^{-1}) trained on the initial set for *cis-trans* isomerization of ***trans*-1** with 4961 data points.

	Energies, Forces		NACs[a]	
Hyperparameters	NN1	NN2	NN1	NN2
Activation function	leaky soft plus		leaky soft plus	
Hidden layers	3	5	5	3
Neurons/layer	400	300	300	600
Batch size	64	128	128	128
Epochs	2700	2700	1600	1600
Learning rate $1\cdot10^{-3}$	1500	1500	900	900
Learning rate $1\cdot10^{-4}$	1000	1000	500	500
Learning rate $1\cdot10^{-5}$	200	200	200	200
Dropout	Not use	Not use	0.005	0.005
L2	$1\cdot10^{-9}$	$1\cdot10^{-9}$	Not use	Not use
MAE	0.028(0.16)[b]	0.024(0.14)[b]	0.188	0.144
R^2	0.9978(0.9324)[b]	0.9985(0.9409)[b]	0.6660	0.7865

[a]Here the NAC only represents the interstate coupling term.[b] The MAE and R^2 of forces are shown in the parenthesis.

Table S3. The NNs hyperparameters and the mean absolute errors with R^2 of predicted energies (eV), forces (eV Å^{-1}), and interstate coupling term of NACs (eV Å^{-1}) trained on the initial set for the 4π-electrocyclic ring-closing of **3** with 3349 data points.

	Energies, Forces		NACs[a]	
Hyperparameters	NN1	NN2	NN1	NN2
Activation function	leaky soft plus		leaky soft plus	
Hidden layers	3	4	3	4
Neurons/layer	400	300	500	300
Batch size	128	128	128	128
Epochs	2700	2700	1600	1600
Learning rate $1\cdot10^{-3}$	1500	1500	900	900
Learning rate $1\cdot10^{-4}$	1000	1000	500	500
Learning rate $1\cdot10^{-5}$	200	200	200	200
Dropout	Not use	Not use	0.005	0.005

L2	$1\cdot10^{-9}$	$1\cdot10^{-9}$	Not use	Not use
MAE	0.028(0.11)[b]	0.026(0.12)[b]	0.062	0.062
R^2	0.9993(0.9932)[b]	0.9995(0.9922)[b]	0.8014	0.8095

PyRAI^{2}MD trains two sets of NNs for energies+forces and NACs (4 NNs in total) in parallel with 28 CPUs, where the *cis-trans* isomerization model takes 24 minutes and the 4π-electrocyclic ring-closing model takes 32 minutes. The MAE of energies is notably below the chemical accuracy threshold (1 kcal mol^{-1}, 0.043 eV). The MAE of forces is comparable to the recently reported values of 0.10–0.15 in SO_2 and $CH_2NH_2^+$.[3]

S5. Adaptive sampling

The initial training set provided preliminary information on the conformational space. The trained NNs potential becomes less reliable when the trajectory frequently encounters the conformations outside the data set. The NNs need to learn these new data to improve the fitted potential. We used adaptive sampling[4, 12] to expand the initial training set iteratively with two sets of NNs. In each iteration, we propagate 250 trajectories from the S_1-FC points of the initial conditions sampled by Wigner distribution at 300K. In each trajectory, two sets of NNs predict the energies, forces, and NACs for the same geometry. The SD measures the prediction uncertainty. Note the energy SD is computed per state, and the SD of forces and NACs are evaluated per atom per axis. Thus, we chose the maximum SD as the prediction uncertainty for each property. We used three thresholds in energies, forces, and NACs to determine the uncertain geometry. When a trajectory completes without exceeding all thresholds, it will be marked as "converged." Otherwise, the trajectory stops and records the geometries exceeding the thresholds. This procedure helps the trajectory avoid running into the unphysical region. The collected geometry would fail the quantum chemical calculations or even poison the training data if the quantum chemical calculations return meaningless data. After collecting all uncertain geometries, the adaptive sampling workflow automatically performs the quantum chemical calculations to compute the energies, forces, and NACs explicitly and restart the NNs training, including the new data. To accelerate the adaptive sampling, we created 20 threads for trajectories propagation and distributed the single quantum chemical calculations on 56 Intel(R) Xeon(R) CPU E5-2680v4@2.40GHz. The training of four NNs sharing the 56 CPUs at the same time.

Westermayr *et al.* have suggested some thresholds of adaptive sampling.[4] We followed similar choices and set the energy threshold to 0.03 Hartree, forces and interstate coupling to 0.25 Hartree Bohr^{-1}. The QC trajectories have shown that most of the *cis-trans* isomerization of **1** completed in less than 300 fs. Thus, the adaptive sampling propagates the NNs trajectories in 500 fs. Figure S6a shows the adaptive sampling progress of the *cis-trans* isomerization of ***trans*-1** in terms of numbers of converged trajectories and new geometries found in each iteration.

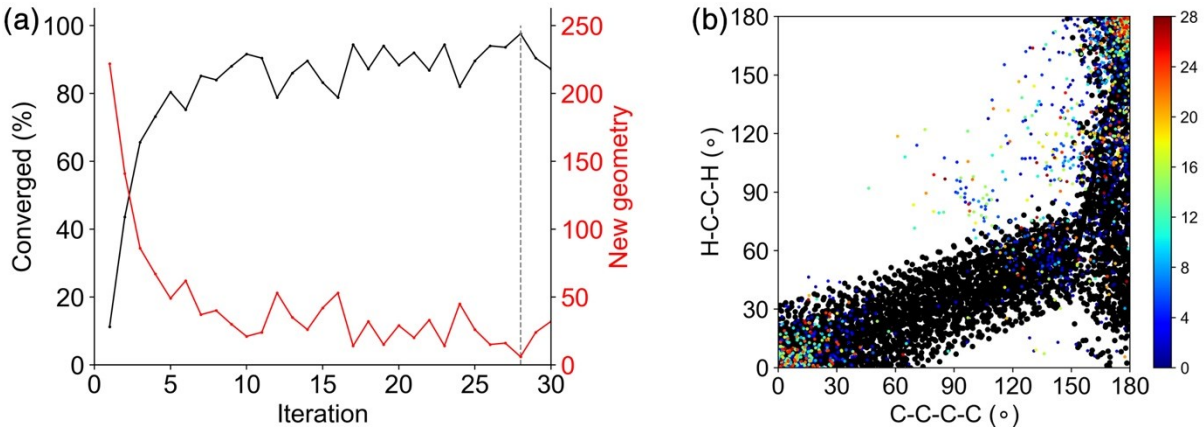


Figure S6. (a) The adaptive sampling progress of the *cis-trans* isomerization of ***trans*-1** in terms of the numbers converged trajectory (black) and new geometries found in each iteration (red). The grey dotted line marks iteration 28. (b) The geometry distribution of the initial training set and adaptively sampled data of the *cis-trans* isomerization of ***trans*-1**. The initial data points are in black.

The adaptive sampling immediately improved the NNs potential for the *cis-trans* isomerization reaction. 80% of the trajectories have converged in 5 iterations, which found 565 new geometries. Up to 28 iterations, 98% of the trajectories have converged and collected in an overall 1516 new geometries. The subsequent iterations did not further improve the NNs. Thus, we decided to use the data at iteration 28 as the final set, which remains a compact data size.

We characterized the adaptively sampled geometries to learn the NNs' perceptions in the conformational space. Figure S6b plots the geometry distribution of the initial and sampled data. At the beginning of adaptive sampling, the NNs explored a broad region in the conformational space. The sampled points were largely discrete and even approached the plot center (∠H-C-C-H = 90° and ∠C-C-C-C = 90°). As the sampling proceeds, the exploration starts to focus on the ***trans*-1** and ***cis*-1**, which suggests our initial training set has effectively sampled the essential conformational space of the photoreaction paths. A few data collected at ∠H-C-C-H = 0–30° and ∠C-C-C-C = 150–180° suggests the final set is sufficient to fit the PES at that region.

Since the 4π-electrocyclic ring-closing of **3** has a more complex excited- and ground-state PES, we have run the adaptive sampling in 100 iterations with 1000 fs trajectories. We increased the simulation temperature to 1200K to speed up the exploration on the PESs, ensuring that NNs visit the conformational space of all possible reactions. Figure S7 shows the adaptive sampling progress of the 4π-electrocyclic ring-closing of **3**.

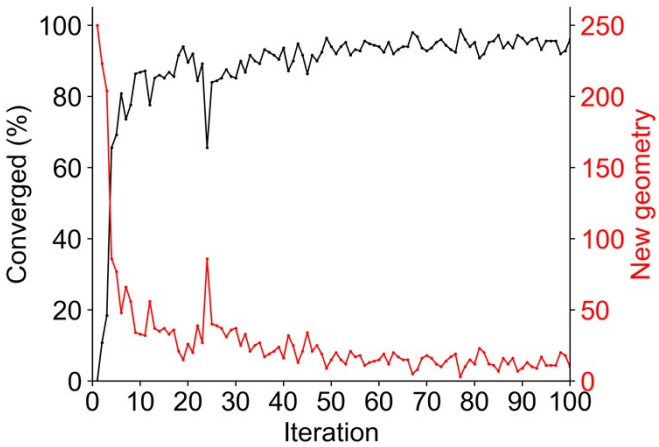


Figure S7. The adaptive sampling progress of the 4π-electrocyclic ring-closing of **3** in terms of the numbers converged trajectory (black) and new geometries found in each iteration (red).

The adaptive sampling of the 4π-electrocyclic ring-closing of **3** shows a slower improvement of the NNs potentials than that of the *cis-trans* isomerization of ***trans*-1**. The convergence first exceeded 80% after 6 iterations. However, it fluctuated in the next 18 iterations. The convergence became higher than 90% after 50 iterations and reached 96% at iteration 100. Figure S8 plots the geometry distribution of the initial and sampled data.

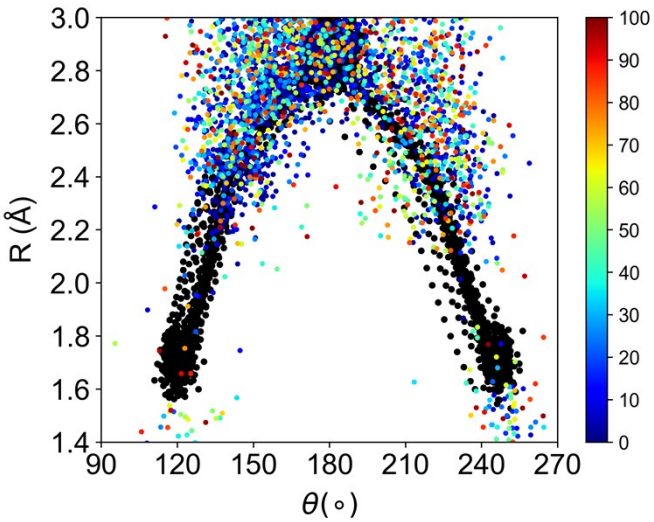


Figure S8. The geometry distribution of the initial training set and adaptively sampled data of the 4π-electrocyclic ring-closing of **3**. The initial data points are in black.

In the first 40 iterations, the adaptive sampling only focused on the regions with R > 2.2 Å, which correspond to the unexpected intermediates **3a** and **3b**. The adaptive sampling turned to visit the conformational space around ***syn*-4** (R < 1.6 Å and θ < 150°) after 30 iterations. The relatively slow progress suggests the low yield of the ring-closing products even at 1200K. We choose the data at iteration 100 as the final set because the NNs have sufficiently explored the conformational space of all possible reactions.

S6. The computational cost of ML-NAMD

The ML-NAMD simulations of the *cis-trans* isomerization of ***trans*-1** and the 4π-electrocyclic ring-closing of **3** used 1 Intel(R) Xeon(R) E5-2680v4@2.40GHz CPU per trajectory. Table S4 collects the wall time of energies, forces, and NACs calculations in one step and the entire trajectory.

Table S4. The computational cost of energies, forces, and NACs in a single point calculation and a trajectory by quantum chemical and neural networks.

	Energies, Forces, and NACs, seconds	Trajectory, seconds
$QC_{cis\text{-}trans}$	335.763[a]	143331.733[b]
$ML_{cis\text{-}trans}$	0.00991[c]	22.775(12.642)[d]
$QC_{ring\text{-}closing}$	2497.999[e]	1436077.054[f]
$ML_{ring\text{-}closing}$	0.0118[g]	56.027 (37.598)[h]

[a]401 CASSCF(2,2)/cc-pVDZ single point calculations. [b]243 trajectories with CASSCF(2,2)/cc-pVDZ in 500 fs. [c]1000 NNs predictions. [d]5573 NNs FSSH trajectories and 5820 NNs ZNSH trajectories (shown in parathesis) in 500 fs. [e]3350 CASSCF(4,3)/ANO-S-VDZP single point calculations. [f]240 trajectories with CASSCF(4,3)/ANO-S-VDZP in 1000 fs. [g]1000 NNs predictions. [h]3910 NNs FSSH trajectories and 3954 NNs ZNSH trajectories (shown in parathesis) in 1000 fs.

The 10 ns ML-NAMD simulations records concise information every 0.5 fs in the first 1 ps. After that, it only outputs every 50 fs for saving disk space. The average cost using the FSSH method over 31 trajectories is 86.2 hours (310318.194 seconds). When using the ZNSH method, the average cost over 89 trajectories is 50.0 hours (179913.449 seconds).

The 1 ns ML-NAMD simulation of the 4π- ring-closing of **3** records the trajectories every 0.5 fs in the first 1 ps, and then it only checkpointed the trajectories every 1 ps in the 1–1000 ps. The average cost of the 1 ns ML-NAMD simulations over 200 FSSH trajectories is 10.9 hours (39361.350 seconds). For comparison, the 1 ns ML-NAMD simulations over 984 ZNSH trajectories spent 4.8 hours (17151.537 seconds).

S7. Data and code availability

PyRAI^{2}MD uses the JSON format to store data. We have converted the training data in plain text and saved them in cis-trans-data6207.txt and ring-closing-data6267.txt. The data includes Cartesian coordinates of nuclear positions in Angstrom, electronic energies for S_0 and S_1 state in Hartree, forces for S_0 and S_1 state in Hartree Bohr^{-1}, and interstate coupling term of NACs in Hartree Bohr^{-1}. We recorded a video for the 10 ns ML-NAMD simulations of the *trans*-to-*cis* reaction of ***trans*-1**, named trans-to-cis-10ns-720p.mp4. The video is fast-forwarded in 10 fs per step in the 0–1000 fs, 1 ps per step in the 1–1000 ps, and then 10 ps per step in the 1–10 ns. The PyRAI^{2}MD code is open-source and free of charge for non-commercial scientific research. The NAMD kernel is available on GitHub https://github.com/lopez-lab/PyRAI2MD. The ML kernel is available upon request; please contact Pascal Friederich (pascal.friederich@kit.edu).

S8. Quantum chemical calculations of *trans*-hexafluoro-2-butene

S8.1 CASSCF calculations

We prepare the reference quantum chemical calculations for ***trans*-1**, ***cis*-1**, **MECP-*trans*-1**, and **2** using the complete active space self-consistent field (CASSCF) theory implemented in OpenMolcas 19.11.[13] We chose the cc-pVDZ basis set[14] for all atoms. We did not use the diffuse functions for two reasons. 1) the S_1 state in ***trans*-1** and ***cis*-1** are mainly $\pi\pi^*$ valence excited-states. 2) the diffuse functions are overwhelmingly expensive as we have performed 67133 single points calculations to prepare the smoothly continued training data for NACs phase correction. We constructed a (2,2) active space with 2 π-electrons and 2 π-type orbitals from the C=C bond. Figure S9 shows the converged orbitals in ***trans*-1** and ***cis*-1**.

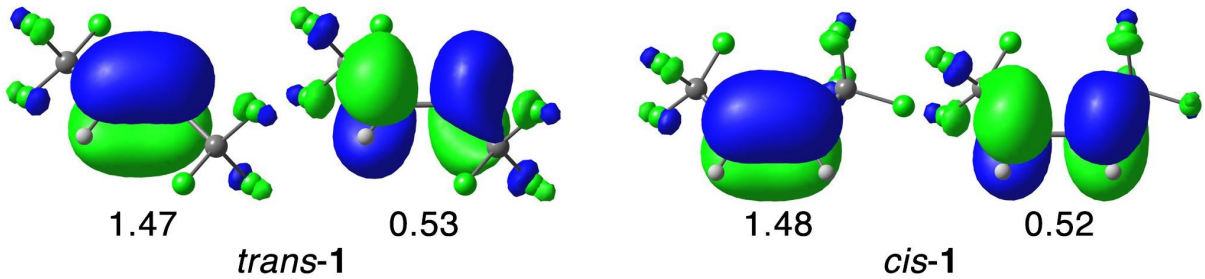


Figure S9. The (2,2) active space of ***trans*-1** and ***cis*-1** with CASSCF(2,2)/cc-pVDZ. The occupation numbers are averaged over two states. Isosurface value: 0.03.

We optimized the geometries of ***trans*-1** and ***cis*-1** with CASSCF(2,2)/cc-pVDZ. Figure S10 shows the geometries of ***trans*-1** and ***cis*-1**.

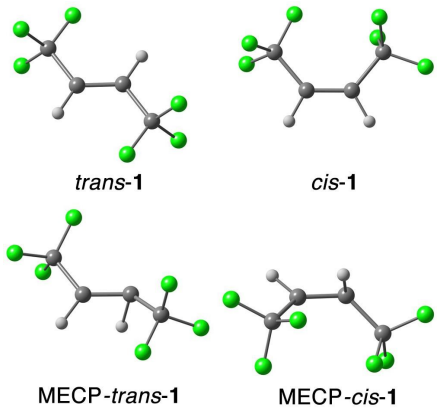


Figure S10. The equilibrium geometries of ***trans*-1** and ***cis*-1** and the minimum energy crossing points **MECP-*trans*-1** and **MECP-*cis*-1**, optimized with the CASSCF(2,2)/cc-pVDZ. Atom colors are white: H; grey: C; green: F.

The geometries are the local minimum, as their vibrational frequencies are positive with the CASSCF(2,2)/cc-pVDZ. From the S_1 Franck-Condon point of ***trans*-1** and ***cis*-1**, we performed minimum energy path (MEP) calculations to search for possible S_1 minimum. Instead of stationary points, we have located two distinct MECPs, **MECP-*trans*-1,** and **MECP-*cis*-1** at the end of the MEP. The C-H, C-C, C=C bonds in ***trans*-1** and ***cis*-1** are 1.08, 1.50, and 1.34 Å. The ∠H-C-C-H and ∠C-C-C-C angles are 180° and 0° in ***trans*-1** and ***cis*-1**, respectively. In **MECP-*trans*-1** and **MECP-*cis*-1**, the C=C bond increased to 1.38 Å and showed a pyramidalized C-H bond. The resulting angle of ∠H-C-C-H became 59° and 109° in **MECP-**

***trans*-1** and **MECP-*cis*-1**, respectively. The ∠C-C-C-C angles are 152° and 38° in **MECP-*trans*-1** and **MECP-*cis*-1**, which are less changed than ∠H-C-C-H angles.

S8.2 Comparison between CASSCF and CASPT2

The CASSCF calculation often overestimates the vertical excitation energies due to the lack of dynamic electron correlations. To reveal the influence of dynamic electron correlations on the PESs of the *cis-trans* isomerization of ***trans*-1**, we compared the CASSCF(2,2)/cc-pVDZ and CASPT2(2,2)/cc-pVDZ using the interpolated reaction coordinate diagram from ***trans*-1** to ***cis*-1** via **MECP-*trans*-1**, shown in Figure S11.

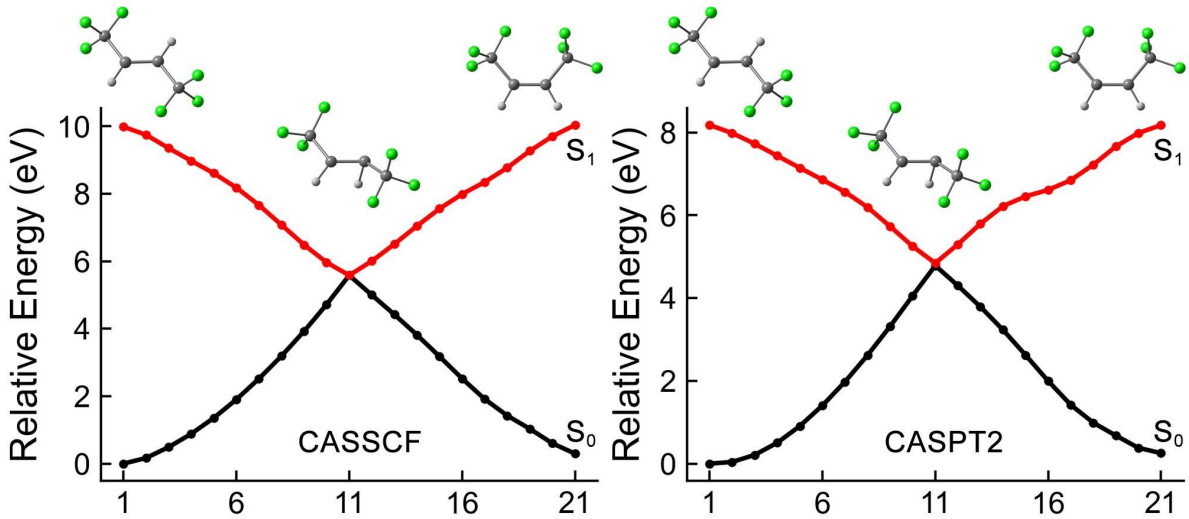


Figure S11. The reaction coordinate diagram from ***trans*-1** to ***cis*-1** via **MECP-*trans*-1** with the CASSCF(2,2)/cc-pVDZ and CASPT2(2,2)/cc-pVDZ. The S_1 and S_0 energies are shown in red and black.

The reaction coordinates are interpolated based on the optimized geometries of ***trans*-1** and ***cis*-1**, and **MECP-*trans*-1** by linearly varying the Z-matrix parameters with equal-spaced increments. Figure S11 shows identical PES topology with CASSCF(2,2)/cc-pVDZ and CASPT2(2,2)/cc-pVDZ. Thus, we expect that the CASSCF calculations are reliable for computing the training data and reference NAMD trajectories of the *cis-trans* isomerization of ***trans*-1**.

S8.3 The influence of the NVE ensemble on the NAMD trajectories of *trans*-1

In the NAMD simulation, we used the thermostat to equilibrate the trajectories at a constant temperature (300K). Here we want to address the importance of the thermostat on obtaining physically meaningful dynamics results, especially CASSCF dynamics. Without a thermostat, the trajectories conserved the total energy representing an NVE ensemble. It is reasonable to describe the excited-state dynamics of simple molecules in a short simulation time, where one can consider the energy transfer has not happened. However, the overestimated excitation energy of CASSCF trajectory would result in excess kinetic energy when the NVE trajectory lands on the ground-state, overcoming the thermal-forbidden barrier. To justify this, we obtained 604 NAMD trajectories for ***trans*-1** in NVE with CASSCF(2,2)/cc-pVDZ. Figure S12 plots the trajectories of the *trans*→*cis* isomerization and *trans*→*trans* reversion of ***trans*-1**. The ∠C–C–C–C and ∠H–C–C–H dihedral angles are tracked during the 500 fs simulations. The angle changes accumulated in the consecutive clockwise or anti-clockwise rotations.

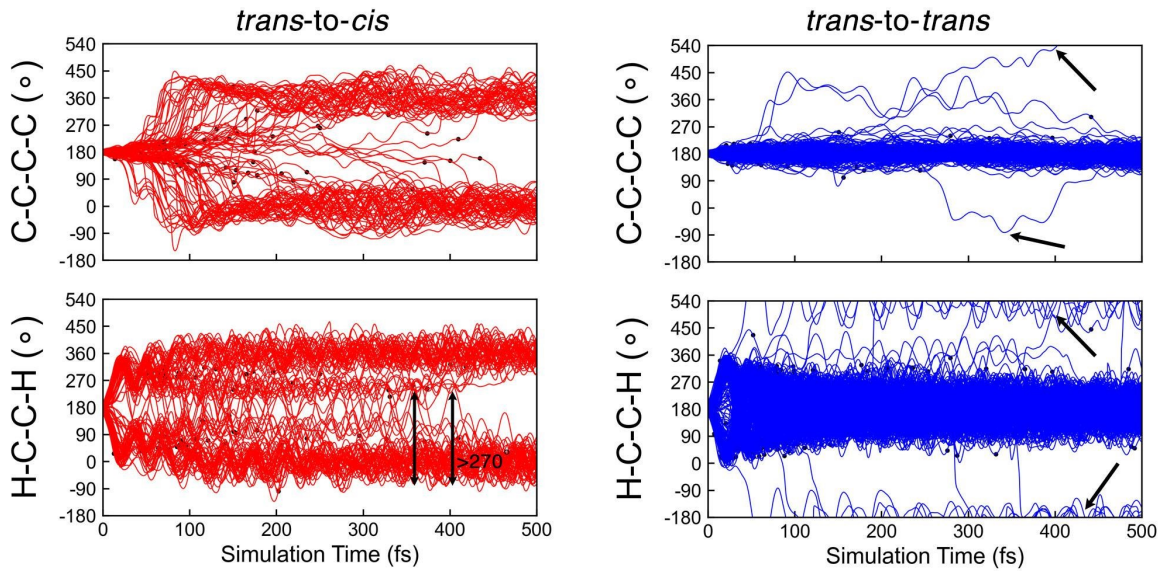


Figure S12. The ∠C–C–C–C and ∠H–C–C–H dihedral angles in 604 NVE trajectories in 500 fs simulations with CASSCF(2,2)/cc-pVDZ. The angle changes accumulated in the consecutive clockwise or anti-clockwise rotations.

In the *trans*→*cis* isomerization, the ∠C–C–C–C angle vibration is similar to that with a thermostat; however, the ∠H–C–C–H angle shows stronger vibrations with increased kinetic energy. We noted a few trajectories show accumulated ∠H–C–C–H rotations larger than 270°, which seems not very physical (one would expect a rotation about 180° in the *cis-trans* isomerization). We observed a 360° CF_3 group rotation in a *trans*→*trans* trajectory; the ∠C–C–C–C angle changed from 180° to 540° (marked with arrows). In another trajectory, the ∠C–C–C–C angle first decreased to 0° then returned to 180° (marked with arrows). These suggest the kinetic energy in the NVE trajectories unphysically broke the C=C bond in the ground-state. We can see a stronger effect on the ∠H–C–C–H vibration that more trajectories showed a 360° rotation of C–H bonds.

We expect the overestimation of kinetic energy would be smaller at a higher level of theory (*e.g.*, XMS-CASPT2). Thus, it might not be a big issue in the NVE ensemble. In general, using a thermostat can avoid the excess or overestimated kinetic energy, especially for the ground-state portion of the NAMD trajectories.

S8.4 Additional information about CASSCF(2,2)/cc-pVDZ trajectories of *trans*-1

Here we summarize some additional results from the CASSCF(2,2)/cc-pVDZ trajectories of ***trans*-1**. We plot the number of hops during the simulation time to compare the QC and NNs predicted surface hopping events (Figure S13).

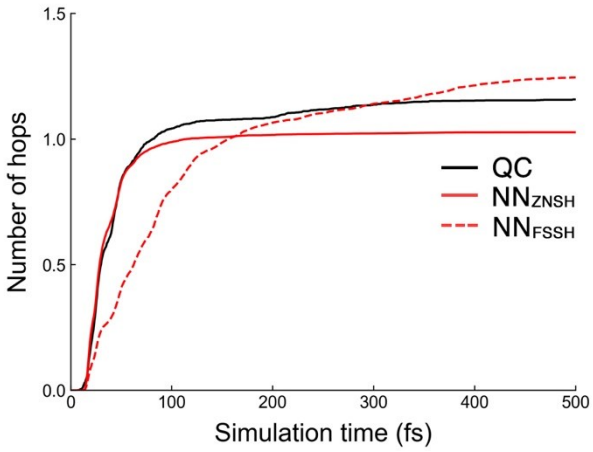


Figure S13. The average number of hops of the QC and NNs trajectories of ***trans*-1** over 500 fs simulation at 300K. The curves average over 1371 trajectories with CASSCF(2,2)/cc-pVDZ, 5573 NNs FSSH trajectories, and 5820 NNs ZNSH trajectories.

The QC and NNs ZNSH trajectories show similar surface hopping rate as they have close S_1 half-life time. Because the ZNSH requires a local minimum of the state energy gap under the threshold (0.5 eV), it tends to avoid the surface hopping if the degenerate region does not satisfy the condition, thus most of ZNSH trajectories undergo a single hop to the ground-state rather than being hopping back and forth between the states. The final average number of ZNSH is 1.03, slightly smaller than the QC trajectories (1.16). The NNs FSSH trajectories show a notably smaller slope due to the overestimated S_1 lifetime. The errors in predicted NACs result in the more frequent surface hopping, thus leading to a larger number of hops (1.25).

To compare the NNs predicted geometrical changes with the CASSCF(2,2)/cc-pVDZ dynamics, we computed the average value of the ∠C–C–C–C and ∠H–C–C–H dihedral angles during the simulation, shown in Figure S14.

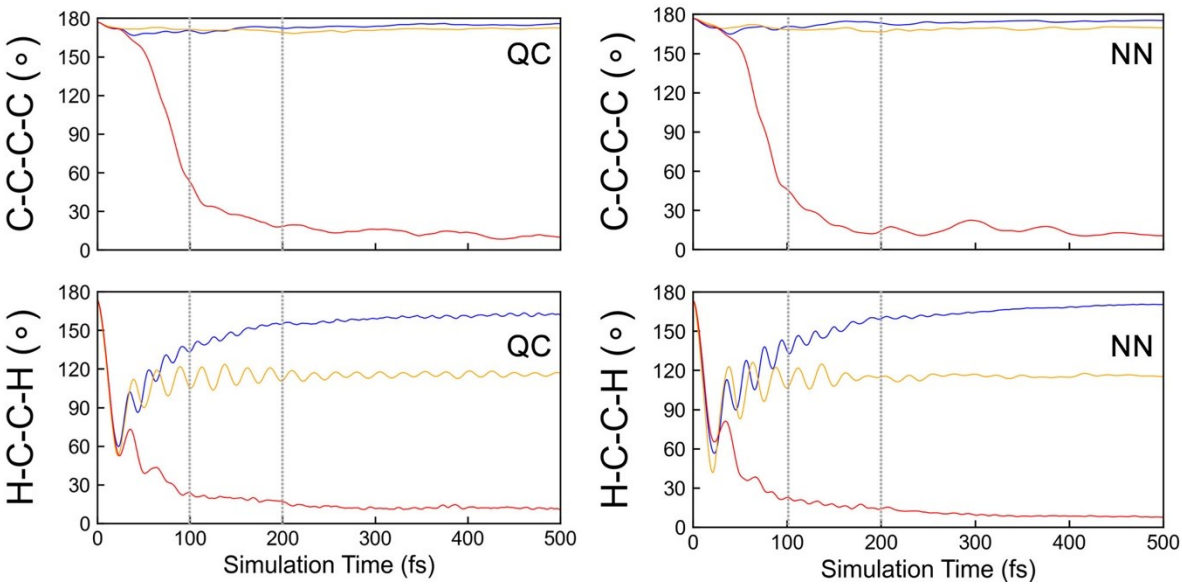


Figure S14. The expectation values of the ∠C–C–C–C and ∠H–C–C–H dihedral angles over time by averaging 1371 QC trajectories and 5820 NNs ZNSH trajectories. The plots color-coded the *trans*→*cis*, *trans*→*carbene*, and *trans*→*trans* pathways in red, yellow, and blue.

The NNs trajectories demonstrate similar geometrical changes to the QC trajectories in all three pathways in the 500 fs simulation, which agree with the individual trajectories plot. Specifically, the NNs trajectories show a slightly faster reduction of the ∠C–C–C–C dihedral than QC trajectories at 100 fs. After that, the NNs and QC trajectories show very similar angles.

S9. Quantum chemical calculations of norbornyl cyclohexadiene

S9.1 CASSCF calculations

The CASSCF calculations for **3**, **MECP-*syn*-3**, **MECP-*anti*-3, 3a**, **3b**, ***syn*-4**, and ***anti*-4** used the ANO-S-VDZP basis set.[15-18] This basis set has shown good performance in our previous study on the 4π-electrocyclic ring-closing reaction of another similar compound. Since we used the phase-less loss to train the NNs, we only performed 3349 single point calculations to generate the initial training data. Thus, the ANO-S-VDZP with diffuse functions does not make the training data calculation a lot more expensive. We chose an active space of 4 π-electrons and 3 π-type orbitals. We removed the π^*-orbital in A_2 symmetry to ensure consistent CASSCF state ordering with CASPT2 calculations. Figure S15 illustrates the natural orbitals and averaged occupation of **3**.

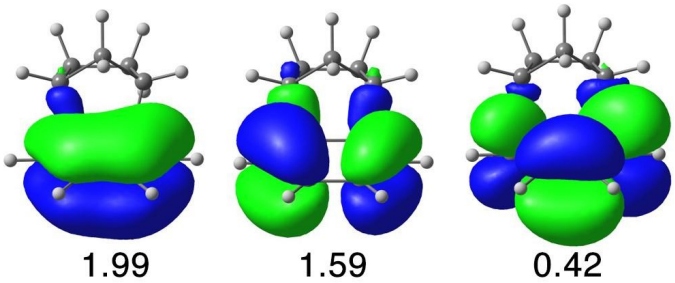


Figure S15. The (4,3) active space of **3** with CASSCF(4,3)/ANO-S-VDZP. The occupation numbers are averaged over two states. Isosurface value: 0.03.

We optimized the geometries of **3, MECP-*syn*-3**, **MECP-*anti*-3, 3a**, **3b**, ***syn*-4***, and ***anti*-4*** with CASSCF(4,3)/ANO-S-VDZP, shown in Figure S16.

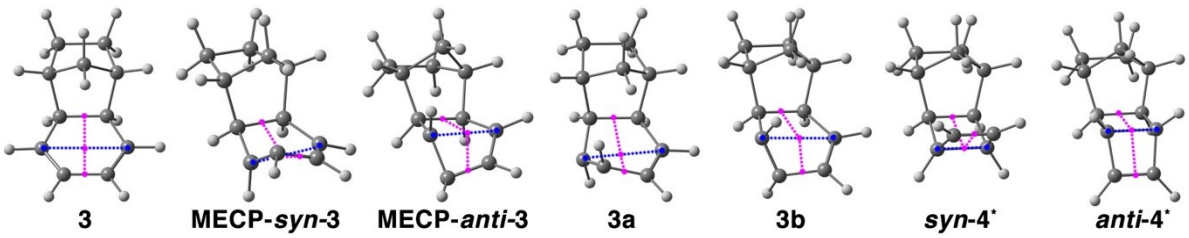


Figure S16. The optimized geometries of **3, MECP-*syn*-3**, **MECP-*anti*-3, 3a**, **3b**, ***syn*-4'**, and ***anti*-4'** with CASSCF(4,3)/ANO-S-VDZP. The blue and pink dotted lines highlight the distance R between the carbon atoms closing the cyclohexadiene ring and the bending angle θ of the cyclohexadiene plane. θ = 0–180° represents a *syn*-configuration and θ = 180–360° represents an *anti*-configuration. ***syn*-4***, and ***anti*-4*** are optimized with R = 1.7 Å. Atom colors are white: H; grey: C.

The geometry of **3** is optimized to a local minimum (R = 2.86 Å, θ=0°), as the vibrational frequencies are positive with the CASSCF(2,2)/ANO-S-VDZP. From the S_1 Franck-Condon point, we slightly perturb **3** toward ***syn*-4** and ***anti*-4** to search possible S_1 minima and MECPs along with the MEP. The last MEP geometries are then optimized to **MECP-*syn*-3** (R = 2.32 Å, θ=138°)**,** and **MECP-*anti*-3** (R = 2.31 Å, θ=225°). The intermediate **3a** (R = 2.60 Å, θ=142°) and **3b** (R = 2.61 Å, θ=220°) are optimized from the non-equilibrium geometry in the NAMD simulations. The frequency calculations confirm them as true minima without imaginary frequency. The newly formed σ-bond in ***syn*-4** and ***anti*-4** become less active as the occupation is getting closer to 2.00. Consequently, the CASSCF(4,3)/ANO-S-VDZP geometry optimizations did not converge well. We constrain the σ-bond with R = 1.7 Å to optimize nearby geometries, ***syn*-4*** and ***anti*-4*****,** and use them in the Wigner sampling.

S9.2 Comparison between CASSCF and CASPT2

We have compared the CASSCF(4,3)/ANO-S-VDZP and CASPT2(4,3)/ANO-S-VDZP to determine the reliability of the CASSCF dynamics for the 4π-electrocyclic ring-closing of **3**. The reaction coordinate diagram from **3** to ***syn*-4*** and ***anti*-4*** are shown in Figure S17.

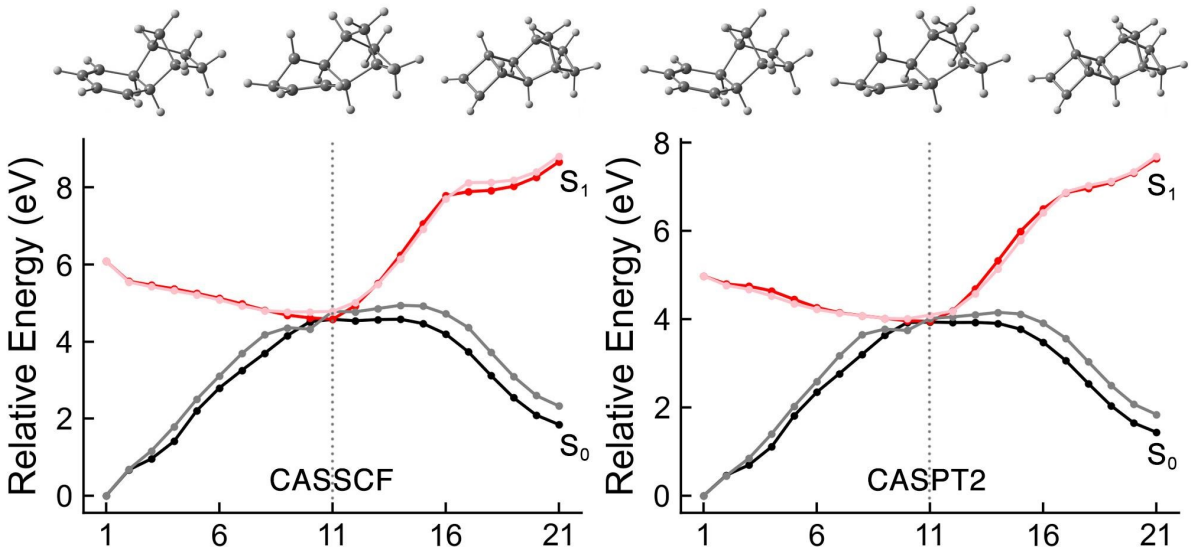


Figure S17. The reaction coordinate diagram from **3** to ***syn*-4'** and ***anti*-4'** via **MECP-*syn*-3** and **MECP-*anti*-3** with the CASSCF(4,3)/ANO-S-VDZP and CASPT2(2,2)/ANO-S-VDZP. The S_1 and S_0 energies for the syn-pathway are shown in pink and grey; the S_1 and S_0 energies for the syn-pathway are shown in red and black.

The CASSCF(4,3)/ANO-S-VDZP and CASPT2(4,3)/ANO-S-VDZP calculations consistently show the *syn*-pathway is slightly steeper descent than the *anti*-pathway from the S_1-FC point. When it is close to the S_1/S_0 intersection (Step 11), the *anti*-pathway has lower energy than the *syn*-pathway, leading to a 0.20 eV lower crossing point. Beyond the intersection, both methods agree that the S_0 PES of the *anti*-pathway is flat while the *syn*-pathway shows a barrier. Thus, we would expect consistent photodynamics of the 4π-electrocyclic ring-closing of **3** with CASSCF and CASPT2.

S9.3 Additional information about CASSCF(4,3)/ANO-S-VDZP trajectories of 3

Here we summarize some additional results from the CASSCF(4,3)/ANO-S-VDZP trajectories of **3**. We plot the number of hops during the simulation time in Figure S19 to compare the QC and NNs predicted surface hopping events

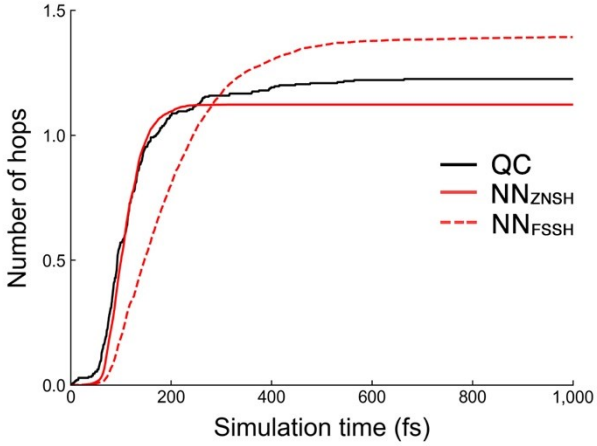


Figure S18. The average number of hops of the QC and NNs trajectories of **3** during 1000 fs simulation at 300K. The curves average over 240 trajectories with CASSCF(4,3)/ANO-S-VDZP, 3910 NNs FSSH trajectories, and 3954 NNs ZNSH trajectories.

We observed similar surface hopping rate in the QC and NNs ZNSH trajectories since their S_1 half-life are close. As we have discussed in Section **S8.4**, the NNs ZNSH trajectories predict less frequent surface hopping (1.12) than the QC trajectories (1.23). The number of hops in the NNs FSSH trajectories is largely deviated from the QC trajectories. The surface hopping rate is slower due to the overestimated S_1 lifetime. Because of the errors in the predicted NACs, the NNs FSSH trajectories show frequent back and forward hops, thus resulting in larger number of hops (1.39) than the QC trajectories.

The NAMD simulations for the 4π-electrocyclic ring-closing of **3** predicted low-yields of ***syn*-4** and ***anti*-4**. The majority of the trajectories reverted to **3** after surface hopping in both QC calculations and NNs predictions. Figure S19 plots the non-productive trajectories obtained from the 1000 fs simulations with CASSCF(4,3)/ANO-S-VDZP and NNs.

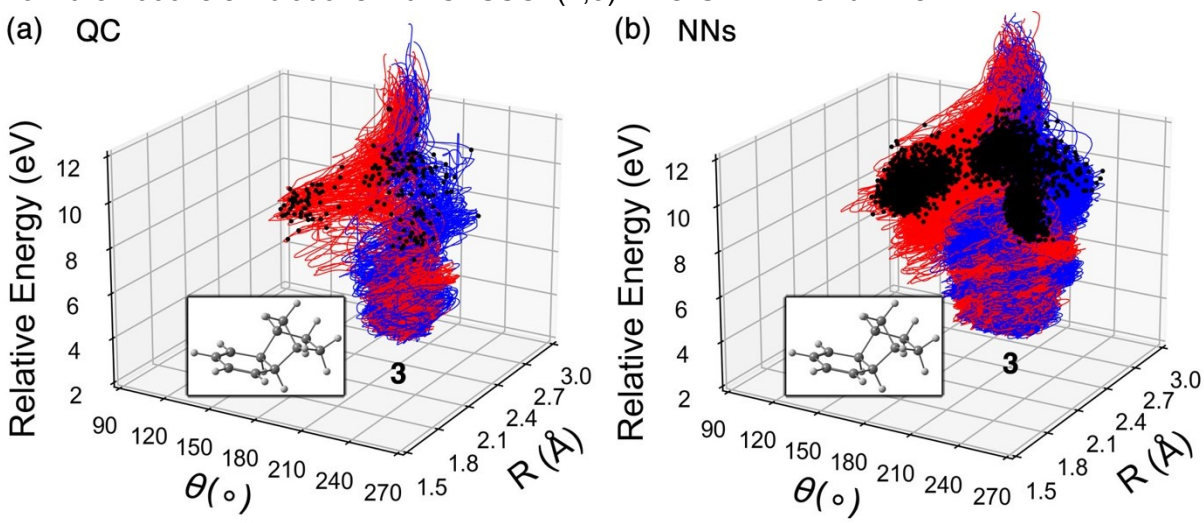


Figure S19. The non-productive NAMD trajectories of electrocyclic ring-closing of **3** computed with (a) CASSCF(4,3)/ANO-S-VDZP and (b) NNs in 1000 fs simulations at 300K. The black dots represent the last surface hopping point in each trajectory.

Figure S19a shows the 80% QC trajectories bifurcated from the S_1-FC regions of **3** to the *syn*- and *anti*-pathway. 44% of the trajectories moved along with the *syn*-pathway, and 35% of the trajectories went to the *anti*-pathway. The NNs trajectories in Figure S19b agree that the *syn*-pathway is preferred (*syn*: 53% *vs. anti*: 31%).

We then compare the QC and NNs trajectories by plotting the yield of **3a**, **3b**, ***syn*-4**, and ***anti*-4** during the 1 ps NAMD simulation at 300K, shown in Figure S20.

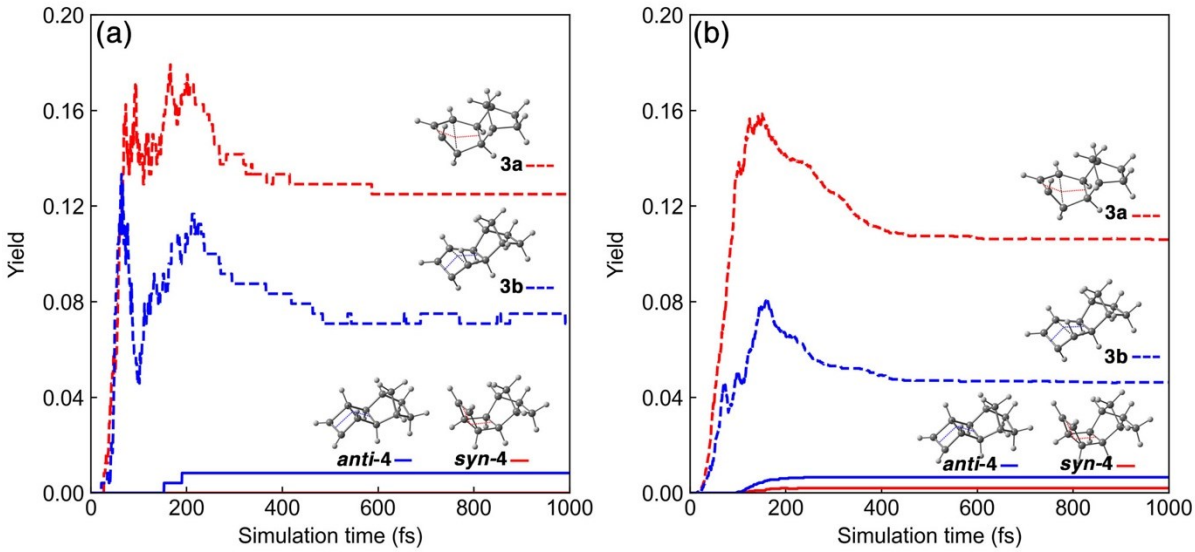


Figure S20. (a) QC and (b) NNs predicted yield of **3a** (dashed red), **3b** (dashed blue), ***syn*-4** (solid red) and ***anti*-4** (solid blue) in 1 ps NAMD simulation at 300K for **3**.

The QC trajectories (Figure S20a) show quick conversions from **3a** and **3b** to **3** in the first 100 fs. The yields of **3a** and **3b** reach the first maximum value then dramatically drop around 100 fs. The second peak of the yield curve of **3a** and **3b** appears at 200 fs, then gradually reduce in 200–500 fs. After that, the yields are nearly unchanged to the end of simulation. The formation of ***anti*-4** starts around 150 fs. The NNs trajectories show rather smooth curves in because of more trajectories are included (Figure S20b). The NNs predicted yields of **3a** and **3b** show similar topology to QC trajectories, which approach the maximum values around 150 fs and reduce to steady values after 500 fs. The formation of ***syn*-4** and ***anti*-4** begin at 100 fs. The yield curves of ***syn*-4** and ***anti*-4** in the NNs trajectories suggest the ring-closing reactions complete at 200 fs since the yield has not changed in the rest of simulation time, which agree with the QC trajectories.

We zoom in the 100-200 fs region to compare the yield of ***syn*-4** and ***anti*-4** in NNs trajectories and the experiment. The yields are normalized with respect to ***anti*-4** and the simulation and experiment time windows are rescaled to 0–1, shown in Figure S21. The detail experiment data are collected in Figure S23.

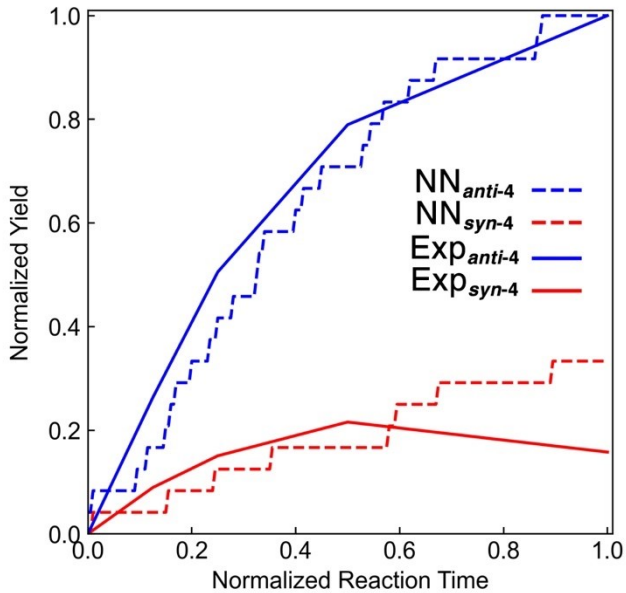


Figure S21. Comparison of the yield of ***syn*-4** and ***anti*-4** in the NNs trajectories and experiment. The simulation time 100–200 fs and the experiment time 0–4 hours are normalized to 0–1. The predicted and measured yields are normalized with respect to ***anti*-4** ranging from 0 to 1.

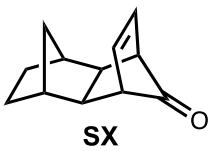
S10. Experimental details for the 4π-electrocyclic ring-closing of norbornyl cyclohexadiene.

S10.1 Synthetic methods

All reagents were obtained from commercial vendors and used as received unless otherwise noted. Dry solvent was collected after passing through a bed of activated alumina in a JC Meyer Solvent System. Flash column chromatography was performed using F60 silica gel (40-63 μm, 230-400 mesh, 60Å) purchased from Silicycle. Analytical thin-layer chromatography (TLC) was carried out on 250 μm 60-F254 silica gel plates purchased from EMD Millipore, and visualization was affected by observation of fluorescence-quenching with ultraviolet light and staining with potassium permanganate as a developing agent.

Solution-state ^{1}H NMR and ^{13}C NMR were recorded on Varian Inova 600, Varian Inova 500, or Varian Mercury 400 spectrometers operating respectively at 600, 500, and 400 MHz for ^{1}H and at 126, and 101 MHz for ^{13}C. Chemical shifts are reported in parts per million (ppm) relative to residual protonated solvent for ^{1}H ($CHCl_3$ = δ 7.26, CD_3CN = δ 1.94) and relative to carbon resonances of the solvent for ^{13}C ($CDCl_3$ = δ 77.16, CD_3CN = δ 118.26). Peak multiplicities are annotated as follows: br = broad, s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet.

S10.2 Materials



SX

Ketone **SX** was prepared from norbornene according to a modified protocol by McCulloch and coworkers.[19] After silica gel column chromatography (3:7 to 6:4 dichloromethane/hexanes), the NMR spectra and melting point of **SX** agreed with the literature. For convenience, we have included the spectral data in $CDCl_3$ below.

Physical properties: white solid, mp = 43 – 44 °C;

R_f = 0.35 (silica gel, 4:6 dichloromethane/hexanes, visualized with UV and permanganate stain);

^{1}H NMR (500 MHz, $CDCl_3$): δ 6.36 (t, J = 2.5 Hz, 2H), 3.03 (p, J = 2.4 Hz, 2H), 2.25 (s, 2H), 2.17 (t, J = 2.0 Hz, 2H), 2.06 (dt, J = 10.4, 2.3 Hz, 1H), 1.44 (dt, J = 7.6, 2.4 Hz, 2H), 1.13 (dd, J = 7.3, 2.5 Hz, 2H), 0.72 (d, J = 10.3 Hz, 1H);

^{13}C NMR (126 MHz, $CDCl_3$): δ 201.3, 130.8, 51.7, 43.0, 40.0, 35.8, 31.9.

3

A solution of ketone **SX** (38 mg, 0.22 mmol, 1.0 equiv.) in hexanes (2.0 mL, 0.11 M) in a microwave tube at room temperature under an atmosphere of nitrogen was sealed and heated to 73 °C for 20 hours. The reaction was allowed to cool to room temperature and the colorless solution was passed through a plug of silica gel. The eluted solution was concentrated by rotary evaporation and then high vacuum for 2 mins to yield **3** as a colorless oil (23 mg, 72%). The NMR spectra of **3** agreed with the literature.[19] For comparison later, we have included the spectral data in CD_3CN below.

Physical properties: colorless oil;

R_f = 0.73 (silica gel, hexanes, visualized with UV and permanganate stain);

^{1}H NMR (500 MHz, CD_3CN): δ 5.59 – 5.53 (m, 2H), 5.43 – 5.39 (m, 2H), 2.48 (s, 2H), 2.02 (t, *J* = 2.3 Hz, 2H), 1.71 (dd, *J* = 9.5, 2.1 Hz, 1H), 1.55 (d, *J* = 7.4 Hz, 2H), 1.36 (dd, *J* = 7.2, 2.2 Hz, 2H), 1.25 (dt, *J* = 9.6, 1.6 Hz, 1H);

^{13}C NMR (126 MHz, CD_3CN): δ 129.2, 122.0, 46.8, 43.8, 35.2, 30.9.

***anti*-4**

anti-**4** was obtained as the major product of and inseparable mixture of isomers after irradiation of **3** (see below for conditions).

Physical properties:

^{1}H NMR (600 MHz, CD_3CN): δ 6.17 (d, *J* = 1.6 Hz, 2H), 2.78 (s, 2H), 2.09 (s, 2H), 1.91 (s, 2H), 1.76 (dt, *J* = 10.0, 2.2 Hz, 1H), 1.46 (dt, *J* = 7.2, 2.4 Hz, 2H), 1.15 (d, *J* = 10.0 Hz, 1H), 1.05 (dd, *J* = 7.2, 2.3 Hz, 2H).

^{3}C NMR (126 MHz, CD_3CN): δ 140.3, 48.5, 45.7, 40.2, 33.1, 28.6.

3'

3' was prepared according to a protocol by Rye and Wege.[20] **3** (5.0 mg, mmol, 1 equiv.) was sealed in a pyrex tube at 1 mmHg. The tube was then heated in an oven at 300 ^{o}C for 1 hour, after which time the tube was removed and allowed to cool to room temperature. The tube was opened, and deuterated acetonitrile was used to rinse the contents and a ^{1}H NMR spectrum in CD_3CN determined that both **3** as well as **3'** were present in a ratio of 3.6 to 1. These two isomers were found to be inseparable by chromatography, and thus the annotated ^{1}H NMR spectrum of **3'** was obtained from the mixture and agreed with the literature.[20] For comparison later, we have included the spectral data in CD_3CN below.

Physical properties:

^{1}H NMR (600 MHz, CD_3CN): 5.74 (d, J = 9.6 Hz, 2H), 5.36 (d, J = 9.6 Hz, 2H), 2.71 (s, 2H), 2.29 (s, 2H), 1.60 – 1.51 (m, 3H), 1.32 – 1.22 (m, 2H), 1.20 (d, J = 9.3 Hz, 1H).

S10.3 Photochemical reaction of norbornyl cyclohexadiene 3

A solution of **3** (20 mg total, 0.14 mmol, 1.0 equiv.) was prepared in deuterated acetonitrile (2.0 mL, 64 mM) and divided into five separate quartz NMR tubes (0.4 mL per tube). Irradiation took place in a Luzchem photoreactor equipped with ten 254 nm low-pressure mercury lamps. Each reaction tube was irradiated for a given time (0, 30, 60, 120, or 240 minutes), after which was added 0.2 mL of a stock solution of internal standard (1,4-dinitrobenzene) in deuterated acetonitrile and a ^{1}H NMR spectrum was obtained. Yields were determined by relative integration compared with the internal standard signal. The products of the reaction were obtained as an inseparable mixture, and thus our analysis was performed on the mixture by COSY and ROESY NMR spectroscopy.

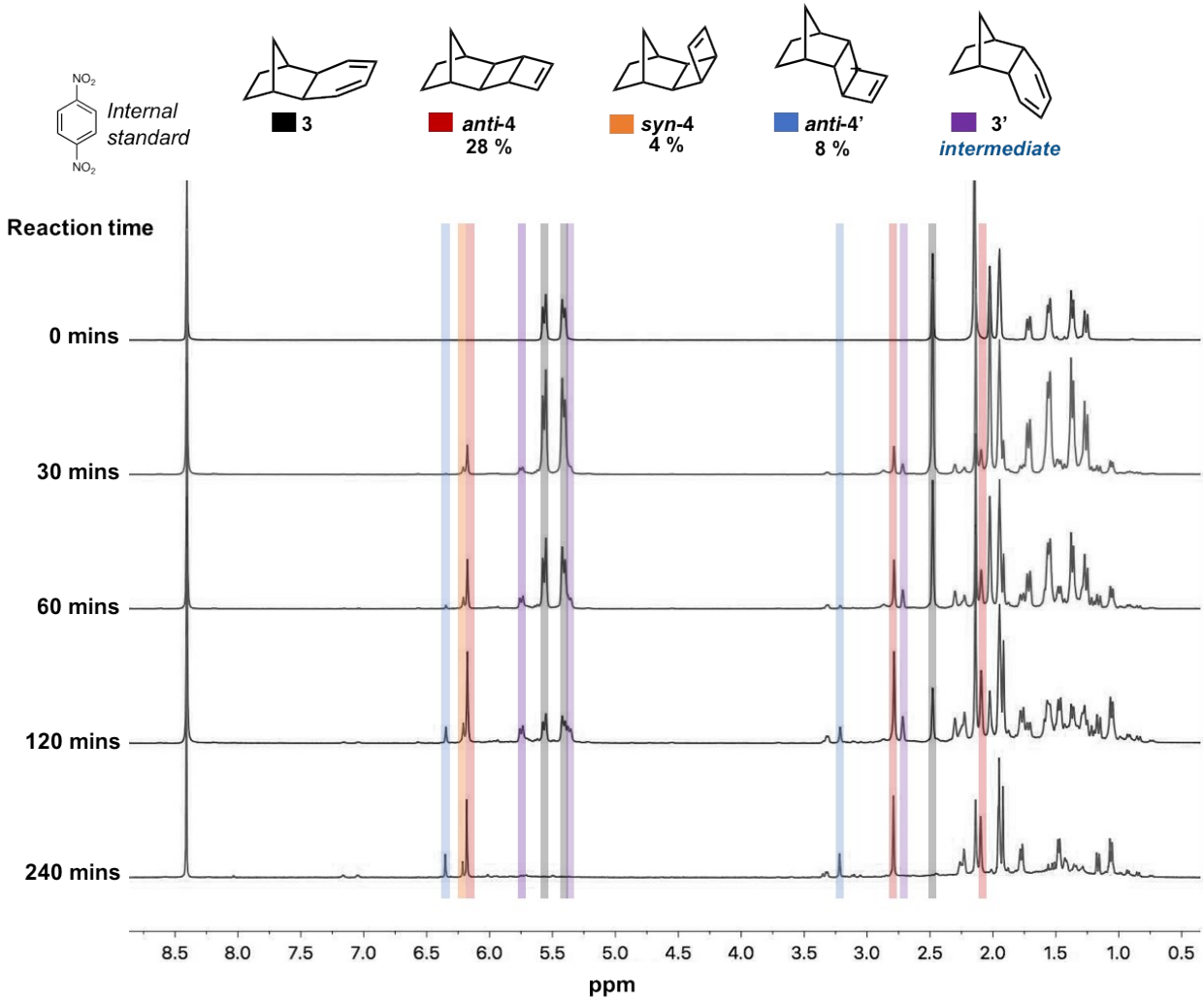


Figure S22. Stacked ^{1}H-NMR spectra after 0, 30, 60, 120 and 240 mins of irradiation of **3** with 254 nm light. Yields were determined by comparison with internal standard (1,4-dinitrobenzene). Colored rectangles were added to indicate signals associated with compounds **3**, *anti*-**4**, *syn*-**4**, *anti*-**4'** and **3'**.

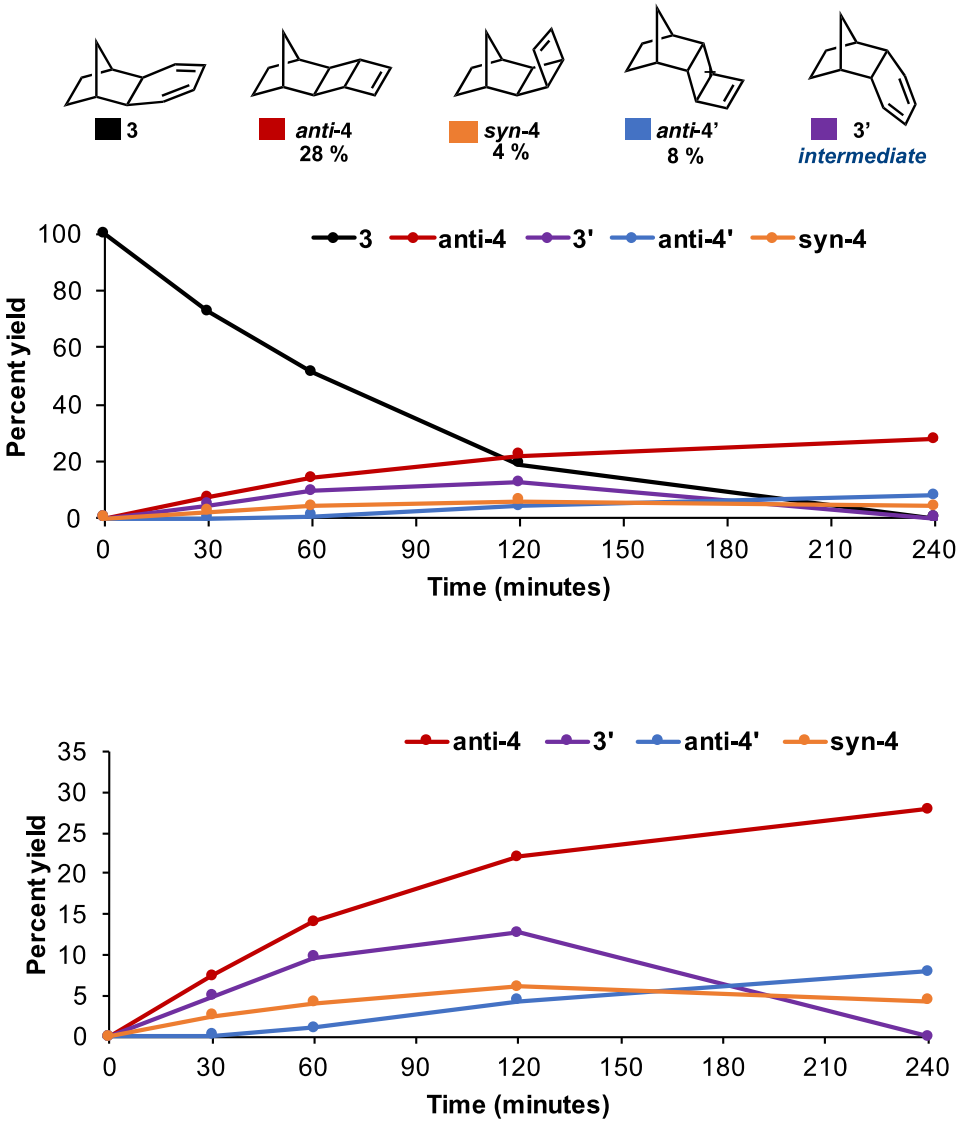


Figure S23. Comparison of yields of **3**, *anti*-**4**, *syn*-**4**, *anti*-**4'** and **3'** obtained from ^{1}H-NMR spectra integration relative to internal standard after 0, 30, 60, 120, or 240 mins of irradiation. The top graph includes the decay of **3**, and the bottom graph is identical except without **3** for ease of the reader.

S10.4 Structural assignment

Due to the products of the irradiation of **3** producing an inseparable mixture of hydrocarbon products, characterization was performed using NMR spectroscopy on the crude reaction mixture after 240 minutes of irradiation.

Evidence for the *anti*-stereochemistry of *anti*-**4** and *anti*-**4'** was obtained from the ^{1}H-NMR spectra. A much larger splitting of the allylic bridgehead protons was measured for the *syn* isomer (J = 6.4 Hz), while an almost nonexistent splitting was measured for the *anti*-isomers, a phenomenon previously observed by Dauben in a similar ladderene system.[21] This other system by Dauben, which features a similar electrocyclic closure of a cyclohexadiene to produce ladderenes, forms the *anti* and *syn* ladderenes as the major photoproducts in 28% and 7% yields, respectively. The relative chemical shifts of their ^{1}H spectra are similar to our system.

We then performed a COSY experiment (see below) to fully assign the major *anti*-**4** product as well as key correlations within *syn*-**4** and *anti*-**4'**. The data we obtained suggest all products are ladderenes due to a correlation between alkene protons (6 – 6.5 ppm) and those at the bridgehead position (2.8 – 3.2 ppm).

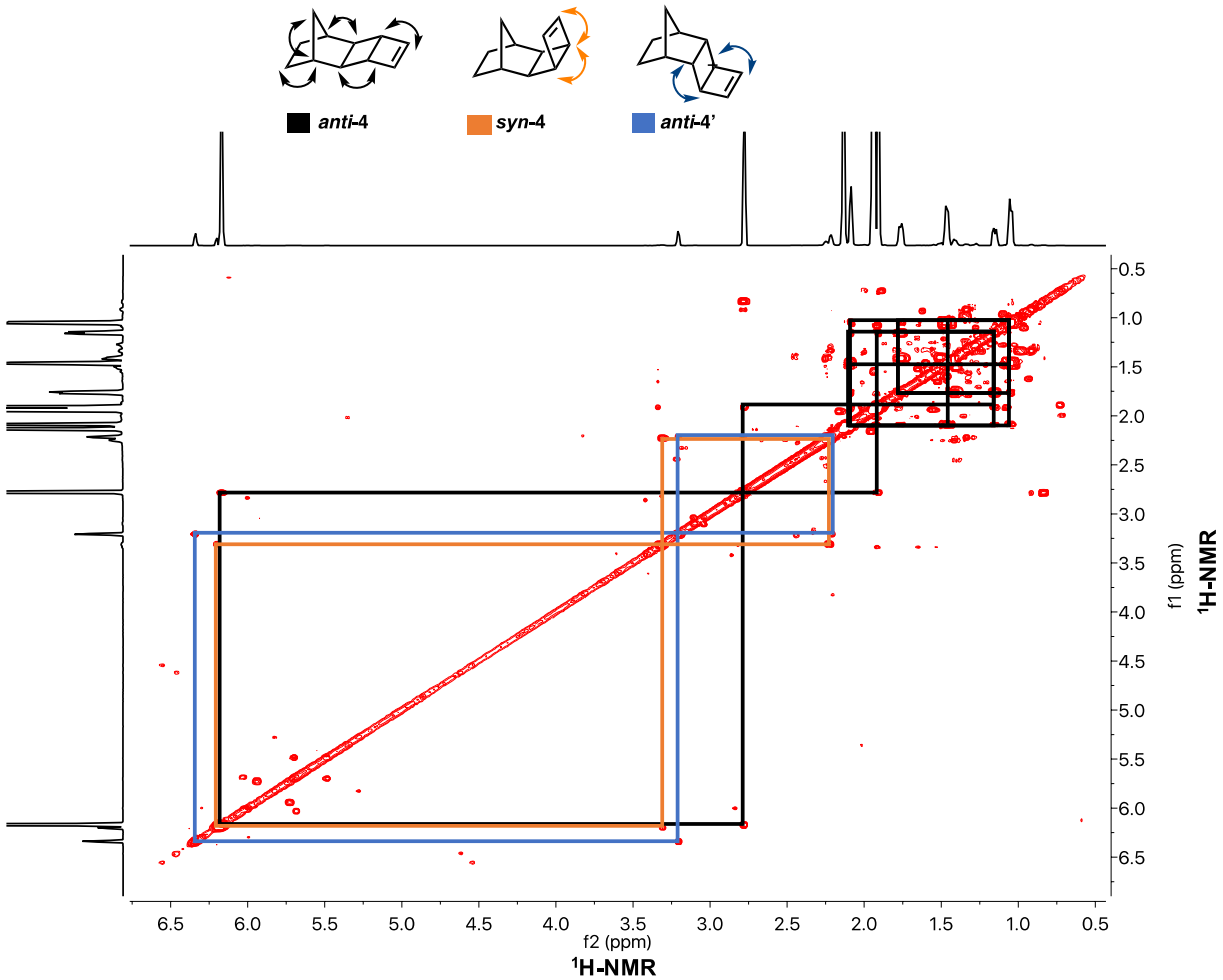


Figure S24. COSY NMR spectrum after 240 minutes of irradiating diene-**3**. Cross-peaks have been added as colored boxes: Black for *anti*-**4**, orange for *syn*-**4**, and blue for *anti*-**4'**. The two axes are both the ^{1}H-NMR spectra in CD_3CN using a 600 MHz instrument.

Finally, to further confirm which ladderene isomer is the major product, we performed a ROESY experiment (see below). We observed a key correlation between the allylic bridgehead signal at 2.78 ppm and a nearby norbornyl proton at 1.76 ppm. This interaction would not be expected for the other two ladderene isomers (*syn*-**4** or *anti*-**4'**).

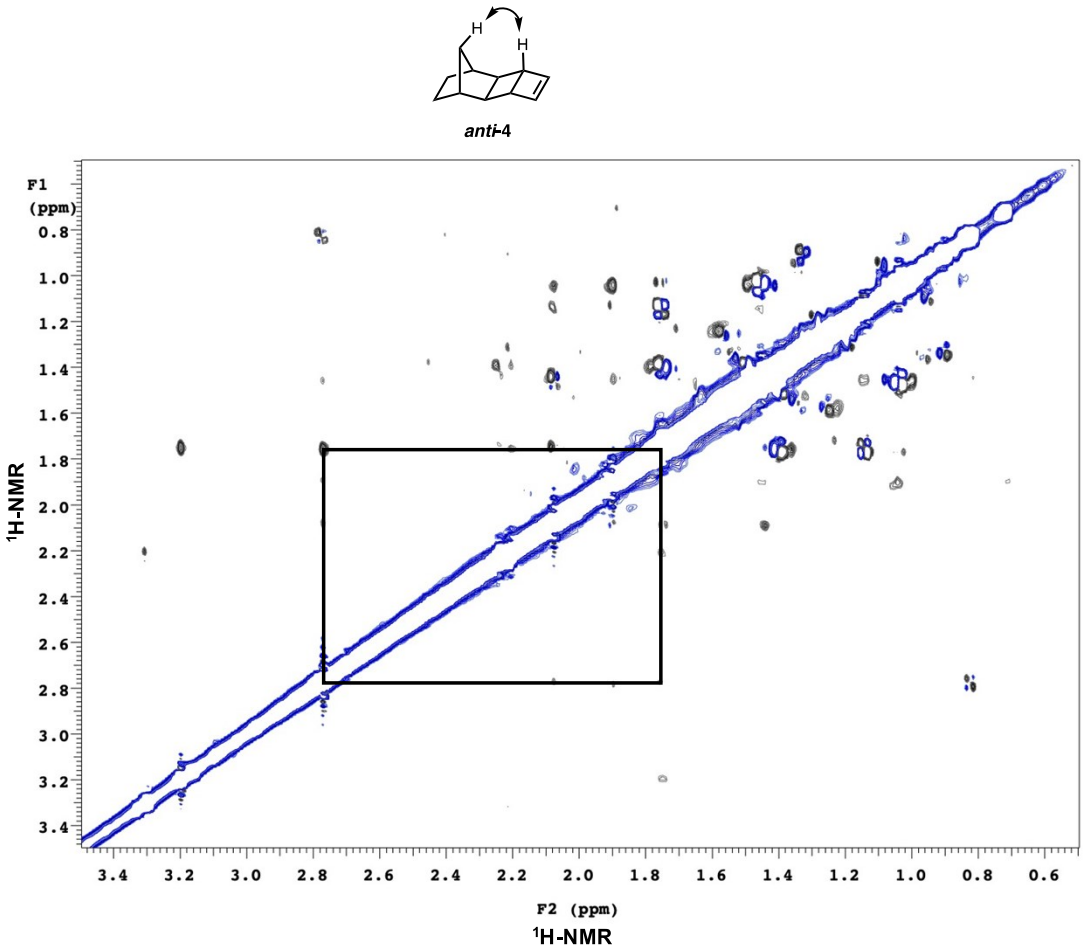


Figure S25. ROESY NMR spectrum after 240 minutes of irradiating **3**. The key cross-peaks have been added as a black box for *anti*-**4**. The two axes are both the ^{1}H-NMR spectra in CD_3CN using a 600 MHz instrument.

S11. Cartesian coordinates of optimized geometries

***trans*-1**

C	0.14501571	0.02837541	-0.67316180
C	-0.16110798	-0.02810034	0.62722014
H	0.49164356	-0.83378859	-1.22301781
H	-0.50888331	0.83376479	1.17682133
C	0.03146785	1.28645873	-1.47301940
F	1.20063722	1.61681414	-1.98773603
F	-0.81011654	1.12429511	-2.47628136
F	-0.39308986	2.30984668	-0.75746665
C	-0.04859882	-1.28642117	1.42682882
F	-1.21778423	-1.61412548	1.94327935
F	0.37241108	-2.31069787	0.71044254
F	0.79487293	-1.12642037	2.42881618

***cis*-1**

C	0.04077144	0.07479198	-0.73458499
C	-0.32346501	-0.06483044	0.54932647

H	0.22569559	-0.81462113	-1.32066341
H	-0.41825273	-1.06367973	0.94860682
C	0.24010371	1.33312220	-1.52154169
F	1.47063953	1.37107529	-1.99945496
F	-0.58315558	1.35177692	-2.55376173
F	0.04451971	2.43412746	-0.83353707
C	-0.63347314	1.01143242	1.55789245
F	0.39787299	1.79057523	1.80121482
F	-1.64699566	1.76842909	1.19761536
F	-0.96519998	0.43456900	2.70231000

MECP-*trans*-1

C	-0.20470118	-0.02115260	-0.76726274
C	-0.27703368	-0.07213838	0.60936430
H	-1.35920281	0.01662226	-0.58408741
H	-0.74122071	0.68702702	1.25086473
C	0.03751109	1.22419398	-1.54256950
F	1.33065798	1.45047591	-1.66880676
F	-0.49018495	1.15226179	-2.74634509
F	-0.47526727	2.32561139	-0.97375776
C	0.08290008	-1.31783927	1.40445187
F	-0.99503077	-2.05725682	1.57389168
F	1.00493590	-2.03870598	0.83668560
F	0.51385686	-0.94387291	2.59495659

MECP-*cis*-1

C	0.13225747	-0.08322303	-0.76022994
C	-0.13460097	-0.14343669	0.58836939
H	-0.99476058	0.24253131	-0.70409076
H	-0.20691252	-1.10312961	1.10470912
C	0.82730624	0.98846628	-1.51986816
F	2.12551584	0.75046285	-1.55202885
F	0.38612679	1.04763684	-2.76011549
F	0.69589620	2.22018683	-1.01653305
C	-0.54678522	1.00328225	1.52551866
F	0.53896638	1.57651208	1.99099651
F	-1.28854119	1.90318407	0.93078885
F	-1.22471886	0.52278557	2.54846739

3

C	-2.33436914	-0.78047160	-0.34159816
C	-0.99767719	-1.12503456	0.33922090
C	-0.99874383	1.12483095	0.33759272
C	-2.33501538	0.77714084	-0.34285531
H	-2.40805247	-1.20291855	-1.34293629
H	-3.17033854	-1.16978114	0.23948855
H	-2.40837183	1.19724278	-1.34518679
H	-3.17232655	1.16610873	0.23653020

C	-0.87727499	0.00093649	1.37699687
H	0.07302203	0.00253151	1.90531683
H	-1.69194025	0.00027925	2.10194503
C	0.15812983	-0.78394227	-0.63127913
H	-0.07638436	-1.16331139	-1.62665403
C	0.15989729	0.78791809	-0.63075217
H	-0.07765868	1.16219431	-1.62732624
H	-0.94977217	-2.14294073	0.72311354
H	-0.95314559	2.14280625	0.72095327
C	1.46019703	-1.42498616	-0.21511052
C	2.54387790	-0.72717188	0.12130256
H	1.49086325	-2.50752935	-0.20379063
C	2.56553750	0.73545018	0.12957206
H	3.44673100	-1.25557102	0.40418177
C	1.46336345	1.43089089	-0.21286233
H	3.47627320	1.24184393	0.41557931
H	1.48274848	2.51355448	-0.20452202

MECP-*syn*-3

C	-2.29126947	-0.84332691	-0.26896012
C	-0.87600213	-1.11889337	0.27955618
C	-0.99759264	1.13958534	0.29644912
C	-2.35048175	0.71511885	-0.31507047
H	-2.46510275	-1.30093983	-1.24218318
H	-3.03995567	-1.23833823	0.41683001
H	-2.48353456	1.09779324	-1.32699204
H	-3.17675563	1.09169416	0.28672277
C	-0.75071718	0.01444728	1.31295026
H	0.21851583	0.05667789	1.78787432
H	-1.53231886	-0.04260102	2.07201973
C	0.09331503	-0.70883237	-0.85285081
H	-0.26845574	-1.02881006	-1.82857984
C	0.10678349	0.86028574	-0.73881557
H	-0.13461198	1.28604349	-1.71067361
H	-0.72521171	-2.12894435	0.65228984
H	-0.97791094	2.15767918	0.68317782
C	1.55156714	-0.97152010	-0.74379494
C	2.27716850	-0.84303754	0.51281959
H	2.06626242	-0.94551373	-1.71122655
C	2.48154052	0.52880925	0.25042601
H	3.15230627	-1.44043771	0.72057096
C	1.51219202	1.31452079	-0.36841496
H	3.44133433	1.00447238	0.44903488
H	1.74911495	2.35364247	-0.56094926

MECP-*anti*-3

C	-2.34807064	-0.81250925	-0.35629347
C	-1.10950356	-1.08252692	0.51870755

C	-1.11224876	1.16029066	0.33441582
C	-2.33898053	0.73897874	-0.49874152
H	-2.29281362	-1.32309091	-1.31654918
H	-3.25212148	-1.15375643	0.14693521
H	-2.26108391	1.06362017	-1.53517358
H	-3.24820019	1.17423136	-0.08546571
C	-1.14833568	0.12634524	1.46877591
H	-0.30155167	0.19015480	2.14522123
H	-2.06947596	0.17073835	2.04985209
C	0.12676471	-0.79424320	-0.34836764
H	0.09321508	-1.27388228	-1.32682264
C	0.15456022	0.76670825	-0.45391735
H	0.08947548	1.13168886	-1.48266058
H	-1.10133611	-2.06784797	0.98348028
H	-1.12030778	2.20753777	0.63187123
C	1.46330549	-1.06598467	0.24853917
C	2.66325550	-0.74478668	-0.52808140
H	1.51628036	-1.32201784	1.30618792
C	2.66197959	0.56783867	-0.02496773
H	3.56538537	-1.32830931	-0.41396690
C	1.45314233	1.23923248	0.16807568
H	3.57098276	1.05111015	0.33399890
H	1.46919824	2.21536143	0.63619827

3a

C	-2.31094468	-0.84930322	-0.20853672
C	-0.90256656	-1.13878339	0.34656685
C	-0.95033884	1.12001429	0.24954077
C	-2.30310235	0.70164288	-0.36796833
H	-2.50759208	-1.36909356	-1.14472704
H	-3.07145807	-1.15997710	0.50756783
H	-2.38972090	1.01393004	-1.40818349
H	-3.12932008	1.15534657	0.17831331
C	-0.76055315	0.04192246	1.32646761
H	0.19207441	0.10664020	1.83664355
H	-1.55609259	0.04961195	2.07219699
C	0.08437918	-0.82193597	-0.80270002
H	-0.35850471	-1.12391452	-1.74917208
C	0.18005379	0.77070875	-0.74290334
H	-0.02832848	1.19468647	-1.72679912
H	-0.78782238	-2.13575678	0.76695728
H	-0.92553215	2.15292528	0.59352586
C	1.53804074	-1.22629197	-0.77190159
C	2.24793904	-0.81851502	0.34978111
H	2.00409324	-0.91087324	-1.70256684
C	2.57991547	0.62804673	0.17278693
H	1.89762055	-1.09436469	1.33649111
C	1.53984102	1.32817687	-0.29829843

H	3.49044461	1.10997428	0.50953704
H	1.61964362	2.41071092	-0.34395612

3b

C	-2.34771640	-0.69763107	-0.47070676
C	-1.12173082	-1.13257185	0.35608952
C	-0.98693745	1.12205413	0.38061618
C	-2.26197145	0.85719612	-0.44334557
H	-2.32361408	-1.10154078	-1.48208773
H	-3.26799744	-1.04484573	-0.00171181
H	-2.20823392	1.29411645	-1.43959583
H	-3.13095350	1.28512333	0.05600123
C	-1.06321968	-0.01374441	1.41178485
H	-0.20133550	-0.06453416	2.07060064
H	-1.96546789	0.03460455	2.02192683
C	0.14556993	-0.87493573	-0.47767348
H	0.05360875	-1.25486740	-1.49449072
C	0.23143027	0.72309566	-0.48495868
H	0.08184484	1.10433906	-1.49821162
H	-1.18373051	-2.15114778	0.73517321
H	-0.92822128	2.13275759	0.78042692
C	1.44884821	-1.29795469	0.15409007
C	2.54141869	-0.76260327	-0.51599962
H	1.46997421	-1.13519820	1.22375559
C	2.73102588	0.62999585	-0.00015659
H	2.57744568	-0.85498144	-1.59906200
C	1.57304574	1.30172380	-0.02000005
H	3.67393244	1.09845490	0.25524327
H	1.57185823	2.36297971	0.20980206

S12. Reference

1. Schutt, K. T.; Sauceda, H. E.; Kindermans, P. J.; Tkatchenko, A.; Muller, K. R., SchNet - A deep learning architecture for molecules and materials. *J. Chem. Phys.* **2018,** *148* (24), 241722.
2. Zhang, Y.; Maurer, R. J.; Jiang, B., Symmetry-Adapted High Dimensional Neural Network Representation of Electronic Friction Tensor of Adsorbates on Metals. *The Journal of Physical Chemistry C* **2019,** *124* (1), 186-195.
3. Westermayr, J.; Gastegger, M.; Marquetand, P., Combining SchNet and SHARC: The SchNarc Machine Learning Approach for Excited-State Dynamics. *J Phys Chem Lett* **2020,** *11* (10), 3828-3834.
4. Westermayr, J.; Gastegger, M.; Menger, M.; Mai, S.; Gonzalez, L.; Marquetand, P., Machine learning enables long time scale molecular photodynamics simulations. *Chem. Sci.* **2019,** *10* (35), 8100-8107.
5. Akimov, A. V., A Simple Phase Correction Makes a Big Difference in Nonadiabatic Molecular Dynamics. *J Phys Chem Lett* **2018,** *9* (20), 6096-6102.
6. Plasser, F.; Ruckenbauer, M.; Mai, S.; Oppel, M.; Marquetand, P.; Gonzalez, L., Efficient and Flexible Computation of Many-Electron Wave Function Overlaps. *J. Chem. Theory Comput.* **2016,** *12* (3), 1207-1219.

7. Westermayr, J.; Marquetand, P., Machine learning and excited-state molecular dynamics. *Machine Learning: Science and Technology* **2020,** *1* (4).
8. Westermayr, J.; Marquetand, P., Machine Learning for Electronically Excited States of Molecules. *Chem. Rev.* **2020**.
9. 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 CH2NH2+: From single-state to multi-state representations and multi-property machine learning models. *Machine Learning: Science and Technology* **2020,** *1* (2).
10. 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 (2)A' states of LiFH. *Phys. Chem. Chem. Phys.* **2019,** *21* (26), 14205-14213.
11. Kuhn, H. W., The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly* **1955,** *2* (1-2), 83-97.
12. Gastegger, M.; Behler, J.; Marquetand, P., Machine learning molecular dynamics for the simulation of infrared spectra. *Chem. Sci.* **2017,** *8* (10), 6924-6935.
13. Fdez Galvan, I.; Vacher, M.; Alavi, A.; Angeli, C.; Aquilante, F.; Autschbach, J.; Bao, J. J.; Bokarev, S. I.; Bogdanov, N. A.; Carlson, R. K.; Chibotaru, L. F.; Creutzberg, J.; Dattani, N.; Delcey, M. G.; Dong, S. S.; Dreuw, A.; Freitag, L.; Frutos, L. M.; Gagliardi, L.; Gendron, F.; Giussani, A.; Gonzalez, L.; Grell, G.; Guo, M.; Hoyer, C. E.; Johansson, M.; Keller, S.; Knecht, S.; Kovacevic, G.; Kallman, E.; Li Manni, G.; Lundberg, M.; Ma, Y.; Mai, S.; Malhado, J. P.; Malmqvist, P. A.; Marquetand, P.; Mewes, S. A.; Norell, J.; Olivucci, M.; Oppel, M.; Phung, Q. M.; Pierloot, K.; Plasser, F.; Reiher, M.; Sand, A. M.; Schapiro, I.; Sharma, P.; Stein, C. J.; Sorensen, L. K.; Truhlar, D. G.; Ugandi, M.; Ungur, L.; Valentini, A.; Vancoillie, S.; Veryazov, V.; Weser, O.; Wesolowski, T. A.; Widmark, P. O.; Wouters, S.; Zech, A.; Zobel, J. P.; Lindh, R., OpenMolcas: From Source Code to Insight. *J. Chem. Theory Comput.* **2019,** *15* (11), 5925-5964.
14. Dunning, T. H., Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989,** *90* (2), 1007-1023.
15. Pierloot, K.; Dumez, B.; Widmark, P.-O.; Roos, B. r. O., Density matrix averaged atomic natural orbital (ANO) basis sets for correlated molecular wave functions. *Theor. Chim. Acta* **1995,** *90* (2-3), 87-114.
16. Pou-Amérigo, R.; Merchán, M.; Nebot-Gil, I.; Widmark, P.-O.; Roos, B. O., Density matrix averaged atomic natural orbital (ANO) basis sets for correlated molecular wave functions. *Theor. Chim. Acta* **1995,** *92* (3), 149-181.
17. Widmark, P.-O.; Malmqvist, P.-k.; Roos, B. r. O., Density matrix averaged atomic natural orbital (ANO) basis sets for correlated molecular wave functions. *Theor. Chim. Acta* **1990,** *77* (5), 291-306.
18. Widmark, P.-O.; Persson, B. J.; Roos, B. r. O., Density matrix averaged atomic natural orbital (ANO) basis sets for correlated molecular wave functions. *Theor. Chim. Acta* **1991,** *79* (6), 419-432.
19. McCulloch, R. K.; Rye, A. R.; Wege, D., Preparation and thermal decarbonylation of some sterically compressed 5,6-endo-Disubstituted Norborn-2-en-7-ones. *Aust. J. Chem.* **1974,** *27* (9).
20. Rye, A. R.; Wege, D., Preparation and thermolysis of exo-and endo-Tricyclo[6,2,1,02,7] undeca-3,5,9-triene. *Aust. J. Chem.* **1974,** *27* (9).
21. Dauben, W. G.; Kellogg, M. S., Photochemistry of cis-fused bicyclo[4.n.0]-2,4-dienes. Ground state conformational control. *J. Am. Chem. Soc.* **1980,** *102* (13), 4456-4463.