# Machine-learning photodynamics simulations uncover the role of substituent effects on the photochemical formation of cubanes

Jingbai Li[1], Rachel Stein[1], Daniel M. Adrion[1], and Steven A. Lopez[1]*
[1] Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA 02115, USA
* Correspondence to s.lopez@northeastern.edu

**Abstract**
Photochemical [2+2]-photocycloadditions are used to efficiently access strained organic molecular architectures, storing solar energy in chemical bonds. Functionalized [3]-ladderenes have been shown to undergo [2+2]-photocycloadditions to afford cubanes, an energy-dense class of organic molecules. The substituents (e.g., methyl, trifluoromethyl, and cyclopropyl) affect the overall reactivities of these cubane precursors leading to a yield from 1% to 48%. We now integrate single and multireference calculations and our machine-learning-accelerated non-adiabatic molecular dynamics (ML-NAMD) to understand how substituents affect the mechanistic photodynamics of [2+2]-photocycloadditions. Our calculations show that steric clashes destabilize the 4π-electrocyclic ring-opening pathway and minimum energy conical intersections by 0.72–1.15 eV and reaction energies by 0.68–2.34 eV. In contrast, favorable dispersive interactions stabilize the [2+2]-photocycloaddition pathway, lower the conical intersection energies by 0.31–0.85 eV and cubane reaction energies by 0.03–0.82 eV. The 2 ps ML-NAMD trajectories reveal that closed-shell repulsions block a 6π-conrotatory electrocyclic ring-opening pathway with increasing steric bulk. 57% of the methyl-substituted [3]-ladderene trajectories proceed through the 6π-conrotatory electrocyclic ring-opening, whereas the trifluoromethyl- and cyclopropyl-substituted 3-ladderenes chemoselectively proceed through [2+2]-photocycloaddition pathways. The predicted cubane yields (H: 0.4% < $CH_3$: 1% < $CF_3$: 14% < cPr: 20%) match the experimental trend; these substituents pre-distort the reactants to resemble the conical intersection leading to cubane.

## Introduction

Light-driven organic reactions have become increasingly important in natural product synthesis,[1] materials chemistry,[2] solar thermal storage,[3-4] and photopharmacology.[5] The use of light as a green reagent is appealing in industry and academia because it is features mild reaction conditions and efficient syntheses due to their high atom economy and stereospecific outcomes. The spatiotemporal activation produces complex molecular architectures that often require extensive heating and or transition metal catalysts or sensitizers.[6] The [2+2]-photocycloaddition is one of the most versatile reactions to synthesize diverse and strained four-membered mono- and bicyclic compounds,[6-7] including cubanes (**2**). The highly-strained cubic structure of **2** is dense ($\rho$ = 1.29 g·cm$^{-3}$),[8] strained (161.5 kcal·mol$^{-1}$),[9] and kinetically stable at room temperature[10-11] and under high pressures.[12] Eaton and Cole published the first practical synthetic protocol of cubane in 1964.[13]

Most recent efforts by Tsanaktsidis achieved pilot-scale production of **2**.[14] The intramolecular [2+2]-photocycloaddition of **1** sidesteps the current laborious synthesis of cubane (Figure 1a). The rigidity of **1** enhances overlap between the π-bonds and red-shifts the ππ$^*$ absorbance

spectrum peak compatible with common light sources (≥200nm). The strained structure prevents the *cis-trans* isomerization, which would interfere with the desired photocycloaddition. Early computational studies by Gleiter suggested the [2+2]-photocycloaddition of **1** is symmetry-forbidden.[15] The strong through-bond interactions in **1** (Figure 1b) raise the in-phase π-orbital ($\pi_{SS}$) to the highest occupied molecular orbital (HOMO). The resulting excited-state configuration ($\pi_{SS}\pi^*_{SA}$) break the orbital-symmetry ($\pi_{AS}\pi^*_{SA}$) of [2+2]-photocycloaddition according to the Woodward-Hoffmann rules.[16] Gleiter also demonstrated that one could achieve the [2+2]-photocycloaddition of **1** by shortening the intramolecular π-π distance ($\Delta r$ = 0.3 Å) with tethered propyl bridges (-$CH_2CH_2CH_2$-), which lowers the $\pi_{SS}$ to HOMO–1, thus restoring the requisite orbital-symmetry to allow the [2+2]-photocycloaddition.[15]

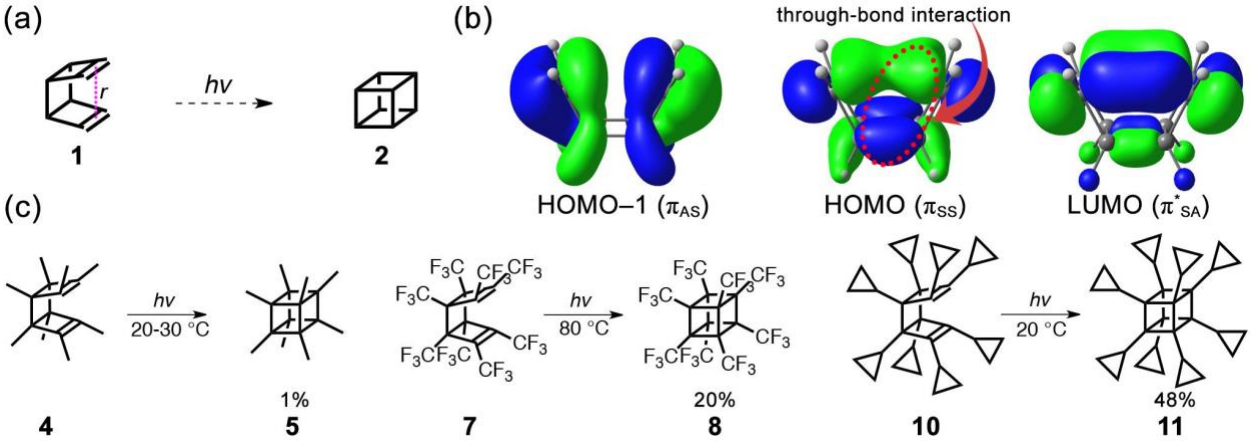


**Figure 1**. (a) Theoretical [2+2]-photocycloaddition of [3]-ladderene (**1**) to cubane (**2**). (b) Frontier molecular orbitals of **1**, computed at ωB97XD/cc-pVTZ//PBE0/cc-pVDZ. (c) Photochemical formation of octamethylcubane (**5**),[17] octa(trifluoromethyl)cubane (**8**),[18] and octacyclopropylcubane (**11**).[19]

The correlation diagrams for the [2+2]-cycloadditions of substituted [3]-ladderenes (i.e., **4**, **7**, and **10**) are qualitatively similar and suggest that this pathway is symmetry-forbidden. In contrast, experiments demonstrate that **5**, **8**, and **11** undergo [2+2]-photocycloadditions in the presence of light (Figure 1c). We interpret this disagreement as an indication of significant dynamic effects that enable this photoreaction. Our previous computational study on the photochemistry of **1** showed that a small portion of energetically accessible non-equilibrium structures of **1** featured relatively short intramolecular π-π distances and led to productive nonadiabatic molecular dynamics (NAMD) trajectories to cubane.[20] These non-equilibrium structures invert the order of the HOMO and HOMO–1 orbitals to allow the [2+2]-photocycloaddition. However, the vast majority of the NAMD trajectories led to irreversible 4π-disrotatory electrocyclic ring-opening to bicyclo[4,2,0]octa-2,4,7-triene (**3**) due to the release of the strain energy.

We hypothesized that the substituents were influencing the photodynamics; to date, there has been an absence of computational or experimental studies explaining the origin of the substituent effects. The requisite thousands of multiconfigurational electronic structure calculations are prohibitively expensive for the NAMD simulations. A single 1 picosecond trajectory with a 0.5 femtosecond time step for **4**, **7**, and **10** requires $10^{3-4}$ wall-clock hours. Recently machine learning (ML) excited-state potentials achieved nanosecond NAMD simulations for organic reactions (e.g., $CH_2NH_2^+$) with reliable accuracy compared to multireference configuration interactions (MRCI).[21-

[22] Our group adapted this methodology based on neural network (NN) to discover the origin of stereoselectivity of a photo-torquoselective isomerization of norbornyl cyclohexadienes.[23]

We now integrate static single and multireference calculations and ML-NAMD techniques to understand how substituents affect the photochemical reactivity and selectivity of [3]-ladderenes. The manuscript first describes how substituents alter the excited-state structures and potential energy surfaces of competing pathways (i.e., [2+2]-photocycloaddition and 4π-disrotatory electrocyclic ring-opening). Second, we discuss the origins of varying quantum yields of cubane systems with ML-NAMD trajectories. Collectively, our findings serve as a guide for the ongoing development of substituent groups assisted [2+2]-photocycloaddition of [3]-ladderene toward cubane systems. This work also demonstrates the value of the ML-NAMD method for designing and discovering novel molecular frameworks.

## Results and Discussion

**4, 7,** and **10** are octa-substituted and feature two substituent sites: vinyl ($R_1$) and bridgehead ($R_2$). The substituents at the bridgehead carbons have considerably less conformational freedom than at the alkenyl position. We first sought to understand how the location and steric bulk of the substituents affect the reactivities of the substituted [3]-ladderenes relative to **1**. We computed the vertical excitation energies, reaction energies, and conical intersection energies for octa- and tetra-substituted [3]-ladderenes.

*Substituent effects on excitation*

Our calculations show that $S_0 \rightarrow S_1$ oscillator strengths of **1, 4**, **7**, **10** range from 0.05–0.19 and a HOMO→LUMO transition to a $\pi\pi^*$-state (Figure 1b). The oscillator strengths are 1–5 orders of magnitude lower in the higher singlet excited states (Table S3), indicating considerably lower transition probabilities. Therefore, the subsequent discussions focus on the $S_1$-state. Figure 2 summarizes the vertical excitation energies for 10 [3]-ladderenes to probe the substituent effects.

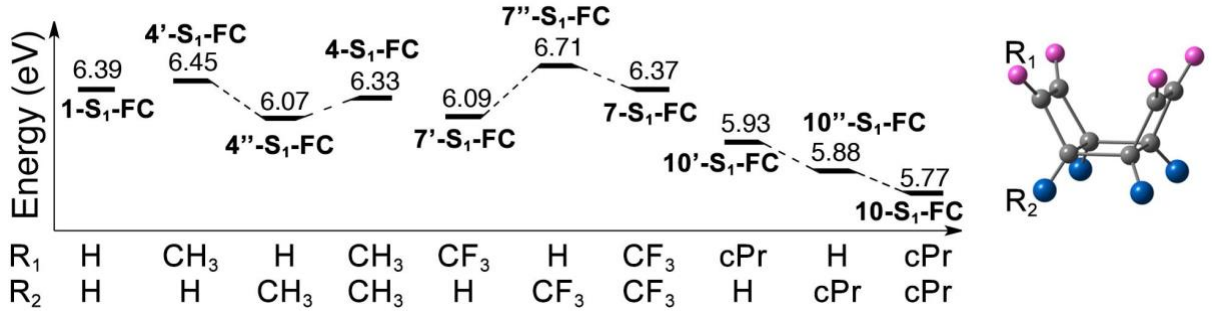


**Figure 2**. Excitation energies of tetra- and octa-methyl ($CH_3$), trifluoromethyl ($CF_3$), and cyclopropyl (cPr) [3]-ladderene, computed with ωB97XD/cc-pVTZ//PBE0/cc-pVDZ. The pink and blue terminals in the 3-dimensional molecular structures denote the substituent groups at the vinyl ($R_1$) and bridgehead ($R_2$) carbon atoms, respectively.

The Franck-Condon (FC) point energy of **1-$S_1$-FC** is 6.39 eV (Figure 2). The $CH_3$ is an inductive electron-donor and slightly increases the energy of **4'-$S_1$-FC** to 6.45 eV at $R_1$ but decreases the energy of **4''-$S_1$-FC** to 6.07 eV at $R_2$. **4-$S_1$-FC** (6.33 eV) and **1-$S_1$-FC** are nearly degenerate; this suggests that the competing effects nearly cancel out with octa-$CH_3$-substitution. The **7'-$S_1$-FC** energy (6.09 eV) is lower at $R_1$ but increases the **7''-$S_1$-FC** energy (6.71 eV) at $R_2$

because the $CF_3$ is a powerful inductive electron-withdrawing group. The opposing effects of tetra-substitution at $R_1$ and $R_2$ nearly cancel in octa-substituted **7-$S_1$-FC** (6.37 eV). The cPr is an electron-donating group as $CH_3$, but the $\sigma_{C\text{-}C}$-orbitals of the cPr at $R_1$ enlarge the anti-bonding overlap with the $\pi_{C=C}$-orbitals in the HOMO (Figure S4). This effect minimizes the HOMO-LUMO gap, thus reducing the energy of **10'-$S_1$-FC** to 5.93 eV. The steric clash of the cPr at $R_2$ compresses the intramolecular π-π distance that increases the orbital overlap in the LUMO. It decreases the energy of **10''-$S_1$-FC** to 5.88 eV. In line with the previous results, the octa-cPr-substitution combines the effects of tetra-cPr-substitutions that further reduces the vertical excitation energy of **10** to 5.77 eV.

*Substituent effects on reactant and product structures*

These substituents can substantially perturb the [3]-ladderene geometries relative to **1** to minimize closed-shell repulsions. We defined two parameters, an intramolecular distance, *r,* and a folding angle *θ,* to quantify the geometrical distortions (Figure 3). The parameter *r* is the distance between the centroids of the two π-bonds, and *θ* is the angle between the planes of the four-membered rings. These parameters are *r* = 2.98 Å and *θ* = 118° in **1;** *r* = 1.56 Å and *θ* = 90° in **2**; *θ* = 114° in **3**, optimized with PBE0/cc-pVDZ.

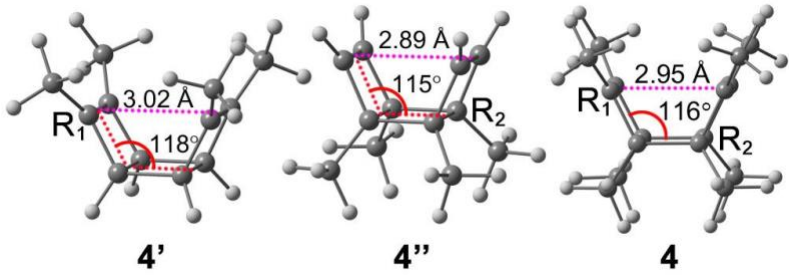


**Figure 3**. Intramolecular π-π distance *r* (pink) and folding angle *θ* (red) show geometric distortions caused by the $CH_3$-substitutions at the vinyl ($R_1$) and bridgehead ($R_2$) position.

$R_1$ tetra-$CH_3$-substitutions increase the *r* to 3.02 Å. The $CH_3$ at $R_2$ show opposite effects; the *r* and *θ* values decrease to 2.89 Å and 115°, maximizing the distance between adjacent substituents (Figure 3). The $CH_3$ at $R_2$ bends the [3]-ladderenes 3° more than the $CH_3$ at $R_1$, introducing additional strain at the bridgehead carbons, and **4''** is 0.46 eV higher than **4'** (Table S5). The tetra-$CF_3$- and cPr-substitution lead to similar geometrical distortions: the *r* and *θ* increase to 2.98–3.02 Å and 120° when substituting $R_1$, whereas the *r* and *θ* values reduce to 2.86–2.87 Å and 114–116° when substituting $R_2$. The energies of **7'** and **10'** are 0.21 and 0.59 eV lower than **7''** and **10''**, respectively (Table S5). Compared to **1**, the *r* and *θ* parameters in the octa-substituted [3]-ladderenes, **4**, **7**, and **10** decreased to 2.89–2.95 Å and 116–118°. It suggests the repulsive steric interaction at $R_2$ generates the dominant geometrical distortions.

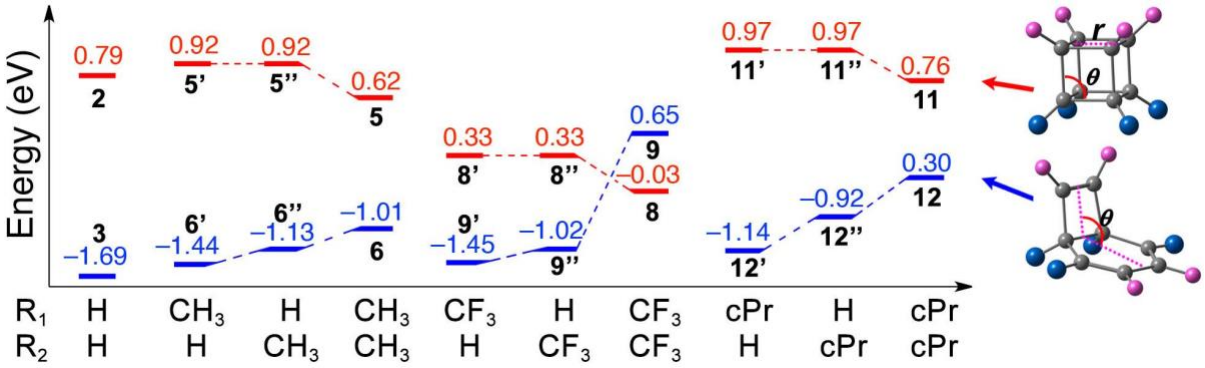


**Figure 4**. [2+2]-photocycloaddition (red) and 4π-electrocyclic ring-opening (blue) reaction energies of tetra- and octa-methyl ($CH_3$), trifluoromethyl ($CF_3$), and cyclopropyl (cPr) [3]-ladderene, computed with ωB97XD/cc-pVTZ//PBE0/cc-pVDZ. The pink dotted lines and red curves in the 3-dimensional molecular structures mark the intramolecular π-π distance *r* and folding angle *θ* in the cubanes and bicyclooctatrienes.

The repulsive steric interactions at $R_1$ enlarges the angle in the 4π-disrotatory electrocyclic ring-opening products to 115–117°, compared to **3** ($\theta$ = 114°); the steric clash at $R_2$ reduces the angle to 106–112°. Figure 4 shows that the substituents at $R_1$ and $R_2$ destabilize the reaction energies by 0.24–0.55 and 0.56–0.77 eV, respectively. The greater destabilization at $R_2$ suggests stronger steric repulsions due to the inflexibility of the substituent at the bridgehead position. The octa-substitution additively increases the destabilizing effects of the $R_1$ and $R_2$ tetra-substitution. The reaction energy of **6**, **9**, and **12** is 0.68–2.34 eV higher than **1**. The octa-substitutions with $CF_3$ and cPr switch the reaction energies from exergonic to endergonic (0.65 and 0.30 eV), indicating unfavorable 4π-disrotatory electrocyclic ring-opening to **9** and **12**.

In the tetra-substituted cubanes, the intramolecular π-π distance in **5'** and **11'** are 0.01–0.02 Å longer than **2** due to the steric repulsion between the substituents. Figure 4 shows that the $CH_3$ and cPr destabilize the reaction energies of the tetra-substituted cubanes, **5'** and **11'** by 0.13 and 0.18 eV compared to **2**, respectively. The cubane framework of the octa-$CH_3$- and cPr-substituted structures are nearly identical to those that are tetra-substituted, but the energies of **5** and **11** decrease to 0.62 and 0.76 eV, respectively. To understand the origin of these more favorable reaction energies, we computed the non-covalent interaction plots for **5** and **11** (Figure 5).

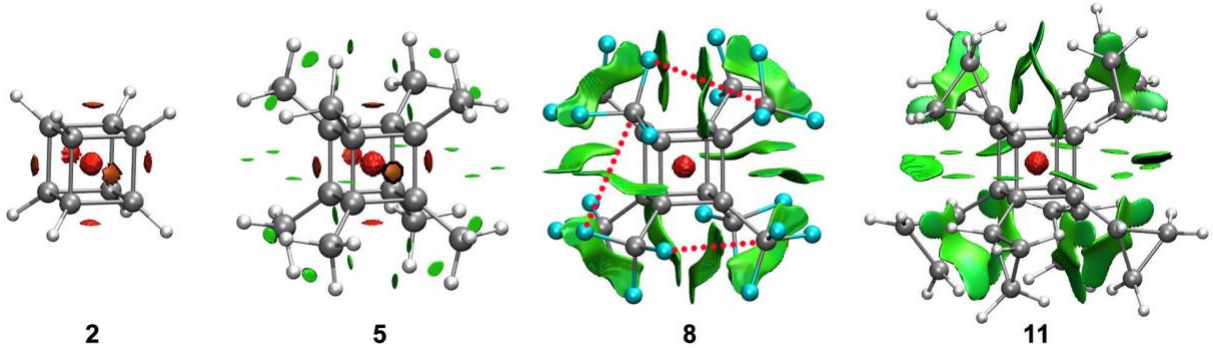


**Figure 5**. Non-covalent interaction plots of octa-substituted cubanes, computed at ωB97XD/cc-pVTZ//PBE0/cc-pVDZ. Green surfaces render the van der Waals interactions; red surfaces render

the steric clashes. The red dotted line in **8** highlights the directions of $n_F \rightarrow \sigma^*_{C-F}$ interaction. The isovalue is 0.7 in **2** and **5** and 0.5 in **8** and **11**.

The decrease in energies results from the attractive dispersion interactions, which have recently been reported in molecules functionalized with bulky charge-neutral groups.[24-25] The $CF_3$ has a Lewis acidic carbon center and can participate in a through-space $n_F \rightarrow \sigma^*_{C-F}$ dispersive interaction with the electronegative fluorine on the adjacent group in **8'** and **8**. The C···F distances in **8'** and **8** range from 3.30–3.48 and 3.14–3.18 Å that falls into the range of the non-covalent $n_F \rightarrow \sigma^*_{C-F}$ interaction,[26] shown in Figure 5. Our calculations also suggest a through-space $n_F \rightarrow \sigma^*_{C-C}$ hyperconjugations in the new-formed $\sigma_{C-C}$ bond that provides secondary cubane stabilization (Figure S6). These synergistic effects substantially lower the reaction energies of **8'** and **8** by 0.66 and 0.82 eV, respectively (Figure 4). These findings suggest that the dispersive interactions are additive and stabilize the cubanes as steric bulk increases. The reaction energy of **8** is negative, thus flipping the thermodynamic preference towards [2+2]-photocycloaddition.

*Substituent effects on conical intersections*

The $S_1/S_0$ conical intersections (CI) of **1** determine the relaxation directions in the [2+2]-photocycloaddition and 4π-disrotatory electrocyclic ring-opening pathways.[20] The $S_0$ PES lifts the energetic degeneracy on the branching plane spanning over the gradient difference vector *g* and the nonadiabatic coupling vector *h* (Figure S9). Figure 6a illustrates the branching planes of the [2+2]-photocycloaddition and 4π-disrotatory electrocyclic ring-opening CIs. Although substituents affect the CI energies (Figure 6b), the PES topologies are virtually identical. The relaxation pathway bifurcates at the [2+2]-photocycloaddition CI to the reactant and product; the 4π-electrocyclic ring-opening CIs are sloped and only proceed to the product direction.

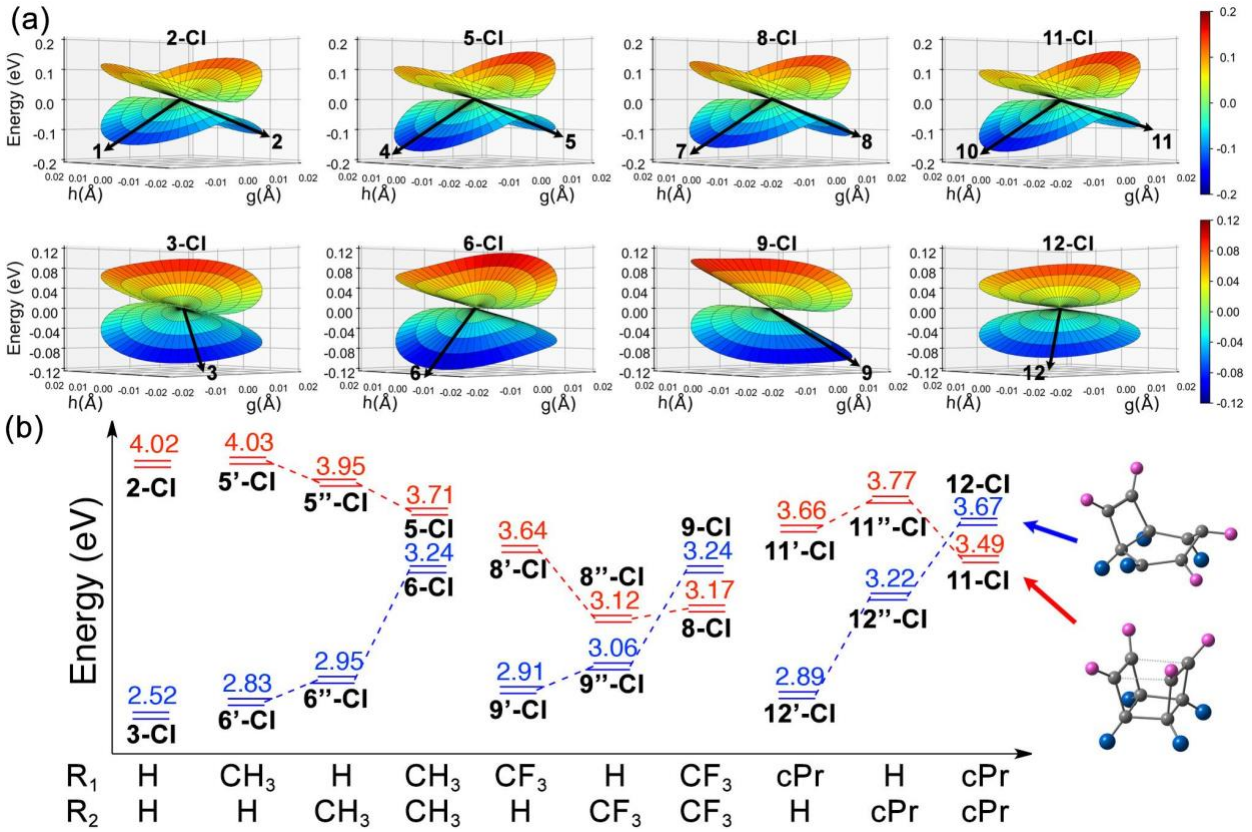


**Figure 6**. (a) Branching planes and (b) conical intersection energies of tetra- and octa-methyl ($CH_3$), trifluoromethyl ($CF_3$), and cyclopropyl (cPr) [3]-ladderene, computed with CASPT2(8,7)/ANO-S-VDZP+ANO-S-MB(R)//CASSCF(8,7)/ANO-S-VDZP+ANO-S-MB(R).

The 4π-disrotatory electrocyclic ring-opening of **1** is energetically more favored than the [2+2]-photocycloaddition; **3-CI** is 1.50 eV lower than **2-CI** (Figure 6b). The tetra-substitutions at $R_1$ increase the energies of the 4π- disrotatory electrocyclic ring-opening CIs by 0.31–0.39; the tetra-substitutions at $R_2$ raise the CI energies by 0.43–0.70 eV. The octa-substitutions demonstrate that the effect is additive and further destabilizes the CIs by 0.72–1.50 eV. The substituents disfavor the 4π-electrocyclic ring-opening CIs because closed-shell steric repulsions weaken the $\sigma_{C-C}$-bonds of the cyclohexadiene ring (Table S6). Figure 7 shows the $CH_3$ elongates the $\sigma_{C-C}$-bond length *c* by 0.06 Å from **3-CI** (*c* = 1.48–1.50 Å) to **6-CI** (*c* = 1.52–1.54 Å). The $CF_3$ and cPr show similar effects; the $\sigma_{C-C}$-bonds length, *c*, increases by 0.03–0.09 Å (Table S6).

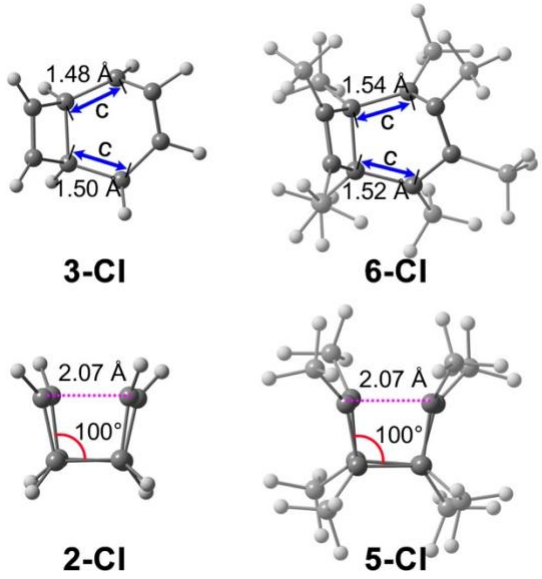


**Figure 7**. $S_1/S_0$ conical intersection of unsubstituted and octa-$CH_3$-substituted [3]-ladderenes, optimized with CASSCF(8,7)/ANO-S-VDZP+ANO-S-MB(R). The blue arrows denote the $\sigma_{C-C}$-bond distances *c*.

The tetra-$CH_3$-substituted [2+2]-photocycloaddition CIs slightly increase the intramolecular π-π distance *r* to 2.09 and 2.08 Å in **5'-CI** ($R_1$) and **5''-CI** ($R_1$), respectively. The octa-$CH_3$-substitution has a similar *r* in **5-CI** (2.07 Å) to **2-CI** (Figure 7). The energies of **5'-CI** (4.03 eV) and **5''-CI** (3.95 eV) approach that of **2-CI** (4.02 eV); the energy of **5-CI** is 3.71 eV (Figure 6b). These energies are in line with more attractive dispersion interactions in the octa-$CH_3$-subsitution. The $CF_3$ has stronger dispersion interactions that lower the CI energies by 0.38–0.90 eV, while the intramolecular π-π distances are 2.12 Å in **9'-CI**, 2.07 Å in **9''-CI**, and 2.09 Å in **9-CI**. The CI energies decrease as the *r* values decrease because these resemble the [2+2]-ophotcycloddition CIs and may promote σ-bond formation. In the tetra- and octa-cPr-substitutions, *r* in **11'-CI** (2.06 Å), **11''-CI** (2.07 Å), and **11-CI** (2.07 Å) are similar to **2-CI** but are lower in energy by 0.25–0.53 eV (Figure 6b). These attractive dispersive interactions lower the energy of the [2+2]-photocycloaddition CIs, similar to $CH_3$ and $CF_3$. These static calculations suggest that the substituent effects favor cubane formation because attractive dispersion interactions lower the energies of the [2+2]-photocycloaddition CIs and the repulsive steric interactions disfavor the 4π-disrotatory electrocyclic ring-opening pathway and CIs.

*Substituent effects on nonadiabatic molecular dynamics*

We performed 395, 372, and 382 ML-NAMD trajectories for **4**, **7**, and **10**, respectively. These trajectories enumerate the possible excited-state relaxation pathways for **4**, **7**, and **10** to evaluate the role of the substituent effects on the photodynamics. The ML-NAMD simulations employ the NN potential for on-the-fly prediction of energies and gradients. The trajectories are propagated from the $S_1$-FC points in 2 ps with a 0.5 fs time step. The technical details of the NAMD simulations and NN training are available in Computational Methods and Supporting Information. 6% and 8% of the trajectories of **4** and **7** (8%) remain on the $S_1$-surfaces at the end of the 2 ps simulations,

respectively. All trajectories of **10** finish on the ground states. The following discussions use the finished trajectories to study the substituent effects on the excited-state dynamics of **4**, **7**, and **10**.

We computed the $S_1 \rightarrow S_0$ relaxation rate in each reaction pathway to understand the substituent effects on the excited-state dynamics of [3]-ladderenes. Figure 8a shows four $S_1$ relaxation pathways we observed among the trajectories of **4**, corresponding to the [2+2]-photocycloaddition, 4π-disrotatory electrocyclic ring-opening, $\sigma_{C-C}$ cleavage, and 6π-conrotatory electrocyclic ring-opening.

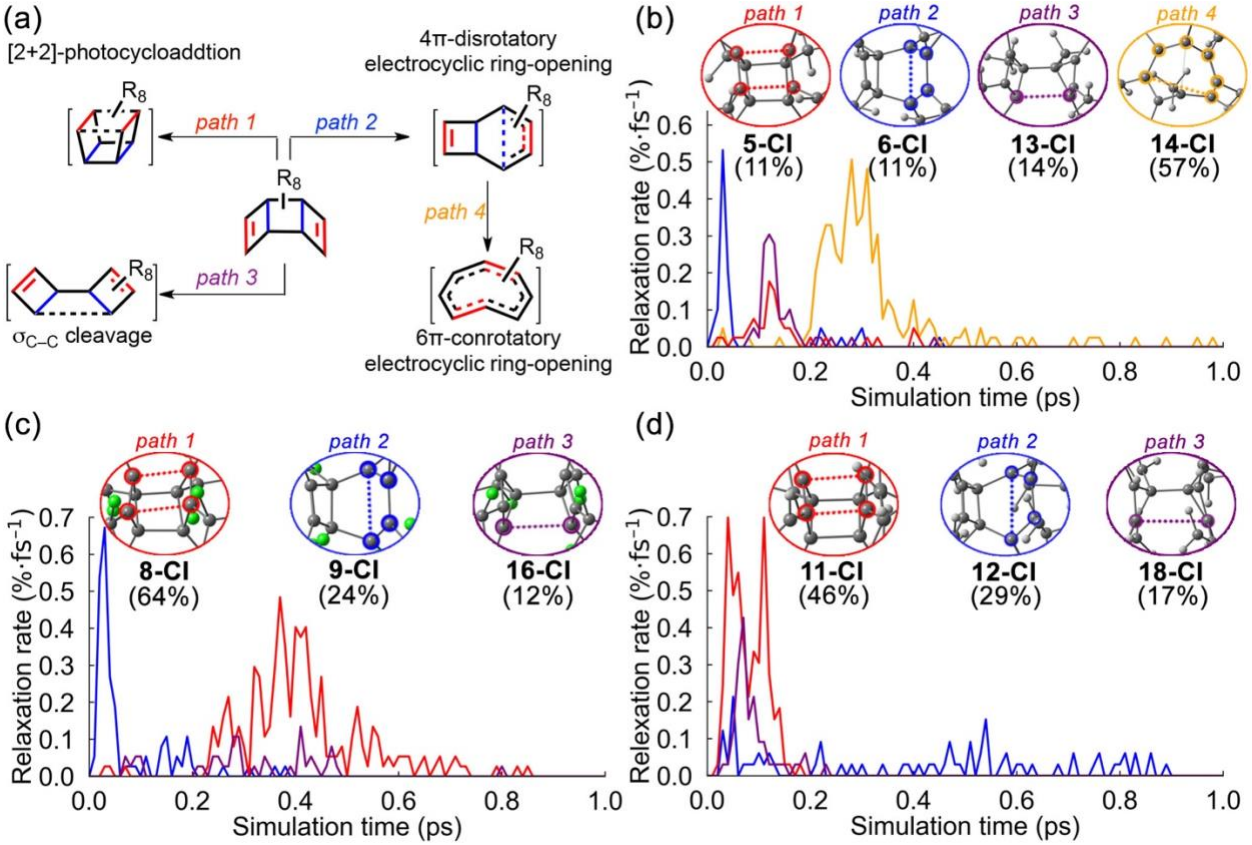


**Figure 8**. (a) Four $S_1/S_0$ conical intersections in the $S_1$ relaxation pathways of the substituted [3]-ladderenes. NN simulated relaxation rates of (b) $CH_3$, (c) $CF_3$, and (d) cPr substituted [3]-ladderene. The relaxation rate unit is defined as the percentage of trajectories passing through the surface hopping in each pathway per femtosecond. The rate curves of the [2+2]-photocycloaddition, 4π-disrotatory electrocyclic ring-opening, $\sigma_{C-C}$ cleavage, and 6π-conrotatory electrocyclic ring-opening are red, blue, purple, and orange, respectively. The inset legends are snapshots of surface hopping structures. The area under the relaxation rate curve determines the reaction selectivity, shown in the parentheses.

The 4π-disrotatory electrocyclic ring-opening is the fastest relaxation from **4-$S_1$-FC**, where 11% of the trajectories hop to the ground state showing a $S_1 \rightarrow S_0$ half-life, $t_{1/2}$ of 36 fs. 57% of the trajectories move to the crossing region of the 6π-conrotatory electrocyclic ring-opening with a longer half-life of 295 fs (Figure 8b). The remaining trajectories undergo the [2+2]-photocycloaddition (11%, $t_{1/2}$ = 130 fs) and $\sigma_{C-C}$ cleavage (11%, $t_{1/2}$ = 130 fs). The percentage of trajectories of **4** proceeding to the [2+2]-photocycloaddition pathway is 10 times larger than **1**

(0.9%[20]), suggesting that the $CH_3$ substantially increases the selectivity of [2+2]-photocycloaddition. However, the substituent effects on the ring-opening selectivity are negligible (68% in **4** *vs.* 69%[20] in **1**).

The 6π-conrotatory electrocyclic ring-opening pathway is not observed in the ML-NAMD simulations of **7** and **10** (Figure 8c and 8d). It suggests that the ring-opening pathways are less favorable, consistent with the higher energies of **9-Cl** and **12-Cl** due to the steric repulsion between substituent groups (Figure 8b). The ring-opening reactions involve the 4π-electrocyclic reaction, and the selectivity in **7** and **11** decreased to 24% and 29%, respectively. The [2+2]-photocycloaddition of **7** increases to 64%, but the half-life is 403 fs, which is longer than the 4π-disrotatory electrocyclic ring-opening ($t_{1/2}$ = 38 fs, Figure 8c). That implies the $CF_3$ enhances the [2+2]-photocycloaddition indirectly via turning down the selectivity of 6π-conrotatory electrocyclic ring-opening. Figure 8d shows the [2+2]-photocycloaddition selectivity of **10** is 46%. Moreover, the cPr reduces the half-life of the [2+2]-photocycloaddition to 79 fs. In contrast, the cPr slows down the 4π-disrotatory electrocyclic ring-opening showing a significantly longer half-life of 548 fs.

The $S_1$ relaxation pathway to the $S_0$ PES through a state-crossing spawns diverse subchannels to products. To quantify the substituent effects on the quantum yields of the products, we monitor the product distributions after the trajectories land on the $S_0$ surface. We defined three geometrical parameters to track the products of **4**, **7**, and **10** (Figure 9a).

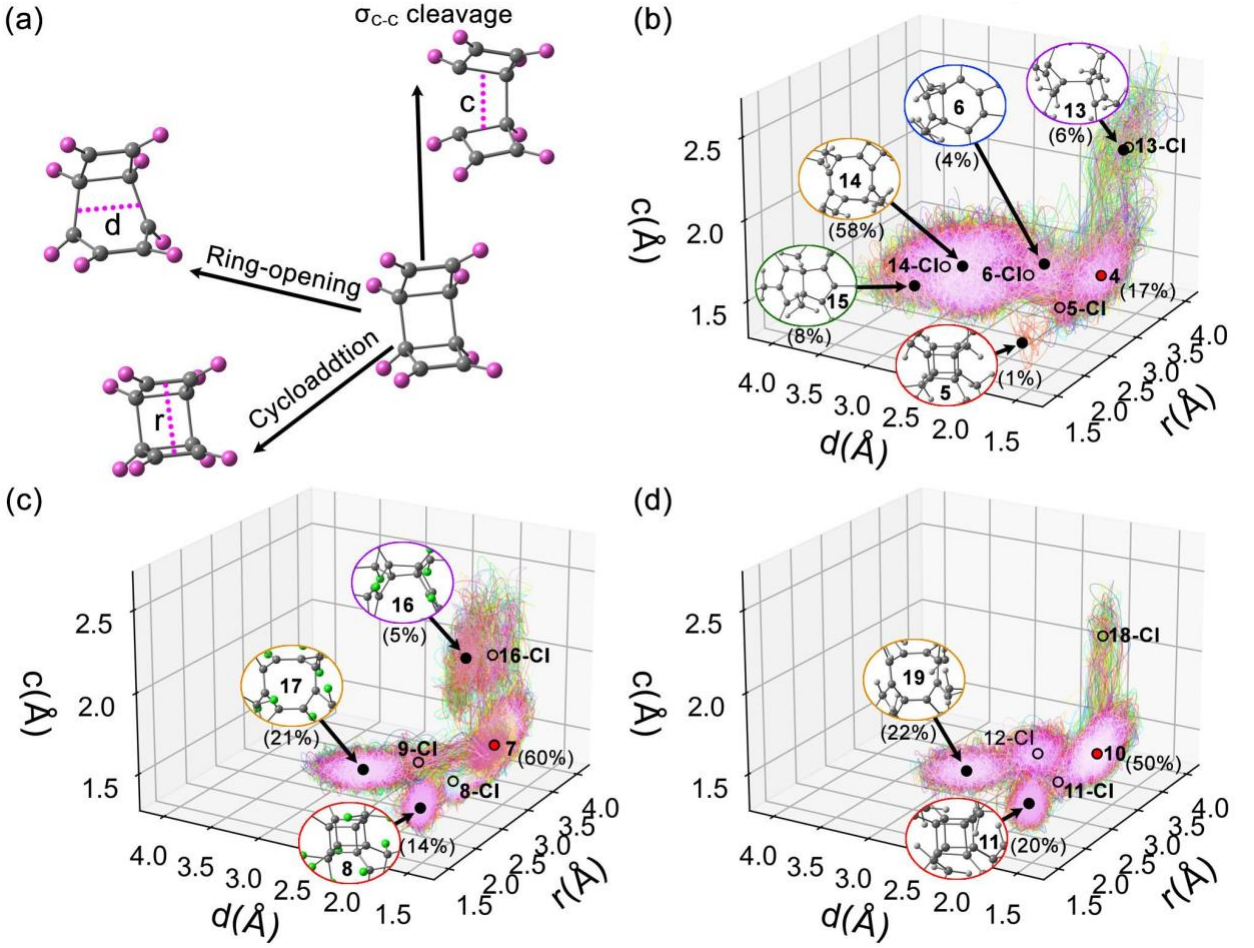

**Figure 9**. (a) Three geometrical parameters describe the reaction space of the substituted [3]-ladderene derivatives. The pink terminals denote the substituent groups. Plots of NN trajectories of (b) $CH_3$, (c) $CF_3$, and (d) cPr substituted [3]-ladderene. The averaged starting points, surface hopping points, and products are marked with red dots, black circles, and black dots. The inset legends illustrate the snapshot of the last trajectory point as the product with predicted quantum yields.

Figure 8b shows the trajectories of **4** beginning from the $S_1$-FC region (average values are *r* = 3.00 Å, *d* = 1.60 Å, *c* = 1.60 Å). The trajectories spread over the three coordinates leading to six products. The major product (58%) is **14** and is accessed via two channels: 1) a 6π-conrotatory electrocyclic ring-opening (49%) or 2) a 4π-disrotatory electrocyclic and 6π-disrotatory electrocyclic cascade reaction (9%). The 6π-conrotatory electrocyclic ring-opening can also promote the cascade reaction to **15** (8%). The bicyclic structure of **15** is a head-to-tail addition of the π-bonds (Figure S10). Considering the structural similarities, we expected **15** to be intermediate to octamethylsemibullvalene and other experimentally observed side products (Figure S10).[17] There is no trajectory reverting to **4** after the 4π- and 6π-disrotatory and conrotatory electrocyclic ring-opening. It agrees with the single-path branching plane of **6-CI** (Figure 6a). The $\sigma_{C-C}$ cleavage pathway bifurcates to **4** (8%) and **13** (6%). The biradical intermediate **13** connects two cyclobutenyl rings through a C–C bond; thus, the averaged distance *c* is elongated to 2.33 Å. The cyclobutenyl ring rotation of **13** possibly leads to the rearrangement of **4** or isomerization to the *anti*-configuration of **4** (Figure S10). The [2+2]-photocycloaddition shows 9% of the trajectories return to **4**, 1% of the trajectories produce **5**, which agrees with the experiments.[17] The resulting cubane formation ratio is 0.09. It suggests the $CH_3$ induces a limited effect shifting the reaction to **5**, although it increases the selectivity of [2+2]-photocycloaddition.

The trajectories of **7** show an averaged starting point at *r* = 3.01 Å, *d* = 1.60 Å, *c* = 1.63 Å (Figure 9c). In contrast to the $CH_3$, the $CF_3$ introduces stronger intramolecular closed-shell repulsions between the fluorines that disfavor the 6π-conrotatory electrocyclic ring-opening pathways. The ring-opening pathways prefer to proceed to **17** in 21% through a 4π-disrotatory and 6π-disrotatory electrocyclic ring-opening cascade reaction. The ratio between the yield of **17** and the percentage of the trajectories undergoing the ring-opening process (24%, Figure 8c) decreases to 0.88. It suggests the steric repulsions in the ring-opening of **7** perturb the PES to revert the reaction. The intramolecular π-π distances in **17** (*r* = 2.81 and *d* = 2.85 Å) is notably shorter than **14** (r = 3.06 and *d* = 3.09 Å), because the $CF_3$ compress the core to minimize the steric repulsion. The $\sigma_{C-C}$ cleavage shows reformation to **7** (7%) and rearrangement to **16** (5%). The [2+2]-photocycloaddition pathway leads to 50% of the trajectories returning to **7** and 14% of the trajectories leading to **8**. The cubane formation ratio thus increases to 0.22 because the $CF_3$ exhibits a stronger driving force to form cubanes than $CH_3$. (Figure 5).

The substitution with cPr shifts the trajectory origin point to *r* = 2.90 Å, *d* = 1.61 Å, *c* = 1.60 Å (Figure 9d). In the 4π-disrotatory electrocyclic ring-opening pathway, 22% of the trajectories produce **19** and 6% of the trajectories reform **10**. The conversion ratio decreases to 0.78, indicating that the cPr notably lowers the preference of the 4π-disrotatory electrocyclic ring-opening. The cPr also restrain the intramolecular π-π distances in **19** to *r* = 2.75 Å and *d* = 2.88 Å. In the $\sigma_{C-C}$ cleavage pathway, 17% of the trajectories all revert to **10**. It suggests the steric repulsions of the cPr further reduces the side reaction channels of **10**. In the [2+2]-

photocycloaddition pathway, 20% of the trajectories produce **11,** and 26% of the trajectories return to **10**. The cubane formation ratio increases to 0.43, consistent with $CF_3$; attractive dispersive interactions slant the potential energy surface to favor [2+2]-cycloaddition.

The trajectories of **4**, **7**, and **10** predict the yield of cubane following the trend of H<$CH_3$<$CF_3$<cPr. Thus, our results suggest that 1) the steric effects of a bulky group increase the preference of [2+2]-photocycloaddition by bringing down the ring-opening process; 2) the attractive dispersion interactions stabilize the cubane skeleton, which drives the [2+2]-photocycloaddition reaction toward cubanes.

## Concluding Remarks

The origin of the chemoselectivities was previously unknown because the experimental resolution of excited-state molecular structures on the sub picosecond timescale has not yet been realized. 1 picosecond multiconfigurational NAMD simulations are still prohibitively costly, but our ML-NAMD technique has enabled the photodynamics simulations for these octa-substituted [3]-ladderenes for the first time. We have integrated the static single and multireference calculations and ML-NAMD simulations to understand how substituents control the competing reaction pathways of [3]-ladderenes. Our static calculations show that intramolecular repulsive steric clashes distort the geometries and raise the the conical intersection energies by 0.72–1.15 eV and reaction energies by 0.68–2.34 eV in the 4π-disrotatory electrocyclic ring-opening pathways. The attractive dispersion interactions lower the energies of the [2+2]-photocycloaddition conical intersections by 0.31–0.85 eV and cubanes 0.03–0.82 eV.

The ML-NAMD photodynamics simulations of **4**, **7**, and **10** have enumerated the bifurcated photochemical pathways and validated the role of the substituents on the chemoselectivities. The ML-NAMD results show that the 4π-disrotatory electrocyclic ring-opening has the shortest $S_1$ half-life in **4** and **7**, whereas the [2+2]-photocycloaddition is faster in **10**. The 6π-conrotatory electrocyclic ring-opening (57%), $\sigma_{C-C}$ cleavage (14%), and 4π-disrotatory electrocyclic ring-opening (11%) of **4** significantly reduced the selectivity of the [2+2]-photocycloaddition (11%). In **7** and **10**, the steric effects of $CF_3$ and cPr groups completely block the 6π-conrotatory electrocyclic ring-opening pathways; 64% and 46% of the trajectories proceed to the [2+2]-photocycloaddition products, respectively. The conversion ratio of [2+2]-photocycloaddition to cubane increased from 0.09 in **4** to 0.22 in **7** and 0.43 in **10**. The predicted cubane yields increase in the order of H (0.4%) < $CH_3$ (1%) < $CF_3$ (14%) < cPr (20%), which reaches excellent agreement with the experiment. We demonstrate that bulky substituents lower the cubane energy and or raise the energies of the electrocyclic ring-opening products. These effects directly and or indirectly increases the cubane yields. This finding provides a solid theoretical base for the ongoing experimental development of substituent-assisted [2+2]-photocycloaddition of [3]-ladderene toward cubanes.

## Computational Methods

*Electronic structure calculations*

The electronic structure calculations of the reactants and products use the (time-dependent) density functional theory (DFT) methods implemented in ORCA 4.2.0.[27] The geometry optimization is performed with PBE0 functional[28-29] and cc-pVDZ basis set.[30] Frequency

calculations confirm the local minimum with no imaginary frequencies. The excitation and electronic energies calculations utilize the range-separate ωB97XD functional[31] and cc-pVTZ basis set.[30] The DFT calculations are accelerated with the Resolution of Identity (RI) approximation using the RIJCOSX method for electron integration.[32]

The conical intersection optimization and NN potential training data calculations use the complete active space self-consistent field (CASSCF) theory implemented in OpenMolcas 19.11.[33] The active space includes 8 electrons and 7 orbitals (4 electrons from two σ- and two $\sigma^*$-orbital of the bridge C–C bond and 4 electrons from two π- and one $\pi^*$-orbital of the vinyl bond. The highest out-of-phase $\pi^*$-orbital is removed to ensure consistent state ordering to CASPT2 level.[20] To reduce the computational cost, we describe the 8 carbon atoms in the [3]-ladderene core with the ANO-S-VDZP basis set.[34-37] The other atoms in the substituent groups use the ANO-S-MB basis set,[34-37] denoted as ANO-S-VDZP+ANO-S-MB(R), where R means for the substituent groups. The orbital optimization averages the first 5 states. Note the reactant geometries are re-optimized to compute the relative energy of the conical intersection. The energies of the reactant and conical intersections are corrected by the complete active space second-order perturbation theory (CASPT2) with 0.2 Hartree energy shift to avoid intruder state.[38-39]

*Machine learning methods*

We built the feedforward neural network potential using the multiple perception layers of the TensorFlow/Keras API for Python. The NN potential is implemented in an open resource code, Python Rapid Artificial Intelligence Ab Initio Molecular Dynamics (PyRAI$^2$MD).[23] The NN potential takes inverse distance as input and predicts the energies and forces for nonadiabatic molecular dynamics. The NN permutes the atom index to exploit the molecular symmetry, reducing the size and computational cost of training data. The predicted potential energy and its first-order derivatives are used to train energies and forces together. The loss function weights the energies and forces in a ratio of 5: 1. The optimal NN hyperparameter is optimized using grid-search. The validation means absolute energy errors are 0.040–0.046, 0.032–0.034, and 0.037–0.039 eV for R = $CH_3$, $CF_3$, and cPr, respectively, satisfying the standard of chemical accuracy (0.043 eV). Details of NN construction, optimization, training, and validation errors are available in Supporting Information.

The training set contains 9303, 6659, and 7697 data points for R = $CH_3$, $CF_3$, and cPr, respectively. The training data generation includes an initial set and adaptive sampling.[21, 23, 40] In the first step, an initial set is prepared by geometrical interpolations from the optimized reactant, conical interaction, and product of the [2+2]-photocycloaddition reaction. Next, we apply the geometrical distortions sampled from the molecular vibrations in the Wigner distribution at the zero-point energy level to the interpolated geometries, which further expands the visited configurations space. The second step trains prior NN potentials using the initial set. It runs 125 trajectories to search out-of-sampled structures by querying a committee of two distinct NN potentials, determined by the standard deviations of the predicted energies and forces. Finally, the collected structures are re-computed with CASSCF calculation and merged into the initial set. The new NN potentials continue to search undersampled structures in the next iteration until the configuration space is adequately sampled. More information on the training data generation is available in the Supporting Information.

*Nonadiabatic molecular dynamics simulations*

The NAMD simulations propagate the trajectories from the $S_1$-FC points. The initial conditions (nuclear positions and momenta) are sampled using the Wigner distribution at the zero-point energy level. The simulation time is 2 ps, and the time step is 0.5 fs. The trajectories are in the microcanonical ensemble (NVE) as the total energy is conserved. We choose the Zhu-Nakamura theory of surface hopping (ZNSH) to compute the surface hopping probability. The ZNSH does not require nonadiabatic coupling and shows consistent performance to Tully's fewest switches surface hopping (FSSH) in our previous study.[23] An energy gap of 1.5 eV is used to determine surface hopping events. The ML-NAMD method is implemented in the PyRAI$^2$MD.[23]

**Acknowledgment**

J.L., R.S., D.A., and S.A.L. acknowledge the Office of Naval Research (ONR N00014-18-1-2659) for funding. J.L., R.S., D.A., and S.A.L. acknowledge the National Science Foundation (NSF-OAC-1940307 and NSF DBI-2031778) for funding this research. J.L., R.S., D.A., and S.A.L. appreciate the assistance from the Northeastern Research Computing Team and access to the computing resources of the Discovery cluster.

**Supporting Information**

Neural network architectures and training; training data generation and adaptive sampling; optimized geometries and structural parameters; frontier molecular orbitals; TD-DFT excitation energies; active spaces; conical intersection branching planes; electronic energies and Cartesian coordinates.

**Reference**

1. Karkas, M. D.; Porco, J. A., Jr.; Stephenson, C. R., Photochemical Approaches to Complex Chemotypes: Applications in Natural Product Synthesis. *Chem. Rev.* **2016,** *116* (17), 9683-747.
2. Cambie, D.; Bottecchia, C.; Straathof, N. J.; Hessel, V.; Noel, T., Applications of Continuous-Flow Photochemistry in Organic Synthesis, Material Science, and Water Treatment. *Chem. Rev.* **2016,** *116* (17), 10276-341.
3. Calbo, J.; Weston, C. E.; White, A. J.; Rzepa, H. S.; Contreras-Garcia, J.; Fuchter, M. J., Tuning Azoheteroarene Photoswitch Performance through Heteroaryl Design. *J. Am. Chem. Soc.* **2017,** *139* (3), 1261-1274.
4. Saydjari, A. K.; Weis, P.; Wu, S., Spanning the Solar Spectrum: Azopolymer Solar Thermal Fuels for Simultaneous UV and Visible Light Storage. *Adv. Energy Mater.* **2017,** *7* (3).
5. Hull, K.; Morstein, J.; Trauner, D., In Vivo Photopharmacology. *Chem. Rev.* **2018,** *118* (21), 10710-10747.
6. Poplata, S.; Troster, A.; Zou, Y. Q.; Bach, T., Recent Advances in the Synthesis of Cyclobutanes by Olefin [2+2] Photocycloaddition Reactions. *Chem. Rev.* **2016,** *116* (17), 9748-815.
7. Boswell, B. R.; Mansson, C. M. F.; Cox, J. M.; Jin, Z.; Romaniuk, J. A. H.; Lindquist, K. P.; Cegelski, L.; Xia, Y.; Lopez, S. A.; Burns, N. Z., Mechanochemical synthesis of an elusive fluorinated polyacetylene. *Nat Chem* **2021,** *13* (1), 41-46.
8. Eaton, P. E., Cubanes: Starting Materials for the Chemistry of the 1990s and the New Century. *Angew. Chem. Int. Ed.* **1992,** *31* (11), 1421-1436.

9. Biegasiewicz, K. F.; Griffiths, J. R.; Savage, G. P.; Tsanaktsidis, J.; Priefer, R., Cubane: 50 years later. *Chem. Rev.* **2015,** *115* (14), 6719-45.
10. Li, Z.; Anderson, S. L., Pyrolysis Chemistry of Cubane and Methylcubane: The Effect of Methyl Substitution on Stability and Product Branching. *The Journal of Physical Chemistry A* **2003,** *107* (8), 1162-1174.
11. Martin, H.-D.; Urbanek, T.; Pföhler, P.; Walsh, R., The pyrolysis of cubane; an example of a thermally induced hot molecule reaction. *J. Chem. Soc., Chem. Commun.* **1985,** (14), 964-965.
12. Huang, H. T.; Zhu, L.; Ward, M. D.; Chaloux, B. L.; Hrubiak, R.; Epshteyn, A.; Badding, J. V.; Strobel, T. A., Surprising Stability of Cubane under Extreme Pressure. *J Phys Chem Lett* **2018,** *9* (8), 2031-2037.
13. Eaton, P. E.; Cole, T. W., The Cubane System. *J. Am. Chem. Soc.* **1964,** *86* (5), 962-964.
14. Falkiner, M. J.; Littler, S. W.; McRae, K. J.; Savage, G. P.; Tsanaktsidis, J., Pilot-Scale Production of Dimethyl 1,4-Cubanedicarboxylate. *Org. Process Res. Dev.* **2013,** *17* (12), 1503-1509.
15. Gleiter, R.; Pfeifer, K.-H.; Koch, W., Propella[34] prismane and its congeners: A MO-theoretical study. *J. Comput. Chem.* **1995,** *16* (1), 31-36.
16. Woodward, R. B.; Hoffmann, R., The Conservation of Orbital Symmetry. *Angew. Chem. Int. Ed.* **1969,** *8* (11), 781-853.
17. Gleiter, R.; Brand, S., Photochemistry of Bridged and Unbridged Octaalkyl-Substituted syn-Tricyclo[4.2.0.02,5]octa-3,7-diene Derivatives. *Chem. Eur. J.* **1998,** *4* (12), 2532-2538.
18. Pelosi, L. F.; Miller, W. T., Syntheses from perfluoro-2-butyne. 2. Perfluorooctamethylcubane, perfluorooctamethylcuneane, and perfluorooctamethylcyclooctatetraene. *J. Am. Chem. Soc.* **1976,** *98* (14), 4311-4312.
19. de Meijere, A.; Redlich, S.; Frank, D.; Magull, J.; Hofmeister, A.; Menzel, H.; Konig, B.; Svoboda, J., Octacyclopropylcubane and some of its isomers. *Angew. Chem. Int. Ed. Engl.* **2007,** *46* (24), 4574-6.
20. Li, J.; Lopez, S. A., Multiconfigurational Calculations and Nonadiabatic Molecular Dynamics Explain Tricyclooctadiene Photochemical Chemoselectivity. *J. Phys. Chem. A* **2020,** *124* (38), 7623-7632.
21. 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.
22. 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.
23. Li, J.; Reiser, P.; Boswell, B. R.; Eberhard, A.; Burns, N. Z.; Friederich, P.; Lopez, S. A., Automatic discovery of photoisomerization mechanisms with nanosecond machine learning photodynamics simulations. *Chem. Sci.* **2021,** *12* (14), 5302-5314.
24. Wagner, J. P.; Schreiner, P. R., London dispersion in molecular chemistry--reconsidering steric effects. *Angew. Chem. Int. Ed. Engl.* **2015,** *54* (42), 12274-96.
25. Grimme, S.; Schreiner, P. R., Steric crowding can stabilize a labile molecule: solving the hexaphenylethane riddle. *Angew. Chem. Int. Ed. Engl.* **2011,** *50* (52), 12639-42.
26. Garcia, L. X.; Bauza, A.; Seth, S. K.; Frontera, A., Importance of R-CF3...O Tetrel Bonding Interactions in Biological Systems. *J. Phys. Chem. A* **2017,** *121* (28), 5371-5376.
27. Neese, F., The ORCA program system. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012,** *2* (1), 73-78.
28. Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996,** *77* (18), 3865-3868.

29. Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized Gradient Approximation Made Simple [Phys. Rev. Lett. 77, 3865 (1996)]. *Phys. Rev. Lett.* **1997,** *78* (7), 1396-1396.
30. 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.
31. Chai, J. D.; Head-Gordon, M., Systematic optimization of long-range corrected hybrid density functionals. *J. Chem. Phys.* **2008,** *128* (8), 084106.
32. Weigend, F., Accurate Coulomb-fitting basis sets for H to Rn. *Phys. Chem. Chem. Phys.* **2006,** *8* (9), 1057-65.
33. 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.
34. 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.
35. 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.
36. 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.
37. 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.
38. Roos, B. O.; Andersson, K., Multiconfigurational perturbation theory with level shift — the Cr2 potential revisited. *Chem. Phys. Lett.* **1995,** *245* (2-3), 215-223.
39. Roos, B. O.; Andersson, K.; Fülscher, M. P.; Serrano-Andrés, L.; Pierloot, K.; Merchán, M.; Molina, V., Applications of level shift corrected perturbation theory in electronic spectroscopy. *J. MOL. STRUC-THEOCHEM* **1996,** *388*, 257-276.
40. Gastegger, M.; Behler, J.; Marquetand, P., Machine learning molecular dynamics for the simulation of infrared spectra. *Chem. Sci.* **2017,** *8* (10), 6924-6935.