

Exploring the Global Reaction Coordinate for Retinal Photoisomerization: A Graph Theory-Based Machine Learning Approach

Goran Giudetti,[†] Madhubani Mukherjee,[†] Samprita Nandi, Sraddha Agrawal, Oleg V. Prezhdo, and Aiichiro Nakano*



Cite This: *J. Chem. Inf. Model.* 2024, 64, 7027–7034



Read Online

ACCESS |



Metrics & More

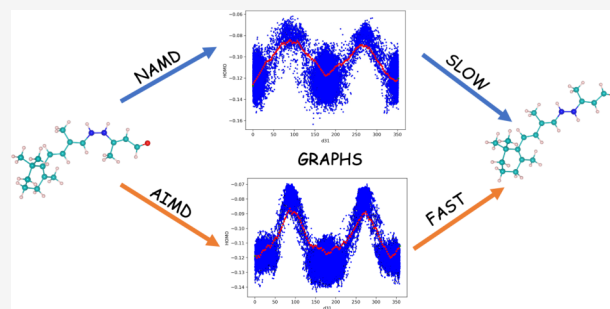


Article Recommendations



Supporting Information

ABSTRACT: Unraveling the reaction pathway of photoinduced reactions poses a great challenge owing to its complexity. Recently, graph theory-based machine learning combined with nonadiabatic molecular dynamics (NAMD) has been applied to obtain the global reaction coordinate of the photoisomerization of azobenzene. However, NAMD simulations are computationally expensive as they require calculating the nonadiabatic coupling vectors at each time step. Here, we showed that ab initio molecular dynamics (AIMD) can be used as an alternative to NAMD by choosing an appropriate initial condition for the simulation. We applied our methodology to determine a plausible global reaction coordinate of retinal photoisomerization, which is essential for human vision. On rank-ordering the internal coordinates, based on the mutual information (MI) between the internal coordinates and the HOMO energy, NAMD and AIMD give a similar trend. Our results demonstrate that our AIMD-based machine learning protocol for retinal is 1.5 times faster than that of NAMD to study reaction coordinates.



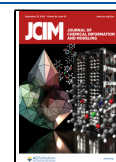
1. INTRODUCTION

To develop a clear understanding of a reaction pathway, we need the potential energy surface (PES)^{1,2} of molecules. PESs form a central concept in the application of electronic structure methods to the study of the structures, properties, and reactivities of a molecular system. We can study the dynamics of the molecule, under the Born–Oppenheimer approximation (BO),^{3–5} on a single potential energy surface with the aid of the ab initio molecular dynamics (AIMD)^{6,7} method. In the AIMD method, the potential energy and its derivatives are evaluated “on the fly”, as needed for the integration of the equations of motion of the system. Since BO approximation breaks down whenever two or more electronic states have a small energy gap, AIMD cannot be used for simulating nonadiabatic processes.^{8,9} Among such processes, we find radiationless decay, intramolecular energy and charge transfer and most photochemical reactions. One of the most frequently used techniques to simulate the dynamics of complex photochemical reactions is nonadiabatic molecular dynamics (NAMD) using the fewest switches surface hopping (FSSH).^{10,11} In this technique, the forces are computed as the gradients of single BO PESs and the electronic wave function is expanded in terms of adiabatic eigenfunctions. The expansion coefficients are evaluated by solving the time-dependent Schrödinger equation using NA coupling terms. The hopping probability between the electronic adiabatic states depends on the electronic amplitudes as well as

nonadiabatic coupling terms. Due to the multistate and multidimensional nature of photoinduced reactions, computation of the PESs and of the reaction coordinate is highly demanding.

The knowledge of the reaction coordinates¹² provides the fundamental details of the underlying mechanisms of a given chemical transformation. Very often, the definition of a reaction path is based on intuitive considerations. Only recently, attention was attracted to systematic approaches for selecting appropriate variables and mapping them onto multistep kinetics. The search for reaction coordinates usually involves obtaining and analyzing large amounts of data from molecular dynamics simulations. This approach considers every possible internal coordinate as a candidate for the reaction coordinate. However, accounting for all of the internal coordinates in the global reaction coordinate is not feasible except for small molecules. Machine learning-based tools can be very helpful in determining the most important internal coordinate involved in the reaction mechanism. Tavazde et

Received: February 26, 2024
Revised: September 3, 2024
Accepted: September 4, 2024
Published: September 11, 2024



al.¹³ have used a graph-based technique in the search for the reaction coordinate of cis–trans isomerization of azobenzene and shown that the C–N=N–C dihedral angle corresponds to the reaction coordinate.

In this work, we model the photoinduced isomerization of retinal, the chromophore present in the human photoreceptor, upon excitation to the lowest and brightest excited states. Photoreceptors consist of a transmembrane protein and a special light-absorbing chromophore that is covalently attached to the associated photosensory domain in the protein.^{14,15} The primary step in visual transduction is initiated when the chromophore in the photoreceptor absorbs light of specific wavelengths.¹⁶ The energy provided by the photon causes the chromophore to undergo photoisomerization, leading to structural changes in the protein through allosteric interactions. The conformation changes in the photoreceptors trigger a chemical signaling cascade that initiates visual phototransduction by the human brain.¹⁷ In the initial stage of vision, retinal is found in the 11-cis-retinal conformation, which isomerizes to an all-trans-retinal isomer upon capturing a photon^{14,15,18} (see Figure 1). Photoexcitation causes an

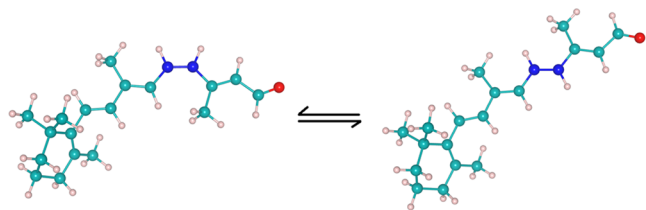


Figure 1. Retinal phototransformation: transition from 11-cis-retinal conformation to all-trans-retinal conformation, C in cyan (except C11=C12 in blue), H in pink, and O in red.

electronic transition that opens the double bond, creating a temporary single bond that can rotate freely around its axis. Once the excited retinal decays to the ground state, the double bond reforms and locks the molecule into the trans configuration. This photoisomerization process is an ultrafast event that occurs in a matter of femtoseconds,^{16,19,20} making it one of the fastest reactions in nature. As a consequence of isomerization, the chromophore activates the opsin protein which eventually sends a chemical signal to the visual cortex.¹⁷ Thus, understanding the mechanism of retinal photoisomerization is fundamental and crucial for the development of therapies for various visual disorders. In this contribution, we aim to model the pathway of the photoisomerization reaction of retinals with the aid of concepts from machine learning such as MI and graph representations. Photoisomerization in retinal was simulated using AIMD and NAMD. We attempt to gain insights into the reaction mechanism using an alternative route to the conventional *ab initio* methods. The goal of the current study is to compare the reaction path obtained with computationally efficient AIMD simulations to that obtained with more compute-intensive NAMD simulations. Thus, we chose a system for which the reaction path is already known. Application of the proposed method to more general cases remains a future study.

Herein, we report on a global reaction coordinate (containing all internal coordinates of retinal) to describe the reaction mechanism of the photoisomerization of retinal. Our global coordinate comprises all internal coordinates (bonds, angles, and dihedral angles) of a retinal molecule. Density

functional theory (DFT) and molecular dynamics provide the electronic and dynamic properties of a molecule at the atomic scale. We obtained the internal coordinates from AIMD and NAMD simulations and calculated the MI between the energy of the highest occupied molecular orbital (HOMO) and the internal coordinates collected from the trajectories. Thus, we rank order all of these coordinates and quantify their contribution to the reaction mechanism of retinal photoisomerization. We construct a graph-based network, where each node is represented by the HOMO energy and the corresponding internal coordinates. Nodes are then connected with edges whose weight is determined using an expression that takes into account energy difference and the MI of the internal coordinates (eq 1). The global reaction coordinates are determined along the network following the paths of the least action. Here, we showed that the most important internal coordinate and the reaction path computed with the graph do not depend on the sampling scheme.

2. THEORETICAL METHODS

We generated configurations with AIMD and NAMD methods, which were passed for calculating the mutual information between the internal coordinates and the HOMO energy at each time step of the simulations. Mutual Information (MI)²¹ is a quantity that measures the relationship between two random data sets that are sampled simultaneously. It also captures the nonlinear correlation between two random data sets. When there is a big data set with a large range of features, MI can help to select a subset of the most crucial features in order to discard the irrelevant ones. Thus, it can be used as an important tool for feature selection.²² For example, it has been used to determine the structure–property relationship in nanomaterials.^{23,24} To calculate the MIs, we first built the data sets from the molecular dynamics simulations and defined the internal coordinates as the features of our system. The data sets include information on all the internal coordinates such as bonds, angles, and dihedrals, at each time step.

To understand MI²⁵ mathematically, let us consider any two random variable sets, X and Y , each with its own probability distribution. In order to evaluate the correlation between these random variable sets, one begins by measuring how similar the joint distribution $p(X, Y)$ is to the factored distribution $p(X)p(Y)$. If X and Y are two independent sets, that is, $p(x, y) = p(x)p(y)$ where $x \in X$ and $y \in Y$, then the MI is zero. MI between two random variable sets X and Y is given by

$$MI(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (1)$$

To check for the correlation between the internal coordinates in our data set during the course of the reaction, we have built the correlation matrix, $X^T X$, where X is a matrix of all the features. We adapted the idea of graph theory and applied it to find the shortest path between the nodes. Each configuration along the trajectories is represented as a node in the graph. We used the Networkx python package²⁶ to find the shortest path with the help of Dijkstra's algorithm.^{27,28}

3. COMPUTATIONAL DETAILS

We optimized the geometry of the trans-retinal at the ω B97X-D/6-31+G* level of theory. We generated two sets of

configurations of retinal. In one set, configurations are extracted from AIMD trajectories, while in the other set, configurations are obtained from NAMD trajectories using Tully's scheme of FSSH.¹⁰

We plotted the HOMO and LUMO in trans-retinal values calculated with the PBE0/STO-3G level of theory (Figure 2).

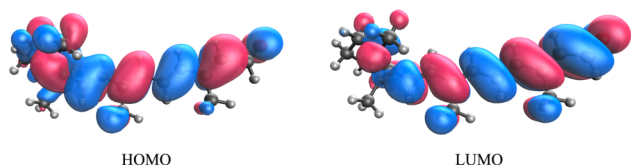


Figure 2. HOMO and LUMO of trans-retinal calculated with the PBE0/STO-3G level of theory.

We can see that the HOMO is a π orbital, whereas the LUMO is of π^* character. The natural transition orbitals (PBE0/STO-3G) of the S_1 and S_2 transitions (Figure 3) show that the S_1

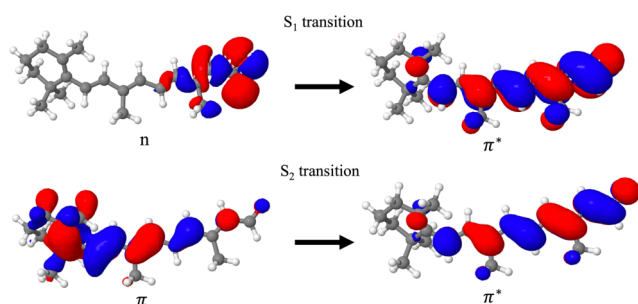


Figure 3. Natural transition orbitals representing transitions from the ground state to the S_1 and the S_2 state in trans-retinal. Holes are shown on the left, and particles are on the right.

transition is the $n\pi^*$ type and the S_2 transition is the $\pi\pi^*$ type. In the photoisomerization of retinals, the S_2 state is the optically active excited state. Thus, we performed both our NAMD and AIMD simulations on the S_2 surface for the final analysis. The graph-based approach requires the calculation of the MI between internal coordinates and the energy gap between the orbitals involved in the transition that allows the isomerization of retinal. With a minimal basis set, these orbitals correspond to the HOMO and LUMO as we show by comparing them to natural transition orbitals. Using a larger basis with polarization and diffuse functions would change the nature of the virtual orbitals, that is, the transition is not between HOMO and LUMO. This poses an additional challenge in identifying the proper unoccupied orbital at each time step in trajectories. Another reason for the choice of the basis relates to the computational cost of simulations. Since many geometries need to be generated/sampled, we opted for the smallest standard basis available in Q-Chem. At the same time, the accuracy of the ab initio calculations does not affect the quality of the reaction path. While the minimal basis set was sufficient for the chosen prototypical photoisomerization, we do recommend using basis sets of appropriate size after careful inspection of the orbitals associated with the system at hand.

We have taken the two transition states for the isomerization, with d31 (the main dihedral angle along which the photoisomerization occurs, shown in Figure S1 and Table S1, equal to 90° and 270° , respectively, as starting points for

AIMD simulations. We ran 125 simulations for each configuration (for a total of 250). We computed 200 AIMD steps for each trajectory using a step size of 40 au (~ 1 fs) with TDDFT at the PBE0/STO-3G level of theory with Q-Chem.²⁹ These initial geometries allowed us to sample the TSs in retinal isomerization while ensuring that half of them end up in the cis basin and the other half in trans. We observed that such a short time was sufficient for the molecule to reach the cis or trans basins. During the time evolution of the system, propagated with the velocity-Verlet algorithm, atoms were allowed to exchange energy with a Nose–Hoover thermostat at 300 K. All AIMD trajectories were computed on the second excited state of retinal for which the central double bond of d31 opens due to a $\pi\pi^*$ transition.

The NAMD simulations were run with TDDFT at the PBE0/STO-3G level of theory using the FSSH scheme without including decoherence with the PySurf³⁰ software package. To validate the robustness of the graph analysis, we ran NAMD starting at both S_1 and S_2 states, with the nonadiabatic coupling (NAC) vectors being computed with Q-Chem. The results of NAMD at the S_1 state are shown in Supporting Information. A total of 50 simulations (25 from cis and 25 from trans) were performed with the maximum initial population in the lowest singlet excited state. Since we did not observe any photoisomerization happening during the runs (1000 steps with a time step of 40 au), we sampled the TS in the same way it was sampled for AIMD, that is, by starting 200 additional simulations from TS geometries and propagating for 30 steps with a time step of 40 au. We ended up with 50000 data points from AIMD and 56000 from NAMD simulations, which were used to calculate the MI between the energy of the HOMO and all the internal coordinates. Overall, the NAMD data points comprise 25257 configurations with d31 values between $0^\circ \pm 45^\circ$ (cis isomer) and 25460 configurations with d31 values between $180^\circ \pm 45^\circ$ (trans isomer). We constructed a graph with these data using the values of HOMO energy and all the internal coordinates as attributes of each node. In this way, each node represents a step in the collection of trajectories generated by the dynamics. Mutual information was used as the weight of the paths connecting the nodes in the graph. While MI can be used as such, one needs to eliminate potential redundancies. To better understand how the internal coordinates are related to one another, we plotted the correlation matrices using three data sets.

We constructed the graphs in two steps. First, we connected the nodes for which the d31 internal coordinate differs by 3° and the HOMO energy increases while moving from one node to the next one. We could join two sections in our graph with these criteria, that is, the nodes between 0° and 90° and 180° and 270° . In the next step, we kept the constraint on the dihedral the same as in the last step but followed the nodes where the difference in HOMO energy decreases on going from one node to the other. Thus, we constructed the entire graph in the range 0° to 360° . After the graph was constructed, we followed the expression of the weight in eq 2 to find the shortest path from the cis conformer to the trans conformer using the Dijkstra algorithm. Figure 4 shows a flowchart of our workflow. In eq 2, the summation runs over the top N internal coordinates ranked by their MI value. Δf_i is the difference in internal coordinate value between the connected nodes, and m_i is the MI value associated with the internal coordinate considered.

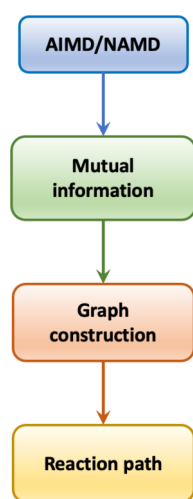


Figure 4. Flowchart of the computational protocol described in this text.

$$w_{pq} = |\Delta E_{pq}| \sum_{i=1}^N m_i |\Delta f_i^{pq}| \quad (2)$$

4. RESULTS AND DISCUSSION

Figure 5 shows the energies of HOMO and LUMO as a function of d31 from the NAMD simulations. The gap between HOMO and LUMO decreases at the transition state, that is, when the dihedral angle d31 reaches 90° and 270°. This is typical for molecules undergoing photoisomerization reaction, and our NAMD simulation corroborates this pattern.

We used the data from AIMD and NAMD simulations to calculate the mutual information between the internal coordinates and the HOMO energy. With the NAMD data, we compared two cases where we considered (i) the full NAMD data set and (ii) only data points that are close to the transition state. Thus, we could investigate the effect of oversampling the cis and trans basins with NAMD simulations.

Figure 6 (top) shows the mutual information between the energy of HOMO and internal coordinates obtained from the AIMD simulation. We have chosen internal coordinates that are in proximity to the main dihedral responsible for

photoisomerization. From Figure 6 (bottom), one can see that all the internal coordinates we have chosen are highly correlated with the energy of the HOMO at each time step of the NAMD simulations with d31 and d13 being associated with the highest mutual information value. The MI plot with an AIMD data set also tells us that the dihedral angles closer to the reaction center play a more important role than the other internal coordinates. d13 and d31 represent dihedral C–C11=C12–C and H–C11=C12–C respectively, which rotate during the photoisomerization process. Figure 6 (top) describes the mutual information with the AIMD data set. We can see that in this case, all the internal coordinates close to d31 are contributing similarly. A similar situation arises for the truncated NAMD data set. We can explain the difference in the mutual information between these three data sets by considering the oversampling of the cis and trans configurations in the NAMD simulations. To generate the AIMD data set and the selected NAMD data set, we started the simulations from the initial configurations close to the transition state. Thus, the resemblance between the mutual information plots of the truncated NAMD data set and the AIMD data set is expected. We are interested in identifying which other internal coordinates may be involved in the isomerization mechanism, apart from d31/d13. For this reason, we build the correlation matrix from our AIMD data set. This is done by normalizing the initial data and calculating the matrix of the covariances between each internal coordinate. Performing this procedure on the NAMD data would show that all features, including d13 and d31, are uncorrelated due to the fact that the majority of the data points are for structures of the cis or trans isomer instead of the TS. We found that avoiding oversampling the cis and trans isomers during NAMD (Figure 7, middle) results in a correlation matrix similar to the one obtained from AIMD simulations (Figure 7, top). If we want to compare the correlation matrices obtained from AIMD and NAMD simulations, we need to change our NAMD data set in such a way that it contains the majority of data points close to the transition state.

The correlation matrix (Figure 7, top) obtained from AIMD data shows different levels of correlation and anticorrelation between pairs of internal coordinates. d13 and d31 are anticorrelated to each other which is expected as they are alternate angles. In a similar manner, r12 and r13 are

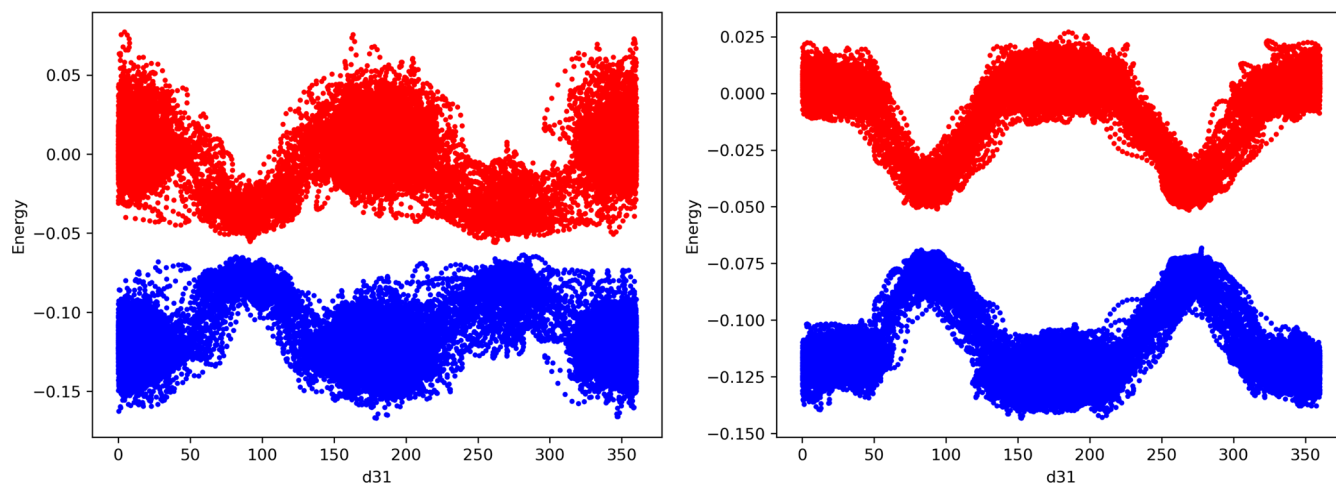


Figure 5. Energy in eV of HOMO (blue) and LUMO (red) as function d31 from NAMD (left) and AIMD (right) simulations.

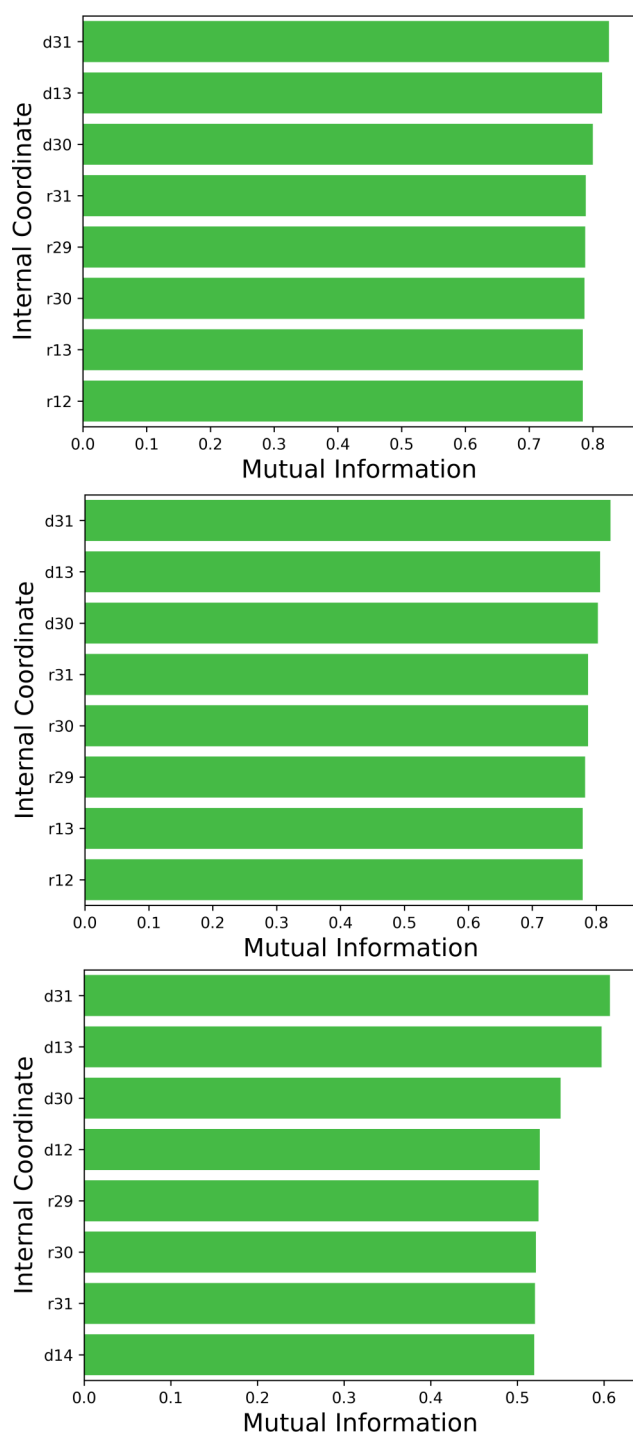


Figure 6. Mutual information between internal coordinates and HOMO energy from the AIMD data set (top), selected NAMD data set (middle), and full NAMD data set (bottom).

anticorrelated, which is coherent with an asynchronous stretching of these bonds. The positive correlation between r30 and r31 suggests a synchronous vibration of the C11–H30 and C12–H31 bonds.

We verified our AIMD results after thoroughly comparing the mutual information and correlation matrices between AIMD and NAMD data sets and decided to perform the graph analysis using the mutual information from the AIMD data. The graphs were constructed with both AIMD and NAMD data sets. Five internal coordinates with the highest MI values

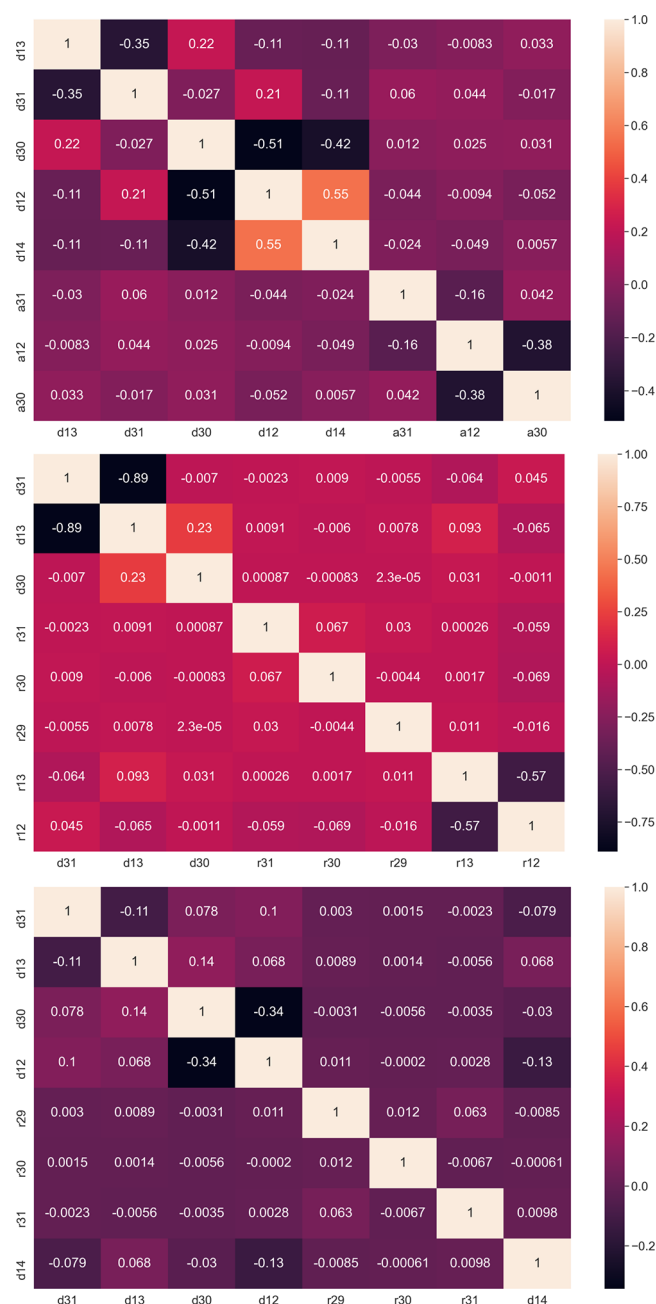


Figure 7. Correlation matrix obtained from AIMD data (top), selected NAMD data (middle), and full NAMD data (bottom). The color bar denotes the correlation between internal coordinates.

from the AIMD data set were chosen for the construction of the graph. We have plotted the HOMO energy as a function of the dihedral angles obtained from the graph with NAMD data (Figure 8). The HOMO energy increases in the beginning and reaches its highest when d31 is close to 90° and starts decreasing, which is consistent with the mechanism of the photoisomerization. If we compare Figures 5 and 8, we can see that the trend in the HOMO energy agrees well between the NAMD simulation and the graph analysis. We obtained a similar trend with the AIMD data, that is, maxima at d31 = 90° and 270°. If we compare NAMD on S1 and S2 surfaces and AIMD on the second excited (S2) state, we can show that graph analysis can correctly identify the most important internal coordinate, and it is independent of the method of

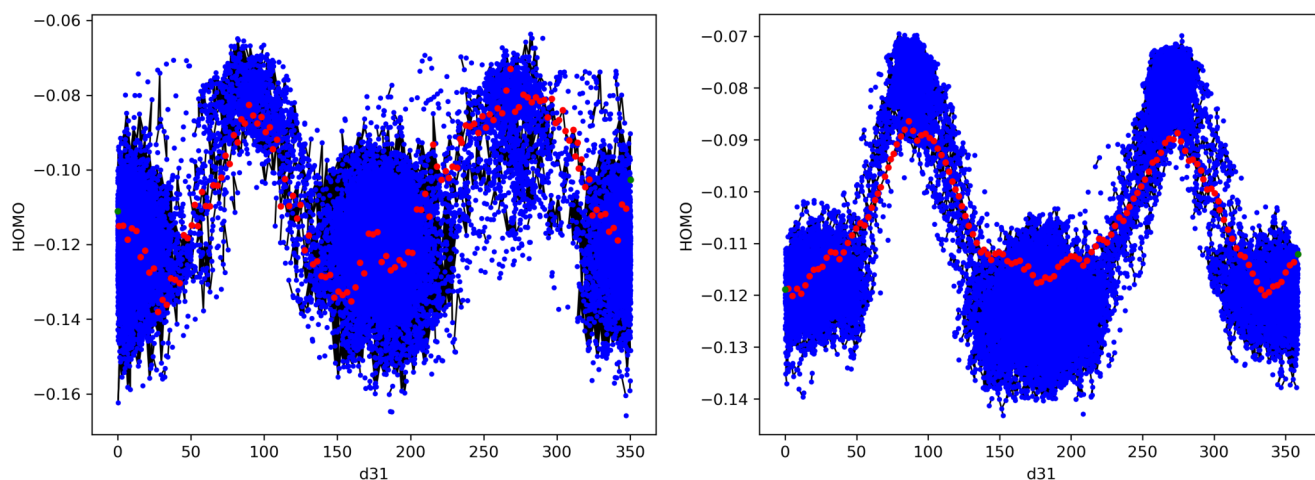


Figure 8. 2D graph representations (blue nodes with black edges) of the data sets produced with NAMD (left) and AIMD (right) simulations. Nodes are displayed as a function of HOMO energies versus torsional angles of d31. Red nodes show one possible shortest path that can be found with the Dijkstra algorithm, resembling a potential energy surface for the isomerization of retinal.

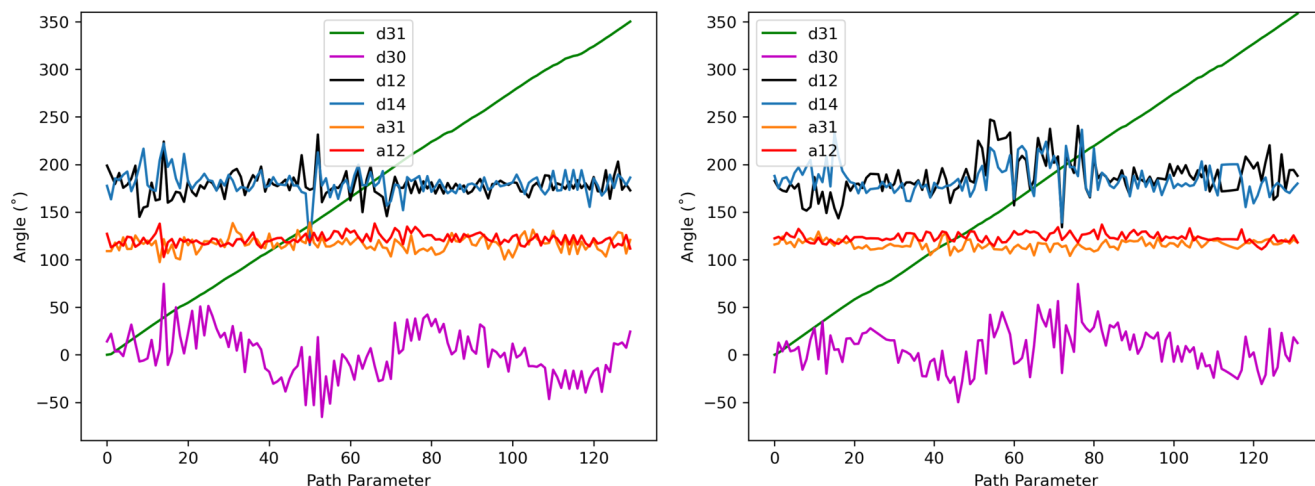


Figure 9. Internal coordinates from the NAMD (left) and AIMD (right) simulation along the reaction path found from the graph analysis.

sampling. Figure 9 shows the trend in internal coordinates as a function of the path parameter, that is, along the shortest path (red nodes in Figure 8). We observe excellent agreement between the AIMD and the NAMD simulations. Note that we have removed d13, as it is the equivalent dihedral angle of d31. To further verify the effect of the basis sets, we performed AIMD simulations with the PBE0/3-21G level of theory. As can be seen from Figures S6–S9, with the increase in the basis, there is no significant change in the quality of the graph. We still find d31 to be the main internal coordinate contributing to the photoisomerization reaction. The plot of the internal coordinates along the path parameter (Figure S9) resembles the one we obtained with the minimal basis set. With this, we can safely conclude that the change in basis does not have a drastic effect on the sampling we used and subsequently on the graph. If one wants to compute the reaction path relying only on ab initio methods, then both the level of theory and the basis set size will affect the accuracy of the calculations. As our goal is to model the isomerization of retinal through a graph analysis and NAMD/AIMD calculations are used only for sampling purposes, we can afford to use a low-quality basis set to save computational cost. AIMD in combination with a graph analysis provides the potential analysis faster with the same

accuracy as NAMD with graph theory. We would also like to mention that for the cases where the nonadiabatic coupling vectors are unavailable, AIMD can be used to obtain the reaction path if it is used in conjunction with graphs. On comparing the computational cost of AIMD and NAMD, we found that AIMD trajectories are computed 1.5 times faster (CPU time) than NAMD ones, using Q-Chem software on an Intel(R)Xeon(R) CPU with the clock speed of 3.07 GHz. Thus, AIMD can be used as a more affordable alternative to NAMD while simulating the reaction pathway in photoinduced reactions.

5. CONCLUSION

In summary, we modeled the photoisomerization reaction with machine learning and graph-theory tools in a biologically important system: retinal. From the analysis of the correlation matrix, we do not notice any relevant internal coordinate involved in the isomerization apart from the main dihedral angle (d13/d31). The shortest path constructed from the graph is consistent with the expected PES associated with the reaction, with transition states located around the d31 dihedral angle values of 90° and 270°. Although this computational protocol does not aim to provide any quantitative measure-

ment concerning the reaction, it can be used for gaining insights into a complete description of the reaction coordinates of chemical processes. We have also shown that the reaction paths obtained through AIMD and NAMD simulations agree well with each other, which enables one to model the reaction paths without involving data from expensive NAMD simulations. Furthermore, additional improvements to the protocol can be incorporated by choosing a better method of assigning weights to the edges in the graph. While we followed Tavadze et al.¹³ and calculated the mutual information between the HOMO energy and the internal coordinates, we do not exclude the possibility of using other descriptors in more general cases. As an example, the HOMO–LUMO gap during the sampling step can be readily used as a descriptor. Last, we believe that the protocols described in this work can be applied to more complex processes other than isomerization reactions.

■ ASSOCIATED CONTENT

Data Availability Statement

The following data files are available free of charge at <https://zenodo.org/records/4280446>: (i) optimized geometry of the cis and trans retinal and (ii) figure labeling the internal coordinates of retinal.

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.4c00325>.

Optimized geometry of the cis and trans retinal, the figure labeling the internal coordinates, and results of NAMD starting at the S1 state and AIMD with the 3-21G basis (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Aiichiro Nakano – Department of Physics and Astronomy, Department of Computer Science, and Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, California 90089, United States; orcid.org/0000-0003-3228-3896; Email: anakano@usc.edu

Authors

Goran Giudetti – Department of Chemistry, University of Southern California, Los Angeles, California 90089, United States; orcid.org/0000-0003-3851-670X

Madhubani Mukherjee – Department of Chemistry, University of Southern California, Los Angeles, California 90089, United States

Sampriya Nandi – Department of Physics and Astronomy, University of Southern California, Los Angeles, California 90089, United States

Sraddha Agrawal – Department of Chemistry, University of Southern California, Los Angeles, California 90089, United States

Oleg V. Prezhdo – Department of Chemistry and Department of Physics and Astronomy, University of Southern California, Los Angeles, California 90089, United States; orcid.org/0000-0002-5140-7500

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jcim.4c00325>

Author Contributions

[†]G.G. and M.M. contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

O.V.P. acknowledges the support of the U.S. National Science Foundation, grant CHE-2154367. A.N. was supported by the National Science Foundation, Award OAC 2118061. Calculations were performed at the Center for Advanced Research Computing (CARC) of the University of Southern California.

■ REFERENCES

- (1) Schlegel, H. B. Exploring potential energy surfaces for chemical reactions: An overview of some practical methods. *J. Comput. Chem.* **2003**, *24* (12), 1514–1527.
- (2) Bernardi, F.; Olivucci, M.; Robb, M. A. Potential energy surface crossings in organic photochemistry. *Chem. Soc. Rev.* **1996**, *25*, 321–328.
- (3) Combes, J. M.; Duclos, P.; Seiler, R. *The Born-Oppenheimer Approximation*; Springer US: Boston, MA, 1981; pp. 185–213.
- (4) Woolley, R. G.; Sutcliffe, B. T. Molecular structure and the born–Oppenheimer approximation. *Chem. Phys. Lett.* **1977**, *45* (2), 393–398.
- (5) Essén, H. The physics of the born–oppenheimer approximation. *Int. J. Quantum Chem.* **1977**, *12* (4), 721–735.
- (6) Ifthimie, R.; Minary, P.; Tuckerman, M. E. *Ab initio* molecular dynamics: Concepts, recent developments, and future trends. *Proc. Nat. Acad. Sci.* **2005**, *102* (19), 6654–6659.
- (7) Tuckerman, M. E.; Ungar, P. J.; Rosenvinge, T.; Klein, M. L. *Ab Initio Molecular Dynamics Simulations*. *J. Phys. Chem.* **1996**, *100* (31), 12878–12887.
- (8) Butler, L. J. Chemical reaction dynamics beyond the Born–Oppenheimer approximation. *Annu. Rev. Phys. Chem.* **1998**, *49* (1), 125–171.
- (9) Nelson, T. R.; Fernandez-Alberti, S.; Tretiak, S. Modeling excited-state molecular dynamics beyond the Born–Oppenheimer regime. *State. Comput. Sci.* **2022**, *2* (11), 689–692.
- (10) Tully, J. C. Molecular dynamics with electronic transitions. *J. Chem. Phys.* **1990**, *93* (2), 1061–1071.
- (11) Wang, L.; Akimov, A.; Prezhdo, O. V. Recent Progress in Surface Hopping: 2011–2015. *J. Phys. Chem. Lett.* **2016**, *7* (11), 2100–2112.
- (12) Li, W.; Ma, A. Recent developments in methods for identifying reaction coordinates. *Mol. Simul.* **2014**, *40* (10–11), 784–793.
- (13) Tavadze, P.; Avendaño Franco, G.; Ren, P.; Wen, X.; Li, Y.; Lewis, J. P. A Machine-Driven Hunt for Global Reaction Coordinates of Azobenzene Photoisomerization. *J. Am. Chem. Soc.* **2018**, *140* (1), 285–290.
- (14) Filipek, S.; Stenkamp, R. E.; Teller, D. C.; Palczewski, K. G. Protein-Coupled Receptor Rhodopsin: A Prospectus. *Annu. Rev. Physiol.* **2003**, *65* (1), 851–879.
- (15) Palczewski, K. G. Protein–Coupled Receptor Rhodopsin. *Annu. Rev. Biochem.* **2006**, *75* (1), 743–767.
- (16) Schoenlein, R. W.; Peteanu, L. A.; Mathies, R. A.; Shank, C. V. The First Step in Vision: Femtosecond Isomerization of Rhodopsin. *Science* **1991**, *254* (5030), 412–415.
- (17) Wang, W.; Geiger, J. H.; Borhan, B. The photochemical determinants of color vision. *BioEssays* **2014**, *36* (1), 65–74.
- (18) Kandori, H.; Shichida, Y.; Yoshizawa, T. Photoisomerization in Rhodopsin. *Biochemistry* **2001**, *40* (11), 1197–1209.
- (19) Palczewski, K.; Kiser, P. D. Shedding new light on the generation of the visual chromophore. *Proc. Nat. Acad. Sci.* **2020**, *117* (33), 19629–19638.
- (20) Johnson, P. J. M.; Halpin, A.; Morizumi, T.; Prokhorenko, V. I.; Ernst, O. P.; Miller, R. J. D. *Local Vib. Coherences Drive Primary Photochem. Vision* **2015**, *7* (12), 980–986.

- (21) Kraskov, A.; Stögbauer, H.; Grassberger, P. Estimating mutual information. *Phys. Rev. E* **2004**, *69*, 066138.
- (22) Peng, H.; Long, F.; Ding, C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27* (8), 1226–1238.
- (23) Mangan, S. M.; Zhou, G.; Chu, W.; Prezhdo, O. V. Dependence between Structural and Electronic Properties of CsPbI₃: Unsupervised Machine Learning of Nonadiabatic Molecular Dynamics. *J. Phys. Chem. Lett.* **2021**, *12* (35), 8672–8678.
- (24) Zhou, G.; Chu, W.; Prezhdo, O. V. Structural Deformation Controls Charge Losses in MAPbI₃: Unsupervised Machine Learning of Nonadiabatic Molecular Dynamics. *ACS Energy Lett.* **2020**, *5* (6), 1930–1938.
- (25) Cover, T. M.; Thomas, J. A. *Elements of Information Theory*; Wiley & Sons: New York, 1991.
- (26) Hagberg, A. A.; Schult, D. A.; Swart, P. J. *Proceedings of the 7th Python in Science Conference*; Varoquaux, G.; Vaught, T.; Millman, J., Eds.; SciPy Pasadena: Pasadena, CA USA, 2008; pp. 1115.
- (27) Dijkstra, E. W. A note on two problems in connexion with graphs. *Numer. Math.* **1959**, *1* (1), 269–271.
- (28) Viswanath, S.; Kreuzer, S. M.; Cardenas, A. E.; Elber, R. Analyzing milestone networks for molecular kinetics: Definitions, algorithms, and examples. *J. Chem. Phys.* **2013**, *139* (17), 174105.
- (29) Epifanovsky, E.; Gilbert, A. T. B.; Feng, X.; Lee, J.; Mao, Y.; Mardirossian, N.; Pokhilko, P.; White, A. F.; Coons, M. P.; Dempwolff, A. L.; et al. Software for the frontiers of quantum chemistry: An overview of developments in the Q-Chem 5 package. *J. Chem. Phys.* **2021**, *155* (8), 084801.
- (30) Menger, M. F. S. J.; Ehrmaier, J.; Faraji, S. PySurf: A Framework for Database Accelerated Direct Dynamics. *J. Chem. Theory Comput.* **2020**, *16* (12), 7681–7689.