Supplemental Information: A universal reaction coordinate for azobenzene photoisomerization

Pedram Tavadze[1], Guillermo Avendaño Franco[1], Pengju Ren[2,3], Xiaodong Wen[2,3], Yongwang Li[2,3], and James P. Lewis[1,*]

1. Department of Physics and Astronomy, West Virginia University, Morgantown, WV, 26506-6315
2. State Key Laboratory of Coal Conversion, Institute of Coal Chemistry, Chinese Academy of Sciences, Taiyuan, Shanxi 030001, China
3. Synfuels China Co. Ltd., Huairou, Beijing 101407, China

Supplemental Information

Computational Methodology. All the calculations presented in this work were performed using a local-orbital density functional theory code, called FIREBALL.[1] We employ non-adiabatic molecular dynamics (NAMD) with Tully's fewest switches surface hopping (FSSH) algorithm, which was fully implemented into the FIREBALL code to investigate non-trivial electron-phonon dynamics.[2-5] In surface hopping methods, an ensemble of trajectories sampling a series of initial conditions is averaged. At any given time the nuclei in any given trajectory are evolving classically on a single potential energy surface (PES). The nuclei then make probabilistic hops from one PES to another. Energy conservation is maintained by re-scaling the velocities along the directions of the non-adiabatic coupling vectors.[6] Surface hopping methods are more equipped to handle systems where changes in the electronic populations also change the nuclear motion, as is the case with photo-induced isomerization.

In our implantation of NAMD simulations (summarized in Ref. 6-8 and references therein),[6-8] atoms evolve along classical trajectories $\{\mathbf{R}_\alpha(t)\}$, but the interaction between the nuclear motion and the electronic quantum state is explicitly accounted for. In time-dependent Kohn-Sham (KS) theory, the time evolution of a single-particle KS orbitals is found from the time-dependent Schrödinger-like equation:

$$\widehat{H}_{KS}\varphi_p(\boldsymbol{r},t) = i\text{ђ}\frac{\partial\varphi_p(\boldsymbol{r},t)}{\partial t} \qquad (1)$$

where $\widehat{H}_{KS}$ is the single-particle KS Hamiltonian, and $\varphi_p(\boldsymbol{r},t)$ are the time-evolving KS orbitals. It is convenient to expand the KS orbitals in the basis of the instantaneous adiabatic eigenstates $\psi_i(\boldsymbol{r},\boldsymbol{R})$:

$$\varphi_p(\boldsymbol{r},t) = \sum_i a_i(t)\psi_i(\boldsymbol{r},\boldsymbol{R}) \qquad (2)$$

By incorporating Eq. (2) into Eq. (1) we can obtain the time-evolution of the coefficients $a_i(t)$.

$$i\hbar\frac{\partial a_i(t)}{\partial t} = a_i(t)\varepsilon_i(R) - i\hbar\sum_j a_j(t)\boldsymbol{d}_{ij}\cdot\boldsymbol{V} \qquad (3)$$

In Eq. (3), the coupling between the classical motion of the nuclei and the electronic quantum state is shown in the nonadiabatic coupling term, $\boldsymbol{d}_{ij}\cdot\boldsymbol{V}$:

$$\boldsymbol{d}_{ij}\cdot\boldsymbol{V} \overset{\text{def}}{=} \sum_\alpha d_{ij}^\alpha\boldsymbol{V}_\alpha, \qquad (4)$$

where $\boldsymbol{V}_\alpha = \partial\boldsymbol{R}_\alpha/\partial t$ is the atomic velocity α and $\boldsymbol{d}_{ij}$ are the nonadiabatic coupling vectors between single particle KS states.

$$\boldsymbol{d}_{ij}^{\alpha} \equiv \left\langle \psi_i \middle| \frac{\partial \psi_j}{\partial \boldsymbol{R}_\alpha} \right\rangle. \qquad (5)$$

The non-adiabatic coupling vectors for a local-orbital basis set are analytically derived by the method described in our previous works.[2]

$$\boldsymbol{d}_{ij}^{\alpha} = \left\langle \psi_i \middle| \frac{\partial \psi_j}{\partial \boldsymbol{R}_\alpha} \right\rangle = \sum_{\mu\nu} c_{i\mu}^{*} \frac{\partial c_{j\nu}}{\partial \boldsymbol{R}_\alpha} \langle \phi_\mu | \phi_\nu \rangle + \sum_{\mu\nu} c_{i\mu}^{*}\, c_{j\nu} \left\langle \phi_\mu \middle| \frac{\partial \phi_\nu}{\partial \boldsymbol{R}_\alpha} \right\rangle, \qquad (6)$$

The FSSH algorithm proposed by Tully[9] is then implemented to obtain the correct statistical distribution of the state populations at all times. The probability of a hop between states *i* and *j* within given time interval, Δt, is:

$$g_{ij} = \max\left(0, \frac{b_{ij}\Delta t}{\left|a_j(t)\right|^2}\right) \qquad (6)$$

where the coefficient b_{ij} is defined as:

$$b_{ij} = -2\mathcal{R}e\left[a_j^{*} a_k \boldsymbol{d}_{ij} \cdot \boldsymbol{V}\right]. \qquad (7)$$

The hopping probability depends explicitly on the nonadiabatic coupling defined in Eqs. 4 and 5. The surface hopping probability is set to zero when g_{ij} is determined to be negative. This ensures that a hop from state *i* to state *j* occurs only if the occupation of state *i* decreases, and the occupation of state *j* increase. The calculated probabilities are then compared with a random number to determine if an electronic transition will occur. When a transition takes place, energy conservation is imposed by re-scaling velocities along the direction of the nonadiabatic couplings, $\boldsymbol{d}_{ij}^{\alpha}$. If a hop up in energy is predicted and there is not enough kinetic energy then the hop is rejected. This velocity rescaling and hop rejection ensures detailed balance between transitions up and down in energy.

We used the Becke exchange[10] with Lee-Yang-Parr correlation[11] (BLYP) for this study. The basis set is made of optimized numerical local atomic orbitals which were confined to regions limited by the corresponding cutoff radii r_c. We have adopted double numerical sp^3 basis sets for C (cutoff radii of r_c (s*)* = 4.4 a.u. and r_c (p*)* = 4.8 a.u.) and N (cutoff radii of r_c *(*s*)* = 4.0 a.u. and r_c *(*p*)* = 4.4 a.u.); whereas we use a minimal basis set is used for H (cutoff radius of r_c *(*s*)* = 4.2 a.u.). The HOMO-LUMO gap is 2.50 eV and 2.68 eV for *trans*- and *cis*-, respectively. A molecular orbital picture of both the *cis*- and *trans*- form of azobenzene is shown in Fig. SI1 and SI2.

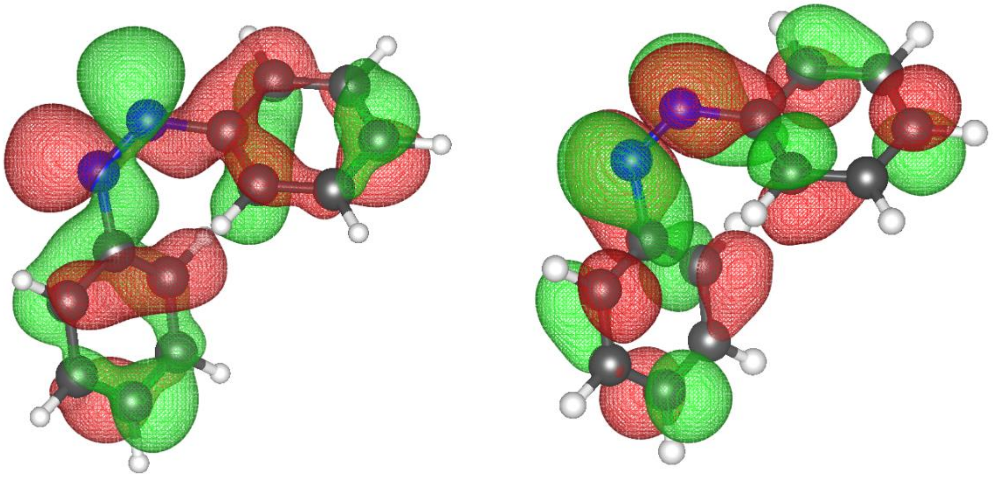

Figure SI1. Frontier orbitals of *cis* isomer of azobenzene (left) HOMO (right) LUMO

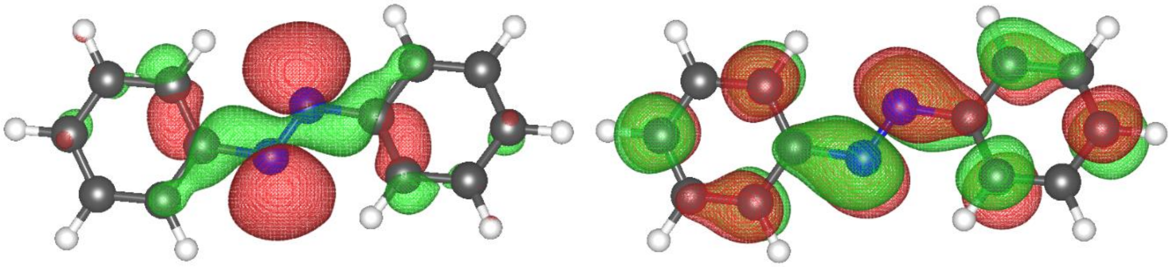

Figure SI2. Frontier orbitals of *trans* isomer of azobenzene (left) HOMO (right) LUMO

From a dataset of 494 molecular dynamics simulations, our review of the data using human-biased analysis and interpretation, shows only 1 simulation undergoes concerted inversion (0.2%), 6 simulations undergo inversion (1.2%), 60 simulations undergo inversion assisted-rotation (12.1%) and 203 undergoes rotation (41.1%) - 223 simulations do not yield isomerization (45.1%).[6]

Mutual Information. Generically consider any two random variable sets of a system, $\boldsymbol{X}$ and $\boldsymbol{Y}$, each with its own probability distribution (schematic for azobenzene shown in Fig. SI3). In order to evaluate the correlation between these random variable sets, then we begin by measuring how similar the joint distribution $p(\boldsymbol{X},\boldsymbol{Y})$ is to the factored distribution $p(\boldsymbol{X})p(\boldsymbol{Y})$. The mutual dependency, or the correlation, of two variables sets is then mapped onto an entropic (information theory) representation. Mathematically, this is represented by

$$MI(\boldsymbol{X};\boldsymbol{Y}) = \sum_x \sum_y p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right) \text{ (1)}$$

If any two variable sets are independent, then the mutual information is zero because $p(x,y) = p(x)p(y)$ as there would be no overlap between the two distributions.

We can understand mutual information by considering a population of a microscopic variable set $\boldsymbol{X}$ with some distribution p. We define the entropy, or measure of uncertainty, of $\boldsymbol{X}$ as $H(\boldsymbol{X}) = -\sum_x p(x)\log p(x)$. Conversely, there exist two additional definitions - the relative entropy (Kullback-Leibler divergence) of two difference probability distributions $p(x)$ and $q(x)$:

$$KL(p\|q) = \sum_x p(x)\log\frac{p(x)}{q(x)}, \text{ (2)}$$

and the cross entropy, $H(p,q) = -\sum_x p(x)\log q(x)$. In essence, MI is the entropy of $\boldsymbol{X}$, $H(\boldsymbol{X})$, reduced by a conditional entropy, $H(\boldsymbol{X}|\boldsymbol{Y})$ (Chapter 2.8 in *Machine Learning, A Probabilistic Perspective* by Murphy offers a very rigorous description).[12] We can interpret the MI between $\boldsymbol{X}$ and $\boldsymbol{Y}$ as the reduction of uncertainty about $\boldsymbol{X}$ after observing $\boldsymbol{Y}$, or, by symmetry, the reduction in uncertainty about $\boldsymbol{Y}$ after observing $\boldsymbol{X}$.

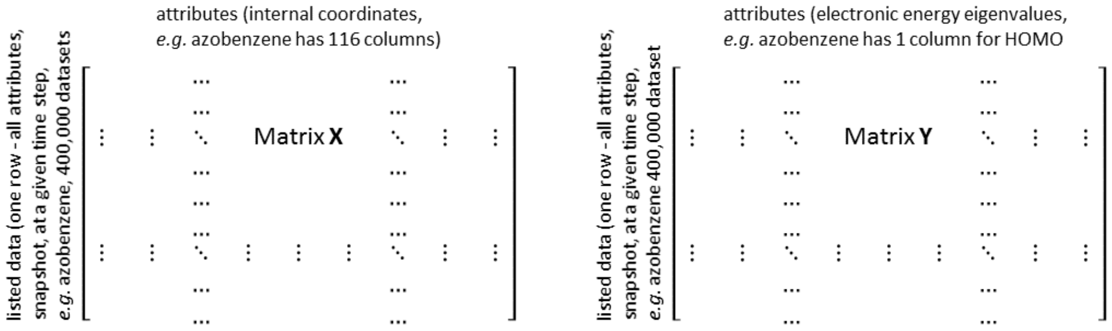
attributes (internal coordinates,
e.g. azobenzene has 116 columns)
listed data (one row - all attributes,
snapshot, at a given time step,
e.g. azobenzene, 400,000 datasets
Matrix X
attributes (electronic energy eigenvalues,
e.g. azobenzene has 1 column for HOMO
listed data (one row - all attributes,
snapshot, at a given time step,
e.g. azobenzene 400,000 dataset
Matrix Y

Figure SI3. Organization of the datasets, $\boldsymbol{X}$ and $\boldsymbol{Y}$, containing the internal coordinates and the electronic energy eigenvalues.

Table SI1.

Internal Coordinate	MI Value	Error
C8-N14=N13-C4	0.320279	0.001639
N14=N13-C4-C5	0.187555	0.001794
C4-N13=N14	0.186897	0.0024
C8-N14=N13	0.182892	0.002261
N13=N14-C8-C7	0.181208	0.001233
N13=N14-C8-C9	0.179811	0.002646
N14=N13-C4-C3	0.179502	0.00276
N14=N13	0.15941	0.003383
N14-C8	0.143024	0.002183
N13-C4	0.141552	0.002861
C7-C8-C9	0.13134884	0.002175
N13-C4-C3-C2	0.134215	0.001948
C3-C4-C5	0.133732	0.001685
C10-C9-C8-N14	0.133509	0.002534
C6-C5-C4-N13	0.132728	0.002356
C3-C4-N13	0.131828	0.002533
N14-C8-C7-C12	0.13149	0.002043
C9-C8-N14	0.131003	0.003095
C8-C7-C12	0.129953	0.002475
C4-C5-C6	0.129112	0.002838
C5-C4-N13	0.128411	0.002483
C7-C8-N14	0.128411	0.002187
N13-C4-C3-H20	0.128009	0.002172
H24-C5-C4-N13	0.127904	0.002918
C2-C1-C6	0.127498	0.001817
C10-C11-C12	0.12703	0.003308
N14-C8-C7-H15	0.126855	0.00251
C8-C9-C10	0.126539	0.002046
C2-C3-C4	0.126372	0.003391
C4-C3	0.126253	0.001159
H19-C9-C8-N14	0.126015	0.002158
C11-C10	0.12597	0.002624
C5-C4	0.125876	0.00246
C8-C7	0.125716	0.001804
C12-C11	0.125231	0.002571
C6-C1	0.125094	0.002516
C9-C8	0.124582	0.001677
C2-C1	0.124472	0.001721
C10-C9-H19	0.124426	0.002442
C3-C2-C1-C6	0.124093	0.002792
C7-C12-C11	0.123866	0.002454
C5-C6-C1-C2	0.123853	0.002177
C2-C1-H22	0.123822	0.002006
C8-C7-H15	0.123817	0.0017
C5-C6-H23	0.123789	0.002214
C5-C4-C3-H20	0.123737	0.002049
C1-C2-C3	0.123597	0.003022
C6-C5-H24	0.123481	0.002242
C3-C2	0.123397	0.00261
C1-C6-C5	0.123369	0.002073
C3-C2-H21	0.123316	0.001376
C1-C2-H21	0.123265	0.001912
H18-C10-C9-C8	0.123228	0.001772
C10-C9-C8-C7	0.123196	0.001831
C9-C10-C11	0.123162	0.002539
C7-C12-C11-C10	0.123133	0.002003
C12-C7	0.123098	0.002598
C2-C3-H20	0.123097	0.002111
C4-C3-H20	0.123022	0.002658
C12-C7-H15	0.123005	0.002096
C9-C8-C7-H15	0.122991	0.002199
H21-C2	0.122991	0.00242
C4-C5-H24	0.12299	0.002483
C12-C11-C10-C9	0.122977	0.002483
C5-C4-C3-C2	0.122959	0.002305
C8-C9-H19	0.122771	0.002618
C1-C6-C5-C4	0.122708	0.002467
H21-C2-C1-C6	0.122682	0.001883
H19-C9-C8-C7	0.122637	0.001325
C11-C10-H18	0.122619	0.001962
H20-C3-C2-C1	0.122616	0.002129
C11-C10-C9-C8	0.122472	0.001574
C9-C8-C7-C12	0.122423	0.001825
C6-C5	0.12242	0.002156
C11-C12-H16	0.12238	0.001907
C6-C5-C4-C3	0.122293	0.002725
H16-C12-C7-C8	0.122274	0.002012
C5-C6-C1-H22	0.122271	0.002202

Internal Coordinate	MI Value	Error
H22-C1	0.122195	0.002114
C10-C9	0.122185	0.002493
C12-C11-H17	0.12215	0.002161
C1-C6-H23	0.122132	0.002665
H23-C6-C5-H24	0.12213	0.00148
H17-C11	0.122095	0.002135
C1-C6-C5-H24	0.122084	0.003711
C11-C10-C9-H19	0.12208	0.001912
C10-C11-H17	0.122075	0.001835
C6-C1-H22	0.122053	0.001918
C7-C12-H16	0.122032	0.002536
H20-C3	0.122019	0.001916
H23-C6-C1-H22	0.121974	0.003352
H16-C12-C11-C10	0.121873	0.001355
C9-C10-H18	0.121861	0.00296
C4-C3-C2-H21	0.121773	0.002156
H24-C5-C4-C3	0.121767	0.002572
C11-C12-C7-C8	0.121759	0.002765
C12-C11-C10-H18	0.121755	0.001787
H23-C6-C5-C4	0.121753	0.002431
C4-C3-C2-C1	0.121729	0.002268
H17-C11-C10-H18	0.121687	0.001793
H17-C11-C10-C9	0.121578	0.00153
C7-C12-C11-H17	0.121542	0.002807
C3-C2-C1-H22	0.121534	0.001926
H20-C3-C2-H21	0.121498	0.002276
H18-C10-C9-H19	0.121468	0.001594
C11-C12-C7-H15	0.121442	0.00241
H16-C12-C11-H17	0.121206	0.00141
H21-C2-C1-H22	0.121037	0.002261
H23-C6-C1-C2	0.121013	0.002239
H16-C12	0.120804	0.002406
H23-C6	0.120744	0.002528
H19-C9	0.120722	0.001684
H16-C12-C7-H15	0.120641	0.002386
H18-C10	0.120065	0.003257
H24-C5	0.119969	0.001422
H15-C7	0.119529	0.002454

Building the dataset matrices. From our datasets including all the bonds, angles, and dihedrals of each time step, we build the matrix $\boldsymbol{X}$ (see Fig. SI3), and subsequently, $\boldsymbol{X}^T\boldsymbol{X}$. Therefore, the columns (attributes) of the data are the internal coordinates for a given time step and each time step defines a row of the matrix $\boldsymbol{X}$. The matrix $\boldsymbol{Y}$ (see Fig. SI3) contains a matrix of the electronic energy eigenvalues; as many or few eigenvalues as we would like to track may be included. For our example in the photoisomerization of azobenzene we choose to track only the HOMO as the attribute of interest. We could also include the LUMO as one of our attributes; however for simplicity we focus only on the HOMO as an attribute in the matrix $\boldsymbol{Y}$.

Now that we have datasets of attributes representing the internal coordinates ($\boldsymbol{X}$) and datasets of attributes representing the HOMO ($\boldsymbol{Y}$), we proceed to calculating the mutual information of $\boldsymbol{X}$ and $\boldsymbol{Y}$. Table SI1 shows all the mutual information calculated data, the entropic correlations between an internal coordinate and the HOMO.

Principal Component Analysis (PCA) is a very common statistical procedure for reducing large datasets by examining correlations between data and generally selecting the uncorrelated variables, called "principal components." We use PCA in a *slightly different approach* - to remove redundancies or correlated variables within the dataset $\boldsymbol{X}$ after sorting through correlations between $\boldsymbol{X}$ and $\boldsymbol{Y}$ which we obtained via mutual information; we would like to remove any redundant variables $\boldsymbol{X}$ which do not add any new information to our final analysis. To find the principal components, one builds a matrix of the correlations which is the matrix $\boldsymbol{X}^T\boldsymbol{X}$ (the columns of $\boldsymbol{X}$ represent the attributes, or degrees of freedom, of the dataset; each dataset is stacked in a row). This matrix may be diagonalized (brute-force approach), utilizing a singular value decomposition approach.[*] The eigenvectors represent the principal components of the data. The eigenvector corresponding to the largest eigenvalue will contain the largest variance in the data, the second eigenvector corresponding to the second largest eigenvalue will contain the next largest variance in the data, *etc*. The eigenvectors are linear combinations of the original attributes of the dataset. Hence, one can take the eigenvector and determine the source of the variance from the original datasets and remove the highly correlated features which correspond to the smallest variance. Effectively, through PCA, we filter out repetitive attributes or degrees of freedom within the data. This approach has been applied often in molecular dynamics simulations where researchers are trying to focus in on specific degrees of freedom such as in transition path sampling[14] and protein folding.[15]

[*] Building the correlation matrix $\boldsymbol{X}^T\boldsymbol{X}$ is the brute-force approach, there are algorithms for by-passing building the matrix and by-passing the diagonalization. All of the necessary algorithms are built into the Scikit-learn package.

Potential Energy Surface Network. To build the network, we represent a series of nodes where each node, $\boldsymbol{q}_i$, represents a (reduced dimensionality) single data point with a "distance", an edge with weight w_{ij}, between each of the nodes (see Fig. SI4).

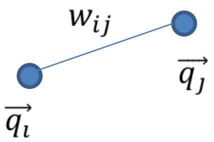

Figure SI4. Two nodes in a potential energy surface network.

Consider that the number of nodes, n, will actually be on the order of 10^5-10^6 points (as seen in Fig. SI5 for the azobenzene photoisomerization network). The distance matrix will have $n(n-1)/2$ elements, so will be on the order of 10^{10}-10^{12} elements; this large number of elements will require large amounts of memory as well as computational time. To improve computational feasibility, we use the fast algorithms provided by the Scikit-learn package[16] in Python to find the k nearest neighbors (k_{NN}). Using this method we get a sparse connectivity matrix which can be used to create a network using the Networkx package in Python.[13] At some regions of our space the data points might be clumped and by using k nearest neighbors we might have missed some possible edges. To avoid this problem, we use a high number for k, such as 50, and then remove the undesired edges using a threshold for distance.

The weights of the network are assigned with a modified gradient.

$$w_{ij} = \exp\left[\frac{\delta E_{ij}}{\left\|\delta q_{ij}\right\|_{2_{MI}} sign\left(MI \cdot \delta q_{ij}\right)}\right]$$

In this equation δE_{ij} is the energy difference between two nodes, $\left\|\delta q_{ij}\right\|_{2_{MI}}$ is the Euclidian distance weighted with MI between two nodes and $sign(MI \cdot \delta q_{ij})$ is the weighted sign of the distance between two nodes. The modified gradient is exponentiated to avoid negative weights.

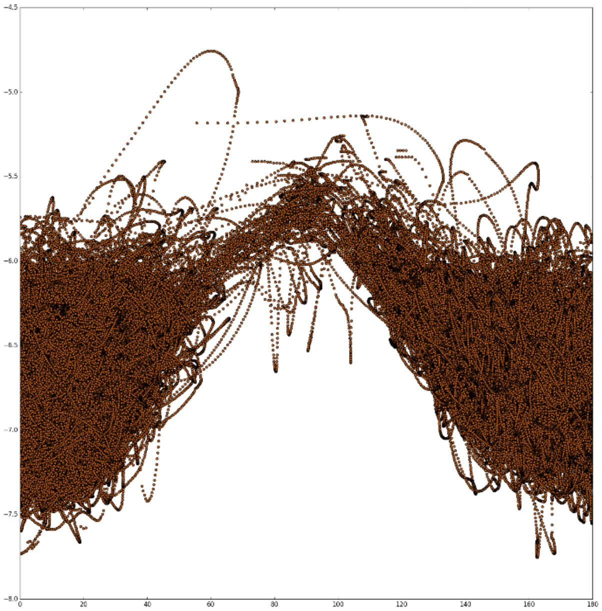

Figure SI5. Two-dimensional representation of the multi-dimensional nodal network for the example of azobenzene photoisomerization.

References

1. Lewis, J. P.; Jelínek, P.; Ortega, J.; Demkov, A. A.; Trabada, D. G.; Haycock, B. J.; Wang, H.; Adams, G.; Tomfohr, J. K.; Abad, E.; Wang, H.; Drabold, D. A., Advances and applications in the fireball ab initio tight-binding molecular-dynamics formalism. *Phys. Stat. Sol. B.* **2011,** *248* (9), 1989-2007.

2. Abad, E.; Lewis, J. P.; Zobac, V.; Hapala, P.; Jelinek, P.; Ortega, J., Calculation of non-adiabatic coupling vectors in a local-orbital basis set. *J. Chem. Phys.* **2014,** *138* (15), 154106-8.

3. Neukirch, A. J.; Shamberger, L. C.; Abad, E.; Haycock, B. J.; Wang, H.; Ortega, J.; Prezhdo, O. V.; Lewis, J. P., Nonadiabatic ensemble simulations of cis-stilbene and cis-azobenzene photoisomerization. *J. Chem. Theory Comput.* **2014,** *10* (1), 14-23.

4. Mendieta-Moreno, J. I.; Walker, R. C.; Lewis, J. P.; Gómez-Puertas, P.; Mendieta, J.; Ortega, J., Fireball/amber: An efficient local-orbital dft qm/mm method for biomolecular systems. *J. Chem. Theory Comput.* **2014,** *10* (5), 2185-2193.

5. Vladimír, Z.; James, P. L.; Enrique, A.; Jesús, I. M.-M.; Prokop, H.; Pavel, J.; José, O., Photo-induced reactions from efficient molecular dynamics with electronic transitions using the fireball local-orbital density functional theory formalism. *J. Phys. Condens. Matter* **2015,** *27* (17), 175002.

6. Neukirch, A. J.; Shamberger, L. C.; Abad, E.; Haycock, B. J.; Wang, H.; Ortega, J.; Prezhdo, O. V.; Lewis, J. P., Nonadiabatic ensemble simulations of cis-stilbene and cis-azobenzene photoisomerization. *Journal of Chemical Theory and Computation* **2013,** *10* (1), 14-23.

7. Abad, E.; Lewis, J. P.; ZobaC, V.; Hapala, P.; Jelinek, P.; Ortega, J., Calculation of non-adiabatic coupling vectors in a local-orbital basis set. *J Chem Phys* **2013,** *138* (15), 154106-8.

8. Mendieta-Moreno, J. I.; Trabada, D. G.; Mendieta, J.; Lewis, J. P.; Gómez-Puertas, P.; Ortega, J., Quantum mechanics/molecular mechanics free energy maps and nonadiabatic simulations for a photochemical reaction in DNA: Cyclobutane thymine dimer. *J. Phys. Chem. Lett.* **2016,** *7* (21), 4391-4397.

9. Tully, J. C., Molecular dynamics with electronic transitions. *J. Chem. Phys.* **1990,** *93* (2), 1061-1071.

10. Becke, A. D., Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988,** *38* (6), 3098-3100.

11. Lee, C.; Yang, W.; Parr, R. G., Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988,** *37* (2), 785-789.

12. Murphy, K. P., *Machine learning: A probabilistic perspective*. The MIT Press: 2012; p 1096.

13. Hagberg, A. A.; Schult, D. A.; Swart, P. J. In *Exploring network structure, dynamics, and function using networkx*, SciPy 2008, Varoquaux, G.; Vaught, T.; Millman, J., Eds. 2008; pp 11-15.

14. Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L., Transition path sampling: Throwing ropes over rough mountain passes, in the dark. *Annu. Rev. Phys. Chem.* **2002,** *53* (1), 291-318.

15. Onuchic, J. N.; and, Z. L.-S.; Wolynes, P. G., Theory of protein folding: The energy landscape perspective. *Ann. Rev. Phys. Chem.* **1997,** *48* (1), 545-600.

16. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, J.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E., Scikit-learn: Machine learning in python. *J. Mach Learn. Res.* **2011,** *12*, 2825-2830.