# Supporting Information for:

# The Principal Component Analysis of the Ring Deformation in the Nonadiabatic Surface Hopping Dynamics

Yifei Zhu,[†,‡] Jiawei Peng,[‡,¶] Xu Kang,[†,‡] Chao Xu,[†,‡] and Zhenggang Lan[*,†,‡]

†*SCNU Environmental Research Institute, Guangdong Provincial Key Laboratory of Chemical Pollution and Environmental Safety, School of Environment, South China Normal University, Guangzhou 510006, P. R. China.*

‡*MOE Key Laboratory of Environmental Theoretical Chemistry, South China Normal University, Guangzhou 510006, P. R. China.*

¶*School of Chemistry, South China Normal University, Guangzhou 510006, P. R. China.*

E-mail: zhenggang.lan@m.scnu.edu.cn

# 1 Methods

## 1.1 Redundant internal coordinates

Different sets of internal coordinates may be obtained by different construction ways. Therefore, it is suitable to construct the internal coordinates with chemical meanings. In the current work, we simply used the redundant internal coordinates that are used as default by Gaussian package.[1,2] More discussions on the construction of the chosen redundant internal coordinates are given in the references. For simplicity, we generated all internal coordinates by using the Gaussian 16 package. In implementation, we created the pseudo-optimization input files, ran the Gaussian calculation and read the redundant internal coordinates of the initial structures.

In this analysis step, most geometries in the dynamics give the same set of redundant internal coordinates, while only a few geometries do not ($<2\%$). This implies that the basic connectivity of the ring moiety is not destroyed in the nonadiabatic dynamics of the current system. Therefore, the current descriptor sets are acceptable.

## 1.2 Principal component analysis (PCA)

The standard PCA algorithm[3–8] is briefly given as follows.

Firstly, it is necessary to centralize the dataset matrix $\mathbf{X}$ by subtracting the mean, and a centered matrix $\mathbf{X}'$ is obtained:

$$\mathbf{X}' = \mathbf{X} - \langle \mathbf{X} \rangle. \tag{1}$$

Secondly, the covariance matrix $\mathbf{C}$ is calculated according to

$$\mathbf{C} = \mathbf{X}'^{T}\mathbf{X}', \tag{2}$$

where the superscript $T$ denotes the matrix transpose operation.

Thirdly, the eigenvalues and eigenvectors are obtained by solving the following eigenvalue problem

$$\mathbf{C}\nu_i = \lambda_i \nu_i \tag{3}$$

Here the eigenvalues $\{\lambda_i\}$ describe the independent variances along each principal component, which are sorted by the decreasing order. Thus, $\lambda_1$ denotes the largest eigenvalue and the corresponding eigenvector $\nu_1$ coincides with the direction of maximum variance, and so on.

Fourthly, the $k$ eigenvectors with the largest $k$ eigenvalues are picked up to form the transformation matrix $\mathbf{U}'$. Then the transformation from original high-dimensional dataset $\mathbf{X}$ to the reduced one $\mathbf{Y}$ is completed by

$$\mathbf{Y} = \mathbf{X}'\mathbf{U}' \tag{4}$$

namely, the PCA representation is ultimately obtained.

## 1.3 Clustering methods

DBSCAN[9] is a density-based clustering algorithm, in which clusters are defined as the regions with the high density of data points, and outliers are referred to the points lying in low-density regions. The DBSCAN algorithm basically requires two parameters, namely a distance measurement ($\epsilon$) and a minimum number of neighbors (*MinPts*). If the data densities in several clusters are rather different or the configuration of clusters are rather complex (such as the hierarchical configuration, that is, some clusters with several sub-clusters), DBSCAN cannot work well.[10,11]

Agglomerative clustering[12,13] is a partitioning scheme that seeks to build a bottom-up cluster tree according to the linkage criterion. It treats each data point as a stand-alone cluster and then successively merges them until the chosen resolution level (*Threshold*) is reached. Since this method needs all pairwise distance between data points, its use is limited by the sensitivity to outliers and the complexity of the calculation.[13–15]

## 1.4 Trajectory surface hopping (TSH) dynamics

TSH is a mixed quantum-classical approach that incorporates the nonadiabatic transition between electronic states into molecular dynamics. In the TSH approach, the nuclear part is treated in the classical framework, and the electronic motion is described by the time-dependent quantum mechanics. The trajectories may undergo sudden "hops" between different electronic states to account for nonadiabatic transitions.

The hopping probability is determined by Tully's FSSH algorithm[16] in this work. After hops, the momentum is rescaled along the direction of the nonadiabatic coupling vector to ensure the energy conservation. For frustrated hops, the velocity component is reversed according to the direction of the nonadiabatic coupling vector.

The ground state (GS) minimum and vibrational normal modes of keto isocytosine were obtained at the B3LYP/6-31G$^*$ level with the standard electronic-structure package Gaussian 16. The initial conditions, *i.e.*, geometries and velocities, are sampled by Wigner distribution function of the lowest vibrational level in the electronic ground state. Then we vertically placed them into the second excited state ($S_2$) to start the TSH dynamics.

In the on-the-fly TSH dynamics simulation, the time step of the nuclear motion and election evolution is 0.5 fs and 0.005 fs, respectively. In total, we considered 1000 trajectories starting from $S_2$ with a total evolution time of 1.5 ps.

# 2 Nonadiabatic dynamics

## 2.1 Time-dependent occupations of electronic states

Because the decoherence correction was used, the time-dependent electronic populations and the fractional occupation numbers are highly consistent in the evolution (see Figure S1). Here, we noticed that some trajectories may sudden re-hops from $S_1$ to $S_2$ states at around 30 - 60 fs. To clarified what happened in this situation, we optimized the $S_1$-$S_2$ CI and constructed the linear interpolated PES from the $S_0$ minimum to $S_1$-$S_2$ CI in Figure S2. In the early stage of the nonadiabatic dynamics, the system moves along the reaction coordinate on the $S_2$ PES to access $S_1$-$S_2$ CI quickly, and performs the hops before 30 fs. Afterwards, some trajectories move along the same reaction coordinates as shown in Figure S2, and return to the $S_1$-$S_2$ CI again due to the profile of the $S_1$ PES. Consequently, the hops back to $S_2$ state may take place, resulting the slight rising of the $S_2$ population between 30 fs and 60 fs. However, this population recurrence is rather minor because the vibrational relaxation and the energy flow to other degrees are dominant in the high dimensional system.
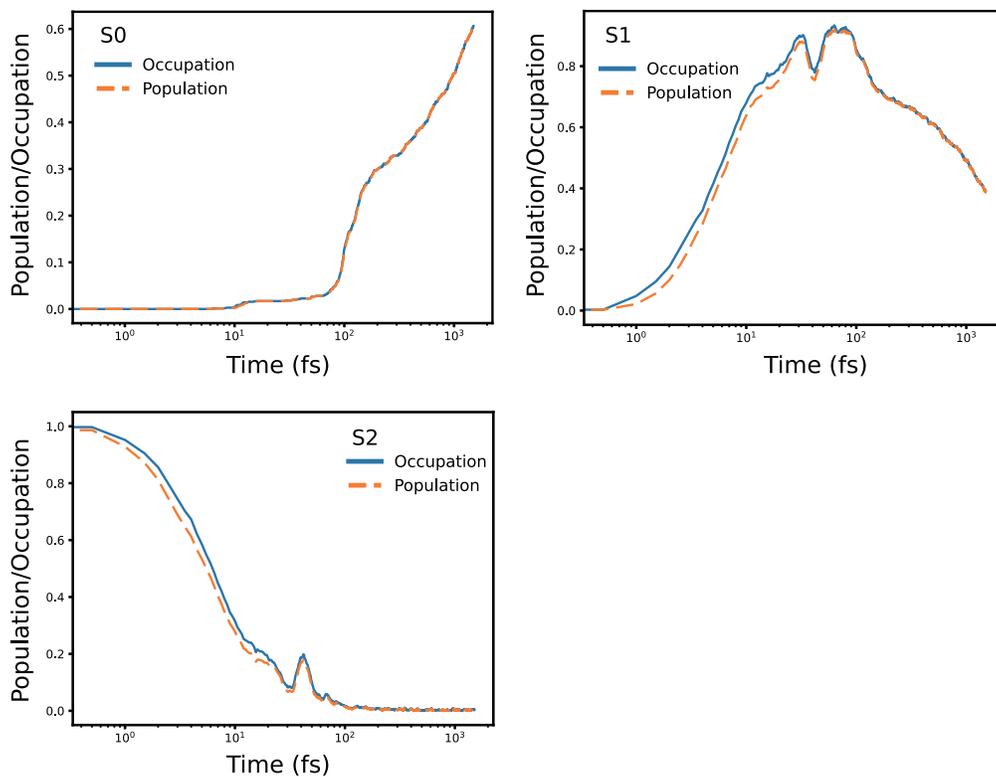
Figure S1. Time-dependent average electronic populations and the fractional occupations of adiabatic states in the nonadiabatic dynamics.
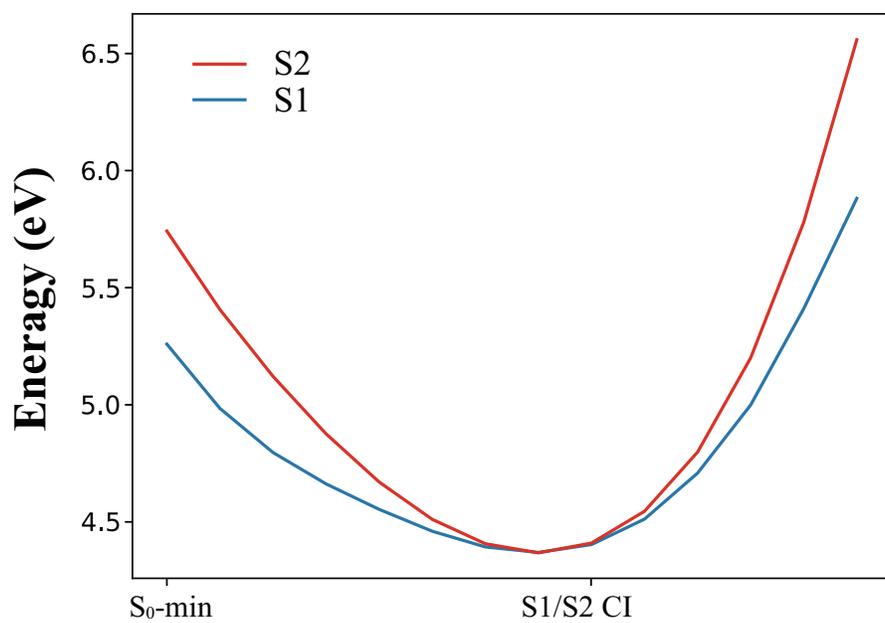
Figure S2. Linear interpolated reaction paths from the ground-state minimum to $S_1$-$S_2$ CI at the SA3-CASSCF level.

## 2.2 Consistency between SA-CASSCF and B3LYP levels

In this work, the initial conditions for the TSH dynamics at the SA3-CASSCF(12,9)/6-31G* level are based on a sampling of a Wigner distribution constructed at the B3LYP/6-31G* level. As illustrated in the following, the perturbation results from the switch from one electronic-structure method to the other can be neglected.

In the simulation of nonadiabatic dynamics, the SA-CASSCF method is widely used to run the on-the-fly TSH dynamics, because this method can give the proper description on the CIs between the ground and excited states. However, this method often suffers from the problem with the non-smooth wavefunction along a reaction path. Therefore, we sometimes do not recommend running the ground-state optimization or normal mode calculations at the SA-CASSCF level. It is proper to choose MP2 and DFT to do these ground-state calculations in the initial sampling if no state crossing is involved.

To further address this question, we also optimized the ground-state geometry at both SA3-CASSCF(12,9)/6-31G* and the B3LYP/6-31G* levels. As shown in Figure S3, these two resulting geometries are very similar, except the minor difference in the $NH_2$ group. Their bond lengths are given in Table S1. At the same time, the normal-mode frequencies at these two levels are comparable (see Figure S4).
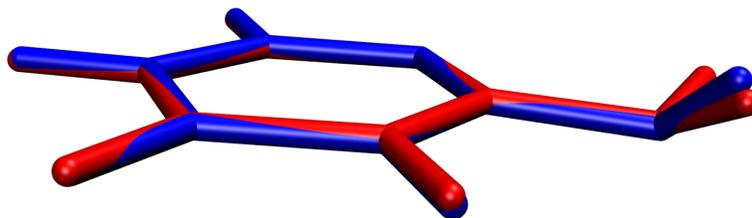


Figure S3. Geometries of ground-state minimum optimized at the B3LYP (blue) and SA3-CASSCF (red) levels, respectively.
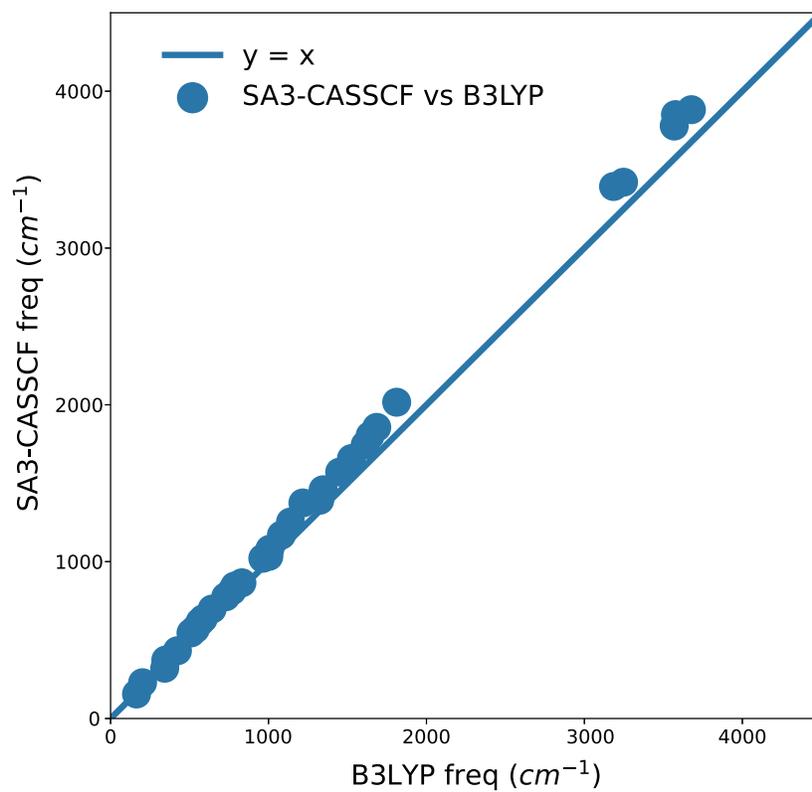
Figure S4. The frequencies obtained at SA3-CASSCF level with respect to these counterparts at B3LYP level.

Table S1: Bond lengths (Å) of the ground-state minimum optimized at the SA3-CASSCF(12,9)/6-31G* and the B3LYP/6-31G* levels.

| Bond | SA-CASSCF | B3LYP |
| --- | --- | --- |
| R(1,4) | 1.3633 | 1.3649 |
| R(1,5) | 1.2823 | 1.3121 |
| R(1,6) | 1.3828 | 1.3743 |
| R(2,3) | 1.3531 | 1.3661 |
| R(2,10) | 1.457 | 1.444 |
| R(2,11) | 1.0718 | 1.0829 |
| R(3,5) | 1.384 | 1.3679 |
| R(3,12) | 1.0733 | 1.0882 |
| R(4,10) | 1.4009 | 1.4262 |
| R(4,13) | 0.9971 | 1.0146 |
| R(6,7) | 0.9987 | 1.0116 |
| R(6,8) | 0.9982 | 1.0122 |
| R(9,10) | 1.2005 | 1.2226 |

# 3 Additional results of the PCA
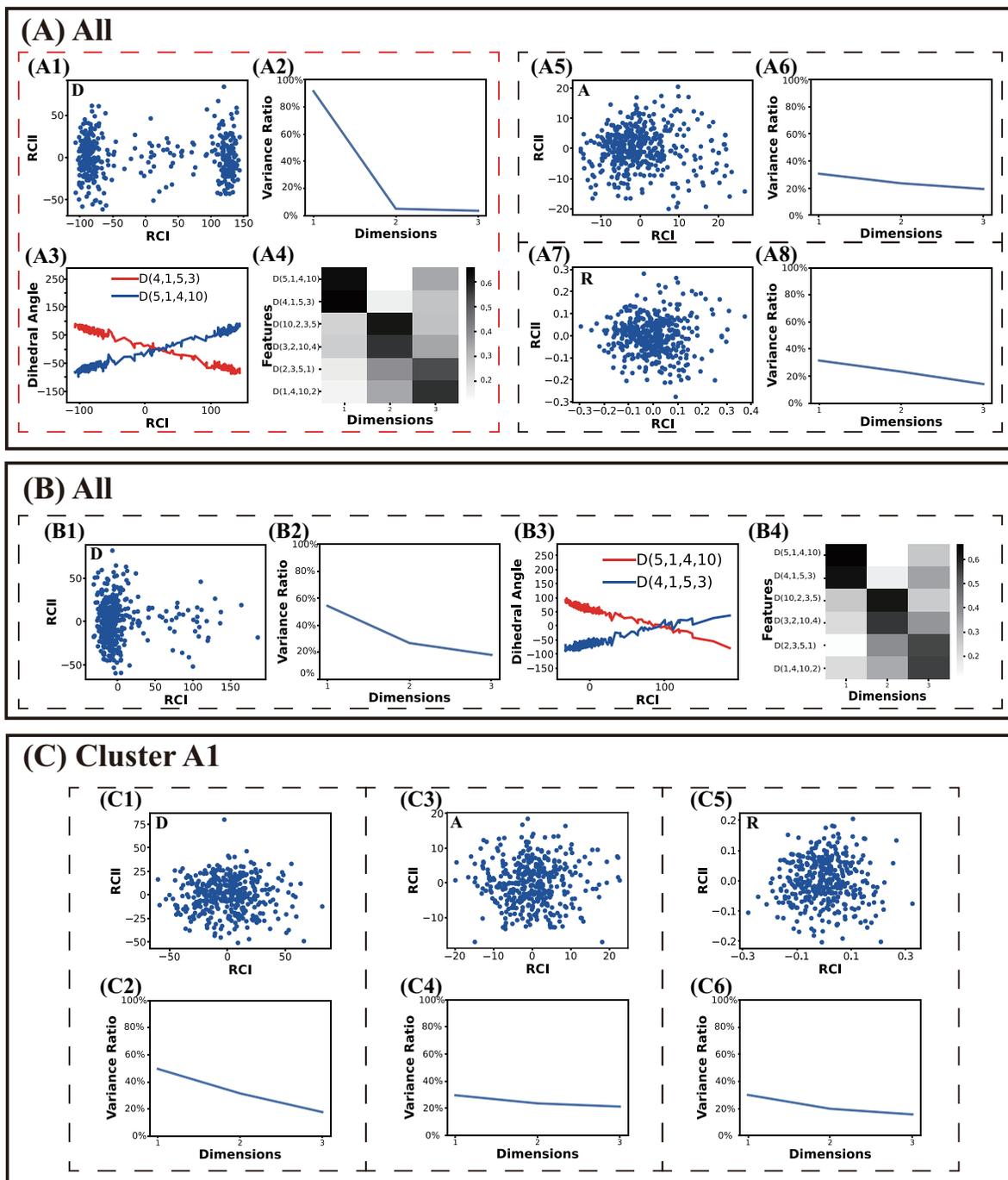
Figure S5. The PCA results of the ring part analysis. (A1)-(A4), (A5)-(A6) and (A7)-(A8) represent the PCA results of all hopping geometries based on $\mathbf{D}_{ring}$, $\mathbf{A}_{ring}$ and $\mathbf{R}_{ring}$, respectively. (B) demonstrates the PCA results of all hopping geometries based on $\mathbf{D}_{ring}$ after the mirror operation, while the PCA results related to $\mathbf{A}_{ring}$ and $\mathbf{R}_{ring}$ are the same as before since the angles and bond lengths are independent of the chirality. (C) displays the PCA results of Cluster $\mathbf{A1}$, and all these three cases indicate that the Cluster $\mathbf{A1}$ is non-separable.
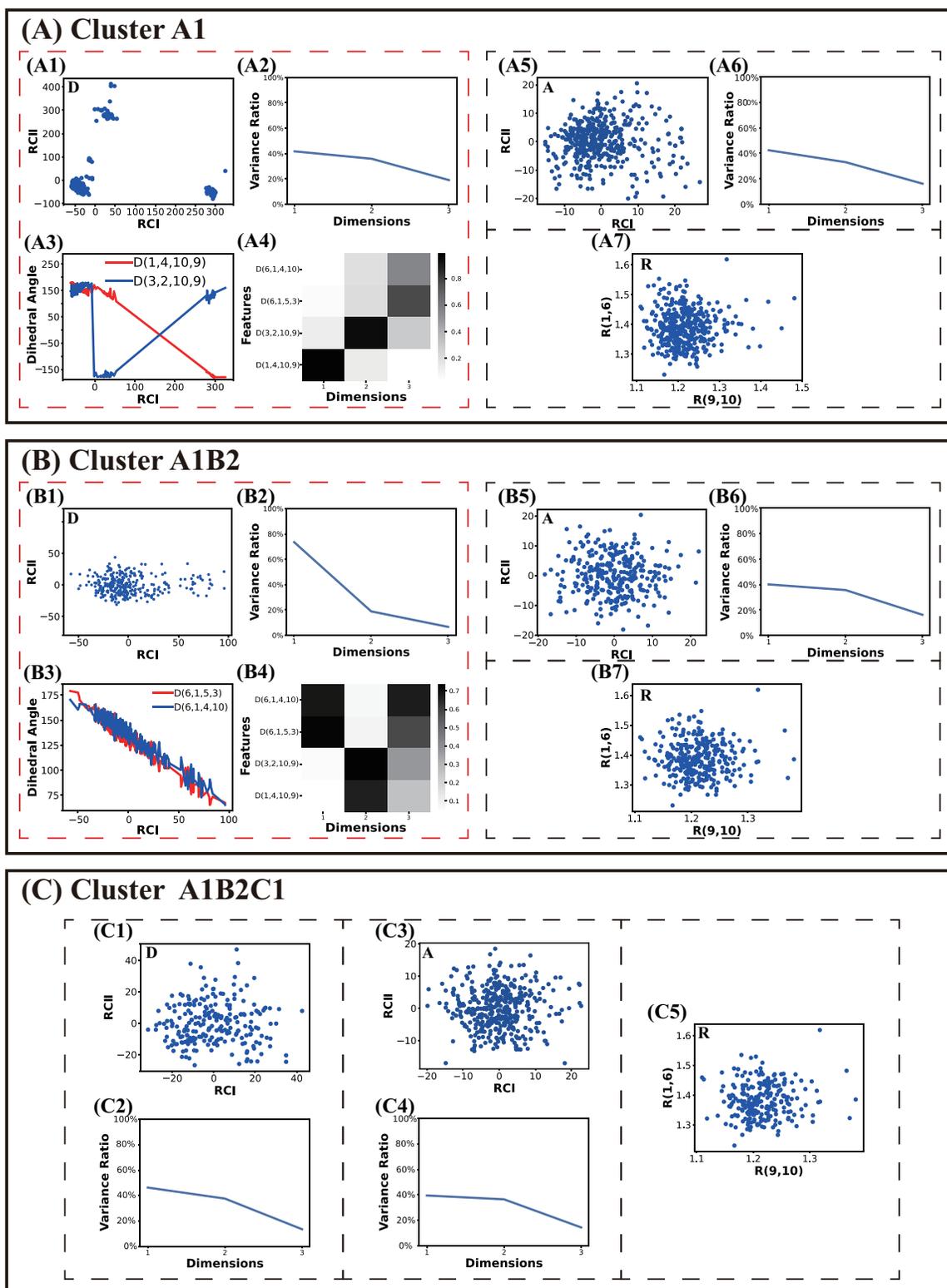
Figure S6. The analysis results of the end-group part. (A), (B) and (C) show the results based on the end-group descriptor sets (that is, $\mathbf{D}_{eg}$, $\mathbf{A}_{eg}$ and $\mathbf{R}_{eg}$) of the Cluster **A1**, **A1B2** and **A1B2C1**, respectively. (C) indicates that the Cluster **A1B2C1** is non-separable.

# 4 Statistical significance of the PCA results

The bootstrapping is a statistical technique used to estimate the model uncertainty by re-sampling the dataset with replacement.[17,18] Here, we employed it to examine the statistical significance of the major molecular motions of each channel, *i.e.*, the robustness of the contribution of geometric features of the first reduced coordinates in each reduced subspace. We took the channel **A1B2C1** as an example. For this channel, we collected the initial and hopping structures to form the data set. Next, we performed the bootstrap resampling (100 times) to give 100 datasets. In each resampled dataset, the initial geometry and corresponding hopping geometry exist or vanish at the same time. Next, we performed PCA in each resampled data set, obtained the first reduced dimension and made the statistical analyses on all results. As shown in Figure S7, all major geometric features in the first reduced dimension remains unchanged in this bootstrap resampling procedure. The average results from the bootstrap analysis are highly consistent with the values obtained from the whole dataset, as the small variances are observed.
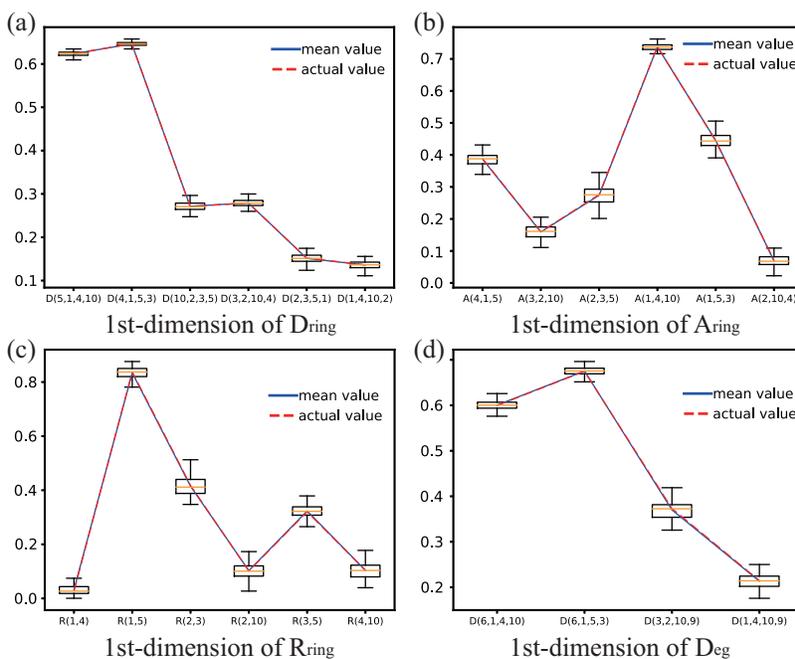


Figure S7. Box plots of the PCA results in four reduced subspaces ($\mathbf{D}_{ring}$, $\mathbf{A}_{ring}$, $\mathbf{R}_{ring}$ and $\mathbf{D}_{eg}$) based on the bootstrap resampling dataset.

15

# 5  Structural information

## 5.1  $S_0$ minimum

Table S2: The Cartesian coordinates of $S_0$ minimum (Unit:Å).

| Cartesian coordinates of $S_0$-min | | | |
|---|---|---|---|
| C | 1.14365000 | -0.17333700 | -0.00453100 |
| C | -1.30214300 | 1.04894800 | 0.01042900 |
| C | -0.11122400 | 1.71814000 | -0.00007100 |
| N | 0.00874200 | -0.93161900 | -0.00512100 |
| N | 1.12788700 | 1.13865300 | 0.00446900 |
| N | 2.34721100 | -0.83390500 | -0.06673100 |
| H | 2.41498300 | -1.74258400 | 0.37272900 |
| H | 3.13808000 | -0.22586300 | 0.10451900 |
| O | -2.26807800 | -1.15785400 | -0.00086100 |
| C | -1.31259000 | -0.39504200 | 0.00438900 |
| H | -2.25464400 | 1.56409600 | 0.01417900 |
| H | -0.09845600 | 2.80623000 | -0.00400100 |
| H | 0.05161400 | -1.94319900 | -0.07017100 |

Table S3: The internal coordinates of $S_0$ minimum. The unit of the bond length is Å, while the units of the bond angles and dihedral angles are degree.

| Internal coordinates of $S_0$-min | | | |
|---|---|---|---|
| R(1,4) | 1.3649 | R(1,5) | 1.3121 |
| R(1,6) | 1.3743 | R(2,3) | 1.3661 |
| R(2,10) | 1.444 | R(2,11) | 1.0829 |
| R(3,5) | 1.3679 | R(3,12) | 1.0882 |
| R(4,10) | 1.4262 | R(4,13) | 1.0146 |
| R(6,7) | 1.0116 | R(6,8) | 1.0122 |
| R(9,10) | 1.2226 | | |
| A(4,1,5) | 123.0596 | A(4,1,6) | 117.4596 |
| A(5,1,6) | 119.4348 | A(3,2,10) | 119.7433 |
| A(3,2,11) | 122.2622 | A(10,2,11) | 117.9921 |
| A(2,3,5) | 125.6008 | A(2,3,12) | 120.0051 |
| A(5,3,12) | 114.3919 | A(1,4,10) | 124.1488 |
| A(1,4,13) | 121.252 | A(10,4,13) | 114.5009 |
| A(1,5,3) | 115.7499 | A(1,6,7) | 118.0821 |
| A(1,6,8) | 112.822 | A(7,6,8) | 114.4402 |
| A(2,10,4) | 111.688 | A(2,10,9) | 129.017 |
| A(4,10,9) | 119.2936 | | |
| D(5,1,4,10) | -0.0257 | D(5,1,4,13) | 176.1668 |
| D(6,1,4,10) | -177.5419 | D(6,1,4,13) | -1.3495 |
| D(4,1,5,3) | -0.627 | D(6,1,5,3) | 176.8425 |
| D(4,1,6,7) | -32.5113 | D(4,1,6,8) | -169.6756 |
| D(5,1,6,7) | 149.8788 | D(5,1,6,8) | 12.7145 |
| D(10,2,3,5) | -1.0769 | D(10,2,3,12) | 179.4877 |
| D(11,2,3,5) | 179.4942 | D(11,2,3,12) | 0.0587 |
| D(3,2,10,4) | 0.3283 | D(3,2,10,9) | -179.2334 |
| D(11,2,10,4) | 179.7815 | D(11,2,10,9) | 0.2198 |
| D(2,3,5,1) | 1.2058 | D(12,3,5,1) | -179.331 |
| D(1,4,10,2) | 0.1831 | D(1,4,10,9) | 179.7926 |
| D(13,4,10,2) | -176.24 | D(13,4,10,9) | 3.3695 |

## 5.2  Internal coordinates of averaged structures

Table S4: The averaged internal coordinates (involving no H atom) of structures in the **A1B1** cluster. The unit of the bond length is Å, while the units of the bond angles and dihedral angles are degree.

| A1B1 | | | |
|---|---|---|---|
| R(1,4) | 1.402 | R(1,5) | 1.4373 |
| R(1,6) | 1.3862 | R(2,3) | 1.4478 |
| R(2,10) | 1.519 | R(3,5) | 1.3206 |
| R(4,10) | 1.4192 | R(9,10) | 1.201 |
| A(4,1,5) | 116.2681 | A(4,1,6) | 120.5751 |
| A(5,1,6) | 117.9529 | A(3,2,10) | 114.8519 |
| A(2,3,5) | 117.2388 | A(1,4,10) | 107.9942 |
| A(1,5,3) | 105.0508 | A(2,10,4) | 111.4593 |
| A(2,10,9) | 124.2974 | A(4,10,9) | 123.2821 |
| D(5,1,4,10) | 57.858 | D(6,1,4,10) | -132.7269 |
| D(4,1,5,3) | -70.2497 | D(6,1,5,3) | 119.8609 |
| D(10,2,3,5) | 29.5103 | D(3,2,10,4) | -40.2258 |
| D(3,2,10,9) | 132.3127 | D(2,3,5,1) | 21.506 |
| D(1,4,10,2) | -7.5297 | D(1,4,10,9) | -173.2045 |

Table S5: The averaged internal coordinates (involving no H atom) of structures in the **A1B2C1** cluster. The unit of the bond length is Å, while the units of the bond angles and dihedral angles are degree.

| A1B2C1 | | | |
|---|---|---|---|
| R(1,4) | 1.3847 | R(1,5) | 1.4635 |
| R(1,6) | 1.3819 | R(2,3) | 1.4629 |
| R(2,10) | 1.481 | R(3,5) | 1.306 |
| R(4,10) | 1.4348 | R(9,10) | 1.2159 |
| A(4,1,5) | 113.0952 | A(4,1,6) | 120.7892 |
| A(5,1,6) | 117.36 | A(3,2,10) | 115.5994 |
| A(2,3,5) | 118.2894 | A(1,4,10) | 106.7593 |
| A(1,5,3) | 105.2174 | A(2,10,4) | 112.6647 |
| A(2,10,9) | 125.7333 | A(4,10,9) | 120.6726 |
| D(5,1,4,10) | 66.7553 | D(6,1,4,10) | -145.2397 |
| D(4,1,5,3) | -69.7947 | D(6,1,5,3) | 141.07 |
| D(10,2,3,5) | 28.5417 | D(3,2,10,4) | -29.8875 |
| D(3,2,10,9) | 154.1494 | D(2,3,5,1) | 18.3939 |
| D(1,4,10,2) | -14.9181 | D(1,4,10,9) | 161.9058 |

Table S6: The averaged internal coordinates (involving no H atom) of structures in the **A1B2C2** cluster. The unit of the bond length is Å, while the units of the bond angles and dihedral angles are degree.

| A1B2C2 | | | |
|---|---|---|---|
| R(1,4) | 1.3921 | R(1,5) | 1.4318 |
| R(1,6) | 1.3986 | R(2,3) | 1.4557 |
| R(2,10) | 1.4978 | R(3,5) | 1.3238 |
| R(4,10) | 1.4099 | R(9,10) | 1.2085 |
| A(4,1,5) | 113.7156 | A(4,1,6) | 125.1102 |
| A(5,1,6) | 119.5453 | A(3,2,10) | 116.0687 |
| A(2,3,5) | 118.0463 | A(1,4,10) | 106.9198 |
| A(1,5,3) | 105.0157 | A(2,10,4) | 113.0124 |
| A(2,10,9) | 122.4846 | A(4,10,9) | 123.573 |
| D(5,1,4,10) | 67.3561 | D(6,1,4,10) | -120.321 |
| D(4,1,5,3) | -70.736 | D(6,1,5,3) | 116.3536 |
| D(10,2,3,5) | 24.1164 | D(3,2,10,4) | -26.7504 |
| D(3,2,10,9) | 150.5567 | D(2,3,5,1) | 20.8707 |
| D(1,4,10,2) | -15.581 | D(1,4,10,9) | 167.8496 |

Table S7: The averaged internal coordinates (involving no H atom) of structures in the **A1B2C3** cluster. The unit of the bond length is Å, while the units of the bond angles and dihedral angles are degree.

| A1B2C3 | | | |
|---|---|---|---|
| R(1,4) | 1.4006 | R(1,5) | 1.3766 |
| R(1,6) | 1.4194 | R(2,3) | 1.4162 |
| R(2,10) | 1.4857 | R(3,5) | 1.3638 |
| R(4,10) | 1.4392 | R(9,10) | 1.1907 |
| A(4,1,5) | 112.2453 | A(4,1,6) | 118.5604 |
| A(5,1,6) | 119.0969 | A(3,2,10) | 120.0945 |
| A(2,3,5) | 116.0868 | A(1,4,10) | 111.693 |
| A(1,5,3) | 111.1105 | A(2,10,4) | 111.8579 |
| A(2,10,9) | 124.2855 | A(4,10,9) | 122.9968 |
| D(5,1,4,10) | 59.0537 | D(6,1,4,10) | -88.0358 |
| D(4,1,5,3) | -65.3207 | D(6,1,5,3) | 81.2665 |
| D(10,2,3,5) | 14.7211 | D(3,2,10,4) | -18.8928 |
| D(3,2,10,9) | 161.3491 | D(2,3,5,1) | 27.0682 |
| D(1,4,10,2) | -17.2491 | D(1,4,10,9) | 163.8875 |

Table S8: The averaged internal coordinates (involving no H atom) of structures in the **A1B3** cluster. The unit of the bond length is Å, while the units of the bond angles and dihedral angles are degree.

| A1B3 | | | |
|------|------|------|------|
| R(1,4) | 1.3952 | R(1,5) | 1.4109 |
| R(1,6) | 1.3998 | R(2,3) | 1.4401 |
| R(2,10) | 1.4766 | R(3,5) | 1.3266 |
| R(4,10) | 1.4697 | R(9,10) | 1.2174 |
| A(4,1,5) | 111.6613 | A(4,1,6) | 121.0451 |
| A(5,1,6) | 119.6842 | A(3,2,10) | 118.8506 |
| A(2,3,5) | 116.9229 | A(1,4,10) | 103.0407 |
| A(1,5,3) | 109.4695 | A(2,10,4) | 113.8025 |
| A(2,10,9) | 125.0355 | A(4,10,9) | 119.3685 |
| D(5,1,4,10) | 71.9437 | D(6,1,4,10) | -112.6557 |
| D(4,1,5,3) | -68.6195 | D(6,1,5,3) | 114.5389 |
| D(10,2,3,5) | 15.2461 | D(3,2,10,4) | -9.8528 |
| D(3,2,10,9) | -171.5635 | D(2,3,5,1) | 25.0008 |
| D(1,4,10,2) | -35.2325 | D(1,4,10,9) | 137.383 |

Table S9: The averaged internal coordinates (involving no H atom) of structures in the **A2** cluster. The unit of the bond length is Å, while the units of the bond angles and dihedral angles are degree.

| A2 | | | |
|---|---|---|---|
| R(1,4) | 1.4063 | R(1,5) | 1.4086 |
| R(1,6) | 1.4132 | R(2,3) | 1.5724 |
| R(2,10) | 1.3246 | R(3,5) | 1.3028 |
| R(4,10) | 1.4029 | R(9,10) | 1.3867 |
| A(4,1,5) | 121.7051 | A(4,1,6) | 116.5984 |
| A(5,1,6) | 119.4355 | A(3,2,10) | 114.0328 |
| A(2,3,5) | 119.3661 | A(1,4,10) | 114.3933 |
| A(1,5,3) | 118.7886 | A(2,10,4) | 124.8 |
| A(2,10,9) | 124.1674 | A(4,10,9) | 108.4868 |
| D(5,1,4,10) | -16.3403 | D(6,1,4,10) | -165.9422 |
| D(4,1,5,3) | 13.1871 | D(6,1,5,3) | 168.6346 |
| D(10,2,3,5) | -12.5521 | D(3,2,10,4) | 10.8199 |
| D(3,2,10,9) | 168.2209 | D(2,3,5,1) | 10.9482 |
| D(1,4,10,2) | 13.0128 | D(1,4,10,9) | 162.2635 |

## 5.3 Cartesian coordinates of the typical structures

Table S10: The Cartesian coordinates of the typical structures of the **A1B1** cluster (Unit:Å).

| | A1B1 | | |
|---|---|---|---|
| C | 1.0413868345 | -0.3070839514 | 0.1159753547 |
| C | -0.9905443110 | 0.8882539190 | -0.7335673091 |
| C | -0.0081772859 | 1.6486562030 | -0.0125405372 |
| N | -0.0345932501 | -1.0883812862 | 0.4437425727 |
| N | 0.8888190206 | 1.0376964937 | 0.6354450029 |
| N | 2.2619148271 | -0.5320635539 | -0.5229254878 |
| H | 3.1276613950 | -0.0032197297 | -0.1418951916 |
| H | 2.2729622337 | -0.8278291643 | -1.4568480217 |
| O | -2.2730631316 | -0.9010270646 | 0.1751950309 |
| C | -1.2186511097 | -0.4853203912 | -0.0733317631 |
| H | -1.6838718711 | 1.3095216329 | -1.4726666699 |
| H | -0.1387034045 | 2.7317575820 | -0.0279058842 |
| H | 0.0206179223 | -2.0378958473 | 0.5680886624 |

24

Table S11: The Cartesian coordinates of the typical structures of the **A1B2C1** cluster (Unit:Å).

| | A1B2C1 | | |
|---|---|---|---|
| C | 0.9957209538 | -0.1994957544 | -0.0509099001 |
| C | -1.1217107234 | 0.9021865963 | -0.5816924609 |
| C | -0.0361624639 | 1.7895343020 | 0.0005242652 |
| N | -0.0274889592 | -0.9060763123 | 0.3813484775 |
| N | 1.0918513015 | 1.1778176365 | 0.5500011025 |
| N | 2.2183385086 | -0.7787725433 | -0.3261751317 |
| H | 2.0518901017 | -1.7059103176 | -0.6686610558 |
| H | 2.8181942521 | -0.0254276595 | -0.6958446092 |
| O | -2.2114105275 | -1.1298945557 | 0.0550844214 |
| C | -1.2286429302 | -0.4821754186 | -0.0131612585 |
| H | -2.0899529622 | 1.2953160957 | -0.620490461 |
| H | 0.0428846503 | 2.8933019513 | 0.1454735844 |
| H | 0.0374897055 | -1.8559318703 | 0.6791520803 |

Table S12: The Cartesian coordinates of the typical structures of the **A1B2C2** cluster (Unit:Å).

| | **A1B2C2** | | |
|---|---|---|---|
| C | 0.9708348860 | -0.0934762073 | 0.2029202195 |
| C | -0.8920574441 | 0.7672074905 | -0.8666924291 |
| C | 0.0739729650 | 1.6996652206 | -0.4703984912 |
| N | -0.0726976688 | -0.8806616737 | 0.7965769225 |
| N | 1.0323622202 | 1.3119581355 | 0.3513628692 |
| N | 2.1352518139 | -0.6779344159 | -0.1450885037 |
| H | 2.8305626524 | -0.3961072858 | -0.8506502676 |
| H | 2.0744189270 | -1.6009677218 | -0.5468788740 |
| O | -2.2968566118 | -1.2020485893 | 0.0936047421 |
| C | -1.2431535494 | -0.5712496877 | 0.0008163889 |
| H | -1.4961223661 | 0.9877334915 | -1.7585583453 |
| H | -0.2652935116 | 2.6656842297 | -0.7586229467 |
| H | 0.0303461670 | -1.0853793066 | 1.7303051695 |

Table S13: The Cartesian coordinates of the typical structures of the **A1B2C3** cluster (Unit:Å).

| A1B2C3 | | | |
|---|---|---|---|
| C | 0.9050719325 | -0.2929361927 | 0.4988235889 |
| C | -0.6251175425 | 0.9558308913 | -0.8961392721 |
| C | 0.3365212594 | 1.7432520503 | -0.1642041534 |
| N | -0.3027205339 | -0.8557718184 | 0.6336679676 |
| N | 1.0045016497 | 1.0750568935 | 0.8616760237 |
| N | 1.8675581891 | -1.0054887075 | -0.206139901 |
| H | 1.9796821979 | -1.8816050025 | 0.2864310304 |
| H | 2.8768887437 | -0.7139074456 | -0.2932925674 |
| O | -2.2193073158 | -0.8858690094 | -0.4763904908 |
| C | -1.1639222942 | -0.3361304420 | -0.2516218127 |
| H | -1.6809914699 | 1.4265243161 | -1.1193510798 |
| H | 0.2930566346 | 2.8346006933 | -0.2850880108 |
| H | -0.4248423375 | -1.5685203481 | 1.2431927806 |

Table S14: The Cartesian coordinates of the typical structures of the **A1B3** cluster (Unit:Å).

| | A1B3 | | |
|---|---|---|---|
| C | 1.0986043426 | -0.1902790124 | 0.2716363242 |
| C | -1.1938955766 | 0.9128111201 | -0.4692701454 |
| C | 0.0937589698 | 1.7711558021 | 0.0871010012 |
| N | -0.0835897156 | -0.8263654037 | 0.7385828462 |
| N | 1.0899298556 | 1.2646415334 | 0.5892611512 |
| N | 1.9567440720 | -0.7560969521 | -0.6425177239 |
| H | 1.6760778377 | -1.3678790940 | -1.3465816776 |
| H | 2.6386277690 | -0.0254570440 | -0.8744018596 |
| O | -2.0000740890 | -1.3619131073 | -0.3944141959 |
| C | -1.2263825559 | -0.4334270400 | -0.0440594228 |
| H | -2.0890132646 | 1.4800749019 | -0.3934355391 |
| H | -0.1995079137 | 2.7774935979 | -0.1423045481 |
| H | -0.0094327899 | -1.8031752876 | 0.9051499589 |

Table S15: The Cartesian coordinates of the typical structures of the **A2** cluster (Unit:Å).

| | **A2** | | |
|---|---|---|---|
| C | 1.1305970189 | -0.2603009267 | -0.1351821600 |
| C | -1.3813607848 | 0.9681435923 | -0.0696961853 |
| C | 0.0294070106 | 1.8149509089 | 0.0690967597 |
| N | -0.0091265466 | -0.9709345943 | -0.0038336372 |
| N | 1.1339510191 | 1.2825089229 | 0.0174327198 |
| N | 2.3375280759 | -0.8913893309 | -0.0389274820 |
| H | 2.2109259547 | -1.6062481799 | 0.4252334189 |
| H | 2.7746303574 | -0.1786492877 | 0.4173529051 |
| O | -2.3027939037 | -1.2569985934 | 0.0379313300 |
| C | -1.2099495269 | -0.2819481071 | 0.0223353940 |
| H | -2.5137870583 | 1.6084475760 | -0.0162152898 |
| H | -0.0484693459 | 3.0575039114 | 0.1132175839 |
| H | -0.1201605429 | -2.0038461129 | -0.2501211240 |

# 6 Conical intersections

## 6.1 The Cartesian coordinates

Table S16: The Cartesian coordinates of the *Ethyl.I* CI (Unit:Å).

| *Ethyl.I* | | | |
|---|---|---|---|
| C | 0.982646 | -0.261045 | 0.757038 |
| C | -0.765071 | 0.958395 | -0.731474 |
| C | 0.295363 | 1.669170 | -0.205710 |
| C | -1.164481 | -0.324923 | -0.115891 |
| N | -0.296511 | -0.787534 | 0.899566 |
| N | 0.955011 | 1.100684 | 0.830000 |
| N | 1.851896 | -0.842078 | -0.178661 |
| O | -2.156085 | -0.940054 | -0.409997 |
| H | 1.747962 | -1.831126 | -0.272666 |
| H | 2.808642 | -0.598386 | -0.028927 |
| H | -1.452659 | 1.392482 | -1.429032 |
| H | 0.465524 | 2.705496 | -0.430420 |
| H | -0.425859 | -1.746043 | 1.147739 |

Table S17: The Cartesian coordinates of the *Ethyl.II* CI (Unit:Å).

| *Ethyl.II* | | | |
|---|---|---|---|
| C | 1.033840 | -0.107553 | -0.067211 |
| C | -1.001759 | 0.917653 | -0.657173 |
| C | -0.037673 | 1.770472 | 0.039862 |
| N | 0.017183 | -0.904932 | 0.451746 |
| N | 1.042721 | 1.220605 | 0.486212 |
| N | 2.256078 | -0.663747 | -0.349755 |
| H | 2.952987 | 0.011072 | -0.586520 |
| H | 2.233567 | -1.428635 | -0.991586 |
| O | -2.129828 | -1.159564 | -0.323290 |
| C | -1.159637 | -0.462665 | -0.170535 |
| H | -1.821985 | 1.329966 | -1.213630 |
| H | -0.258016 | 2.795110 | 0.285585 |
| H | 0.138281 | -1.884716 | 0.593061 |

Table S18: Cartesian coordinates of the C=O stretching CI (Unit:Å).

| C=O stretching | | | |
| --- | --- | --- | --- |
| C | 1.131280 | -0.143784 | 0.105868 |
| C | -1.333251 | 0.993382 | 0.190412 |
| C | -0.061705 | 1.814972 | -0.095676 |
| C | -1.197695 | -0.276015 | -0.052031 |
| N | -0.002075 | -0.981078 | -0.144037 |
| N | 1.123477 | 1.152470 | -0.044085 |
| N | 2.329847 | -0.868614 | 0.057935 |
| O | -1.915261 | -1.466363 | 0.308440 |
| H | 2.394755 | -1.540199 | 0.797794 |
| H | 3.111439 | -0.246529 | 0.095322 |
| H | -2.206112 | 1.435326 | 0.628087 |
| H | -0.091958 | 2.668122 | -0.746006 |
| H | 0.101014 | -1.531621 | -0.979038 |

## 6.2 Branching space

The description of the branching space for CIs is also critical, which is spanned by two vectors, *i.e.*, the gradient difference $\boldsymbol{g}$ vector and the derivative coupling $\boldsymbol{h}$ vector. We calculated the orthogonal $\boldsymbol{g}$ and $\boldsymbol{h}$ vectors for each CI, and these scaled vectors are shown in Figure S8. In addition, we also computed the $\boldsymbol{s}$ vector (after the orthogonalization to both $\boldsymbol{g}$ and $\boldsymbol{h}$ vectors) that characterizes the CI seam. The way to construct these three orthogonal vectors can be found in the work of Yarkony.[19]

These vectors confirm the major molecular motions identified by our analytical protocol. For example, at the *Ethyl.I* CI, the C1-puckering motion was identified as the major one by our protocol, which was confirmed by the $\boldsymbol{g}$ vector as shown in Figure S8.(a). At the *Ethyl.II* CI, the $\boldsymbol{g}$ vector verifies the dominant roles of the C1-puckering motion and the conjugation alteration of the ring part, consistent with our analysis of active coordinates. While the $\boldsymbol{s}$ vector represents that the *Ethyl.II* CI seam is characterized by different $NH_2$ out-the-plane motions, also consistent with the channels **A1B1**, **A1B2C1** and **A1B2C2** are correlated to this CI. At the C=O stretching CI, the $\boldsymbol{g}$ vector includes the C=O elongation motion.
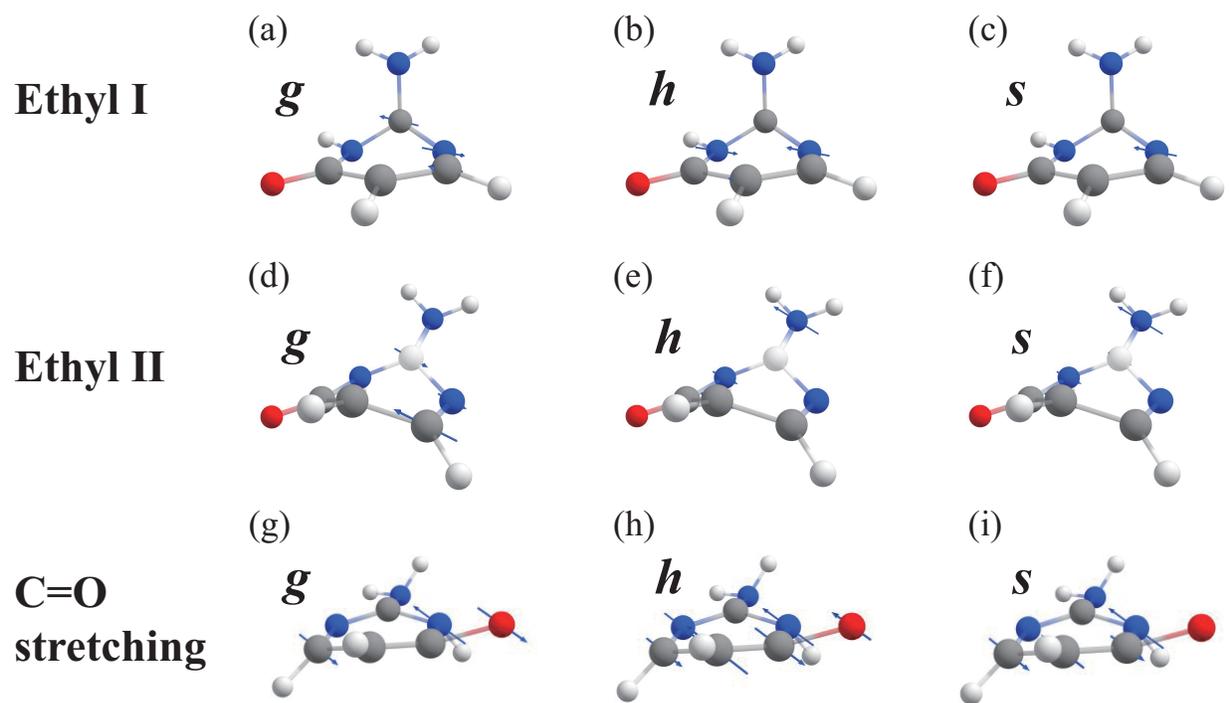
Figure S8. The orthogonal **g**, **h** and **s** vectors of three CIs (*Ethyl.I*, *Ethyl.II* and C=O stretching)

# References

(1) Pulay, P.; Fogarasi, G. Geometry optimization in redundant internal coordinates. *J. Chem. Phys.* **1992**, *96*, 2856–2860.

(2) Peng, C.; Ayala, P. Y.; Schlegel, H. B.; Frisch, M. J. Using redundant internal coordinates to optimize equilibrium geometries and transition states. *J. Comput. Chem.* **1996**, *17*, 49–56.

(3) Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. Chemometr Intell Lab 2 (1–3): 37–52. 1987.

(4) Abdi, H.; Williams, L. J. Principal component analysis. *WIREs Comput. Stat.* **2010**, *2*, 433–459.

(5) Peng, J.; Xie, Y.; Hu, D.; Lan, Z. Analysis of bath motion in MM-SQC dynamics via dimensionality reduction approach: Principal component analysis. *J. Chem. Phys.* **2021**, *154*, 094122.

(6) Capano, G.; Penfold, T.; Chergui, M.; Tavernelli, I. Photophysics of a copper phenanthroline elucidated by trajectory and wavepacket-based quantum dynamics: a synergetic approach. *Phys. Chem. Chem. Phys.* **2017**, *19*, 19590–19600.

(7) Rohrdanz, M. A.; Zheng, W.; Clementi, C. Discovering mountain passes via torchlight: Methods for the definition of reaction coordinates and pathways in complex macromolecular reactions. *Annu. Rev. Phys. Chem.* **2013**, *64*, 295–316.

(8) Glielmo, A.; Husic, B. E.; Rodriguez, A.; Clementi, C.; Noé, F.; Laio, A. Unsupervised learning methods for molecular simulation data. *Chem. Rev.* **2021**, *121*, 9722–9758.

(9) Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X., et al. A density-based algorithm for discovering clusters in large spatial databases with noise. kdd. 1996; pp 226–231.

(10) Kriegel, H.-P.; Kröger, P.; Sander, J.; Zimek, A. Density-based clustering. *WIREs Data. Min. Knowl.* **2011**, *1*, 231–240.

(11) Schubert, E.; Sander, J.; Ester, M.; Kriegel, H. P.; Xu, X. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. *ACM T. Database Syst.* **2017**, *42*, 1–21.

(12) Ward Jr, J. H. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* **1963**, *58*, 236–244.

(13) Kaufman, L.; Rousseeuw, P. J. *Finding groups in data: an introduction to cluster analysis*; John Wiley & Sons, 2009.

(14) Gordon, A. D. A review of hierarchical classification. *J. R. Sata. Soc. A Stat.* **1987**, *150*, 119–137.

(15) Landau, S.; Leese, M.; Stahl, D.; Everitt, B. S. *Cluster analysis*; John Wiley & Sons, 2011.

(16) Tully, J. C. Molecular dynamics with electronic transitions. *J. Chem. Phys.* **1990**, *93*, 1061–1071.

(17) Efron, B.; Tibshirani, R. J. *An introduction to the bootstrap*; CRC press, 1994.

(18) Varian, H. Bootstrap tutorial. *Mathematica J.* **2005**, *9*, 768–775.

(19) Yarkony, D. R. On the adiabatic to diabatic states transformation near intersections of conical intersections. *J. Chem. Phys.* **2000**, *112*, 2111–2120.