

# SCIENTIFIC REPORTS



OPEN

## Direct Learning Hidden Excited State Interaction Patterns from *ab initio* Dynamics and Its Implication as Alternative Molecular Mechanism Models

Fang Liu<sup>1</sup>, Likai Du<sup>1</sup>, Dongju Zhang<sup>2</sup> & Jun Gao<sup>1</sup>

The excited states of polyatomic systems are rather complex, and often exhibit meta-stable dynamical behaviors. Static analysis of reaction pathway often fails to sufficiently characterize excited state motions due to their highly non-equilibrium nature. Here, we proposed a time series guided clustering algorithm to generate most relevant meta-stable patterns directly from *ab initio* dynamic trajectories. Based on the knowledge of these meta-stable patterns, we suggested an interpolation scheme with only a concrete and finite set of known patterns to accurately predict the ground and excited state properties of the entire dynamics trajectories, namely, the prediction with ensemble models (PEM). As illustrated with the example of sinapic acids, The PEM method does not require any training data beyond the clustering algorithm, and the estimation error for both ground and excited state is very close, which indicates one could predict the ground and excited state molecular properties with similar accuracy. These results may provide us some insights to construct molecular mechanism models with compatible energy terms as traditional force fields.

The photophysical or photochemical processes are extremely important for the evolution of life and environments. After the molecule is excited onto higher electronic states, the molecule would undergo a rather complex sequence of dynamics, such as radiative electronic transitions (fluorescence, phosphoresences), nonradiative electronic transitions (internal conversions, intersystem crossings), energy transfers and chemical reactions, etc. Many efforts have been devoted to understand the molecular basis of the possible photophysical or photochemical mechanism in the last decades<sup>1–10</sup>. The reliable theoretical simulation of excited state processes of polyatomic molecules is not so straightforward in most cases, which requires the accurate calculation of electronic excited states for highly non-equilibrium molecular geometries and the nonadiabatic transitions between the electronic and nuclear degrees of freedom<sup>11–15</sup>. The non-equilibrium nature of excited state processes presents some difficulty in establishing a realistic description of these ultrafast processes which are completed in picoseconds or even femtoseconds. Similar problem is also mentioned in the thermally activated ground state reactions, especially, for the transition state structures, the dynamical correlations are known to disrupt the minimal energy path picture<sup>16–18</sup>. Therefore, a dynamic description is much preferred.

*Ab initio* molecular dynamics (AIMD) methods have been extended for the excited states problems<sup>19–24</sup>, such as the on-the-fly surface hopping method<sup>21, 22, 25–27</sup>, for which, the dynamics and electronic structure problems are solved simultaneously. This involves nuclear dynamics to determine the time evolution of the molecular geometry in concert with electronic structure methods capable of computing electronic excited state potential energy surfaces (PES). A large number of trajectories are usually produced from the excited state AIMD simulations. Such simulations directly include all nuclear degrees of freedom, which provide a rather rich picture of the microscopic processes. However, for medium to large size molecules, the trajectories are generally chaotic and becoming

<sup>1</sup>Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan, 430070, P. R. China. <sup>2</sup>Institute of Theoretical Chemistry, Shandong University, Jinan, 250100, P. R. China. Correspondence and requests for materials should be addressed to L.D. (email: [dulikai@mail.hzau.edu.cn](mailto:dulikai@mail.hzau.edu.cn)) or J.G. (email: [gaojun@sdu.edu.cn](mailto:gaojun@sdu.edu.cn))

inscrutable for human to extract the physical insights of interesting. Therefore, it is very necessary to design more intelligent algorithms to depict not only the available reaction channels, but also further dynamics details, such as the hidden meta-stable states and their interaction networks.

To meet the challenge of tackling with the PES complexity over AIMD trajectories, much recent efforts have been devoted to machine learning (ML) algorithms<sup>28–35</sup>. Generally, the number of local minima, and hence, the number of meta-stable states, grows exponentially along with system size. An important method for shrinking the data set is to apply a clustering algorithm to obtain a family of clusters (microstates) of much smaller size than the original data set. In this aspect, the nonlinear dimensionality reduction algorithms have been used to investigate the conical intersection topology of the excited state dynamics<sup>28</sup>. And the Markov state models (MSMs) have been proposed to automatically construct coarse grained representations for biological macromolecular conformation dynamics<sup>36,37</sup>, that are readily humanly understandable. As a practical issue, the prospect of using ML algorithms to tackle the flood of dynamics data to yield statistical significance is indeed very promising.

In recent years, ML algorithms also become a popular and effective tool to improve computational chemistry methods<sup>38–43</sup>. Using ML algorithms to reproduce *ab initio* calculation results would greatly reduce the computational cost without loss of the accuracy<sup>30</sup>. And the ML algorithms have been successfully used to predict various molecular properties on their electronic ground or excited states<sup>30,44,45</sup>. For example, the ML algorithms have been implemented to predict (PES) at QM and QM/MM level successfully using neural networks (NN) model<sup>39,46,47</sup>. The ML algorithms are also reported to accelerate the AIMD simulation of material systems<sup>48,49</sup>. In this regard, it is very interesting to construct new efficient computational methods based on the knowledge of the statistical significance from ML studies.

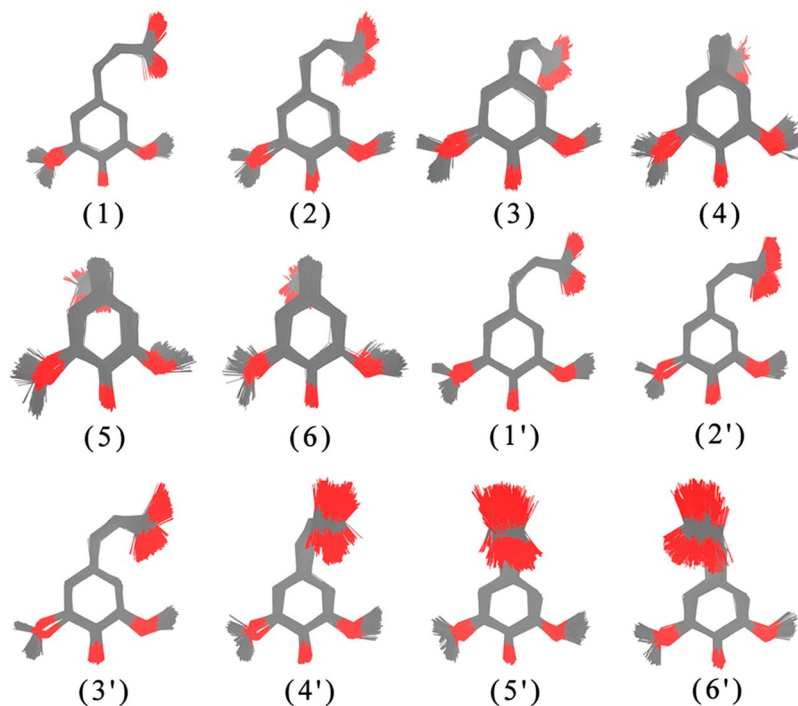
So far, excited state molecular dynamics (MD) are often restricted within direct *ab initio* methods for small to medium size molecules, and the use of parameter-based empirical force field is generally avoided. In contrast, the classical MD with empirical force field has been successfully applied to very large molecule systems in the ground state, such as protein conformational dynamics or protein-ligand interactions<sup>50–57</sup>. A few attempts have been devoted to develop excited state force field for efficiently describing electronic excited states motions. The force fields parameter sets has been developed for a few typical molecules in low-lying electronic excited states based on quantum chemical calculations<sup>58,59</sup>. A few novel model for excited state empirical force field have also been proposed, such as the interpolated mechanics-molecular mechanics (IM/MM)<sup>60</sup>, and electron force field (eFF)<sup>61</sup>. And these progresses may provide more insights for much longer time scale (i.e. nanoseconds) excited state simulation of large molecules in condense phases. The main difficulty in developing excited state empirical force fields is the relative scarcity of the universal mathematical function forms (PESs and their couplings) and the failure to sufficiently characterize excited state motions due to their highly non-equilibrium nature. The excited state AIMD trajectories contain large amount of information about their traveling PES, which can be used intrinsically as data sets to design the new models. We suspect the ML algorithm may provide an idea tool for revealing the coarse-grained representations of the excited state processes, as well as the main transitions between the hidden meta-stable states. And the data mining of the AIMD trajectories may promote the future excited state force field development.

In this work, we present a time series guided clustering algorithm to extract the main features of meta-stable states and their correlations from an ensemble of excited state AIMD trajectories. On the basis of these finite meta-stable patterns, the conformation similarity was explored to build an interpolation scheme, namely, the prediction with ensemble models (PEM), to estimate the ground and excited state properties of the entire dynamics trajectories. The PEM method does not require any training data beyond the clustering algorithm, and we could correctly predict the charge population and excitation energy, in comparable with the DFT/TDDFT calculations. As a test case, the excited states  $S_1$  of sinapic acid (SA) was used as a benchmark system, which is an essential UV-B screening ingredient in natural plants<sup>62</sup>. This work highlights the potential power of ML algorithm in computational chemistry to extract chemical insights or develop the state-of-art theoretical models.

## Results

**Characters of Possible Excited State Meta-stable Patterns.** Here, we focus our attentions on directly extracting the physical insights from excited state dynamics trajectories. To build such kinetic models, it is necessary to map out the dominant long lived, kinetically meta-stable states visited by the molecular system. Thus, we use a conformation clustering algorithm (K-means) to automatically split the time series of dynamics trajectories into geometrically distinct clusters. This also allows us to characterize possible rare events not easily observable in simulations. The detail distribution of molecule structures for each meta-stable pattern is given in Table S1.

Figure 1 shows the overlapped molecular geometries for each pattern. The geometric features of meta-stable patterns were sampled on the basis of K-means clustering algorithm. Some attempts have been performed to vary the number ( $k$ ) of distinct clusters, and finally, the number of  $k = 12$  are adopted. We also provide the meta-stable patterns with the cluster number  $k = 20$  in Figure S1. It is interesting that the conformation space from dynamics trajectories could be split into a limited number of concrete clusters. The clusters  $i$  mainly refers to the anti-clockwise rotation of the collective dihedral angles, meanwhile, the cluster  $i'$  describes the clockwise rotation. The standard deviations of heavy atom coordinates for each pattern after alignments are usually less than 0.5 Å (Table S2). Because the nearly symmetric character of the dihedral angle motions, the clustering results show very similar symmetric features (Fig. 1). Two meta-stable patterns (1/1') are close in the same plane of the aromatic ring (anti-clockwise and clockwise), while some meta-stable patterns (6/6') are perpendicular to the plane of the aromatic ring. Most other patterns show transition characters between these two kind of meta-stable patterns. Since only a few dihedral angles were selected as molecule descriptors, the geometric deviation from the methoxy group is observed. This shows neglectable effects for our qualitative analysis of the excited state *cis*-SA molecule, and thus the clustering results are still very reasonable.



**Figure 1.** Possible meta-stable patterns derived from K-means clustering results. The oxygen and carbon atoms are shown in red and gray, and the position of the hydrogen atom is not shown for clarity.

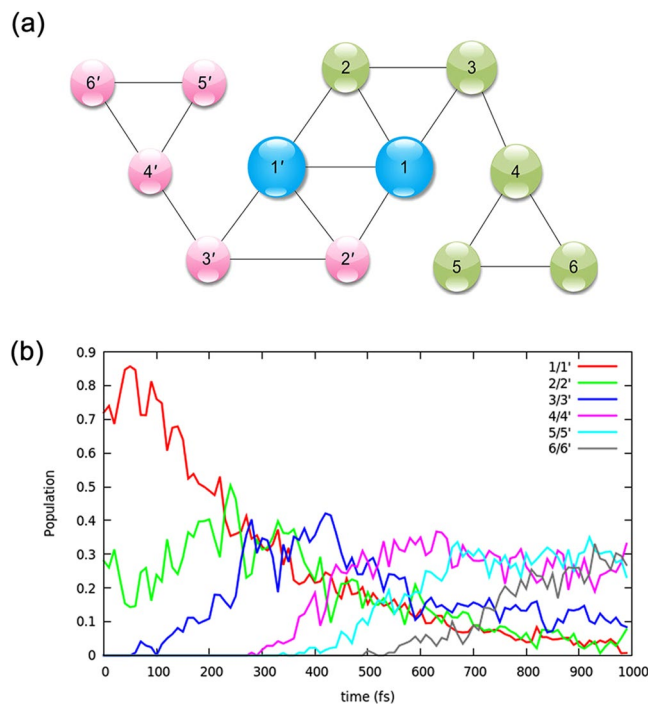
Next, the transition probabilities between these meta-stable patterns are determined, in order to create a kinetic model of the system's conformational dynamics. This can be easily achieved by counting the number of transition times along the time series of the dynamics trajectories. After the transition probabilities between clusters are resolved, one can have a kinetic network, which retains a coarse version of the dynamics. This conformation space network can be drawn as a graph, which can be described by graph theory. Each pattern is represented by a vertex as a colored circle in the graph. And the edge represents the connection among patterns, which could undergo direct transformation between each other.

At first glance, the network is not fully connected (Fig. 2a), which indicates there exists explicit reaction pathway among these meta-stable patterns. The near symmetric rotation character of *cis*-SA molecule is also reflected in the network. The network seems to be divided into three sub-graphs; one is correlated to near in-plane patterns, while the other two are correlated to out of plane patterns (i.e. inward and outward of the aromatic ring plane). The patterns 3/3' and 4/4' are vertex separators, the removal of which would disconnect the remaining network. Therefore, this graph provides a suitable view to directly interpret the excited state trajectories, and an acceptable way to recognize the non-negligible intermediate states. We also try to provide a time dependent description of this kinetic model. Figure 2b shows the population of each meta-stable pattern as a function of time. Note that the pattern  $i$  and  $i'$  (1~6) are combined to provide a simplified representation. Generally, the features of in-plane patterns (1/1' or 2/2') are often found at the beginning of the trajectories, and features of out of plane patterns (i.e. 4/4', 5/5', 6/6') are mainly observed during the evolution of the dynamic trajectories. The initial quick drop of the population of the patterns 1/1' is observed within the first 300 fs. And then, the continuous decay from 20% to 5% takes place along with fast oscillations. The total population of the out of plane patterns (i.e. 4/4', 5/5', 6/6') increases to nearly 80% beyond 600 fs. In this way, the dynamics process could be described using merely a few meta-stable patterns.

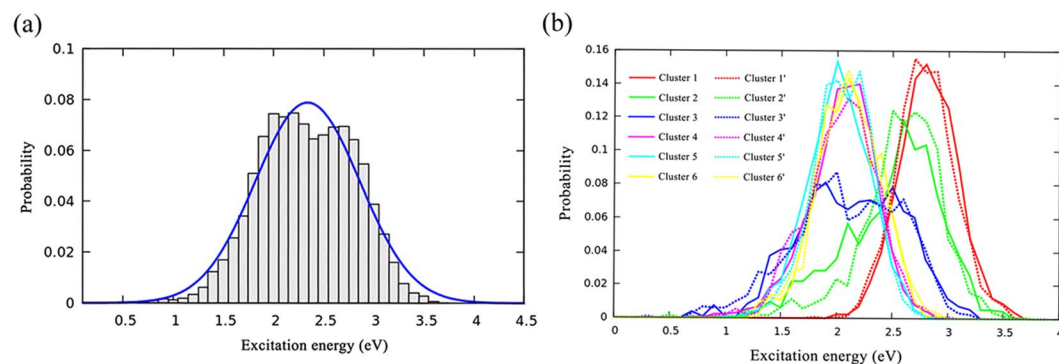
Note that this kinetic network is built from short timescale (~1 ps) dynamics trajectories and an ensemble of uncoupled trajectories were used to extract long timescale dynamics features. This ensemble dynamics approach has been commonly used in the classical MD simulations<sup>63–66</sup>. We have also performed five longer timescale (~6 ps) excited state AIMD simulations and these meta-stable patterns are indeed observed in longer excited state AIMD trajectories. Therefore, our algorithm can successfully tackle the issue of capturing proper meta-stable patterns that faithfully represent excited state dynamics at the timescale of interest.

**The Statistical Analysis of the Meta-stable Patterns.** The clustering algorithm creates a group of networks, which gives us a unique perspective for understanding excited states processes. Generally, each cluster shows low structural bias, while different clusters show high structural variance. This means that each meta-stable pattern holds very similar molecular structures. Frequently, the molecular properties of similar structure are very close. Here, we focus our attention on the distribution of molecular properties for each meta-stable pattern, such as excitation energy and charge population, which is also distinguishable among meta-stable patterns.

The excitation energies are very important to understand the molecular excited states. Figure 3a shows distribution of excitation energies ( $S_1$ ) for all 13226 sampled structures. Significant variation of the excitation energies



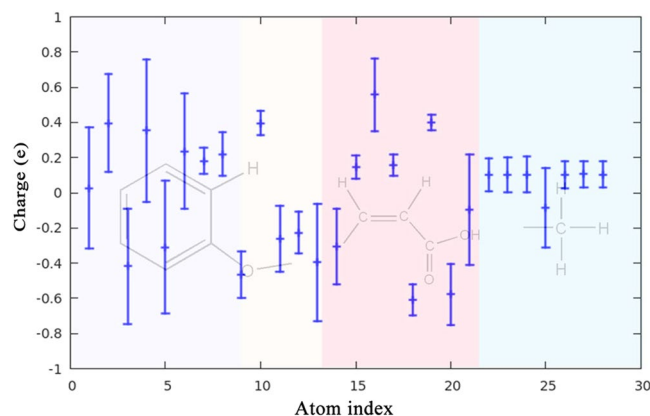
**Figure 2.** (a) The network formed by connecting the meta-stable patterns. The size of the circles reflects the relative number of snapshots in each pattern. (b) The time dependent population of each meta-stable pattern. And the meta-stable patterns could be corresponding to the possible free energy basins.



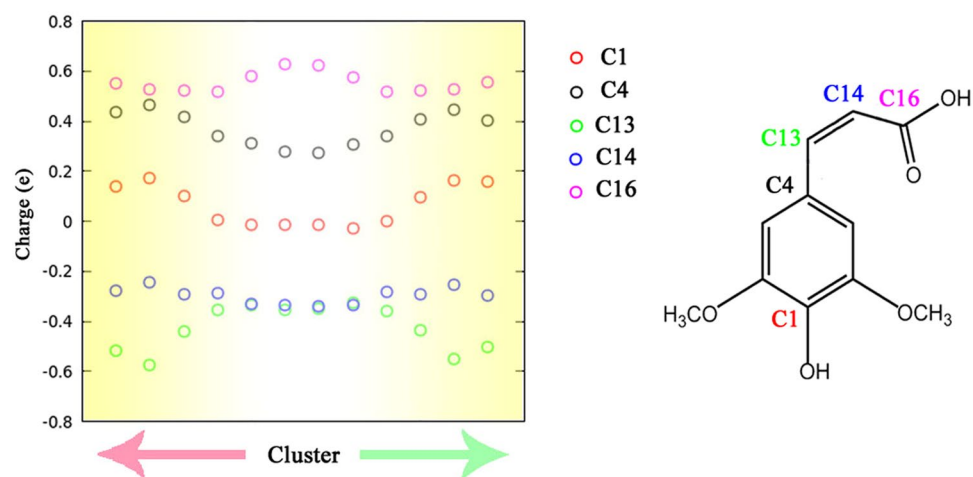
**Figure 3.** (a) The distribution of excited energies for all meta-stable patterns. The fitted Gaussian distribution curve is shown ( $\mu = 2.34$ ,  $\sigma = 0.52$ ). (b) The Gaussian distribution of excited energies for each meta-stable pattern.

is observed. The distribution curve is fitted using Gaussian function with the mean value ( $\mu$ ) of 2.34 eV, and standard deviation ( $\sigma$ ) of 0.52 eV. Furthermore, the distribution of the excitation energies is also analyzed for each meta-stable pattern (Fig. 3b). The mean value and standard deviation of the Gaussian distribution were fitted as indicators. The excitation energies of the patterns 1/1' show the highest mean value of excitation energy ( $\mu = 2.8$  eV). As the carboxyl moiety rotates out of the aromatic ring plane, the value of excitation energy obviously reduces. The patterns 4/4', 5/5', 6/6' show lower excitation energy ( $\mu = 1.9\sim 2.1$  eV). The transition patterns (3/3', 4/4') show boarder excitation energy distribution ( $\sigma = 0.30\sim 0.50$  eV), which may cover the characters of the in-plane and out-of-plane geometries. In summary, each pattern has its distinct characters and property. The standard deviation ( $\sigma$ ) for each pattern is much smaller (lower bias) than the whole meta-stable patterns network.

Similar conclusion is also available for the ESP partial charge population on the excited state. The partial charge distribution for each atom of SA molecule in 13226 structures has been fitted with Gaussian function. Figure 4 shows the distribution features of the partial charge population for each atom of SA molecule. The median of the error bar refers to the mean value of the Gaussian distribution, meanwhile, the maximum and minimum of the error bar refers to  $\mu \pm 2.58\sigma$ , which covers 99% of the distribution probability. The partial charge fluctuation for the hydrogen atom is usually neglectable; however, the partial charge for the aromatic ring or the



**Figure 4.** The distribution of partial charge for each atom of SA molecule in the meta-stable patterns network. The maximum and minimum of the error bars are related to the value of  $\mu \pm 2.58\sigma$  in Gaussian distribution.



**Figure 5.** The average partial charges of each meta-stable pattern for a few selected atoms of SA molecule. The value of the meta-stable patterns  $1/1'$  are shown in the middle of the X axis.

carboxyl motif varies much large. For instance, the partial charge fluctuation is indeed very large, nearly  $\sim 1.0$  e, for a few atoms. Therefore, it is very necessary to consider the geometric dependence of the partial charge population on the excited state.

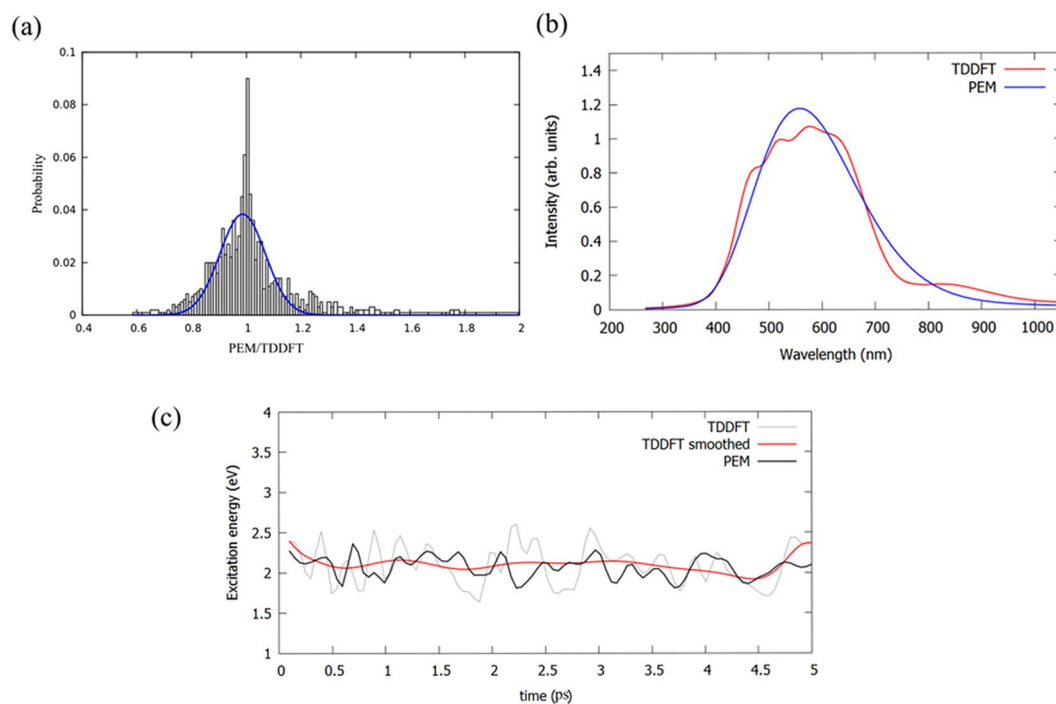
Then, we also summarized the partial charge population for each meta-stable pattern. For simplify, Fig. 5 shows the mean value of partial charges of a few specific atoms with larger variations for each meta-stable pattern. It is obvious that each meta-stable pattern shows distinct partial charge population. The distribution features of the partial charges for each pattern, i.e. mean value ( $\mu$ ) and standard derivation ( $\sigma$ ) are much smaller than that of the entire meta-stable patterns. And each meta-stable pattern shows very different partial charge distribution. In general, the clustering algorithm indeed creates a group of networks, each with low bias and high variance.

**Prediction with Ensemble Models.** In the language of machine learning, the clustering algorithm serves as a classifier to partition the molecular conformations into a few distinct meta-stable patterns. The PEM algorithm requires the number ( $M$ ) of meta-stable patterns for prediction (see Eq. 1). The value of  $M$  used for prediction could be equal to the total number of meta-stable patterns, since there is only a finite set of distinct patterns (i.e.  $M = 12$ ) for SA molecule. If the number of distinct patterns is very large (i.e. a few thousands), we can optionally reduce this number by screening the value of the general distance, which is inversely proportional to the contribution of each pattern to the molecular properties.

The number ( $n$ ) of possible samples in each meta-stable pattern is another critical parameter to be determined in the PEM algorithm. Two possible solutions can be considered.

- 1) All elements in each pattern are used, namely, “batch” option, for which all samples are used for predicting molecular properties. In this case, the number  $n$  may be very large, i.e. a few thousands for SA molecule (Table S1). This would strongly lower the efficiency of the PEM algorithm.
- 2) Only one element in each meta-stable pattern is used, namely, “stochastic” option, for which randomly select only one sample from each pattern, or use the average structure of each pattern. In this case, the





**Figure 6.** (a) The ratio distribution of the excited energy (ratio = PEM *v.s.* TDDFT). The blue curve is fitted with Gaussian function ( $\mu = 0.99$ ,  $\sigma = 0.08$ ). (b) The fluorescence emission spectra roughly estimated from PEM and TDDFT. (c) The time evolution of the excitation energy calculated at PEM and TDDFT level. The red line is the TDDFT result smoothed with Bézier curve, in order to remove the fast oscillations.

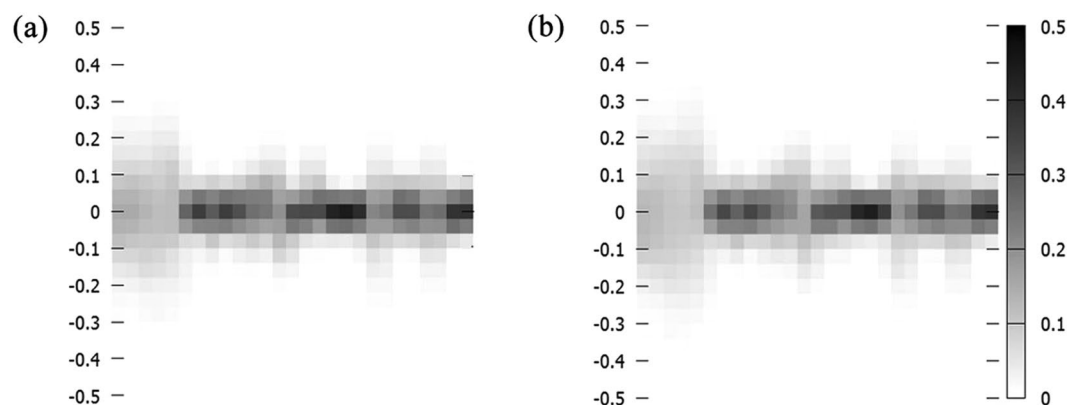
number  $n$  should be equal to 1. This option is much faster, however, the convergence to the optimal value is too oscillation and stochastic.

In order to overcome the defects of both options, we find a trade-off between the efficiency and reliable, for which a small set of the samples (i.e.  $n = 10$ ) in each pattern is used for PEM algorithm, namely “mini-batch” option. This option could reduce the total number of numerical calculations, and also reduce the stochastic behavior. Similar ideas have been commonly used in the realm of machine learning<sup>43, 67</sup>. The PEM algorithm typically scales as  $O(M \cdot n)$ , whereas  $M$  and  $n$  are very small constant value in the “mini-batch” option. Thus, this algorithm should be significantly faster than electronic structure approaches, such as TDDFT.

It is also very important to determine the uncertainty in the prediction, so one can evaluate the confidence of the results. In the following calculations, the conformations of validation sets (1000 geometries) were randomly sampled from dynamics trajectories. The “mini-batch” option is used with  $M = 12$  and  $n = 10$ . This means that all of meta-stable patterns (12 clusters) were used to estimate the molecule properties of any unknown geometry, meanwhile, we randomly select ten geometries ( $n = 10$ ) from each pattern to build parameter sets. Note that no training sets were required for the PEM algorithm beyond the clustering algorithm.

Figure 6a shows the performance of excitation energy prediction for  $S_1$  state. The distribution of the ratio between the PEM and TDDFT excitation energies is given. The distribution curve is fitted to a Gaussian distribution ( $\mu = 0.99$ ,  $\sigma = 0.08$ ), although the distribution is even more sharp than a Gaussian distribution. The derivation of PEM excitation energy from the TDDFT is within 0.1–0.2 eV (99% cases), which is acceptable even for electronic structure calculations. More sophisticated designing of molecular descriptors or clustering algorithm may improve the prediction. Then, the PEM algorithm is used to estimate the fluorescence emission spectroscopy. In order to take into account of dynamical effects (*vibronic*), the emission spectroscopy is calculated with 100 snapshots sampled from the last 5 ps of five excited state AIMD trajectories. Note also that in all the cases fluorescence was considered to happen only from the first excited singlet state following the so called Kasha’s rule<sup>68</sup>. Figure 6b shows the emission spectroscopy obtained from the PEM and TDDFT. The band shape of the emission spectroscopy at PEM and TDDFT level is very close.

Then, the PEM algorithm is used to predict the excitation energies along with a specific excited state AIMD trajectory, which are not used to construct the meta-stable patterns. Figure 6c shows the time evolution for  $S_1$  excitation energy from the PEM and TDDFT results for a  $\sim 5$  ps dynamics trajectory. It seems that the rough trends of time dependent excited energy can be reasonably predicted by the PEM method, however, the details of the excitation energy fails to be predicted. This is reasonable, because our PEM algorithm only includes a few dihedral angles as molecule descriptors. So, the contribution from the fast degree of freedoms is averaged out in this model. More consistent results can be anticipated if more sophisticated molecular descriptors are adopted. In general, excitation energy could be properly described by an ensemble model with a finite set of molecular conformations.



**Figure 7.** Partial charge deviation between PEM and DFT/TDDFT calculations for ground state (a) and excited state (b), respectively. X axis is the atom index for SA molecule.

Figure 7 shows the distribution of the partial charge deviation between the PEM and DFT/TDDFT calculations for the ground state and the excited state (S1). At first glance, the PEM partial charges are improved significantly, and the deviation from the DFT/TDDFT is often less than 0.05 e, in contrast to  $\sim 0.50$  e in Fig. 4. The symmetric fluctuation of the charge deviation around the zero value also suggests that the error cancellation effects may enhance the robustness of the PEM partial charges at long timescale dynamics simulation. Another interesting result of the PEM algorithm is that the excited state partial charge is predicted with similar accuracy as the ground state partial charge, considering the very similar distribution of the partial charge deviation in Fig. 7. This superiority allows us to update the ground and excited state partial charge at the same foot, although the electronic partial charge of the excited state is more complex in comparison with the ground state.

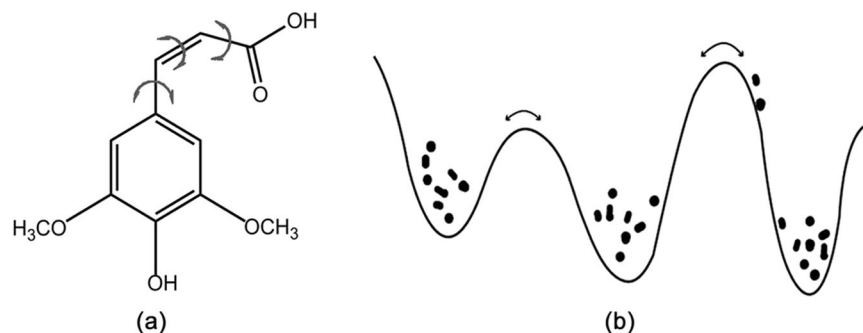
Beyond the symmetric fluctuation of the partial charge deviation, the strength of the charge fluctuation for each atom is not the same, for example, the description of the aromatic ring is not very good. This is expected, since the molecular descriptors mainly focus on a few dihedral angles, which only provide a rough representation of the slow degrees of freedom. Therefore, more molecular descriptors may be required to take into account the fast degrees of freedom in the future clustering algorithm. In contrast to the fixed point charge model, the PEM point charges provide a much more flexible description of the ground and excited state charge fluctuation (Figs 4 and 7). Although the electronic structure calculations are performed at the DFT/TDDFT level in this work, the PEM method can also be applied to any level of electronic structure methods. The use of the PEM method offers a significant improvement in the quality and applicability of electrostatically determined partial charges. In summary, the PEM method provides an alternative choice to directly incorporate the polarization or charge transfer effects on predicting partial charge population.

## Discussions

The PEM method includes a classifier to identify the possible meta-stable patterns as molecular features, and a predictor to construct the molecular properties for any unknown molecule structure. We have found that model ensembles are usually more accurate than any single model, and they are typically more fault-tolerant than single models. Therefore, the performance of PEM can be steadily improved by taking into more reference data or optimizing the kernel function. And more sophisticated and intelligent algorithms should also be helpful, i.e. replace the K-means clustering classifier by an auto-encoder or even an artificial neural network. However, such treatment would sacrifice the clear physical meaning of variables in this work, and thus the dedicated balance between the performance and human interpretability is of paramount interest<sup>43</sup>.

The simple analytical form of the PEM algorithm provides an alternative choice or procedure to take into account the large variation of the excitation state molecular properties, i.e. charge population and excitation energy. The good performance of the PEM method may give us some inspiration on the empirical force field development. Theoretically, the description of both molecular ground and excited states should be possible with empirical force fields, especially, when the potential is intended for an application toward adiabatic dynamics on a single surface. However, the “parameters” in the empirical potentials can be vastly different along the molecular conformation for the excited state. And more sophisticated empirical force field models or complex analytic functions may be required.

We suggest that the “parameters” in empirical force fields can be constructed on top of the PEM algorithm. Practically, the force field parameters can be assigned for each kind of meta-stable pattern. For instance, the widely used point charge model could be applied for both ground state and excited state, and the real dependencies of partial point charges on molecular conformation may be handled by combining several or even thousands of meta-stable patterns into a single, new point charge predicting model (Eq. 1). Therefore, the requirement to incorporate polarization or even charge transfer into the standard pair-wise potentials can be easily achieved in the PEM algorithm, which is compatible with traditional force fields. There are two practical advantages. 1) There is no need to construct a complex analytic model; this procedure is cumbersome and computationally inefficient for large molecules. 2) With use of the PEM algorithm, the computer program is very similar with the available empirical force field. This model is fast enough to allow millions of calculations along the dynamics propagation with adequate accuracy. The performance of this assumption would be reported in our continuing work.



**Figure 8.** (a) The schematic structure of cis-SA molecule and the important dihedral angles are labeled with arrows. (b) Possible distribution of the conformational states along a simple potential energy curve. The points represent the accessible conformational states during dynamics.

The main disadvantage of this approach is the requirements of fully searching the most possible meta-stable patterns. The good news is that the PEM algorithm could easily incorporate more meta-stable patterns of some specific conformation space in a fast iterated procedure, meanwhile, the performance of another part of conformation space would not be affected if the kernel function is properly defined. So, it is highly extensible and flexible. Thanks to the increasing computing power, modern dynamics simulations can easily generate data sets with millions of configurations from an ensemble of uncoupled dynamics trajectories. We also note that similar machine learning based approaches have also been used to accelerate the AIMD simulation of material systems<sup>48, 49</sup>, because such approach avoids the repeated *ab initio* calculations for the same molecule with similar conformations.

## Conclusions

Our results strongly suggest that ensemble models together with a proper classifier for model selection provides a useful research tool to gain insights from time series of *ab initio* dynamics, as demonstrated in the excited state studies of the SA molecule. Data mining of dynamics trajectories could gain a direct view of the possible meta-stable patterns and their relationship on the physical or biological processes of interest, with only a few commonsense rules. We further suggest that the “state-of-art” PEM method shows good performance in predicting ground and excited state molecular properties, in comparable to DFT/TDDFT calculations. The PEM could sufficiently characterize the feature of excited state motions, and naturally form knowledge based data sets. And the performance of PEM method could steadily be improved in a fast iterated procedure. The PEM method may provide us an alternative perspective to construct excited state force field with similar function form as the ground state one, without using much advance knowledge of the molecule details in the excited states. Its generality and ease of implementation should make it useful in various situations. Further work is going on to investigate the dynamic dependence of the inter-atomic potential itself and its realistic applications on the molecular dynamics simulation.

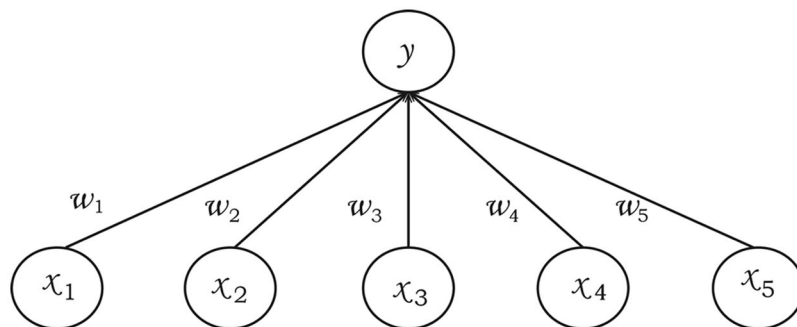
## Models and Methods

**Dynamics Simulation and Data Sets.** The molecular conformation data sets were mainly collected from our previous excited state AIMD simulations of synephrine (SA)<sup>62</sup>. The dynamics details can be found in the supporting information. Here, we only give a brief summary. The ultrafast excited-state dynamics was performed by on-the-fly surface hopping approach as implemented in JADE package<sup>69, 70</sup>. Non-adiabatic transitions between excited states were taken into account via Tully’s fewest switches approach<sup>26</sup>. The initial geometries and velocities of the excited state dynamics simulations were generated from the Wigner distribution function of the first vibrational level of the ground electronic state<sup>71</sup>. Starting from the initial sampling geometries, the molecule is electronically excited to the S<sub>1</sub> state for 100 trajectories. Each trajectory was calculated for 1000 fs. The time step for integration of classical equations was 0.5 fs and of quantum equations, 0.005 fs. The decoherence correction proposed by Granucci *et al.* was taken and the parameter is set to  $\alpha = 0.1$  Hartree<sup>72</sup>. For simplicity, the cis-SA molecule in gas phase is used as a benchmark system (Fig. 8), and we only focus on a single potential surface (S<sub>1</sub>). This is reasonable since the nonadiabatic decay to the ground state is not observed within the simulation time scale, which is very different from the dynamic behavior of the solvated SA molecule<sup>62, 73–75</sup>. Thus, the excited state dynamics of SA molecule in gas phase would stay on a single excited state surface (S<sub>1</sub>) in the limited simulation time. However, it should be noted that our following protocols are mainly based on geometric metrics, which can be easily applied to two or more coupled PES conditions.

The molecule spends most of the trajectory time dwelling in a free energy minimum, “waiting” for thermal fluctuations to push the system over a free energy barrier. Thus, it is extremely difficult to adequately sample the conformation space for complex molecules due to the limited timescales accessible for excited state AIMD simulations. Thus, an ensemble of uncoupled AIMD trajectories was used for our subsequent analysis. Totally, 200,000 snapshots from dynamics trajectories were obtained. To improve the sampling efficiency, we only sample the snapshots of local minima along the each trajectory (13226 structures).

The ground state calculations were performed at B3LYP/6–31 G(d,p) level, while, the excited state calculations were performed at TD-B3LYP/6–31 G(d,p) level. The ground and excited state molecular electrostatic potential





**Figure 9.** The schematic representation of the ensemble averaging, which can be viewed as an interpolation algorithm.

(ESP) is also calculated for subsequent analysis. The adopted DFT/TDDFT methods have been carefully calibrated with known experimental evidence in our previous work<sup>62</sup>, and can be used for thousands to millions energy and gradient calculations. Dynamics treatment with more accurate electronic-structure and advanced dynamical methods at all atoms level should represent the great challenge for future theoretical chemistry. All electronic structure calculations were performed with Gaussian 09 package<sup>76</sup>. The data sets and some scripts can be obtained upon request or downloaded from <https://github.com/dulikai/bidiu>.

**The Clustering Algorithms.** The dynamics of complex systems typically involves various meta-stable intermediates. It is necessary to decompose conformation space into a set of kinetically meta-stable states. Although the meta-stable state is not a static structure, which can be represented as an ensemble of structures with similar geometric features. This can be achieved by various clustering algorithms, such as K-means clustering, mean shift, hierarchical clustering, artificial neural network and etc. Here, the K-means, as a simple and robust algorithm, is used to classify the conformation space into a set of discrete states or clusters, which should be corresponding to basins of attraction of local minima on the PES.

The K-means clustering algorithm aims to partition  $n$  observations into  $k$  clusters, in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. The most critical parameter in the K-means clustering is the number ( $k$ ) of the clusters or the centroids. Thus, we have tested many cluster numbers (i.e.  $k = 4\sim 400$ ). And the selection of cluster number should be an iterative procedure. The geometric RMSD value of each cluster is a very suitable criteria (Table S2). The subsequent analysis also verified that 12 clusters are sufficient and efficient. The molecular internal coordinates are used as molecular descriptors for clustering, especially, a few critical dihedral angles<sup>62,77</sup> are chosen to characterize the excited state motion of the SA molecule (Fig. 8). The protocol is implemented with scikit-learn module<sup>78</sup> in Python.

**The Prediction with Ensemble Models.** The clustering algorithm ensures that similar conformations are grouped into the same meta-stable states or clusters, in a methodical and unbiased fashion. Thus, the dynamics process could be represented by only a limited number of meta-stable patterns. The conformations in the same cluster should have much similar properties (low bias) than among different clusters (high variance). And we can estimated the transition probabilities between meta-stable states by counting the number of transitions along the time series of the trajectories, and thus form a kinetic meta-stable states network, namely time series fusion (TSF) network.

Based on the concept of ensemble averaging<sup>79–81</sup>, we suggest an interpolation scheme, namely Prediction with Ensemble Models (PEM) algorithm, to build a model and predict reliable molecular properties, such as charge population and excitation energy. In the machine learning realm, ensemble averaging is one of the simplest types of committee machines, which provides an ideal technique to combine multiple models or patterns to produce a desired output. Usually, an ensemble of models performs better than any individual model, because the various errors of the models can be averaged out<sup>79</sup>. In summary, ensemble averaging creates a group of networks, each with low bias and high variance, and combines them to a new network with low bias and low variance. It is thus a possible solution of the bias-variance dilemma<sup>79,81</sup>.

The theory of ensemble averaging relies on two properties of artificial neural networks: 1) In any network, the bias can be reduced at the cost of increased variance. 2) In a group of networks, the variance can be reduced at no cost to bias. In this work, a general version of ensemble averaging is defined as a weighted sum of finite number of clusters, with low bias and high variance (Fig. 9). If the molecule descriptors ( $\mathbf{X}$ ) are specified, the estimated result  $V$  can be defined as

$$V(\mathbf{X}) = \sum_{i=1}^M \sum_{j=1}^n \omega_{ij}(\mathbf{X}) \mathbf{T}_{ij}(\mathbf{X}) \quad (1)$$

In the above equation,  $\mathbf{T}_{ij}$  is the property of one element ( $j$ ) in each cluster ( $i$ ), and the distinct clusters are generated by the K-means algorithm from the time series of trajectories. And  $M$  is the number of clusters,  $n$  is the number of possible elements in each cluster, and  $\omega$  is a set of weights. The same sets of geometric based molecular

descriptors in the clustering algorithm are adopted in the ensemble averaging model. The performance of the prediction model with such simple molecular descriptors seems to be sufficient in the scenario of “big data sets”.

Here, we use the kernel function in kernel-based ML methods<sup>29</sup> to directly obtain the weights ( $\omega$ ), which tends to be somewhat easier to set up in practice than the artificial neural networks. The kernel function should have the following features: (1) The formula should be continuous in the input space of molecular descriptors, so any small perturbation of the system does not change the results too much. (2) The  $\mathbf{X}$  far from a specific cluster should have less weight, because such cluster has little similarity on the conformation space.

Thus, the weight/kernel function is defined as a function of general distances between an arbitrary geometry and a set of known geometries.

$$\omega_{ij} = \frac{1/u_{ij}}{\sum_{i=1}^N (1/u_{ij})^4} \quad (2)$$

The general distances is defined as

$$u_{ij}(\mathbf{X}) = \sqrt{\sum_{k=1}^p (\mathbf{X} - \mathbf{X}_k')^2} \quad (3)$$

In Eq. (3),  $p$  is the number of molecular descriptors (internal coordinates), and  $\mathbf{X}_k'$  is known values of molecular descriptors. The above equation is very similar with well-known PES interpolation algorithm, i.e. the Shepard interpolation<sup>82,83</sup>. This is not surprising because much of ML algorithms are just the interpolation between data points, at its core. It should note that the optimization problem of finding the weight  $\omega$  can also be solved through the training of neural networks, if one does not care about the drawback of their interpretability as a data exploration tool<sup>84,85</sup>.

## References

- Stolow, A., Bragg, A. E. & Neumark, D. M. Femtosecond time-resolved photoelectron spectroscopy. *Chem. Rev.* **104**, 1719–1757 (2004).
- Miller, R. J. Femtosecond crystallography with ultrabright electrons and x-rays: capturing chemistry in action. *Science* **343**, 1108–1116 (2014).
- Wang, S. & Wang, X. Multifunctional Metal–Organic Frameworks for Photocatalysis. *Small* **11**, 3097–3112 (2015).
- Liu, Y. H., Lan, S. C., Zhu, C. & Lin, S. H. Intersystem Crossing Pathway in Quinoline–Pyrazole Isomerism: A Time-Dependent Density Functional Theory Study on Excited-State Intramolecular Proton Transfer. *J. Phys. Chem. A* **119**, 6269 (2015).
- Linic, S., Aslam, U., Boerigter, C. & Morabito, M. Photochemical transformations on plasmonic metal nanoparticles. *Nat. Mater.* **14**, 567–576 (2015).
- Padalkar, V. S. & Seki, S. Excited-state intramolecular proton-transfer (ESIPT)-inspired solid state emitters. *Chem. Soc. Rev.* **45**, 169–202 (2015).
- Ruban, A. V. Nonphotochemical Chlorophyll Fluorescence Quenching: Mechanism and Effectiveness in Protecting Plants from Photodamage. *Plant Physiology* **170**, 1903–1916 (2016).
- Chen, M., Zhong, M. & Johnson, J. A. Light-controlled radical polymerization: Mechanisms, methods, and applications. *Chem. Rev.* **116**, 10167–10211 (2016).
- Jara-Cortés, J., Guevara-Vela, J. M., Martín Pendás, Á. & Hernández-Trujillo, J. Chemical bonding in excited states: Energy transfer and charge redistribution from a real space perspective. *J. Comput. Chem.* **38**, 957–970 (2017).
- Zhao, G.-J. & Han, K.-L. Hydrogen Bonding in the Electronic Excited State. *Acc. Chem. Res.* **45**, 404–413 (2012).
- Pisana, S. *et al.* Breakdown of the adiabatic Born–Oppenheimer approximation in graphene. *Nat. Mater.* **6**, 198–201 (2007).
- Middleton, C. T. *et al.* DNA excited-state dynamics: from single bases to the double helix. *Annu. Rev. Phys. Chem.* **60**, 217–239 (2009).
- González, L., Escudero, D. & Serrano-Andrés, L. Progress and challenges in the calculation of electronic excited states. *ChemPhysChem* **13**, 28–51 (2012).
- Akimov, A. V., Neukirch, A. J. & Prezhdo, O. V. Theoretical insights into photoinduced charge transfer and catalysis at oxide interfaces. *Chem. Rev.* **113**, 4496–4565 (2013).
- Zhang, D. H. & Guo, H. Recent advances in quantum dynamics of bimolecular reactions. *Annu. Rev. Phys. Chem.* **67**, 135–158 (2016).
- Ammal, S. C., Yamataka, H., Aida, M. & Dupuis, M. Dynamics-driven reaction pathway in an intramolecular rearrangement. *Science* **299**, 1555–1557 (2003).
- Lourderaj, U., Park, K. & Hase, W. L. Classical trajectory simulations of post-transition state dynamics. *Int. Rev. Phys. Chem.* **27**, 361–403 (2008).
- Sun, L., Song, K. & Hase, W. L. A S(N)2 reaction that avoids its deep potential energy minimum. *Science* **296**, 875–878 (2002).
- Bennun, M. Nonadiabatic molecular dynamics: Validation of the multiple spawning method for a multidimensional problem. *J. Chem. Phys.* **108**, 7244–7257 (1998).
- Bennun, M., Jason Quenneville, A. & Martínez, T. J. Ab Initio Multiple Spawning: Photochemistry from First Principles Quantum Molecular Dynamics. *J. Phys. Chem. A* **104**, 5161–5175 (2000).
- Li, X., Tully, J. C., Schlegel, H. B. & Frisch, M. J. Ab initio Ehrenfest dynamics. *J. Chem. Phys.* **123**, 084106 (2005).
- Bittner, E. R. & Rossky, P. J. Quantum decoherence in mixed quantum-classical systems: Nonadiabatic processes. *J. Chem. Phys.* **103**, 8130–8143 (1995).
- Zhu, C., Jasper, A. W. & Truhlar, D. G. Non-Born–Oppenheimer trajectories with self-consistent decay of mixing. *J. Chem. Phys.* **120**, 5543–5557 (2004).
- Yarkony, D. R. In *Conical Intersections: Theory, Computation and Experiment* (ed David R Yarkony Wolfgang Domcke, Horst Köppel) 4–128 (World Scientific, 2004).
- Miller, W. H. & George, T. F. Semiclassical Theory of Electronic Transitions in Low Energy Atomic and Molecular Collisions Involving Several Nuclear Degrees of Freedom. *J. Chem. Phys.* **56**, 5637–5652 (1972).
- Tully, J. C. Molecular dynamics with electronic transitions. *J. Chem. Phys.* **93**, 1061–1071 (1990).
- Craig, C. F., Duncan, W. R. & Prezhdo, O. V. Trajectory Surface Hopping in the Time-Dependent Kohn–Sham Approach for Electron–Nuclear Dynamics. *Phys. Rev. Lett.* **95**, 163001 (2005).
- Virshup, A. M., Chen, J. & Martínez, T. J. Nonlinear dimensionality reduction for nonadiabatic dynamics: The influence of conical intersection topography on population transfer rates. *J. Chem. Phys.* **137**, 22A519 (2012).

29. Rupp, M. Machine learning for quantum mechanics in a nutshell. *Int. J. Quantum Chem.* **115**, 1058–1073 (2015).
30. Hase, F., Valleau, S., Pyzer-Knapp, E. & Aspuru-Guzik, A. Machine learning exciton dynamics. *Chem. Sci.* **7**, 5139–5147 (2016).
31. Sánchez-Lengeling, B. & Aspuru-Guzik, A. Learning More, with Less. *ACS Central Science* **3**, 275–277 (2017).
32. Singer, A., Erban, R., Kevrekidis, I. G. & Coifman, R. R. Detecting intrinsic slow variables in stochastic dynamical systems by anisotropic diffusion maps. *Proc. Natl. Acad. Sci. USA* **106**, 16090–16095 (2009).
33. Shaw, D. E. *et al.* Atomic-level characterization of the structural dynamics of proteins. *Science* **330**, 341–346 (2010).
34. Maragakis, P., van der Vaart, A. & Karplus, M. Gaussian-Mixture Umbrella Sampling. *J. Phys. Chem. B* **113**, 4664–4673 (2009).
35. Tenenbaum, J. B., De Silva, V. & Langford, J. C. A global geometric framework for nonlinear dimensionality reduction. *science* **290**, 2319–2323 (2000).
36. Sirur, A., De, S. D. & Best, R. B. Markov state models of protein misfolding. *J. Chem. Phys.* **144**, 075101 (2016).
37. Harrigan, M. P. *et al.* MSMBuild: Statistical Models for Biomolecular Dynamics. *Biophys. J.* **112**, 10–15 (2017).
38. Snyder, J. C. *et al.* Orbital-free bond breaking via machine learning. *J. Chem. Phys.* **139**, 224104 (2013).
39. Bartók, A. P., Payne, M. C., Kondor, R. & Csányi, G. Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.* **104**, 136403 (2010).
40. Rupp, M., Tkatchenko, A., Müller, K. R. & von Lilienfeld, O. A. Fast and accurate modeling of molecular atomization energies with machine learning. *Phys. Rev. Lett.* **108**, 058301 (2012).
41. Pozun, Z. D. *et al.* Optimizing transition states via kernel-based machine learning. *J. Chem. Phys.* **136**, 174101 (2012).
42. Handley, C. M. & Popelier, P. L. Potential energy surfaces fitted by artificial neural networks. *J. Phys. Chem. A* **114**, 3371–3383 (2010).
43. Goh, G. B., Hodas, N. O. & Vishnu, A. Deep learning for computational chemistry. *J. Comput. Chem.* **38**, 1291–1307 (2017).
44. Hansen, K. *et al.* Assessment and validation of machine learning methods for predicting molecular atomization energies. *J. Chem. Theory Comput.* **9**, 3404–3419 (2013).
45. Ramakrishnan, R., Dral, P. O., Rupp, M. & von Lilienfeld, O. A. Big data meets quantum chemistry approximations: the  $\Delta$ -machine learning approach. *J. Chem. Theory Comput.* **11**, 2087–2096 (2015).
46. Behler, J. Neural network potential-energy surfaces in chemistry: a tool for large-scale simulations. *Phys. Chem. Chem. Phys.* **13**, 17930–17955 (2011).
47. Shen, L., Wu, J. & Yang, W. Multiscale Quantum Mechanics/Molecular Mechanics Simulations with Neural Networks. *J. Chem. Theory Comput.* **12**, 4934–4946 (2016).
48. Botu, V. & Ramprasad, R. Adaptive machine learning framework to accelerate ab initio molecular dynamics. *Int. J. Quantum Chem.* **115**, 1074–1083 (2015).
49. Li, Z., Kermode, J. R. & De Vita, A. Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Phys. Rev. Lett.* **114**, 096405 (2015).
50. Cornell, W. D. *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).
51. Van Duin, A. C., Dasgupta, S., Lorant, F. & Goddard, W. A. ReaxFF: a reactive force field for hydrocarbons. *J. Phys. Chem. A* **105**, 9396–9409 (2001).
52. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & De Vries, A. H. The MARTINI force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B* **111**, 7812–7824 (2007).
53. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).
54. Xie, W., Orozco, M., Truhlar, D. G. & Gao, J. X-Pol Potential: An Electronic Structure-Based Force Field for Molecular Dynamics Simulation of a Solvated Protein in Water. *J. Chem. Theory Comput.* **5**, 459–467 (2009).
55. Bissantz, C., Kuhn, B. & Stahl, M. A medicinal chemist's guide to molecular interactions. *J. Med. Chem.* **53**, 5061–5084 (2010).
56. Dill, K. A. & MacCallum, J. L. The protein-folding problem, 50 years on. *Science* **338**, 1042–1046 (2012).
57. Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983).
58. Song, C. I. & Rhee, Y. M. Development of force field parameters for oxyluciferin on its electronic ground and excited states. *Int. J. Quantum Chem.* **111**, 4091–4105 (2011).
59. Ando, K. Excited state potentials and ligand force field of a blue copper protein plastocyanin. *J. Phys. Chem. B* **108**, 3940–3946 (2004).
60. Park, J. W. & Rhee, Y. M. Interpolated Mechanics–Molecular Mechanics Study of Internal Rotation Dynamics of the Chromophore Unit in Blue Fluorescent Protein and Its Variants. *J. Phys. Chem. B* **116**, 11137–11147 (2012).
61. Su, J. T. & Goddard, W. A. III Excited electron dynamics modeling of warm dense matter. *Phys. Rev. Lett.* **99**, 185003 (2007).
62. Liu, F., Du, L., Lan, Z. & Gao, J. Hydrogen bond dynamics governs the effective photoprotection mechanism of plant phenolic sunscreens. *Photochem. Photobiol. Sci.* **16**, 211–219 (2017).
63. Voter, A. F., Montalenti, F. & Germann, T. C. Extending the time scale in atomistic simulation of materials. *Annu. Rev. Mater. Res.* **32**, 321–346 (2002).
64. Shirts, M. R. & Pande, V. S. Mathematical analysis of coupled parallel simulations. *Phys. Rev. Lett.* **86**, 4983 (2001).
65. Joshi, K. L., Raman, S. & van Duin, A. C. Connectivity-based parallel replica dynamics for chemically reactive systems: from femtoseconds to microseconds. *J. Phys. Chem. Lett.* **4**, 3792–3797 (2013).
66. Perez, D., Uberuaga, B. P. & Voter, A. F. The parallel replica dynamics method – Coming of age. *Comput. Mater. Sci.* **100**, Part B, 90–103 (2015).
67. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
68. Kasha, M. Characterization of electronic transitions in complex molecules. *Discuss. Faraday Soc.* **9**, 14–19 (1950).
69. Du, L. & Lan, Z. An on-the-fly surface-hopping program jade for nonadiabatic molecular dynamics of polyatomic systems: implementation and applications. *J. Chem. Theory Comput.* **11**, 1360–1374 (2015).
70. Du, L. & Lan, Z. Correction to An On-the-Fly Surface-Hopping Program JADE for Nonadiabatic Molecular Dynamics of Polyatomic Systems: Implementation and Applications. *J. Chem. Theory Comput.* **11**, 4522–4523 (2015).
71. Barbatti, M., Aquino, A. J. & Lischka, H. The UV absorption of nucleobases: semi-classical ab initio spectra simulations. *Phys. Chem. Chem. Phys.* **12**, 4959–4967 (2010).
72. Granucci, G. & Persico, M. Critical appraisal of the fewest switches algorithm for surface hopping. *J. Chem. Phys.* **126**, 134114 (2007).
73. Baker, L. A. *et al.* Ultrafast photoprotecting sunscreens in natural plants. *J. Phys. Chem. Lett.* **7**, 56–61 (2015).
74. Tan, E. M., Hilbers, M. & Buma, W. J. Excited-state dynamics of isolated and microsolvated cinnamate-based UV-B sunscreens. *J. Phys. Chem. Lett.* **5**, 2464–2468 (2014).
75. Stavros, V. G. Photochemistry: A bright future for sunscreens. *Nat. Chem.* **6**, 955–956 (2014).
76. Gaussian 09 (Gaussian, Inc., Wallingford, CT, USA, 2009).
77. Chang, X.-P., Li, C.-X., Xie, B.-B. & Cui, G. Photoprotection mechanism of p-methoxy methylcinnamate: a CASPT2 study. *J. Phys. Chem. A* **119**, 11488–11497 (2015).
78. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
79. Geman, S., Bienenstock, E. & Doursat, R. Neural networks and the bias/variance dilemma. *Neural Comput.* **4**, 1–58 (1992).
80. Naftaly, U., Intrator, N. & Horn, D. Optimal ensemble averaging of neural networks. *Network-Comp Neural* **8**, 283–296 (1997).
81. Clemen, R. T. Combining forecasts: A review and annotated bibliography. *International journal of forecasting* **5**, 559–583 (1989).

82. Higashi, M. & Truhlar, D. G. Electrostatically Embedded Multiconfiguration Molecular Mechanics Based on the Combined Density Functional and Molecular Mechanical Method. *J. Chem. Theory Comput.* **4**, 790–803 (2008).
83. Collins, M. A. Molecular potential-energy surfaces for chemical reaction dynamics. *Theor. Chem. Acc.* **108**, 313–324 (2002).
84. Hashem, S. Optimal linear combinations of neural networks. *Neural Netw.* **10**, 599–614 (1997).
85. Liu, Y. & Yao, X. Ensemble learning via negative correlation. *Neural Netw.* **12**, 1399–1404 (1999).

### Acknowledgements

The work is supported by National Natural Science Foundation of China (Nos 21503249, 21373124), and Huazhong Agricultural University Scientific & Technological Self-innovation Foundation (Program No.2015RC008), and Project 2662016QD011 and 2662015PY113 Supported by the Fundamental Funds for the Central Universities. The authors also thank the support of Special Program for Applied Research on Super Computation of the NSFC-Guangdong Joint Fund (the second phase) under Grant No.U1501501.

### Author Contributions

Fang Liu and Likai Du designed and performed the experiments. Fang Liu, Likai Du, Dongju Zhang and Jun Gao analyzed data. Fang Liu, Likai Du and Jun Gao prepared the manuscript and figures.

### Additional Information

**Supplementary information** accompanies this paper at doi:[10.1038/s41598-017-09347-2](https://doi.org/10.1038/s41598-017-09347-2)

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017