Constructing polyatomic potential energy surfaces by interpolating diabatic Hamiltonian matrices with demonstration on green fluorescent protein chromophore

Jae Woo Park and Young Min Rhee

Citation: The Journal of Chemical Physics **140**, 164112 (2014); doi: 10.1063/1.4872155
View online: http://dx.doi.org/10.1063/1.4872155
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/140/16?ver=pdfcov
Published by the AIP Publishing

Constructing polyatomic potential energy surfaces by interpolating diabatic Hamiltonian matrices with demonstration on green fluorescent protein chromophore

Jae Woo Park and Young Min Rhee[a)]
Center for Self-assembly and Complexity, Institute for Basic Science (IBS), Pohang 790-784, Korea
Department of Chemistry, Pohang University of Science and Technology (POSTECH), Pohang 790-784, Korea

(Received 13 January 2014; accepted 10 April 2014; published online 28 April 2014)

Simulating molecular dynamics directly on quantum chemically obtained potential energy surfaces is generally time consuming. The cost becomes overwhelming especially when excited state dynamics is aimed with multiple electronic states. The interpolated potential has been suggested as a remedy for the cost issue in various simulation settings ranging from fast gas phase reactions of small molecules to relatively slow condensed phase dynamics with complex surrounding. Here, we present a scheme for interpolating multiple electronic surfaces of a relatively large molecule, with an intention of applying it to studying nonadiabatic behaviors. The scheme starts with adiabatic potential information and its diabatic transformation, both of which can be readily obtained, in principle, with quantum chemical calculations. The adiabatic energies and their derivatives on each interpolation center are combined with the derivative coupling vectors to generate the corresponding diabatic Hamiltonian and its derivatives, and they are subsequently adopted in producing a globally defined diabatic Hamiltonian function. As a demonstration, we employ the scheme to build an interpolated Hamiltonian of a relatively large chromophore, *para*-hydroxybenzylidene imidazolinone, in reference to its all-atom analytical surface model. We show that the interpolation is indeed reliable enough to reproduce important features of the reference surface model, such as its adiabatic energies and derivative couplings. In addition, nonadiabatic surface hopping simulations with interpolation yield population transfer dynamics that is well in accord with the result generated with the reference analytic surface. With these, we conclude by suggesting that the interpolation of diabatic Hamiltonians will be applicable for studying nonadiabatic behaviors of sizeable molecules.
 [http://dx.doi.org/10.1063/1.4872155]

I. INTRODUCTION

Potential energy surfaces (PESs) are essential for performing various molecular simulations. Because elucidating nonadiabatic processes that govern the dynamics of excited state species require reliable descriptions of multiple potential surfaces and their couplings, constructing accurate PESs is even more important in studying properties of excited state species produced by light absorption.[1–8] In this situation, fitting analytical models to reference quantum chemical calculations becomes enormously difficult compared to the cases with single electronic states. Thus, PES construction is often completely avoided by directly adopting quantum chemical calculations. This approach, however, normally demands very high computational power. In many cases, if not all, the formidable computational cost forces us to utilize small numbers of trajectories with limited statistical meaning, or to employ low levels of quantum chemical theories in simulations. Although numerous studies have accumulated essential knowledge regarding many nonadiabatic processes within these limitations,[8–20] developing a method for explicitly constructing accurate PESs will still be highly valuable.

In fact, this computational expense problem is an issue in obtaining not only multiple state PESs but also a single state surface to some degree. A "model-less model" using an interpolated PES has been developed to relieve the problem.[13,21–26] This approach was first pioneered by Collins and co-workers,[21] and was initially used for studying gas phase reaction dynamics of small molecular systems.[22,23,27,28] Recently, investigations on condensed phase adiabatic excited state dynamics with interpolated PESs have also been reported.[13,24] These condensed phase studies involved relatively long durations of simulations with comparatively large ensemble sizes. Considering that they involved dynamics simulations on single surfaces, a natural extension will be to account for the nonadiabatic phenomena by engaging multiple surfaces along the trajectories. By incorporating this feature, the applicability of interpolation technique will be greatly expanded especially toward studying biological systems.[13]

For this purpose, nonadiabatic coupling matrix elements near degeneracy should be computed.[1–3,7,29] Also, for performing the interpolation, a smooth function without singularity is needed. This requirement makes the diabatic potential energy matrix a good candidate for the interpolation, over the conventional adiabatic potential energies and derivative coupling vectors. Namely, because the diabatic potential

a)E-mail: ymrhee@postech.ac.kr

energy matrix is smooth near conical intersections, its interpolation will be mathematically more tractable. Although a perfect diabatization is only possible for small systems,[30] various quasi-diabatization methods can successfully generate reliable states without singularity in state-to-state derivative couplings.[30–36]

In fact, there have already been developments in methods for interpolating diabatic Hamiltonians.[17,28,29] In these works, the adiabatic potential energies are diabatized with the requirement that interpolated diabatic potential energy matrix should be smooth. The interpolation of diabatic Hamiltonian is conducted with the adiabatic-to-diabatic transformation (ADT) matrix obtained by this internally imposed diabatization scheme. This method was reported to yield very accurate derivative couplings and electronic state occupation probabilities.[17,29] However, how it will perform with large systems has not been investigated yet, especially with the internal diabatization. Indeed, imposing the continuity condition might require an excessively large number of data points toward interpolation for large systems. On the other hand, many diabatization processes have been proposed for elucidating excited states of large molecules. The diabatic states obtained by these external schemes have given useful information on the chemical and physical properties of many molecules.[30–36] For investigating the nonadiabatic dynamics of large systems, therefore, developing a scheme for interpolating diabatic Hamiltonians through such external diabatization procedures will be useful.

In this article, we present a method for interpolating diabatic Hamiltonian with external diabatization, namely, under the situation where adiabatic-to-diabatic transformation matrix is already given. We will evaluate the reliability of the diabatic Hamiltonian interpolation by applying it to previously defined analytic surface[35] of the green fluorescent protein (GFP) chromophore, *para*-hydroxybenzylidene imidazolinone (*p*HBI). By adopting an analytic model, the convergence of the potential energy matrix can be readily verified. We will show that the present interpolation scheme is capable of describing both the energy landscape and the derivative couplings in large systems, with more than 20 atoms. We will also show nonadiabatic molecular dynamics (MD) simulation results of the gas phase GFP chromophore to demonstrate the reliability of the interpolated diabatic Hamiltonian for studying dynamics. Finally, we will discuss the directions to which we should improve this scheme in future studies when we attempt to investigate the photodynamics of complex systems.

II. THEORY

We are trying to interpolate the diabatic potential energy matrix **D**. In the Shepard interpolation of a potential surface, the energy at an arbitrary molecular geometry **X** is normally formulated as

$$V(\mathbf{X}) = \sum_n w(n)\, V(\mathbf{X};\mathbf{X}(n)). \tag{1}$$

In this expression, $V(\mathbf{X};\mathbf{X}(n))$ is the Taylor expansion of the potential energy from the nth geometrical data point $\mathbf{X}(n)$, and $w(n)$ is the interpolation weighting function. In a similar manner, the interpolation of a matrix **D** can be formulated as

$$\mathbf{D}(\mathbf{X}) = \sum_n w(n)\, \mathbf{D}(\mathbf{X};\mathbf{X}(n)). \tag{2}$$

Then, the subsequent tasks for interpolating **D** will be preparing the Taylor expansions $\mathbf{D}(\mathbf{X};\mathbf{X}(n))$ at an arbitrary point **X** from a set of data points $\{\mathbf{X}(n)\}$, and determining the weights $\{w(n)\}$ for the data points. We will form the Taylor series by directly employing derivatives of **D**. For $w(n)$, we will adopt weighting functions based on Cartesian Euclidean distances. These aspects will be explained in detail in this section, together with other practical issues that need to be handled for interpolation. Hereafter, when it is not ambiguous, $\mathbf{X}(n)$ will be denoted as n for the sake of simplicity.

A. Taylor expansion

The Hamiltonian on the diabatic state basis, namely, the diabatic potential energy matrix **D**, is a unitary transformation of an adiabatic and diagonal potential matrix **V**. With the adiabatic-to-diabatic transformation matrix **R**, it can be expressed as

$$\mathbf{D} = \mathbf{R}^{\mathrm{T}}\mathbf{V}\mathbf{R}. \tag{3}$$

To perform the Taylor expansions, we need to obtain the derivatives of **R** and **V**. From the natural property of the diabatization,[29,31,33] we can write the gradient of **R** as

$$\nabla\mathbf{R} = -\mathbf{F}\mathbf{R}, \tag{4}$$

where the nonadiabatic coupling matrix **F** is composed of vector elements. With this relationship, the first and the second derivatives of $\mathbf{D}(n)$ are obtained by differentiating the elements of diabatic Hamiltonian as

$$\frac{\partial \mathbf{D}(n)}{\partial X_a} = \mathbf{R}^{\mathrm{T}}(n)\left[\frac{\partial \mathbf{V}(n)}{\partial X_a} + \mathbf{F}_a(n)\mathbf{V}(n) - \mathbf{V}(n)\mathbf{F}_a(n)\right]\mathbf{R}(n), \tag{5}$$

$$\begin{aligned}\frac{\partial \mathbf{D}(n)}{\partial X_b \partial X_a} = \frac{\partial \mathbf{D}(n)}{\partial X_a \partial X_b} = \mathbf{R}^{\mathrm{T}}(n)\Bigg[&\frac{1}{2}\frac{\partial^2 \mathbf{V}(n)}{\partial X_b \partial X_a} + \frac{1}{2}\frac{\partial^2 \mathbf{V}(n)}{\partial X_a \partial X_b} + \mathbf{F}_a(n)\frac{\partial \mathbf{V}(n)}{\partial X_b} + \mathbf{F}_b(n)\frac{\partial \mathbf{V}(n)}{\partial X_a} \\ &- \frac{\partial \mathbf{V}(n)}{\partial X_b}\mathbf{F}_a - \frac{\partial \mathbf{V}(n)}{\partial X_a}\mathbf{F}_b - \mathbf{F}_a(n)\mathbf{V}(n)\mathbf{F}_b(n) - \mathbf{F}_b(n)\mathbf{V}(n)\mathbf{F}_a(n) \\ &+ \frac{1}{2}\mathbf{V}(n)\left(\mathbf{F}_a(n)\mathbf{F}_b(n) + \mathbf{F}_b(n)\mathbf{F}_a(n) - \frac{\partial \mathbf{F}_a(n)}{\partial X_b} - \frac{\partial \mathbf{F}_b(n)}{\partial X_a}\right) \\ &+ \frac{1}{2}\left(\mathbf{F}_b(n)\mathbf{F}_a(n) + \mathbf{F}_a(n)\mathbf{F}_b(n) + \frac{\partial \mathbf{F}_a(n)}{\partial X_b} + \frac{\partial \mathbf{F}_b(n)}{\partial X_a}\right)\mathbf{V}(n)\Bigg]\mathbf{R}(n).\end{aligned} \tag{6}$$

Here, for the second derivative, we have employed its symmetrized form. For further details on derivations of these expansions, one should refer to the previous work by Collins and co-workers.[29]

When the use of any arbitrary coordinate system $\mathbf{Q}$ is desired, the Cartesian derivatives are transformed as

$$\frac{\partial \mathbf{D}(\mathbf{Q}(n))}{\partial Q_i} = \sum_a J_{ia} \frac{\partial \mathbf{D}(\mathbf{X}(n))}{\partial X_a}, \tag{7a}$$

$$\frac{\partial^2 \mathbf{D}(\mathbf{Q}(n))}{\partial Q_j \partial Q_i} = \sum_{ba} J_{jb} J_{ia} \frac{\partial^2 \mathbf{D}(\mathbf{X}(n))}{\partial X_b \partial X_a} + \sum_a \frac{\partial J_{ja}}{\partial Q_j} \frac{\partial \mathbf{D}(\mathbf{X}(n))}{\partial X_a} \tag{7b}$$

with the Jacobian factor $J_{ia} = \partial X_a / \partial Q_i$.[24,27] Using these derivatives in $\mathbf{Q}$ coordinates, the diabatic expansion at an arbitrary configuration $\mathbf{Q}$ from a data point $\mathbf{Q}(n)$ is

$$\mathbf{D}(\mathbf{Q};\mathbf{Q}(n)) = \mathbf{D}(\mathbf{Q}(n)) + \sum_i \Delta_i^Q \frac{\partial \mathbf{D}(\mathbf{Q}(n))}{\partial Q_i} + \sum_{ij} \Delta_i^Q \Delta_j^Q \frac{\partial^2 \mathbf{D}(\mathbf{Q}(n))}{\partial Q_j \partial Q_i} \tag{8}$$

with $\mathbf{\Delta}^Q = \mathbf{Q} - \mathbf{Q}(n)$. Any coordinate system can be used for Taylor expansions, including curvilinear internal coordinates $\mathbf{Z}$ and the Cartesian coordinates $\mathbf{X}$, as long as its Jacobian factors are available.

B. Weighting function

As explained earlier, the interpolated $\mathbf{D}$ at an arbitrary geometry $\mathbf{X}$ involves weighting function $w(n)$. As in the Taylor expansion case, there are two possible coordinate systems for this function: internal coordinates[21,27,37] and the Cartesian coordinates.[23,38] When we combine these two possibilities with another two possibilities in the coordinate system choice toward the Taylor expansion, we have in total four possibilities. In this work, we will utilize a Cartesian weighting function in combination with the internal coordinate based Taylor expansions. Although the advantages of employing this scheme have already been discussed quite some time ago,[23] for completeness, we will elaborate them again with comparisons to other alternative approaches for computing the distances.

When the internal coordinates are to be employed for the weight $w(n)$, a combination of various geometrical parameters, such as bond lengths, bending angles, and/or dihedral angles need to be considered. Of the three types, dihedral angles are often required as bond lengths and bond angles do not possess any chiral sense. However, a dihedral can at the same time suffer from singularity in measuring the distance. For example, when we imagine a system composed by four atoms as A-B-C-D, we can define a dihedral angle ϕ_{ABCD}. Now, suppose that the A-B-C angle is close to 180°. In this situation, when one rolls the atom A around the B-C axis from $\phi_{\mathrm{ABCD}} = 0$ to 360°, the distance this atom travels will actually be nearly zero. However, the angle ϕ_{ABCD} still travels from 0° and 360°. This implies a singularity in determining the distances. Even though the singularity itself may be avoided by coupling the dihedral with other metrics (e.g., $\sin(\angle\mathrm{ABC}) \cdot \phi_{\mathrm{ABCD}}$), such a variation tends to add another level of complication to the distance measurements. Furthermore, in the condensed phase molecular dynamics simulations, where high energy fluctuations occur quite often, near-collinearities can take place rather frequently.

This aspect has led us to adopt Cartesians for weighting. Because the Cartesian distance between two conformations can be affected by the molecular rotations in space, a way to implement uniqueness to the weighting function should be devised. For this purpose, the sum-over-the-orbit (SOO) procedure[38] or a simple alignment strategy[23] can be adopted. The SOO approach takes all the possible three-dimensional rotations into account and computes all continuously varying distances toward the weighted average of the Taylor expansion. This process is conventionally matched with Cartesian Taylor expansion for the molecular potential and is mathematically well-defined to generate a smooth potential surface in a rigorous manner. However, averaging over the rotated poses of a molecule may sometimes yield unphysical molecular forces. Let us take an example of an ethylene molecule depicted in Fig. 1(a). When its molecular force is described by SOO with an interpolation data point, whose conformation is exactly the same except a switch between two hydrogen atoms as shown in the figure, one can easily show that the Cartesian distance is invariant to the angle ϕ when the C-C axes of the two conformations overlap with each other. When there is pure C-H stretching (or contraction) force within the interpolation data point, the SOO procedure over the angle ϕ will render this force into an average over a cone, as pictorially represented in Fig. 1(b). Thus, the stretching force becomes contaminated by fictitious bending force after the weighted average.

Such an unphysical distortion can be avoided by measuring the distance with an alignment of the two conformations, paired with Taylor expansions in internal coordinates.[23]

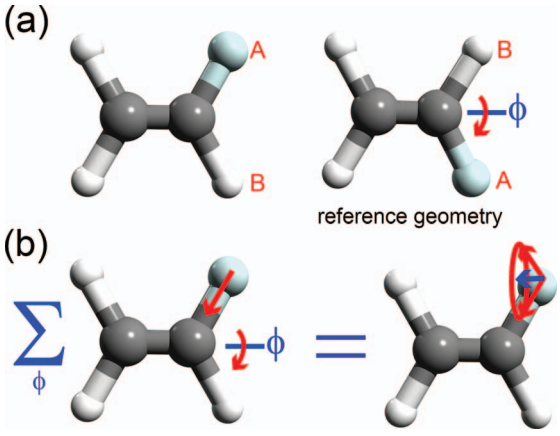


FIG. 1. Example of a case where the sum-over-the-orbit averaged force can become unphysical. (a) For symmetric ethylene, the Cartesian Euclidean distance from the left geometry to the reference on the right is invariant to the ϕ-rotation around the C-C axis when the two conformations differ by an H-atom switch. (b) When there is a pure C-H stretching force at the reference geometry (interpolation data point), the sum-over-the-orbit averaging in ϕ-angle will generate an average over a force cone (red), effectively producing an in-plane bending force vector (blue arrow).

The alignment actually finds the rotation of the interpolation data point that minimizes its distance to the geometry of interest. During molecular dynamics simulations, because the molecule changes its orientation and shape only by a small amount at each time step, newly aligning each interpolation data point can be achieved very fast when an iterative scheme is adopted in the Eulerian angle space.[23,39] Verification of the global minimum nature of the alignment can also be readily attained in conjunction with the iteration. One potential concern against using the alignment scheme may be the ambiguity in the alignment. Namely, for certain molecules, the alignment may lead to a non-unique solution. The ethylene molecule shown in Fig. 1(a) will be a good example, as any angle ϕ with overlapping C-C axes yield degenerate alignment minima. However, this does not pose any difficulty for our scheme because it is paired with Taylor energy/force expansions in internal coordinates. Practically, the iteration will converge to an undefined angle ϕ. However, the conformational distance is constant over all possible ϕ in this case, and the internal gradient exerts exactly the same molecular force no matter where the alignment procedure converges to. The degeneracy with ethylene in our example is related to the intrinsic molecular symmetry. Even in cases, where the distance degeneracy happens for an accidental reason, our alignment procedure will converge to one point in the degenerate space. Because the distance and the weight will be identical within that degenerate space and because the molecular force is dictated by internal coordinates, the force and the potential become uniquely well defined. In addition, the Cartesian weighting function can be readily obtained even for large molecules, and does not neglect or over-represent any degrees of freedom inside any given molecule.[23]

We also note that our scheme of adopting Cartesian weighting and internal Taylor expansion may not be a universally optimal solution. Namely, depending on the properties of the molecular systems, alternative choices may work better. For example, when the molecule does not experience collinearities, adopting redundant internal coordinates for both weighting and Taylor expansion can perform very well. This can even be supplemented with a separate weighting scheme for bonds, angles, and dihedrals toward improved convergence in terms of the dataset size.[17,37]

In any case, the working expression for our interpolated diabatic Hamiltonian $\mathbf{D}(\mathbf{X})$ at an arbitrary geometry $\mathbf{X}$ can be written as

$$\mathbf{D}(\mathbf{X}) = \sum_n w(n)\mathbf{D}(\mathbf{X};n) = \sum_n w(\mathbf{X};n)\mathbf{D}(\mathbf{Z};\mathbf{Z}(n)). \quad (9)$$

The weighting function is determined by the distance between $\mathbf{X}$ and the Cartesian data point geometry $\mathbf{X}(n)$. The Taylor expanded diabatic Hamiltonian $\mathbf{D}(\mathbf{Z};n)$ is defined with the curvilinear internal coordinates $\mathbf{Z}$. This internal coordinate is straightforwardly transformed from the Cartesian coordinate, i.e., $\mathbf{Z} = \mathbf{Z}(\mathbf{X})$. The explicit definition of the Cartesian weighting function that we have adopted is

$$w(n) = w(\mathbf{X};\mathbf{X}(n)) = \frac{\left(1/d_n^2\right)^p}{\sum_m \left(1/d_m^2\right)^p} \quad (10)$$

with $p = 6$, where

$$d_n^2 = \sum_i |(\mathbf{X})_i - (\mathbf{S}\mathbf{X}(n) + \mathbf{T})_i|^2. \quad (11)$$

Here, $(\mathbf{X})_i$ is the Cartesian sub-vector from $\mathbf{X}$ belonging to the ith atom of the system, while $\mathbf{S}$ and $\mathbf{T}$ are the rotational matrix and the translational vector for minimizing the squared Euclidean distance, d_n^2. The fast iterative approach for obtaining $\mathbf{S}$ and $\mathbf{T}$ mentioned earlier follows the formulation reported by Rhee.[23] See Appendix for the form of the weighting function derivative.

By evaluating the interpolated diabatic Hamiltonian along with its gradient, the physical quantities utilized in conventional molecular simulations, such as adiabatic energies and nonadiabatic coupling matrix elements, can be computed in a straightforward manner. The first derivative of $\mathbf{D}(\mathbf{Z})$ is obtained as

$$\frac{\partial \mathbf{D}(\mathbf{Z}(\mathbf{X}))}{\partial X_a} = \sum_n w(\mathbf{X};n)\frac{\partial \mathbf{D}(\mathbf{Z};\mathbf{Z}(n))}{\partial X_a} + \mathbf{D}(\mathbf{Z};\mathbf{Z}(n))\sum_n \frac{\partial w(\mathbf{X};n)}{\partial X_a}. \quad (12)$$

The adiabatic energy and its derivative can be obtained by diagonalizing $\mathbf{D}(\mathbf{Z})$ and then unitary transforming $\partial\mathbf{D}(\mathbf{Z})/\partial X_a$ with the eigenvectors of $\mathbf{D}(\mathbf{Z})$.[29] The set of eigenvectors is naturally obtained as the row vectors of the adiabatic-to-diabatic transformation matrix $\mathbf{R}$ at any given geometry $\mathbf{X}$. Also, after some algebra, one can show that the derivative coupling $\mathbf{F}^{\text{intp}}$ between states k and j is also related to the off-diagonal element of the unitary transform of $\partial\mathbf{D}(\mathbf{Z})/\partial X_a$:

$$\left(\mathbf{F}_{j,k}^{\text{intp}}\right)_a = \frac{\langle j|\left(\partial H_e/\partial X_a\right)|k\rangle}{E_j - E_k} = -\frac{\mathbf{e}_j^{\mathrm{T}}(\partial\mathbf{D}/\partial X_a)\mathbf{e}_k}{E_j - E_k}. \quad (13)$$

Here, H_e denotes the electronic Hamiltonian. In addition, $\mathbf{e}_i$ is the ith eigenvector of $\mathbf{D}(\mathbf{Z})$, whose elements constitute the transformation rule from the diabatic states into the ith adiabatic state.[3,29,33,40]

C. Diabatic sign consistency toward interpolation

Let us suppose a general situation where we obtain the diabatic Hamiltonian directly from quantum chemical calculations. In this case, the adiabatic energies ($\mathbf{V}$), the ADT transform matrix ($\mathbf{R}$), and the derivative coupling ($\mathbf{F}$) are obtained as the raw data, and then subsequently adopted to generate the diabatic information. Because $\mathbf{R}$ and $\mathbf{F}$ are constructed with vector quantities, whose directions are affected by solutions of eigenproblems, care must be taken to keep their directions consistent for successful interpolation. In fact, quantum chemical calculations rely on processes with undetermined signs in various places: for example, in optimizing one-electron orbitals and in obtaining configuration interaction (CI) coefficients.

In the end, this issue affects our formulation through the sign conventions of the adiabatic and the diabatic bases. From Eq. (3), one can easily see that the ADT matrix transforms

these two bases as

$$\Psi_i = \sum_j R_{ij}\Phi_j, \tag{14}$$

where $\{\Psi_i\}$ and $\{\Phi_i\}$ represent adiabatic and diabatic bases, respectively. Thus, flipping the sign of the kth adiabatic state will change the ADT matrix as

$$\mathbf{R}'^k = \mathbf{d}^k\mathbf{R} \tag{15}$$

with a diagonal matrix $\mathbf{d}^k$ whose (k, k)th element is -1 and whose all other diagonal elements are $+1$. However, this change has no effect on interpolation as the components of the derivative coupling are affected in a canceling manner:

$$\mathbf{F}'^k_a = \mathbf{d}^k\mathbf{F}_a\mathbf{d}^k. \tag{16}$$

Namely, because $\mathbf{d}^k$ and $\mathbf{V}$ commute with each other and because $(\mathbf{d}^k)^2 = \mathbf{1}$, Eqs. (3), (5), and (6) are invariant to the sign flips of any adiabatic state functions.

When the sign of a diabatic state is changed, however, the behavior is completely different. Flipping the sign of the kth diabatic state will change the ADT matrix as

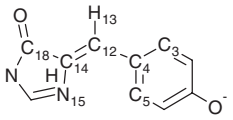
$$\mathbf{R}'^k = \mathbf{R}\mathbf{d}^k. \tag{17}$$

Quite naturally, this leads to the change in the diabatic Hamiltonian as $\mathbf{D}'^k = \mathbf{d}^k\mathbf{D}\mathbf{d}^k$. Thus, consistently keeping the signs of diabatic states on the interpolation data points will be crucial in constructing the database toward interpolation. In this work, because we have the reference analytic model, imposing sign consistency in the dataset can be achieved rather trivially. In future applications with actual quantum chemical calculations, a scheme for building global consistency should be implemented.

III. NUMERICAL TESTS

Ultimately, we aim to apply the present scheme in combination with quantum chemically obtained dataset for interpolation. Before we attempt such applications, however, we need to check how this interpolation approach performs in comparison with the reference original surface. Because adopting quantum chemical calculations for this verification purpose will be extremely time-consuming, we will instead employ an analytical model system. Here, a model potential of the GFP chromophore, *p*HBI,[35] is used as a reference. Its diabatic potential energy matrix, which adopts functions of four important dihedral angles in the molecule (Fig. 2), describes three diabatic states.[35] These states were originally determined with the localizations of the electronic charges after quantum chemical calculations.[35] These functions were then supplemented with the AMBER99SB force field[41] parameters to generate an all-atom model of the *p*HBI anion with the diabatic states. Of course, this diabatic model was converted into adiabatic potential and derivative coupling first before being adopted as inputs for interpolation to mimic the situation with quantum chemical calculations. The definitions of internal coordinates that we adopted for Taylor expansions can be found in the supplementary material.[42]

We have chosen the GFP chromophore for a number of reasons. First, it is a relatively large system compared to previous test cases.[17,28,29] Therefore, it can give a sense of how applicable the diabatic interpolation can be toward large systems. In addition, because the GFP chromophore and its homologues are frequently found in various fluorescent proteins,[13,43–47] the present work may produce valuable preliminary information toward their studies. Of course, using an analytical model surface will expedite any benchmark tests as it avoids any time consuming quantum chemical calculations as noted above. If we can demonstrate that the multi-dimensional surface is successfully reproduced with the present interpolation scheme, we will surely be encouraged to extend the procedure toward interpolations with *ab initio* calculations. In this section, we will first test various criteria for sampling the points toward the interpolation dataset construction. Then, we will adopt a combination of the tested criteria for the actual dataset generation, and analyze the performance of the produced interpolated potential in comparison with the reference analytic surface. We will also present results of nonadiabatic simulations as a simple demonstration of the applicability of the present scheme.



θ_i C_{18}-C_{14}-C_{12}-H_{13}
θ_i' C_4-C_{12}-C_{14}-N_{15}
θ_p C_{14}-C_{12}-C_4-C_5
θ_p' H_{13}-C_{12}-C_4-C_3

$\phi_i = (\theta_i+\theta_i')/2$
$\phi_p = (\theta_p+\theta_p')/2$

FIG. 2. Molecular structure of the green fluorescent protein chromophore *p*HBI anion, along with the definitions of torsional angles for defining the analytic diabatic Hamiltonian.

A. Sampling of the interpolation data points: Preliminary consideration

To conduct the diabatic potential energy matrix interpolation, we need to construct a dataset composed of diabatic Hamiltonian $\mathbf{D}$ and its first and second derivatives. This starts from a small primitive dataset, which will grow in size by adding more points gathered from sampling simulations. For the primitive set, we have generated 71 initial geometries by varying two torsional angles ϕ_p and ϕ_i from the S_0 optimized geometry (Franck-Condon point) with $\phi_p = \phi_i = 0^\circ$ (Fig. 2). This was based on previous observations that the chromophore normally twists through only one angle in the excited state.[14,35,48] Even if this primitive dataset is inappropriate, it can be systematically improved through a construction algorithm.[13] With our algorithm, each time a conformation that was farther from all existing data points than a cutoff distance was reached by gas phase simulation on the S_1 adiabatic surface,[13] it was added to the dataset. The conformation selected with this scheme will be relevant with the dynamics considered. Compared to the PES growing

algorithm[37] suggested by Collins *et al.*,[21,28,29,38] where many candidate conformations are collected first in a similar fashion but then only a small fraction are added based on the interpolation energy correction criteria, our approach will likely exhibit a slower convergence in terms of the number of the data points. This is because our scheme only considers spatial distributions but not the interpolation error itself, and a region with large harmonicity may become over-represented with our distance-based scheme. However, our approach is more easily parallelized. When there are multiple data points to be added in the distance-based algorithm, actually adding one point to the dataset does not change the fates of the other remaining points. Therefore, all the multiple points can be added simultaneously in a parallel fashion. In the energy correction based algorithm, however, adding one point among the multiple candidates will inevitably change the fates of the remaining points, and the parallel addition is not as straightforward. Considering that the interpolation is computationally extremely fast and that the wall time for simultaneously computing quantum chemical information for multiple points can actually cost comparably to single point calculation through parallelization, the slower convergence in terms of the dataset size will not be a serious issue.

To employ the construction algorithm with sampling simulations, we must decide what molecular geometries we will use initially. The first choice that we can think of will be the molecular conformation with the lowest energy on the ground state PES. It is reasonable because the nonadiabatic simulation after excitation will start from the Franck-Condon region, which is distributed near the most stable geometry on the S_0 PES. The S_0 optimized geometry in the model Hamiltonian employed here was with $(\phi_p, \phi_i) = (0, 0)$. (All angles here are denoted in degree units.) In our sampling simulations starting from this geometry, we have observed that structures with $|\phi_i| > 30^\circ$ were not added to the dataset due to the presence of a barrier toward the ϕ_i-twist in the model S_1 PES. To facilitate sampling, therefore, $(\phi_p, \phi_i) = (0, \pm 30)$ were also selected as additional starting geometries. We also attempted to adopt high energy hula-twist geometries with $(\phi_p, \phi_i) = (90, 90)$, $(90, -90)$, $(-90, 90)$, and $(-90, -90)$. By changing the starting conformations in these manners, excess energies of ~10 kJ/mol and ~80 kJ/mol were initially supplied during the sampling simulations.

Progressive convergence of the PES is displayed in Fig. 3 by adopting these initial geometries in an exclusive manner. With the basic samplings starting from only $(\phi_p, \phi_i) = (0, 0)$, the geometries with $|\phi_i| > 30^\circ$ were only rarely sampled as described above. As a result, PES in the unaccessed region is poorly described with interpolation even when the dataset size is large. When additional initial geometries were adopted, the contour lines in this region exhibit improved convergence as shown in Fig. 3(b). Similarly, the thermally inaccessible region, which in fact is not likely to be relevant for the low kinetic energy dynamics simulations, is described well with interpolation after accumulating enough points to the dataset with sampling simulations initiated with larger excess energies (Fig. 3(c)). In this case, however, the region with relatively low energies with small ϕ_p and ϕ_i is not as well converging. Therefore, we can deduce that employing multiple starting structures in a combinatorial manner for sampling simulations is actually desirable for the data construction procedures.

Before proceeding to the actual combined sampling, let us focus on one more aspect that can be important for nonadiabatic situations. In nonadiabatic simulations, reproducing not only the adiabatic energies but also the gaps between different states will be important. In particular, the gap energies near conical intersections need to be properly described. In the sampling technique based on the adiabatic MD simulations, however, the conical intersection region might be considered only rarely as the conformation collection is not affected by the energy gap at all. Indeed, when we inspected the surface gap even in the high energy sampling case, reasonably reliable description could be attained only after a large dataset was established with ~600 data points (Fig. 3(d)). A simplest approach to overcome this limitation will be to add any visited conformation with small enough interpolated gap energy. This process will supply more and more points near the conical intersection to the dataset, and the surface topology around the intersection will naturally improve. Thus, we have tested adding molecular geometries with small interpolated gap energies (smaller than 25.0 kJ/mol). With this, the region with small gap energies ($\Delta E < 6.0$ kJ/mol) became well described even in the early stage with much fewer data points (Fig. 3(e)). Therefore, we can infer that including state-crossing geometries, at least after the interpolated surface becomes stabilized, will efficiently improve the quality of the diabatic Hamiltonian interpolation.

B. Combined approach for sampling and convergence of diabatic Hamiltonian

As discussed above with Fig. 3, the dataset construction with a single criterion will not be sufficient for fulfilling various desired aspects. Therefore, we have combined four conditions listed in Table I in constructing the dataset. The detailed procedures are also described in the table. These sampling procedures were rationalized to expedite the surface convergence as follows. The starting geometries with $(\phi_p, \phi_i) = (0, 0)$, and $(\phi_p, \phi_i) = (0, \pm 30)$ were employed for adding the first 700 geometries. At first, geometries close to the starting geometries were mainly added into the dataset by using smaller cutoff distance ($r = 0.71$ Å). This was to first stabilize the interpolated potential energy. Then, the PES was improved with $r = 1.00$ Å. At this stage, the S_1 surface would be stable and with an intermediate level of reliability. We then intended to add more conformations around the surface crossing region by considering the S_0-S_1 gap energies. Finally, we designed a procedure of obtaining good descriptions in the high energy region of the surface by employing sampling simulations with a large initial kinetic energy. Of course, we have to admit that setting up these procedures had some level of arbitrariness. Potentially, a more systematic approach could be devised without sacrificing the convergence speed. For the present, we will leave it as an open question for future considerations.

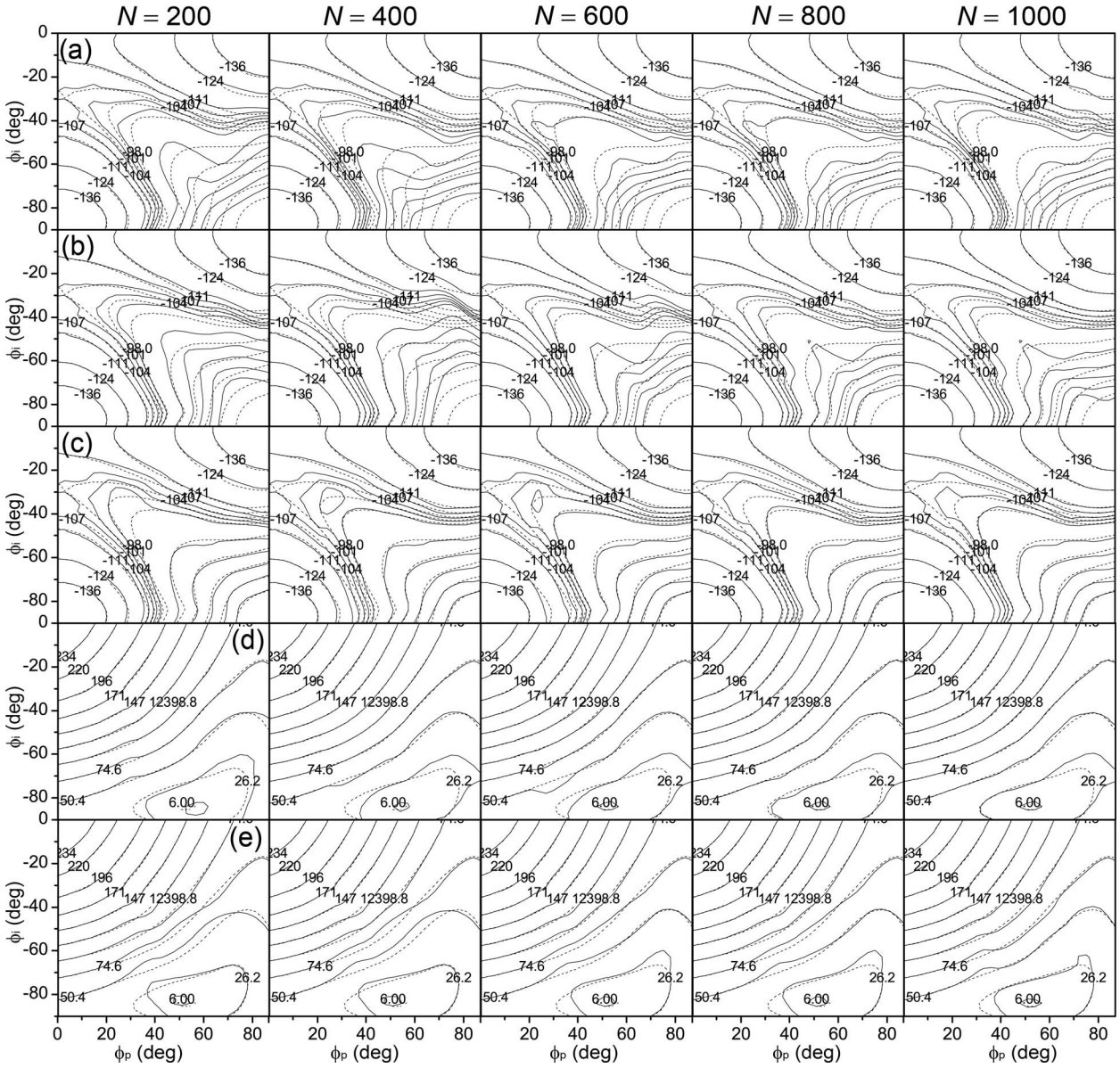


FIG. 3. Convergence comparisons with different data point sampling schemes. Convergence patterns of the interpolated S_1 adiabatic potential when sampling simulations were initiated (a) from $(\phi_p, \phi_i) = (0, 0)$, (b) from $(\phi_p, \phi_i) = (0, \pm 30)$, and (c) from $(\phi_p, \phi_i) = (\pm 90, \pm 90)$. (d) Convergence pattern of the $S_0 - S_1$ gap energy from the dataset collected for (c), without the gap energy criterion. (e) Convergence pattern of the $S_0 - S_1$ gap energy when the data collection was supplemented by the gap energy criterion. Energies are in the kJ/mol unit.

From this point on, the detailed analysis on the diabatic interpolated Hamiltonian convergence will be conducted using the dataset constructed with the procedures presented above in combination with Table I. Figure 4 displays the convergence of the adiabatic S_1 surface with respect to the number of data points collected through the sampling. This figure indeed shows that the adiabatic potential energies converge to the reference potential values. For example, a description in the low energy region (with $E < -104$ kJ/mol), which will likely be important for dynamics on this surface, is better than the description generated with any single collection scheme shown in Fig. 3 at $N = 1000$. PES shape near the hula-twist geometry is also well reproduced after conducting the sampling step with high initial potential energies. As we expected, combining multiple criteria for sampling was helpful in reproducing the PES in terms of the adiabatic energies.

To conduct the nonadiabatic simulations with interpolated diabatic Hamiltonian, for example, with surface hopping, energy gaps, and nonadiabatic coupling matrix elements will often be required.[49] In particular, these quantities become important near degeneracy, because the quantum amplitudes of the electronic states change drastically when the derivative coupling is large.[1–3] As alluded before, naturally, correctly describing the energy difference by means of interpolation is more challenging than describing the energy itself in the single surface interpolation case, especially near the degeneracy point. Figure 5 displays the convergence of the $S_0 - S_1$ energy gap with the increase of the dataset size. The interpolated

TABLE I. Simulation conditions adopted for the interpolation dataset construction.[a]

Round	Number[b]	Distance[c] (Å)	Gap[d]
Primitive	71		
1[e]	229	0.71	No
2[f]	200	1.0	No
3[f]	200	1.0	Yes
4[g]	300	1.0	Yes

[a]Temperature adopted for generating random initial velocity was 300 K for all rounds.
[b]Number of data points obtained from each round of dataset construction.
[c]Distance cutoff from pre-existing data points for deciding the addition of a new data point.
[d]When yes, a point was also added when $E(S_1) - E(S_0) < 25$ kJ/mol condition was met.
[e]In this round, the starting geometries for MD simulations were with $(\phi_p, \phi_i) = (0, 0)$ for the first 129 points, then $(\phi_p, \phi_i) = (0, 30)$ for the next 50 points, and $(\phi_p, \phi_i) = (0, -30)$ for the last 50 points. Angles here are in degree units. The reason for adopting different starting geometries toward the last 100 points was to facilitate the sampling in the span of torsional angle ϕ_i. After inspecting the first 129 added points, we observed no geometries with $|\phi_i| > 30°$ were sampled.
[f]In this round, the starting geometries for MD simulations were with $(\phi_p, \phi_i) = (0, 0)$ for the first 100 points, then $(\phi_p, \phi_i) = (0, 30)$ for the next 50 points, and $(\phi_p, \phi_i) = (0, -30)$ for the last 50 points.
[g]In this round, 50 points were obtained from each of the starting angle combinations of $(\phi_p, \phi_i) = (90, 90)$, $(90, -90)$, $(-90, 90)$, and $(-90, -90)$ for the first 200 points, and then 25 more points were again added to each combination to make the final dataset size 1000.

gap energies converge well to the energy differences calculated from the analytic diabatic Hamiltonian. Even the surface crossing region with the energy gap smaller than 5 kJ/mol is well-described after sampling the surface crossing points, although there is some level of discrepancy in its nearby region. This discrepancy does not necessarily mean that much larger dataset is needed to properly describe the conical intersection. In this test case, in fact, the intersection is located in thermally inaccessible region on the S_1 analytical surface as shown in Figs. 4 and 5. Because we have sampled data points up to $N = 700$ on the S_1 surface with the initial potential energies of -109.1 and -103.5 kJ/mol, the intersection region was not visited frequently. After adopting samplings that started from high potential energy regions, the description in the intersection region improves noticeably.[42]

More detailed aspects of the convergence can be seen from the one-dimensional cut of the energy gap pattern at $\phi_p = 57°$. This line cut possesses the conical intersection at $(\phi_p, \phi_i) = (57, -84)$. Indeed, Fig. 6(a) well displays the pattern the interpolated surfaces converge with the increase in the dataset size. In addition, Fig. 6(b) demonstrates that the singularity pattern of the derivative coupling vector is reproduced well with the interpolation scheme. It should be noted that if adiabatic Hamiltonians were interpolated, the singularity pattern of the coupling would not have been properly captured. By interpolating diabatic Hamiltonians, it is indeed possible to overcome this issue as described by Evenhuis *et al.*[17,28,29]

C. Effect of quasi-diabatization on the quality of interpolation

As mentioned before, the ultimate goal of developing the present method for interpolating diabatic Hamiltonian is

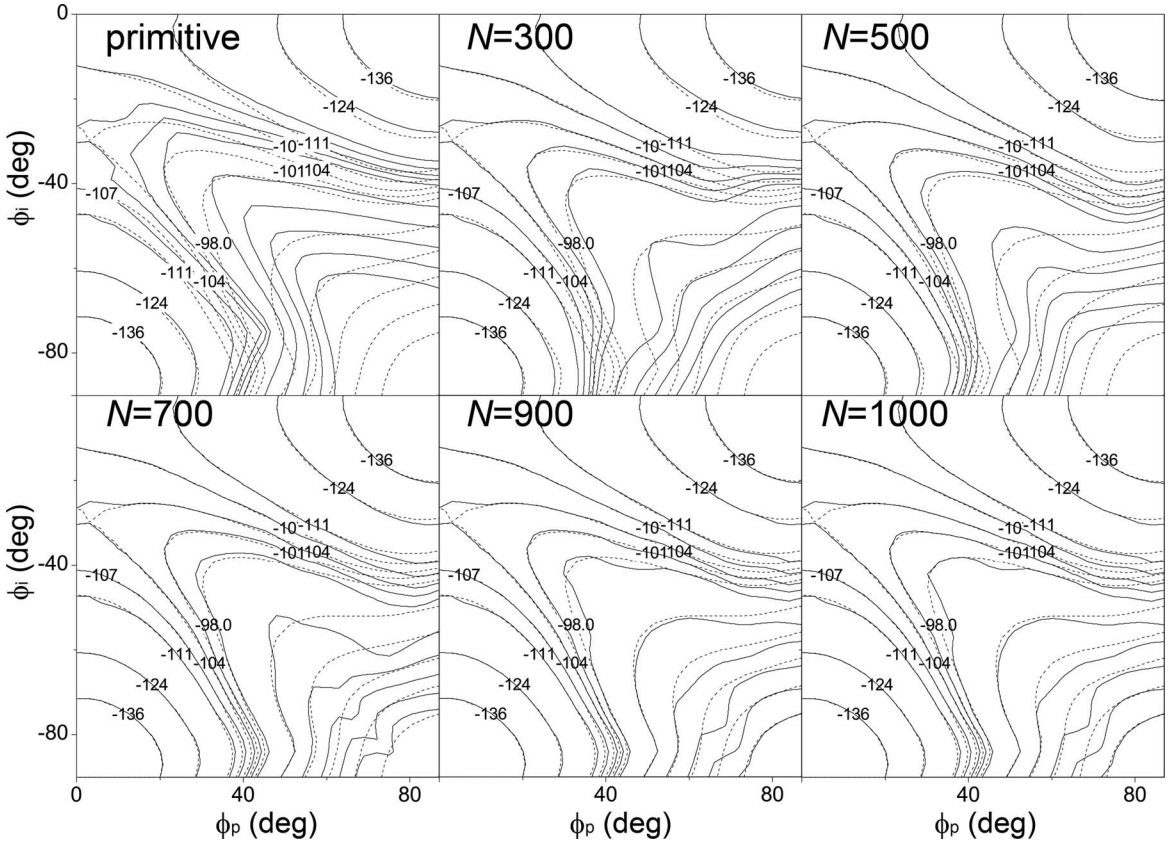


FIG. 4. Convergence pattern of the interpolated S_1 adiabatic potential energy surface, with respect to the number of data points N. The reference analytic adiabatic energies are plotted as dashed contours. Energies are in the kJ/mol unit.

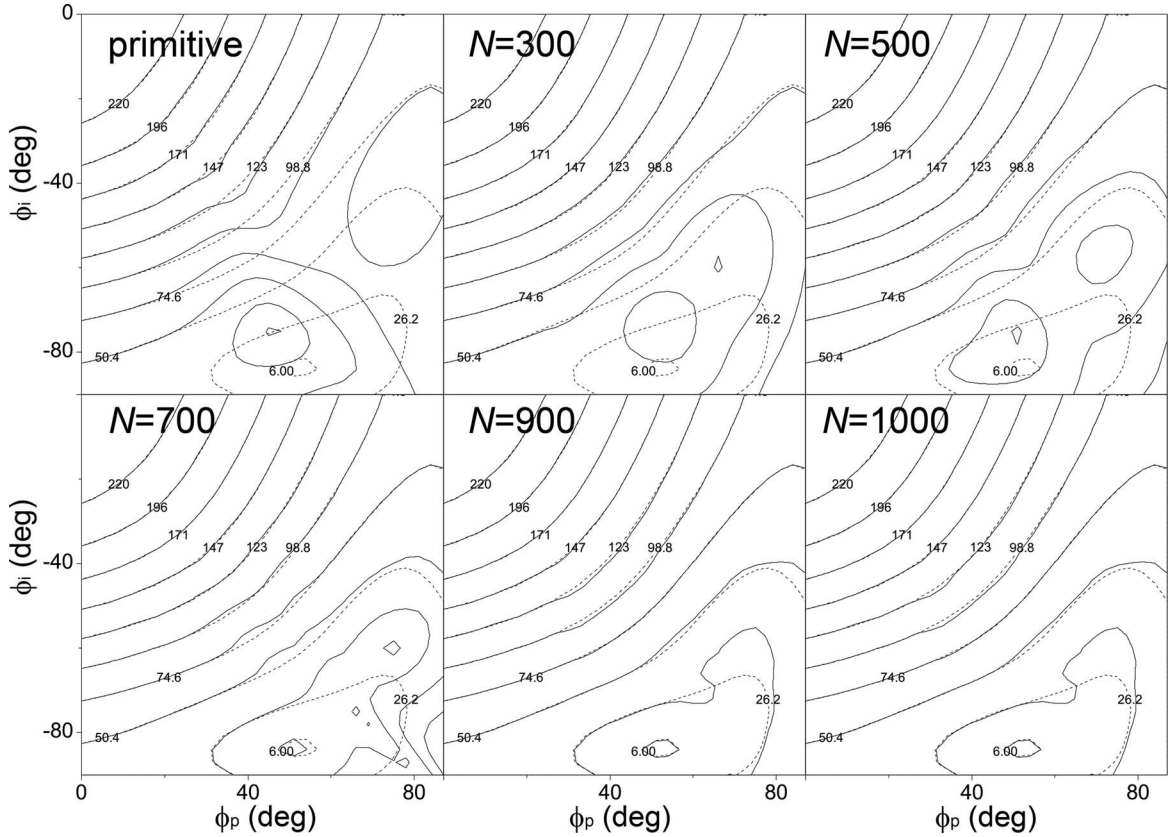


FIG. 5. Convergence pattern of the energy gap between the S_0 and S_1 adiabatic states, with respect to the number of data points N. The adiabatic energy gap with the reference analytic model is also shown with dashed contours. Energies are in the kJ/mol unit.

to apply it to photodynamics studies with potential surfaces which are as reliable as quantum chemically calculated ones. In this sense, our approach has an important limitation as it is based on externally defined diabatization from a finite number of electronic states. In practice, in any practical quantum chemical calculations, a perfect diabatization is impossible except for very small molecules.[30] Indeed, the interactions between active states and non-active states are conventionally omitted during practical diabatization schemes. This leads to the well-known non-removable residual derivative coupling components, which subsequently induce inconsistency between the adiabatic-to-diabatic transform matrix **R** and the nonadiabatic coupling matrix **F**. Namely, the relationship of Eq. (4) becomes only an approximation. Then, a naturally arising question will be how the residual part of the derivative coupling vector and the **F**-to-**R** inconsistency will affect the quality of the interpolation.

In order to bring an answer to this question, we have additionally performed a rather simple but semi-quantitative test on the accuracy of the interpolation by intentionally implementing some **F**-to-**R** inconsistency. For this, we have extended the toy Hamiltonian into 4×4 dimension by adding a fourth diabatic state. Its corresponding nonadiabatic coupling vectors are then trivially obtained with Eq. (13). This **F** will of course be different from the original **F** in the 3×3 space. Therefore, if we now consider the original 3×3 portion of the extended Hamiltonian and take its adiabatic transformations as before, we have **F** and **R** matrices inconsistently defined with each other. Of course, this inconsistency is generated by the neglect of the interactions between the new fourth diabatic state and the original three states. We have designed the Hamiltonian extension such that the additional matrix elements for describing the fourth state and its couplings are linearly independent from the already existing elements.[42] The residual coupling in the quantum chemical calculation case will have a close correspondence to the error in the **F** vector,

$$\delta_{\mathbf{F}} = \frac{|\mathbf{F}_{4\times4} - \mathbf{F}_{3\times3}|}{|\mathbf{F}_{3\times3}|}. \tag{18}$$

The size of the error can be tuned by scaling the off-diagonal elements of the Hamiltonian between the fourth state and the other states. Because $\delta_{\mathbf{F}}$ itself varies with the changes in the molecular geometry, we have taken 3600 mesh point geometries utilized in generating contours in Figs. 3–5 and adopted the root-mean-squared $\langle\delta_{\mathbf{F}}\rangle$ as the representative error. The interpolations were then performed again using the erroneous **F** vectors in the dataset, and the errors in the potential energies were monitored as a function $\langle\delta_{\mathbf{F}}\rangle$. Based on the same 3600 mesh point molecular structures, the root-mean-squared and the maximum errors in the S_1 potential energy were obtained and are displayed in Fig. 7(a). Not so surprisingly, both the maximum and the r.m.s. errors have almost linear relationships with $\langle\delta_{\mathbf{F}}\rangle$. The most important feature to notice is the fact that the potential error associated with $\langle\delta_{\mathbf{F}}\rangle$ is rather small when the coupling error level is below 10%. Figure 7(b) also displays how different the interpolated potential surface and the surface gap become with a ~6% level of coupling error. Indeed, even the most sensitive conical intersection region

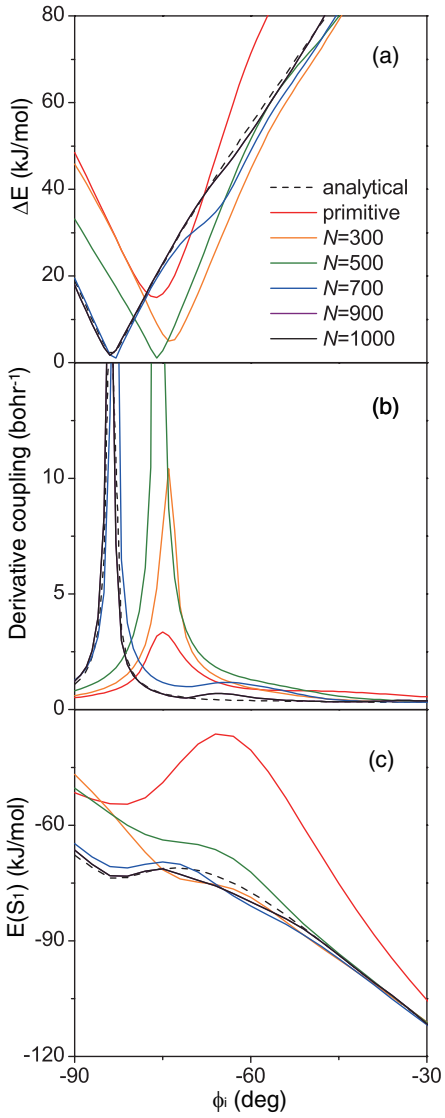


FIG. 6. One-dimensional cut representation of the surface characters at $\phi_p = 57°$, with the interpolated (solid) and analytical reference (dashed) potential surfaces: (a) $S_0 - S_1$ energy gap, (b) norms of derivative coupling vectors, and (c) adiabatic S_1 energy. Lines with $N = 900$ and 1000 are almost exactly overlapping with each other.

is not significantly affected. As it has been discussed that the non-removable couplings can become quite negligible with a proper choice of the diabatization space,[33,36] this result implies that the interpolation based on external diabatization algorithm from quantum chemical calculation can be a reliable approach, even with the limitations associated with the **F**-to-**R** inconsistency. Of course, a relatively good correspondence is not automatically guaranteed, and we admit that care must be taken in applying any diabatization scheme, especially near dynamically important regions.

D. Applications to nonadiabatic dynamics

As a final test of the applicability of the current scheme to molecular simulations, we have performed nonadiabatic trajectory surface hopping calculations using both analytical and interpolated diabatic potential energies. The starting geometries for these simulations were generated with an ensemble from 1.0 ns equilibrium NVT ($T = 300$ K) MD simulation on the S_0 surface.[50] The standard fewest switches surface hopping (FSSH) trajectories[1–3] with analytical or interpolated diabatic potential energies were then propagated for 50 ps. In total, 500 trajectories were propagated.

The resulting nonadiabatic population transfers are displayed in Fig. 8. The S_1 population change in time with interpolated PES with $N > 500$ is similar to the one observed with the analytical surface model. The reason for the incorrect population decay patterns at $N = 71$ and at $N = 300$ will likely be the incorrect location of the conical intersection and the subsequent error in the barrier height toward the conical intersection region. As described in the above, after $N = 700$, the conical intersection is similarly located in comparison to the analytical surface (Figs. 5, 6(a), and 6(b)). In addition, the barrier height toward the intersection becomes quite reliable after reaching this database size as displayed in Fig. 6(c). As a consequence, the population transfer dynamics is in good agreement with the results from the analytical potential. These results imply that current dataset construction scheme is working properly for conducting nonadiabatic MD simulations, at least with initial kinetic

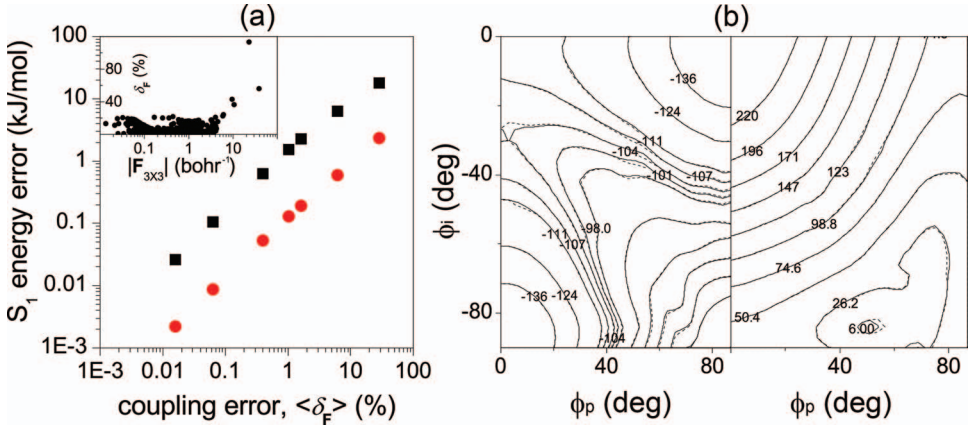


FIG. 7. Effect of coupling errors in the accuracy of interpolated PES. (a) The root-mean-squared (red circles) and the maximum errors (black squares) of the S_1 potential energies as functions of root-mean-squared %-error in the coupling. The inset displays the distribution of absolute coupling errors for the case of 6% average coupling error. (b) Interpolated S_1 potential energies and the $S_1 - S_0$ gap at the 6% level of coupling errors (solid lines). The interpolated potential with exact coupling vectors is also shown (dashed lines) for comparison.

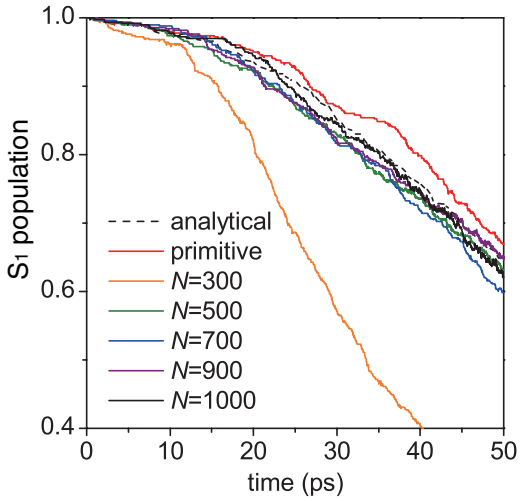


FIG. 8. Decay of S_1 populations from nonadiabatic trajectory hopping simulations with interpolated (solid) and analytical (dashed) potentials.

energies corresponding to $T = 300$ K. Correct population transfer dynamics should be reachable when the conical intersection and the surface shape toward it are well-described with the diabatic Hamiltonian interpolation.

IV. CONCLUSIONS

We have presented a scheme for interpolating diabatic potential energy matrices. By using the Cartesian weighting coordinates, interpolation could be straightforwardly performed even for a relatively large molecule. To evaluate its applicability to actual surface crossing dynamics, we have tested it with a model of the green fluorescent protein chromophore. We observed that the diabatic Hamiltonian interpolation was indeed appropriate for calculating adiabatic energies in the vicinity of the surface crossing point as well as in regions where the surfaces are well separated. Nonadiabatic coupling and its singularity near the conical intersection were also well described. The surface description was systematically improved with the conventional molecular dynamics simulation based approach,[13,24] adapted for the surface crossing situation. This approach correctly generated improving behaviors in predictable ways. Simple nonadiabatic dynamics simulations were also performed for a demonstration purpose, displaying that the interpolation of diabatic potential energy matrices is in fact suitable for computer simulations of chemical events of that class.

Nevertheless, the present approach at the current stage will not be completely adequate for investigating photodynamics of biological systems. In such studies, the nonbonded interactions between the chromophore and the protein environment need to be accurately described. Because interpolation toward condensed phase simulations has only utilized the fixed point charge model,[13,24] and because the fixed charge model cannot describe remixing of diabatic states, it is not yet known how interpolation will behave in such a regime. In addition, systems that are reactive in the excited states may require much larger dataset for interpolation. For example, excited state proton transfer[51,52] occurs quite frequently with fluorescent proteins. Even though interpolation technique, in principle, may describe such reactions, representing large and flexible systems may become practically difficult.

Despite these limitations, the present approach will still be useful in studying nonadiabatic dynamics of relatively large molecules when combined with appropriate quantum chemical calculations, as long as a suitable quasi-diabatization process is provided. For example, many monomethine cyanine dyes and, therefore, fluorescent protein chromophores have similar electronic structures, and diabatization for those systems will likely be well-defined.[35,53] By design, the present method avoids the time-consuming on-the-fly quantum chemical calculations, but can yield nonadiabatic simulations on as much reliable PESs. Most importantly, it can generate a large number of trajectories at a reasonable computational cost. Our next step will be to include the diabatic state mixing on the nonbonded interactions to overcome the first limitation stated in the above paragraph. This will make the combination of the present approach with the computationally less demanding molecular mechanics (MM) description reliable. Namely, by adopting the present scheme for the chromophore part and by employing the MM description for the remainder in a multiscale manner,[54–57] we will be able to tackle large biomolecular systems. In fact, compared to the widely applied quantum mechanics/molecular mechanics (QM/MM) approaches, the interpolation based scheme will have one additional merit besides the benefits described above. Once the interpolation dataset construction procedures are completed, any systems involving the same molecular moiety in the QM or interpolation part can be readily studied without performing any additional quantum chemical calculations. For example, with interpolation, studying protein mutation effects in the MM region can be performed at much reduced cost than with the QM/MM approach. With this in mind, we are currently pursuing simulations with interpolation dynamics based on *ab initio* data and are making progress in developing electrostatic model that will be more relevant for diabatic description of the potential. We hope to report on these in due course in the future.

ACKNOWLEDGMENTS

This work was supported by the Institute for Basic Science (IBS) in Korea. We thank Joseph Subotnik (University of Pennsylvania) for insightful conversations. The supercomputer time from Korea Institute of Science and Technology Information (KISTI) is also gratefully acknowledged.

APPENDIX: ANALYTICAL GRADIENT OF THE WEIGHTING FUNCTION $w(n)$

To obtain globally well-behaved diabatic Hamiltonian $\mathbf{D}(\mathbf{X})$, not only the Taylor expansions from each data point but also the weighting function should be differentiable. In addition, for molecular dynamics simulations, the differentiation should be performed analytically but not numerically. Here, we prove that the Cartesian weighting function is indeed differentiable with analytic formulations, and therefore globally well-behaved.[23] The derivative of $w(n)$ with respect

to a Cartesian coordinate X_a is

$$\frac{\partial w(n)}{\partial X_a} = \frac{1}{\sum_m \left(1/d_m^2\right)^p} \frac{\partial}{\partial X_a} \left(\frac{1}{d_n^2}\right)^p + \left(\frac{1}{d_n^2}\right)^p \frac{\partial}{\partial X_a} \left(\frac{1}{\sum_m \left(1/d_m^2\right)^p}\right) = \frac{1}{\sum_n \left(1/d_m^2\right)^p} \frac{\partial}{\partial X_a} \left(\frac{1}{d_n^2}\right)^p - \left(\frac{1}{d_n^2}\right)^p \times \left[\sum_m \left(1/d_m^2\right)^p\right]^{-2} \frac{\partial}{\partial X_a} \left(\sum_m \left(1/d_m^2\right)^p\right). \quad \text{(A1)}$$

The two key terms can be written as

$$\frac{\partial}{\partial X_a} \left(\frac{1}{d_n^2}\right)^p = p \left(\frac{1}{d_n^2}\right)^{p-1} \frac{\partial}{\partial X_a} \left(\frac{1}{d_n^2}\right) = -p \left(\frac{1}{d_n^2}\right)^{p-1} \left(\frac{1}{d_n^2}\right)^2 \frac{\partial d_n^2}{\partial X_a}, \quad \text{(A2)}$$

$$\frac{\partial}{\partial X_a} \left(\sum_m \frac{1}{d_m^2}\right)^p = p \left(\sum_m \frac{1}{d_m^2}\right)^{p-1} \frac{\partial}{\partial X_a} \left(\sum_m \frac{1}{d_m^2}\right) = -p \left(\sum_m \frac{1}{d_m^2}\right)^{p-1} \sum_m \left(\frac{1}{d_m^2}\right)^2 \frac{\partial d_m^2}{\partial X_a}. \quad \text{(A3)}$$

The remaining task is to differentiate $d_n^2 = \sum_i |(\mathbf{X})_i - (\mathbf{SX}(n) + \mathbf{T})_i|^2$ with respect to X_a. For simplicity, we assume that the centers of masses of the data point $\mathbf{X}(n)$ and the molecular conformation $\mathbf{X}$ coincide (namely, $\mathbf{T} = \mathbf{0}$), which can be trivially enforced during simulations. From the fact that $\mathbf{X}(n)$ is independent of X_a,

$$\frac{\partial d_n^2}{\partial X_a} = 2 \sum_i \left[(\mathbf{X})_i - (\mathbf{SX}(n))_i\right] \sum_i \left(\frac{\partial X_i}{\partial X_a} - \left(\frac{\partial \mathbf{S}}{\partial X_a} \mathbf{X}(n)\right)_i\right). \quad \text{(A4)}$$

Because $\mathbf{S}$ is obtained by minimizing the Euclidean distance with respect to $\mathbf{X}$, $\partial \mathbf{S}/\partial X_a = 0$.[23] Therefore, the final derivative of d_n^2 with respect to X_a simplifies to

$$\frac{\partial d_n^2}{\partial X_a} = 2 \sum_i \left[(\mathbf{X})_i - (\mathbf{SX}(n))_i\right]. \quad \text{(A5)}$$

By combining these expressions, the analytic gradient of $w(n)$ can be readily obtained.

[1] J. C. Tully, J. Chem. Phys. **93**, 1061 (1990).
[2] S. Hammes-Schiffer and J. C. Tully, J. Chem. Phys. **101**, 4657 (1994).
[3] E. Fabiano, T. W. Keal, and W. Thiel, Chem. Phys. **349**, 334 (2008).
[4] M. Ben-Nun, J. Quenneville, and T. J. Martínez, J. Phys. Chem. A **104**, 5161 (2000).
[5] O. V. Prezhdo and P. J. Rossky, J. Chem. Phys. **107**, 825 (1997).
[6] C. Zhu, A. W. Jasper, and D. G. Truhlar, J. Chem. Phys. **120**, 5543 (2004).
[7] M. Desouter-Lecomte, and J. C. Lorquet, J. Chem. Phys. **71**, 4391 (1979).
[8] W. W. Parson and A. Warshel, Chem. Phys. **296**, 201 (2004).
[9] H. W. Kim, A. Kelly, J. W. Park, and Y. M. Rhee, J. Am. Chem. Soc. **134**, 11640 (2012).
[10] A. Ishizaki and G. R. Fleming, J. Chem. Phys. **130**, 234111 (2009).
[11] B. Heggen, Z. Lan, and W. Thiel, Phys. Chem. Chem. Phys. **14**, 8137 (2012).
[12] Z. Lan, E. Fabiano, and W. Thiel, J. Phys. Chem. B **113**, 3548 (2009).
[13] J. W. Park and Y. M. Rhee, J. Phys. Chem. B **116**, 11137 (2012).
[14] A. M. Virshup, C. Punwong, T. V. Pogorelov, B. A. Lindquist, C. Ko, and T. J. Martínez, J. Phys. Chem. B **113**, 3280 (2009).
[15] X. Li, L. W. Chung, H. Mizuno, A. Miyawaki, and K. Morokuma, J. Phys. Chem. Lett. **1**, 3328 (2010).
[16] M. Boggio-Pasqua, C. F. Burmeister, M. A. Robb, and G. Groenhof, Phys. Chem. Chem. Phys. **14**, 7912 (2012).
[17] C. R. Evenhuis and T. J. Martínez, J. Chem. Phys. **135**, 224110 (2011).
[18] M. M. Gallo and D. R. Yarkony, J. Chem. Phys. **86**, 4990 (1987).
[19] L. M. Frutos, T. Andruniów, F. Santoro, N. Ferré, and M. Olivucci, Proc. Natl. Acad. Sci. U.S.A. **104**, 7764 (2007).
[20] G. Cui, Z. Lan, and W. Thiel, J. Am. Chem. Soc. **134**, 1662 (2012).
[21] J. Ischtwan and M. A. Collins, J. Chem. Phys. **100**, 8080 (1994).
[22] K. A. Nguyen, I. Rossi, and D. G. Truhlar, J. Chem. Phys. **103**, 5522 (1995).
[23] Y. M. Rhee, J. Chem. Phys. **113**, 6021 (2000).
[24] J. W. Park, H. W. Kim, C.-I. Song, and Y. M. Rhee, J. Chem. Phys. **135**, 014107 (2011).
[25] M. Higashi and D. G. Truhlar, J. Chem. Theory Comput. **4**, 1032 (2008).
[26] M. Higashi and D. G. Truhlar, J. Chem. Theory Comput. **4**, 790 (2008).
[27] Y. M. Rhee, T. G. Lee, S. C. Park, and M. S. Kim, J. Chem. Phys. **106**, 1003 (1997).
[28] C. R. Evenhuis and M. A. Collins, J. Chem. Phys. **121**, 2515 (2004).
[29] O. Godsi, C. R. Evenhuis, and M. A. Collins, J. Chem. Phys. **125**, 104105 (2006).
[30] C. A. Mead and D. G. Truhlar, J. Chem. Phys. **77**, 6090 (1982).
[31] T. Pacher, L. S. Cederbaum, and H. Köppel, J. Chem. Phys. **89**, 7367 (1988).
[32] H. Nakamura and D. G. Truhlar, J. Chem. Phys. **118**, 6816 (2003).
[33] W. Domcke, D. R. Yarkony, and H. Köppel, *Conical Intersections: Electronic Structure, Dynamics & Spectroscopy* (World Scientific, River Edge, NJ, 2004).
[34] J. E. Subotnik, S. Yeganeh, R. J. Cave, and M. A. Ratner, J. Chem. Phys. **129**, 244101 (2008).
[35] S. Olsen and R. H. McKenzie, J. Chem. Phys. **130**, 184302 (2009).
[36] S. Fatehi, E. Alguire, and J. E. Subotnik, J. Chem. Phys. **139**, 124112 (2013).
[37] C. R. Evenhuis and M. A. Collins, J. Phys. Chem. A **113**, 3979 (2009).
[38] K. C. Thompson, M. J. T. Jordan, and M. A. Collins, J. Chem. Phys. **108**, 564 (1998).
[39] R. Fletcher, *Practical Methods of Optimization* (Wiley, New York, 1987).
[40] B. H. Lengsfield III, P. Saxe, and D. R. Yarkony, J. Chem. Phys. **81**, 4549 (1984).
[41] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, Proteins **65**, 712 (2006).
[42] See supplementary material at http://dx.doi.org/10.1063/1.4872155 for the definitions of the internal coordinates of *p*HBI; for convergence tests with interpolation toward an anharmonic variation of AMBER99SB; and for the actual forms of the Hamiltonian terms related to the fourth diabatic state in checking the **F**-to-**R** inconsistency issue.
[43] R. Y. Tsien, Ann. Rev. Biochem. **67**, 509 (1998).
[44] M. E. Martin, F. Negri, and M. Olivucci, J. Am. Chem. Soc. **126**, 5452 (2004).
[45] J. N. Henderson and S. J. Remington, Physiology **21**, 162 (2006).
[46] J.-Y. Hasegawa, K. Fujimoto, B. Swerts, T. Miyahara, and H. Nakatsuji, J. Comput. Chem. **28**, 2443 (2007).
[47] R. N. Day and M. W. Davidson, Chem. Soc. Rev. **38**, 2887 (2009).
[48] S. Olsen, K. Lamothe, and T. J. Martínez, J. Am. Chem. Soc. **132**, 1192 (2010).
[49] This is not necessarily the case, and all-atom style nonadiabatic simulations can readily be performed from diabatic Hamiltonians at least in an approximate manner. For example, see Ref. 9. Indeed, adopting diabatic Hamiltonian does not pose any difficulty in interpolation-based dynamics.
[50] Even though it may not be rigorously acceptable to put a gas phase molecule in a thermostat, we adopted such an approach simply to generate initial conditions with varying energies. For this purpose, we employed the stable Berendsen thermostat algorithm (see H. J. C. Berendsen, J. P. M.

Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984)) with a relaxation time $\tau_c = 5.0$ ps.
[51]S. R. Meech, Chem. Soc. Rev. **38**, 2922 (2009).
[52]J. J. van Thor, Chem. Soc. Rev. **38**, 2935 (2009).
[53]S. Olsen and R. H. McKenzie, J. Chem. Phys. **134**, 114520 (2011).
[54]J. Åaqvist and A. Warshel, Chem. Rev. **93**, 2523 (1993).
[55]H. Lin and D. G. Truhlar, Theor. Chem. Acc. **117**, 185 (2007).
[56]H. M. Senn and W. Thiel, Angew. Chem., Int. Ed. **48**, 1198 (2009).
[57]L. W. Chung, H. Hirao, X. Li, and K. Morokuma, WIREs Comput. Mol. Sci. **2**, 327 (2012).