JCTC Journal of Chemical Theory and Computation



Reaction Mechanisms

# Reaction path analysis from potential energy contributions using forces: an accessible estimator of reaction coordinate adequacy

Nicolas Oscar Foglia, Mariano C. González Lebrero, Rodolfo Biekofsky, and Dario A Estrin

*J. Chem. Theory Comput.*, **Just Accepted Manuscript** • DOI: 10.1021/acs.jctc.9b01081 • Publication Date (Web): 30 Jan 2020



## Just Accepted



is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

# Reaction path analysis from potential energy contributions using forces: an accessible estimator of reaction coordinate adequacy

Nicolás O. Foglia,[†] Mariano C. Gonzalez Lebrero,[†] Rodolfo R. Biekofsky,[‡] and Dario A. Estrin[*,†]

†*Departamento de Química Inorgánica, Analítica y Química Física/INQUIMAE-CONICET, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pab. II, Buenos Aires (C1428EHA) Argentina*

‡*Moebius Research Ltd., Systems Biomedicine, 24 Chedworth House, West Green Rd, N15 5EH, London, United Kingdom*

E-mail: dario@qi.fcen.uba.ar

**Abstract**

The calculation of potential energy and free energy profiles along complex chemical reactions or rare events processes are of great interest because of their importance for many areas in chemistry, molecular biology, and material science. One typical way to generate these profiles is to add a bias potential to modify the energy surface, which can act on a selected degree of freedom in the system. However, in these cases the quality of the result is strongly dependent on the selection of the degree of freedom over which this bias potential acts. The present work introduces a simple method for the analysis of the degree of freedom selected to describe a chemical process. The proposed methodology is based on the decomposition of contributions to the potential energy

profiles by the integration of forces along a reaction path, which allows evaluating the different contributions to the energy change. This could be useful for discriminating the contributions to the energy arising from different regions of the system, which is particularly useful in systems with complex environments that must be represented using hybrid QM/MM schemes. Furthermore, this methodology allows generating a quick and simple analysis of the degree of freedom used to describe the potential energy profile associated to the reactive process. This is computationally more accessible than the corresponding free energy profile, and can therefore be used as a simple estimator of reaction coordinate adequacy.

# 1 Introduction

Knowledge of the reactivity of a system is an essential concept in chemistry. Most of the phenomena of interest in chemistry are rare events in the sense that they occur on time scales that are orders of magnitude longer than those of the elementary molecular motions. In most of the cases, chemical reactions that involve forming and breaking of chemical bonds entail passing from reactants to products going over a free energy barrier associated to the transition state, and thus are rare events. The theoretical characterization of a chemical reaction requires the determination of a magnitude called reaction coordinat(RC).[1]

An accurate RC is often difficult to identify, because many degrees of freedom may be involved.[2,3] For systems with a normalizable equilibrium distribution, the natural RC is the slowest eigenfunction of the master equation.[3] However, the exact solutions of this high-dimensional equation are extremely difficult to obtain. In this context, approaches like Markov state models (MSMs)[4–7] and diffusion map[8–11] have been proposed. Other approaches are based on the calculation of the committor function, which, at any point in configurational space, gives the probability that a trajectory initiated at this point (with initial velocities sampled from Maxwell–Boltzmann distribution) will reach first the product state rather than the reactant state.[1,3,12–14] Both frameworks can generate adequate RCs, but this kind of analysis is computationally challenging, plus it is hard to obtain a clear physical interpretation of the mechanism. For this, it is necessary to apply extra analysis methods to extract a RC based on limited information from molecular dynamics simulations capable to generate a physical/chemical interpretation of the process.[1,3,15]

Most of chemical reaction trajectories are typically confined to a narrow region of the free energy surface (FES) connecting reactants and products. Since for regions with lower free energy the sampling probability is higher, the exploration of the full free energy landscape in itself is not necessary and the reaction tube typically lies along the minimum free energy path (MFEP).,[15] or in other words the MFEP describes well the committor function[16] The

chemical behavior of these systems depends on the shape of its potential energy surface (PES), and, in principle, all the thermodynamic and kinetic information is contained in the PES.

Some of the schemes to explore PES, such as umbrella sampling,[17] include a bias potential that acts on a selected degree of freedom $\xi$ modifying the PES. A more accurate $\xi$ results in more efficient sampling in free energy calculations.[18,19] In this regard, for cases in which there is a single channel in the free energy surface that connects reactants and products (narrow reaction tube), the best degree of freedom $\xi$ to apply the bias potential is the MFEP as a way to guide the simulation through the true dynamic bottleneck in the configuration space. In spite of the proposal of several mentioned sophisticated schemes to obtain RCs, the choice of $\xi$ based on chemical intuition is still employed, and this typically includes variables related to the covalent bonds that break or form along the transition process.

Because the region of the conformational space to be sampled is strongly dependent on the selected $\xi$ (for a finite sampling time), the choice of $\xi$ has a fundamental role on the reaction path obtained, and therefore on the results obtained.[2,20–22]This kind of analysis has one disadvantage: human intuition in many cases may fail in providing an adequate RC.

A partial solution for improving the description of a chemical reaction in which the selected $\xi$ is not good enough to describe the correct reaction path is to increase the dimensionality of the degrees of freedom to explore the reaction process. For instance, Rosta and collaborators[2] have done a statistical search for discontinuous interatomic distances in Umbrella Sampling/Potential of Mean Force simulations, and they have added those as a new degree of freedom in a 2D free-energy profile. From this 2D free-energy profile, an optimized $\xi$ is obtained. It must be taken into account that all methods relying on increasing the dimensionality usually involve a significant increase in the computational expense. This makes them prohibitive when analyzing large systems, especially under first principles schemes. Because of this, obtaining a good one-dimensional RC is crucial in the analysis of a system using methodologies based on molecular dynamics.

For reactions with a narrow reaction tube, an approximation to study a transition process - which in principle neglects thermal and entropic contributions - is the obtention of a minimum energy path (MEP) between reactants and products states. This strategy is usually the basis for methods of direct dynamics that need as input a description of the reaction path.[23] However, it is necessary to note that MEP generates a limited analysis for systems with an energy surface that is too rough, like for example a reaction in solution, being necessary, in these cases, to model the bulk solvent with a continuum approach or using a microsolvated model that involves explicit water molecules and implicit solvent in the optimization.[24] This eliminates the source of the problem that resides in the solvent's degrees of freedom.

A simple scheme, which has been developed for estimating MEP, performs geometry optimizations in all degrees of freedom except a selected one $\xi$. This procedure is called in the literature potential energy scan or adiabatic mapping (AM) calculation.[25] This methodology obtains a trajectory through the optimization of the geometry of the system modifying the potential energy with a bias potential that acts on a selected degree of freedom $\xi$. However, in these methods, as mentioned above, special care is required in the selection of $\xi$ to be used to describe the reactive process. An alternative more sophisticated method for obtaining MEP is the nudged elastic band (NEB) scheme,[26,27] which describes the reaction path in terms of multiple images that are optimized in the perpendicular direction to the reaction path, allowing the inclusion of a large number of generalized coordinates.

In the present work, an integration of forces (IoF) scheme is presented for evaluating in a simple manner the suitability of a selected $\xi$ degree of freedom to be used to describe a reactive process. The proposed scheme uses low-cost trajectories obtained by AM or NEB calculations as starting points to evaluate the quality of $\xi$ to describe a transition process. Basically, the IoF scheme is a way to decompose the energetic change in analyzable contributions by atoms or by degrees of freedom along a reaction path obtained by AM or by NEB. From an analysis of the atomic level contributions, an improved $\xi$ can be proposed

to describe the reactive process.

Furthermore, the IoF scheme allows discriminating the contributions to the energy arising from different regions of the system, which is particularly useful in complex systems. For instance, in QM/MM studies of complex systems, the IoF scheme allows assessing the effect of the MM environment on energy profiles and changes in atomic contributions.

Example applications of the IoF scheme presented in this study, with small molecule reactions and with reactions in complex environments, demonstrate a practical new estimator of RC adequacy together with an analysis of atomic contributions, which is better compared to free energy schemes in terms of computational cost.

# 2 Methodology

## 2.1 Analysis of reaction path by energy decomposition.

Considering a system of $N$ atoms that interact exclusively by conservative forces described by $n$ degrees of freedom $\{q_1, q_2, ..., q_n\}$ and considering A($\vec{r}_1^A$, $\vec{r}_2^A$, ..., $\vec{r}_N^A$) and B ($\vec{r}_1^B$, $\vec{r}_2^B$, ..., $\vec{r}_N^B$) as two different system configurations, the potential energy difference between both states can be obtained through the path integral of the force of each atom or the generalized force of each degree of freedom.

$$\Delta V = V(B) - V(A) = \begin{cases} -\sum_{j=1}^{N} \int_{\Gamma_{A\to B}} \vec{F}_j \cdot d\vec{r}_j = \sum_{j=1}^{N} W_j \\ -\sum_{k=1}^{n} \int_{\Gamma_{A\to B}} F_{q_k} dq_k = \sum_{k=1}^{n} w_k \end{cases} \quad (1)$$

where $\Gamma_{A\to B}$ is a path in the configuration space that goes from state A to B and $F_{q_k} = -\frac{\partial V}{\partial q_k}$, where $q_k$ is the k-th generalized coordinate that describes the system. In principle, there are endless ways to select the $n$ degrees of freedom to describe the configurational space of a chemical system. However, as mentioned earlier, the description of one $\Gamma_{A\to B}$ in particular is interesting, the MEP, in terms of a single degree of freedom "$\xi$". Therefore, the ideal selection of degrees of freedom to describe the configurational space will include $\xi$ so the progress of the reaction in these coordinates is determined by a single parameter. In this work we propose to use the splitting of energy contributions as a quality criterion for analyzing the degree of freedom $\xi$ as a MEP descriptor and for understanding the molecular determinants of a given process. Using eq. 1 the energy change can be decomposed in the contribution generated by the work in $\xi$ and the work done in all the other degrees of freedom $\xi^{\perp}$.

$$\Delta V = -\int_{\Gamma_{A\to B}} \sum_{j=1}^{n-1} F_{q_j} dq_j - \int_{\Gamma_{A\to B}} F_\xi d\xi = w_{\xi^\perp} + w_\xi \quad (2)$$

The present study is based on a simple idea: if the selection of the $n$ degree of freedom

is made in such a way that the MEP is described only by $\xi$, the work generated by the other degrees of freedom must be null along the reaction path. This approach generates a criterion that allows us to evaluate the quality of $\xi$ to describe the path of minimum energy. However, it should be taken into account that other $\xi$ that do not meet this criterion, but adequately describe reactants, products and TS transition state, will be acceptable too. If the work generated by the degrees of freedom of non $\xi$ ($\xi^{\perp}$) is negligible along the reaction path ($w_{\xi^{\perp}} \approx 0$), then $\Delta V \approx w_{\xi}$, and therefore the degree of freedom used, describes reasonably well the potential energy change along $\Gamma_{A \to B}$ as exemplified in Figure 1A. On the other hand, if the difference between $\Delta V$ and $w_{\xi}$ is large (Figure 1B), then there will be significant contributions to the energy by other degrees of freedom. This will possibly result in discontinuities in the obtained trajectory and hysteresis in the profiles calculated both using restricted optimizations or free energy schemes, when using $\xi$ as a trajectory descriptor.

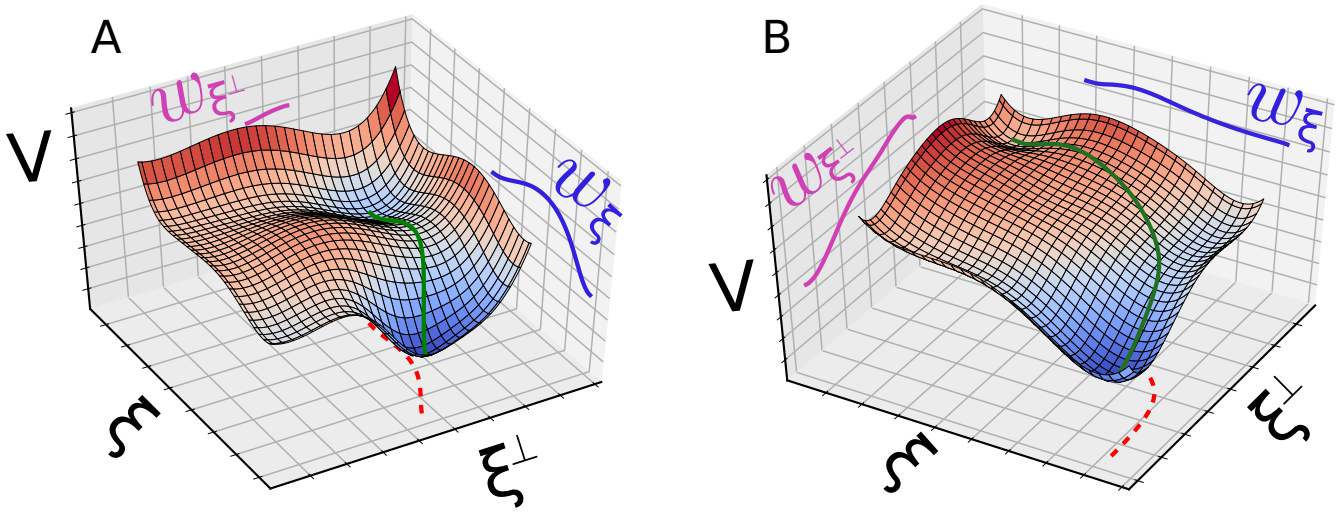



Figure 1: Reaction path over a schematic energy surface. Green and red lines represent the MEP and its projection over the $\xi$, $\xi^{\perp}$ space. Blue and pink lines are the work associated to $\xi$ and $\xi^{\perp}$ respectively.
A: case in which $w_{\xi^{\perp}} \approx 0$ and therefore $\Delta V \approx w_{\xi}$ being $\xi$ a good descriptor of the MEP.
B: case in which $w_{\xi^{\perp}} \neq 0$ and therefore $\Delta V \neq w_{\xi}$ being $\xi$ a bad descriptor of the MEP.

## 2.2 Discrete trajectory calculations and numerical integration.

The first thing to note is that the MEP obtained in any computational atomic level calculation is described by discrete conformations of the system in the configurational space, which does not allow to make an analytical integration of equation 1, being necessary to use a discrete integration.

In this work, two methodologies will be used to estimate those discrete configurations that estimate the MEP. On one hand, the trajectory will be obtained from a NEB calculation minimizing the NEB force with a FIRE[28] algorithm, and on the other hand, the trajectory will be obtained from an AM calculation using a conjugate gradient minimization method on a system to which a $V^{bias}$ is added that acts on a selected degree of freedom $\xi$, according to equation 3.

$$V^{bias}(\xi) = \frac{k}{2}(\xi - \xi_0)^2 \tag{3}$$

In both cases, the estimation of the path integrals in the atomic contribution case was done by a midpoint technique using affine functions among the configurations obtained from the QM - QM/MM calculation (eq. 4) in which $\vec{R}_{i,p}$ and $\vec{F}_{i,p}$ are the position and force of *i-th* atom in the *p-th* configuration obtained by AM or NEB.

$$W_i \approx -\sum_{p=1}^{P} \frac{(\vec{R}_{i,p+1} - \vec{R}_{i,p}) \cdot (\vec{F}_{i,p+1} + \vec{F}_{i,p})}{2} \tag{4}$$

In the AM calculations $w_\xi$ may be estimated easily using the effect of the $V^{bias}$. In each AM optimized configuration $F_\xi + F^{bias}(\xi) \approx 0$ and therefore $w_\xi$ may be obtained by:

$$w_\xi = -\int_{\Gamma_{A\to B}} F_\xi d\xi \approx \int_{\Gamma_{A\to B}} F^{bias}(\xi) d\xi = \int_{\Gamma_{A\to B}} k(\xi - \xi_0) d\xi \tag{5}$$

and finally in the discrete scheme:

$$w_\xi \approx -\sum_{p=1}^{P} \frac{((F_\xi^{bias})_{p+1} + (F_\xi^{bias})_p)(\xi_{p+1} - \xi_p)}{2} \quad (6)$$

The discretization of the integral in both cases (eq. 4 and eq. 6) has associated a quadrature error, which must be analyzed so as not to make a mistake in the interpretation of the results. In this work, this was made including intermediate structures obtained by linear interpolation between the converged structures to verify the correct convergence of the integral with the quadrature, and comparing the total energy variation of the system along obtained by direct calculation of the energy and by IoF of all atoms ($W_{total}$).

## 2.3 Implementation

All the QM and QM/MM calculations were performed using a modified version of the original Hybrid code.[29] Hybrid is a free code (https://github.com/MALBECC/hybrid) designed primarily to obtain potential energy profiles in complex systems.

The new version of Hybrid uses the code LIO[30,31] developed in our group for the computation of energies and forces at QM/MM level. LIO is a very efficient implementation to perform the description of the quantum subsystem based on DFT employing Cartesian Gaussian functions, in which the most expensive parts of the calculation are computed using graphical processing units (GPU). The code has been used for the analysis of different types of systems, in schemes of both nuclear and electronic dynamics.[32–38] A detailed description of the LIO can be found in the article of Marcolongo and collaborators.[31] The description of the classical subsystem together with the Lenard-Jones QM/MM interactions is computed using the AMBER parm99 force field.[39]

In our implementation, using the modified version of the Hybrid code, the potential energy profiles can be obtained by two methodologies: AM and NEB. The implementation based on AM adds to the system a potential of the form of equation 3, in which $\xi$ can be defined in terms of angles, dihedrals, distances or combinations thereof, and then optimizing

the geometry by minimizing the potential energy plus the added $V^{bias}$.

The implementation carried out for the calculations of NEB uses $P + 1$ images of the system to represent the reaction path, with $P$ an integer typically in the 10-100 range. The initial system configuration in each image is obtained by linearly interpolating the positions between the previously optimized images 1 and $P + 1$, being able to use a third intermediate structure proposed for interpolation. Then, images 2 through $P$ are optimized by minimizing the NEB force defined according to:[26,27]

$$\vec{F}_{i,j}^{NEB} = \vec{F}_{i,j}^{\perp} + \vec{F}_{i,j}^{S\parallel} \tag{7}$$

$\vec{F}_{i,j}^{\perp}$ is the component of the force due to the gradient of the potential energy projected in the direction perpendicular to the trajectory (see eq. 8, where "$\cdot$" represents the scalar product between two vectors) and $\vec{F}_{i,j}^{S\parallel}$ is the spring force in the direction parallel to the trajectory $\hat{\tau}_{i,j}$. The sub-index $i$ indicates the image number of the discrete configurations path that describes the transition process (1 - $P+1$) and the subindex $j$ indicates the atom number of the system (1 - $N$).

$$\vec{F}_{i,j}^{\perp} = -\nabla_{i,j}V + (\nabla_{i,j}V \cdot \hat{\tau}_{i,j})\hat{\tau}_{i,j} \tag{8}$$

## 2.4 Calculations details.

All the simulations reported in this study were performed using the Perdew-Burke-Ernzerhof (PBE) generalized gradient approximation functional[40] with DZVP basis sets[41] obtained from the EMSL Basis Set Exchange.[42]

AM calculations were performed exploring the values of the bias potential spring constant and the number of points that were used to perform the integration. No significant variations were obtained in the profiles for spring constants in the range 50-1000 kcal mol$^{-1}$ Å$^{-2}$ and with samples numbers between 50 and 300 (data no shown).

The IoF scheme was performed in all cases by removing the motions associated with translations and rotations of the system (generated by numerical error in forces) in such a way that the position of the center of mass and the directions of the eigenvectors of the inertia tensor were maintained.

# 3 Results and discussion.

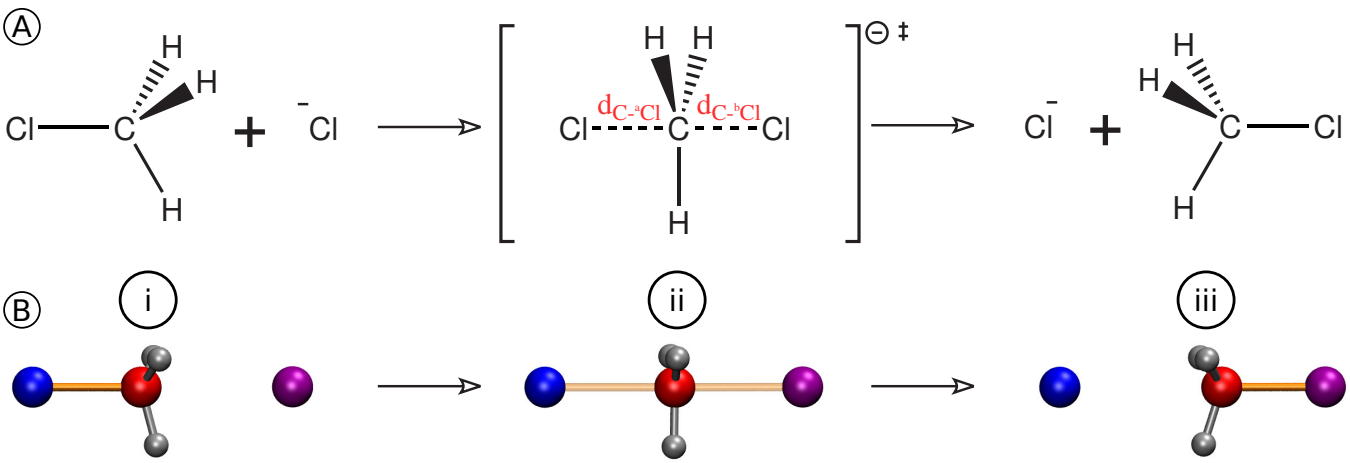
## 3.1 Selection and analysis of a particular degree of freedom to describe a reaction path.

The present work introduces a simple method for the selection and analysis of a particular degree of freedom to describe a reaction path: the energetic variation can be decomposed in contributions (atomic or degree of freedom) obtained by IoF along a reaction path, and from an analysis at the atomic level contributions, an improved reaction coordinate can be proposed to describe the reactive process.

The methodology is exemplified for three possible cases: in subsection 3.1.1 we present an example in which the selected degree $\xi$ is a good descriptor of the trajectory, in subsection 3.1.2 we present an example in which the selected degree of freedom $\xi$ is a less accurate descriptor of the trajectory, and finally in subsection 3.1.3 we present an example in which the selected reaction coordinate $\xi$ does not represent well the trajectory of the system.

### 3.1.1 The selected degree of freedom $\xi$ is a good descriptor of the trajectory.

Firstly, the presented approach is illustrated by using the reaction of $CH_3Cl$ with $Cl^-$, a bimolecular nucleophilic reaction, one of the most studied processes in chemistry (Scheme 1). The reaction is produced by the approach of a $Cl^-$ to the carbon atom in concert with the inversion of $CH_3$ and the increase of the distance of the C atom to the initially bound Cl. This is a nice model system since it has the additional benchmark advantage of being a symmetric reaction.




Scheme 1: A: Mechanism of $S_N2$ reaction of $CH_3Cl$ with $Cl^-$.
B: Obtained structures for the transition process being i), ii) and iii) reactants, transition state and products, respectively.

#### 3.1.1.1 Selection of degree of freedom $\xi$ and AM calculation.

As shown in Scheme 1 for this reaction, one C-Cl bond is formed and another is broken along the reaction. A degree of freedom $\xi$ was selected by combining these two processes, the formation and the breaking of the C and Cl bonds, taking as the reaction coordinate the antisymmetric combination of both distances ($\xi = d_{^aCl-C} - d_{^bCl-C}$).

To model the reaction, AM calculations were performed using a bias potential given by equation 3.The results presented were obtained for a spring constant of 200 kcal mol$^{-1}$ Å$^{-2}$ and a sampling of 100 values in the $\xi = (-2\text{Å}, 2\text{Å})$ range.

Although not presented in this work, the reaction coordinate used is reversible and the system does not show hysteresis in the potential energy profiles.

#### 3.1.1.2 Analysis of $\xi$ as a descriptor for the reaction path by IoF.

The potential energy profile for the reaction of $Cl^-$ with $CH_3Cl$ using AM with $\xi$ is presented in Figure 2A, detailing the profile obtained by direct calculation of the energy (black line), the work done in $\xi$ by the bias potential $w_\xi$ (green line) and the total work $W_{total}$ (orange line).

The superimposition of the profiles obtained by direct calculation of the potential energy (black line) and by the work done by all the atoms (orange line) allows verifying that the numerical integration made is acceptable. For this type of systems, in which all forces are conservative, the integration of the forces along a reaction path, by definition, should be identical to the variation of potential energy computed directly. However, the calculation of the energy and forces has a numerical error, and the integration step is finite, so it is necessary to test the methodology for guaranteeing that the quadrature error is negligible.

The superimposition of the profiles obtained by direct calculation of the potential energy (black line) and the work done in $\xi$ by the bias potential $w_\xi$ (green line) confirms that the chosen degree of freedom $\xi$ is a good descriptor of the reaction path. This reaction has been widely studied and it is known that the antisymmetric combination of the selected distances generates an appropriate reaction path in an AM scheme.[43–45]

#### 3.1.1.3 Verification of $\xi$ as a reaction path descriptor with NEB.

For an additional verification, the trajectory obtained in this calculation (black dots) is used as the initial band for the NEB calculation (pink dots); the result is presented in Figure 2B. It can be shown that (at least in the C$-^{a}$Cl, C$-^{b}$Cl space) both trajectories have practically no differences.

#### 3.1.1.4 The decomposition of the energy into atomic contributions.

In this case, the work associated with the bias potential (green line, Figure 2A) describes well the total energy change of the system (black line, Figure 2A). The decomposition of the energy into atomic contributions (Figure 2C) allows us to understand this fact. In this particular case, the major energy change throughout the reaction is produced by the atoms that are contained in $\xi$, in this case C, $^{a}$Cl and $^{b}$Cl (red, blue and violet lines respectively in Figure 2C), while the contribution of the three H atoms is minor (grey line in Figure 2C).

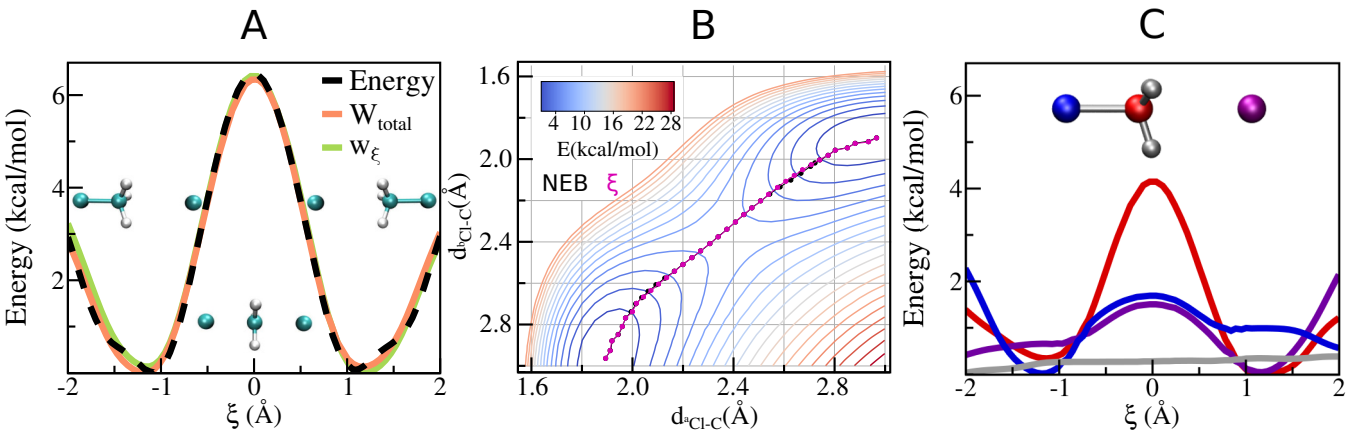



Figure 2: $S_N2$ reaction of $CH_3Cl$ with $Cl^-$
A) Potential energy profile obtained for the $S_N2$ reaction of $Cl^-$ with $CH_3Cl$ using AM with $\xi = d_{^aCl-C} - d_{^bCl-C}$. The results are presented for the profile obtained by direct calculation of the energy (black line), the work done in $\xi$ by the bias potential $w_\xi$ (green line) and the total work (orange line) vs $\xi$.
B) Potential energy surface in the $d_{C-^aCl}$, $d_{C-^bCl}$ space for the $S_N2$ reaction of $Cl^-$ with $CH_3Cl$ in vacuum. Trajectory obtained with AM using $\xi$ (black dots) and NEB (pink dots).
C) Potential energy contributions to $W_{total}$ of each atom for $S_N2$ reaction of $Cl^-$ with $CH_3Cl$ using AM with $\xi$. The result for the C atom is shown with a red line, while the blue and violet lines correspond to $^a$Cl and $^b$Cl atoms respectively. The contribution of the 3 H atoms is very small; the sum of their contributions is shown with a grey line.

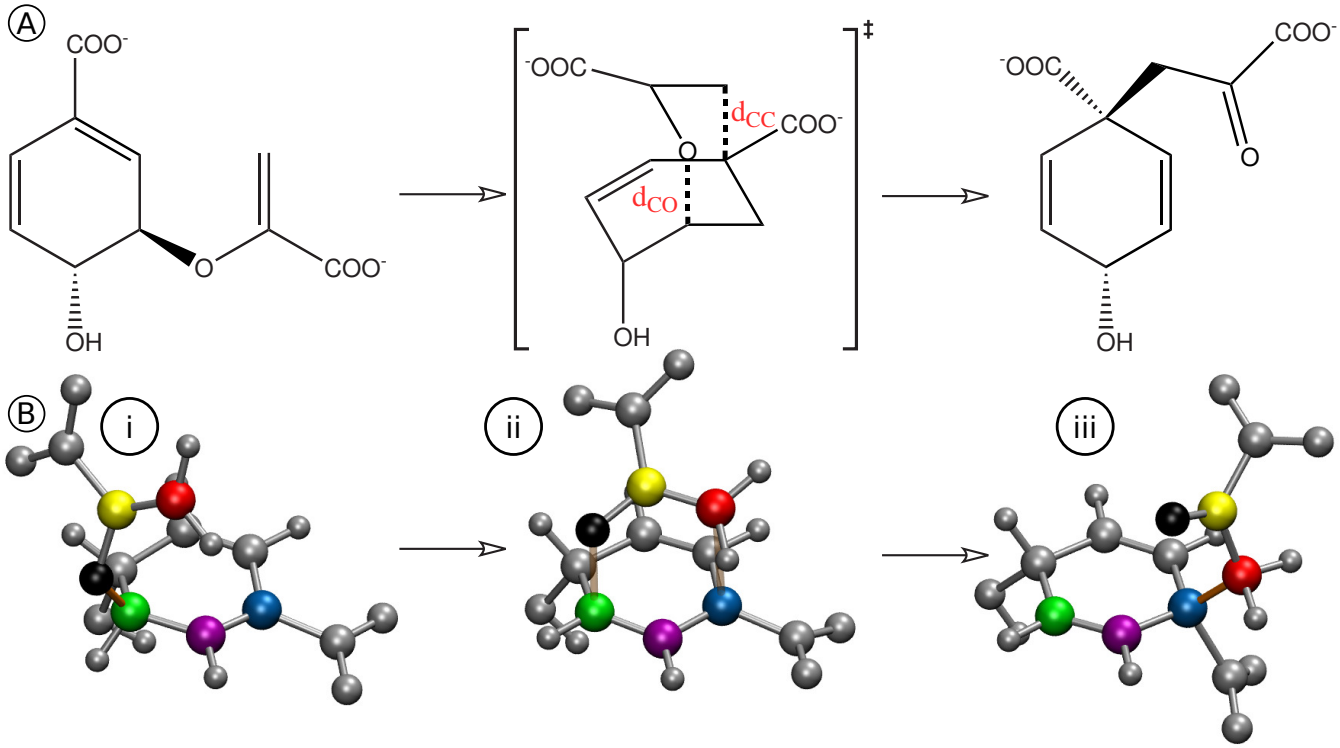
### 3.1.2 The selected degree of freedom $\xi$ is a less accurate descriptor of the trajectory.

Secondly, the methodology presented in this work is illustrated by analyzing the Claisen rearrangement[46] of chorismate to prephenate. The aliphatic Claisen Rearrangement is a [3,3]-sigmatropic rearrangement in which an allyl vinyl ether is converted thermally to an unsaturated carbonyl compound, and illustrated by the equation shown in Scheme 2. It has been found that chorismate adopts several conformations in vacuum and in condensed phase.[47–49] In the reactant conformation, the system adopts a diaxial conformation so that the pericyclic rearrangement occurs to give prephenate.[29]

#### 3.1.2.1 Selection of degree of freedom $\xi$ and AM calculation.

As shown in Scheme 2 for this reaction, a C-C bond is formed and a C-O bond is broken along the reaction. Similar to case 3.1.1), a degree of freedom $\xi$ was selected by combining these two processes, taking as the reaction coordinate the antisymmetric combination of the distances $d_{C-C}$ and $d_{C-O}$ ($\xi = d_{C-C} - d_{C-O}$).

To model the reaction, starting from the optimized diaxial conformer, AM calculations were performed using a bias potential given by equation 3.




Scheme 2: A: Mechanism of conversion between chorismate (left) to prephenate (right). B: Obtained structures for the transition process being i) ii) and iii) reactants, transition state and products, respectively.

#### 3.1.2.2 Analysis of $\xi$ as a descriptor for the reaction path by IoF.

The obtained results are presented in Figure 3A, showing the profiles obtained by direct calculation of the energy (black line), the integration of the forces on all atoms of the system (orange line) and the integration of the force associated with the bias potential (green line).

Unlike the previous system considered in case 3.1.1), in this case the profile obtained by the integration of the force associated with the bias potential (green line) reproduces only qualitatively the values generated by direct calculation of the energy (black line). Therefore, there is a contribution to the energetic change generated by the other degrees of freedom of the system.

#### 3.1.2.3 Verification of $\xi$ as a reaction path descriptor with NEB.

In this case, the re-optimization of the obtained trajectory by a NEB algorithm (Figure 3B) generates a non negligible change in the computed path, especially in the zone of maximum energy. However, this zone is quite flattened area, being the energetic difference practically null in the maximum energy zone for both trajectories.

#### 3.1.2.4 The decomposition of the energy into atomic contributions.

The decomposition of the energetic change in atomic contributions, however, does not present significant values for any particular atom not contained in $\xi$, but the difference is due to the sum of small contributions of several atoms. In this case, the 20 atoms not considered in $\xi$ generate a continuous and small energetic drift along the reaction of $\approx 1$ kcal mol$^{-1}$ Å$^{-1}$ (Figure 3C). In spite of this, the selected reaction coordinate still provides a qualitatively reasonable representation of the reactive path.

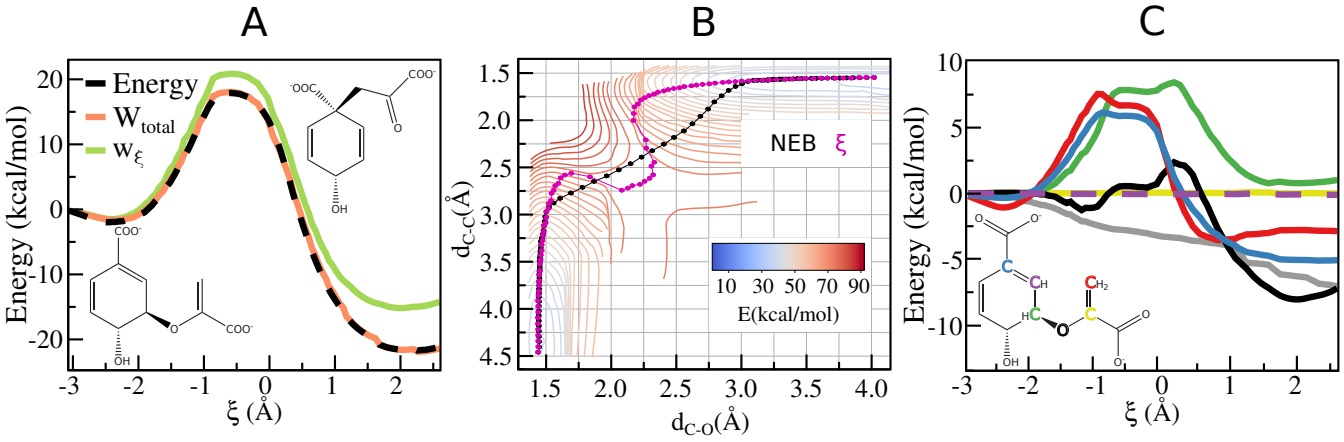



Figure 3: Conversion of chorismate to prephenate
A) Potential energy profile obtained for the conversion reaction of chorismate to prephenate in vacuum using AM with $\xi = d_{C-C} - d_{C-O}$. The results are presented for the profile obtained by direct calculation of the energy (black line), the work done in $\xi$ by the bias potential $w_\xi$ (green line) and the total work (orange line) vs $\xi$.
B) Potential energy surface in the $d_{C-C}$, $d_{C-O}$ space for the conversion reaction of chorismate to prephenate in vacuum. Trajectory obtained with AM using $\xi$ (pink dots) and NEB (black dots).
C) Potential energy contributions to $W_{total}$ of each atom for the conversion reaction of chorismate to prephenate in vacuum using AM with $\xi$. The atoms involved in the pericyclic rearrangements are represented in yellow, green, black, blue, red and purple following the inset color representation. The remainder atoms are depicted using grey lines.

### 3.1.3 The selected degree of freedom $\xi$ does not represent well the trajectory of the system.

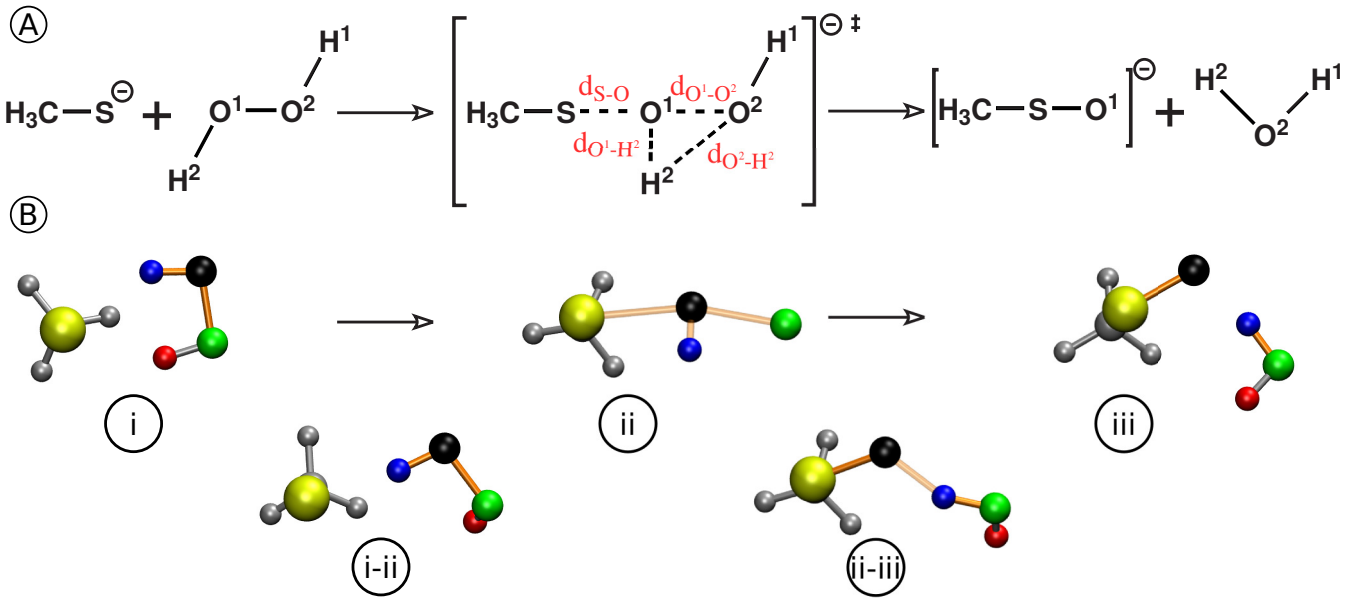
The third system that will be analyzed with the proposed methodology is the reduction of hydrogen peroxide by thiolates, concomitant with thiolates oxidation.[50] The reaction occurs by a $S_N2$ like mechanism in which the sulfur atom attacks one of the oxygen atoms of hydrogen peroxide concomitantly with the breakdown of the O-O bond and the transfer of one of the hydrogen atoms to give water (Scheme 3).

#### 3.1.3.1 Selection of degree of freedom $\xi_1$ and AM calculation.

As shown in Scheme 3 for this reaction, an S-O bond is formed and an O-O bond is broken along the reaction. Similar to case 3.1.1, a degree of freedom $\xi_1$ was selected by combining

these two processes, using the antisymmetric combination of the distances to describe the conversion of reactants to products.

To model this reaction in vacuum, AM calculation was performed using $\xi_1$ to describe the conversion of reactants to products and a bias potential given by equation 3.




Scheme 3: A:Reaction mechanism of methyl thiolate with hydrogen peroxide to give methyl sulphenate and water.
B: Obtained structures for the transition process beeing i), ii) and iii) reactants, transition state and products.i-ii) correspond to an intermediate structure between i) and ii) in which $H^1$ recedes from S atom and rotates modifying the dihedral angle of hidrogen peroxide. ii-iii) correspond to an intermediate structure between ii) and iii) in which the proton transfer occurs from $O^1$ to $O^2$ while the rupture of the $O^1-O^2$ bond and the formation of the $S-O^1$ bond is generated.

#### 3.1.3.2 Analysis of $\xi_1$ as a descriptor for the reaction path by IoF.

The results obtained are presented in Figure 4A, showing the profiles obtained by direct calculation of the energy (black line), the integration of the forces on all atoms of the system (orange line) and the integration of the force associated with the bias potential (green line).

In this case, the difference obtained by direct calculation of the energy (black line) and the IoF associated with the bias potential (green line) is very large, which indicates that the selected reaction coordinate does not represent well the trajectory of the system and orthogonal motions must be considered in the selection of $\xi_1$ to generate a better description of the reaction. The energy profile obtained has a maximum for $\xi_1 \approx -0.6$ Å but also presents a significant discontinuity. This makes it very hard to get a reasonable estimate of

the activation energy of the process.

#### 3.1.3.3 MEP and atomic energy decomposition by IoF over NEB trajectory.

In order to propose a better degree of freedom using the IoF scheme, it is necessary to obtain a trajectory from reactants to products through a path of minimum energy. For this, the initial and final structures of the previous calculation were reoptimized without the bias potential, and a NEB calculation was performed using 50 images generated by linear interpolation with a transition state proposed by chemical criteria. The calculation was performed taking a spring constant of 200 kcal $mol^{-1}$ $Å^{-2}$ and using a convergence criterion of 0.5 kcal $mol^{-1}$ $Å^{-1}$ in NEB forces. Results are presented in Figure 4B.

It can be seen in Figure 4C that along the reaction coordinate the greatest energetic change occurs due to the $H^2$ and $O^2$ atoms (blue and green lines respectively), which are involved in the proton transfer that occurs in the final stages of the reaction. On the other hand, there is a relevant contribution of the work associated with the $H^1$ (red line) in the first stage of the reaction, which is not involved in any bond break/formation. Considering the system structure, it can be seen that the conformational change in this zone is mainly due to the breakdown of the interaction of $H^1$ with the $S^-$, the change in the dihedral angle of the hydrogen peroxide and the separation of $O^2$ and the S atom (as shown in Scheme 3).

#### 3.1.3.4 Selection of an improved degree of freedom $\xi_2$ and AM calculation.

Among the contributions mentioned in previous subsection, the most important omitted contribution is the work associated with atoms $H^2$ and $O^2$.

One simple possibility to include it is to select a new degree of freedom $\xi_2$, according to $\xi_2 = 2d_{O^1-O^2} - d_{S-O^1} - d_{H^2-O^2}$. Following the same protocol as previous sections, AM calculations were performed using $\xi_2$ to describe the conversion of reactants to products.

#### 3.1.3.5 Analysis of $\xi_2$ as a descriptor for the reaction path by IoF.

The energetic profile obtained using this reaction coordinate in an AM scheme is presented in Figure 4D.

It can be seen that the inclusion of the proton transfer in $\xi_2$ generates a much smoother profile, with a well-defined maximum and with a value close to that obtained by NEB. Even so, in this scheme there is an appreciable difference for values of $\xi_2 \approx 1.5$Å related to the breaking of the interaction between $H^1$ and S and the reorganization of hydrogen peroxide to generate a much more aligned state among $O^1$ and S that is not considered in $\xi_2$.

#### 3.1.3.6 Comparison of the trajectory obtained with AM using $\xi_1$, $\xi_2$ and NEB.

Figure 4E shows the differences in the trajectories obtained by $\xi_1$, $\xi_2$ and NEB calculations in the $d_{H^2-O^2} - d_{H^2-O^1}$, $d_{O^1-O^2} - d_{S-O^1}$ space. It can be seen that the $\xi_1$ trajectory exhibits a major discontinuity that corresponds to the proton transfer process, while $\xi_2$ shows a minor discontinuity close to reactive zone ($d_{H^2-O^2} - d_{H^2-O^1} \approx 0.9$ Å, $d_{O^1-O^2} - d_{S-O^1} \approx 1.3$ Å). $\xi_2$ on the other hand goes through the same transition state as NEB at ($d_{H^2-O^2} - d_{H^2-O^1} \approx 0.9$ Å, $d_{O^1-O^2} - d_{S-O^1} \approx 1.1$ Å).

#### 3.1.3.7 Free energy profiles obtained by umbrella sampling using $\xi_1$ and $\xi_2$.

The analysis performed in this case serves to evaluate the correct selection of $\xi$ either to obtain a potential energy profile or as a guide for its use in analysis schemes based on more sophisticated MD based free energy profiles.

Figure 4F shows the free energy profiles obtained by umbrella sampling using $\xi_1$ and $\xi_2$. Due to the fact that both variables are different, we use a normalized variable $\xi_N$ in the x axis defined by the equation 9, in which $i$=1,2 and $R$ and $P$ denotes the value of $\xi_i$ in reactants and products conformations respectively.

$$\xi_N = \frac{\xi_i - \xi_{iR}}{\xi_{iP} - \xi_{iR}} \tag{9}$$

In both cases the profiles have been obtained by molecular dynamics simulations at 300K using a force constant $k$=200 kcal mol$^{-1}$, a sampling time of 5ps with a time step of 0.2fs and $V^{bias}$ separated every 0.1 Å in $\xi_0$ in separate simulation windows.

It can be observed in the profiles obtained when using $\xi_1$ there is a discontinuity in $\xi_N \approx 0.6$. It is generated by the impossibility of combining the results of adjoining umbrella sampling windows. The latter cannot be corrected by the inclusion of additional simulation windows, nor by modifying the value of k, since the proton transfer process is not contained in $\xi$. On the other hand, when using $\xi_2$ the profile obtained is smooth, sampling correctly the nucleophilic attack of S as well as the transfer of $H^1$.

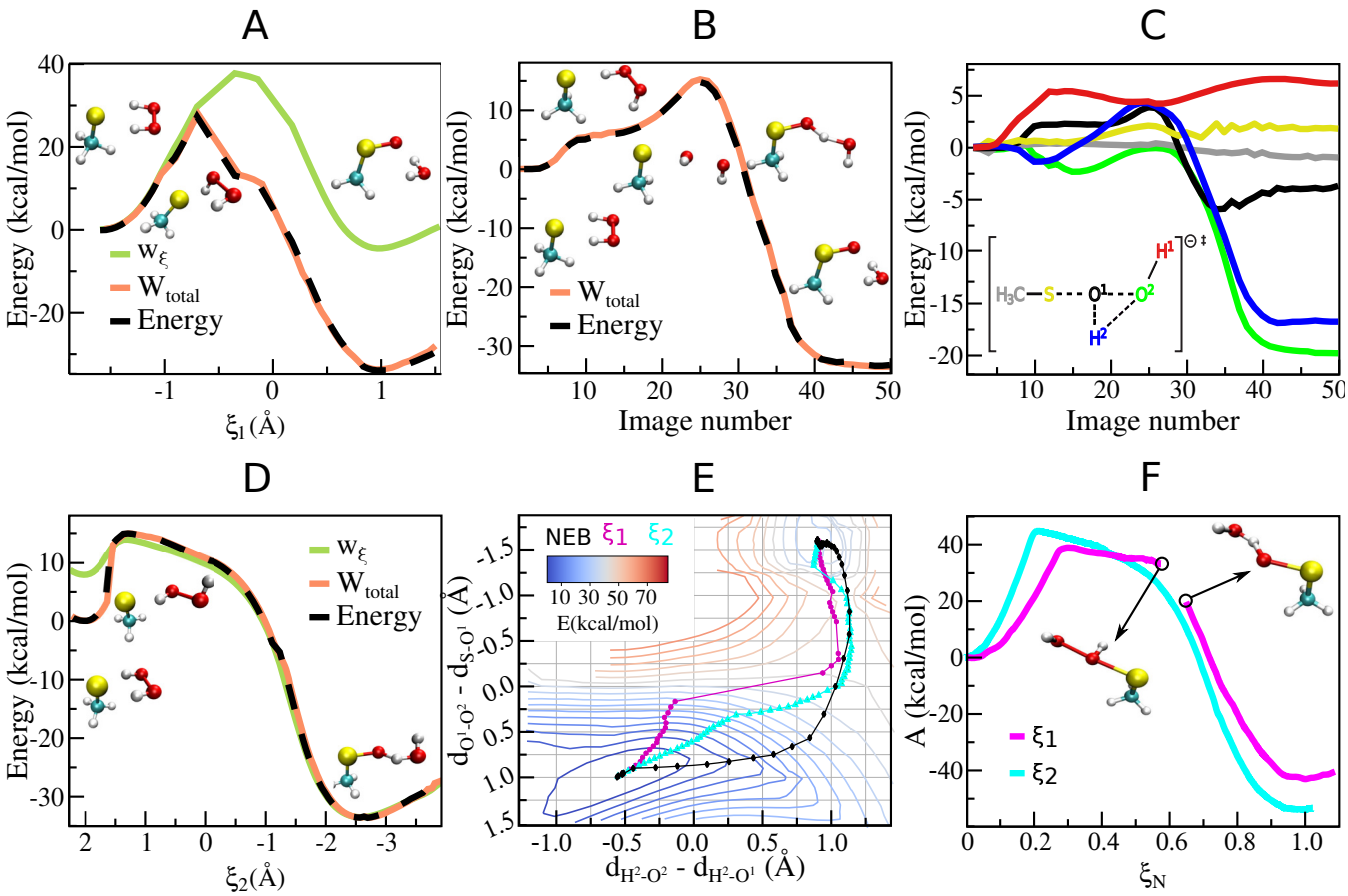



Figure 4: A) Potential energy profile obtained for the reaction of methyl thiolate with hydrogen peroxide using AM with $\xi_1 = d_{O^1-O^2} - d_{S-O^1}$. The results are presented for the profile obtained by direct calculation of the energy (black line), the work done in $\xi_1$ by the bias potential $w_{\xi_1}$ (green line) and the total work (orange line) vs $\xi_1$.
B) Potential energy profile obtained by direct calculation (black line) and by the integration of the forces on all the atoms (orange line) for the energy profile obtained for reaction of methyl thiolate with hydrogen peroxide obtained using NEB.
C) Contribution to $W_{total}$ of each atom is presented for the energy profile obtained for reaction of methyl thiolate with hydrogen peroxide obtained using NEB. The colors correspond to those in the inset, S, $O^1$, $O^2$, $H^2$ and $H^1$ contributions are depicted using yellow, black, green, blue, and red lines, respectively. Because the contribution of the 4 atoms of the $CH_3$ group is much smaller than the rest throughout the calculation, the sum was plotted as a grey line for simplicity.
D) Potential energy profile obtained for the reaction of methyl thiolate with hydrogen peroxide using AM with $\xi_2 = 2d_{O^1-O^2} - d_{S-O^1} - d_{H^2-O^2}$. The results are presented for the profile obtained by direct calculation of the energy (black line) and the work done in $\xi_2$ by the bias potential $w_\xi$ (green line) vs $\xi_2$.
E) Potential energy surface in the $d_{H^2O^2} - d_{H^2O^1}$ , $d_{O^1O^2} - d_{SO^1}$ space for the reaction of methyl thiolate with hydrogen peroxide in vacuum. Trajectory obtained with AM using $\xi_1$ (pink dots), $\xi_2$ (green triangles) and NEB (black dots).
F) Free energy profiles for the reaction of methyl thiolate with hydrogen peroxide obtained by Umbrella Sampling using $\xi_1$ and $\xi_2$ vs a normalized degree of freeom $\xi_N$ defined in eq. 9.

### 3.1.4 Analysis of reaction path conclusions.

In the three cases, the IoF scheme allowed analyzing the quality of the trajectories obtained when selecting a degree of freedom to describe the process, and it allows improving the degree of freedom by a simple analysis of decomposition in atomic contributions.

In the case of the reaction of $Cl^-$ with $CH_3Cl$ (case 3.1.1), $w_\xi$ contains practically all the energy change throughout the reaction ($\Delta V \approx w_\xi$) and therefore the path obtained by AM does not differ from the MEP obtained by NEB, being the selected degree of freedom $\xi$ a good descriptor of the trajectory.

In the case of the reaction of chorismate to prephenate (case 3.1.2), there is a minor contribution, but not negligible, of degrees of freedom other than $\xi$ ($\xi^\perp$), which leads to the trajectory obtained being different from the MEP, however, the energetic barrier difference is small because the potential energy surface is very flat in the region near the state of transition.

In the case of the reaction of $CH_3S^-$ with $H_2O_2$ (case 3.1.2), the work associated with $\xi_1$ differs significantly from the total energy change, the trajectory obtained being very different from the MEP and presenting a strong discontinuity. This degree of freedom can be improved by a simple analysis of decomposition in atomic contributions.

## 3.2 The study of reactions in complex environments.

Analyzing the reactivity of a complex system through optimized structures is a significant advance compared to isolated species analysis.[20] However, correctly describing transition states and evaluating the activation energy in a complex system usually requires an exhaustive conformational analysis. Much work has gone into this endeavor especially when calculating reaction mechanisms in complex systems, such as proteins.[51,52] This section focuses on the methodological analysis of obtaining a single potential energy profile and does not take into account the conformational analysis necessary for exploring relevant initial

configurations.

To exemplify this methodology, two cases are presented, a) one in which it can be seen that the effect of the MM environment of the reactive system does not have an appreciable contribution through work, b) another in which the MM environment has a contribution to the global energy by performing non-negligible work throughout the reactive process.

Two of the previously used reactions in vacuum will be used as models, but in this case, catalyzed by proteins: 3.2.1 the interconversion of chorismate to prephenate in the chorismate mutase of *Bacillus subtilis* (corresponding to the reaction in vacuum shown in section 3.1.2) and 3.2.2 the reduction of hydrogen peroxide by the cystein methyl thyolate in the Alkyl hydroperoxide reductase E (AhpE) of *Mycobacterium tuberculosis* (corresponding to the reaction in vacuum shown in section 3.1.3). The effect of the MM environment on the energy profiles will be evaluated by direct calculation and by decomposing contributions using IoF.

### 3.2.1 The MM environment does not have a contribution to the global energy.

As a model for this case, one of the reactions in vacuum used in the previous section will be used, but in this case, catalyzed by a protein: the conversion of chorismate to prephenate in the chorismate mutase of *Bacillus subtilis* (corresponding to the reaction in vacuum shown in section 3.1.2). This intramolecular transformation is of special interest because it is catalyzed by the enzyme chorismate mutase, part of the biosynthetic pathway of aromatic amino acids in bacteria, fungi and plants leading to the synthesis of tyrosine and phenylalanine.[53]

The conversion reaction of chorismate to prephenate catalyzed by the chorismate mutase of *Bacillus subtilis* was used as a benchmark system. The X-ray crystal structure of the wild type enzyme (PDB ID 1COM)[54] was used as a starting point for the calculations, consistently with previous work.[29,34,55] The protocol used by Ramirez et al. for thermalization and equilibration of the system was followed.[34]

A simulated annealing was carried out before relaxing the system to 0K. The system was

limited to a 30Å environment around the reactive zone and then the geometry was optimized with a conjugate gradient algorithm keeping frozen MM regions further away than 10Å from the reactive zone. On the final structure, the potential energy profile was obtained by AM using the same $\xi$ as in the section 3.1.2.

The obtained result is presented in Figure 5. In Figure 5A, the total variation of energy is presented together with the discriminated contributions of QM and MM atoms by IoF scheme (orange, light blue and pink lines respectively). In Figure 5B, the decomposition of the energy obtained for the most relevant QM atoms is presented.

It can be seen that the energetic barrier associated to the reactive process is smaller in the protein than in vacuum, consistently with the expected protein catalytic effect.

The QM and MM contribution decomposition shows that, in this case, the MM region does not generate an appreciable energetic change to the system throughout the reaction via atomic work (pink line in Figure 5A) so the effect of the protein environment in the catalytic process is due to the substrate-protein interaction.

The differences in the activation barrier can be analyzed in terms of atomic work decomposition from only the QM atoms:

i) The energetic barrier associated with all atoms involved in bond breakage/formation (C(blue), C(red), C(green) and O(grey)) is smaller in the protein case (Figure 5B) compared to the vacuum system (Figure 3C), and the other 2 atoms involved in the pericyclic rearrangement (C(yellow) and C(purple)) have a negligible contribution in both cases.

ii) Finally, in the vacuum system all atoms not involved in the pericyclic rearrangement (black line Figure 3C) generate a small energetic stabilization of the system along the reaction ($\approx 1$ kcal mol$^{-1}$Å$^{-1}$), but in the protein case (black dotted line in Figure 5B) the same atoms generate a small destabilization because in this case the reactant molecule has a rigid protein environment (as shown in Figure 5C).

It can be seen that the effect of the environment of the reactive system does not have an appreciable contribution through work, however long-range interactions generate an signifi-

cant reduction of the energy barrier associated with the 4 QM atoms more relevant in the process by stabilizing the formation of the C-C bond and the breaking of the C-O bond. It is also interesting to note that in the case of the other 20 atoms of the reactive molecule, the small displacements throughout the reaction have a positive contribution to the global energy of the system, which is due to the rigid protein environment.

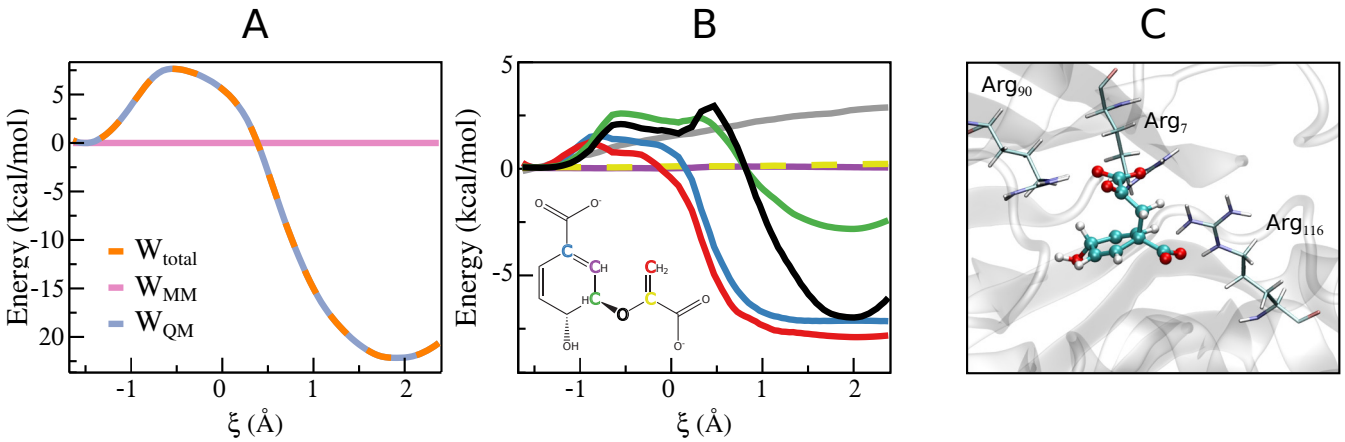



Figure 5: potential energy profile obtained for the reaction from chorismate to prephenate catalyzed by chorismate mutase of *Bacillus subtilis* using AM with $\xi = d_{C-C} - d_{C-O}$.
A) Total potential energy profile (orange) and contributions of QM (light blue) and MM subsystems (pink).
B) Contribution of the atoms involved in the pericyclic rearrangement are represented in yellow, green, black, blue, red and purple following the inset color representation. The remaining QM atoms depicted as grey lines.
C) Respresentative conformation of the reaction site.

### 3.2.2 The MM environment has a contribution to the global energy.

As a model for this case, another of the reactions in vacuum used in the previous section will be used, but in this case, catalyzed by a protein: the reduction of hydrogen peroxide by the cystein methyl thyolate in the Alkyl hydroperoxide reductase E (AhpE) of *Mycobacterium tuberculosis* (corresponding to the reaction in vacuum shown in section 3.1.3). The molecular basis of the reaction mechanism has been analyzed by Zeida and collaborators in aqueous solution[50] and in 1-Cys Prx AhpE from Mt.[35] The initial structure used for the calculations was the dimer of Mt AhpE in its reduced state (PDBid: 1XXU).[56] The preparation and equilibration of the system was done following the protocol presented in reference 35. The equilibrated structure was treated in the same way as in the previous case to obtain the initial structure of the reactants.

The structures of the products to perform the NEB calculation were obtained by a first geometry optimization with a quadratic potential using $\xi = d_{O^1-O^2} - d_{S-O^1}$ and subsequent

optimization by removing the bias potential. The initial band was generated with these two structures and an intermediate state similar to the transition state obtained in vacuum. Finally, the NEB calculation was performed on the system with a spring constant of 200 kcal $mol^{-1}$ $Å^{-2}$ and a convergence criterion in to NEB forces of 0.5 kcal $mol^{-1}$ $Å^{-1}$ allowing only the atoms that were less than 8Å from the QM region to move.

The energetic contributions obtained by the IoF scheme are presented in Figure 6. The total variation of energy is presented In Figure 6A, together with the discriminated contributions of QM and MM atoms (orange, light blue and pink lines respectively). In this case, it can be observed that the main energy change is produced by the QM atoms (grey line), while the MM environment exhibits a non-negligible contribution via atomic work (pink line) along the reaction that destabilizes the system at the onset of the reaction and stabilizes the products state.

The decomposition of the energy obtained for the most relevant QM atoms is presented in Figure 6B. If the atomic contributions of the QM sub-system are compared with the previous calculation in vacuum (Figure 4C), several differences can be observed:

i) In the first place, the $H^2$ contribution (blue line) is lower in the protein case for the stabilization of products, while the $O^2$ (green line) is negligible along the reaction, which suggests that the process of $H^2$-$O^2$ bond formation has a minor contribution in the protein reaction.

ii) On the other hand, the $O^1$ contribution has a similar barrier, and S and $O^1$ have a greater stabilization of the products in the protein case being in this way the formation of the S-O bond more important in the stabilization of products for the protein case.

iii) Finally, the hydrogen $H^1$ in this case does not present an appreciable contribution to the energy change at the onset of the reaction. This is due to the initial conformation of the reactants generated by the protein environment, in which the hydrogen peroxide does not establish the $H^1$-S interaction (inset of Figure 4D).

In this model case, the MM environment has a contribution to the global energy by

performing non-negligible work throughout the reactive process. This generates an increase in the total energy of the system in the zone of the transition state and a stabilization of the products.

The destabilizing effect in this case is strongly compensated by a significant decrease in the work associated with the QM atoms of the system. The reaction in protein occurs without any positive contribution to the energy by atoms S and $H^1$. The energy of the transition state (Figure 6B) is $\approx$ 6 kcal/mol less with respect to the reactant state compared to gas phase reaction (Figure 4C). A differentiated stabilization in the state of products can also be observed, in the vacuum system, the majority of the contribution being that of the $H^2$ and $O^2$ atoms corresponding to the proton transfer process, while in the protein-catalyzed reaction, that of the $O^1$ atom involved in the formation of the S-O bond and on the breaking of the $O^1$-$O^2$ bond are more prominent.

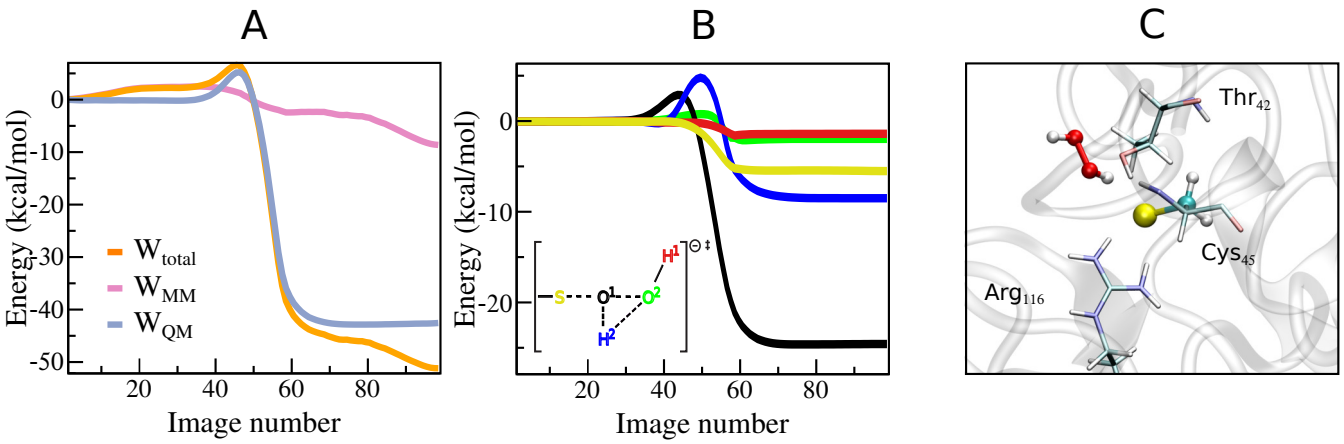



Figure 6: potential energy profile obtained for the reaction of hydrogen peroxide with 1-Cys Prx alkyl hydroperoxide reductase E from *Mycobacterium tuberculosis* obtained using NEB. A) Total potential energy profile (orange) and contributions of QM (light blue) and MM subsystems (pink).
B) Contribution of the most relevant QM atoms in the reaction being, S, $O^1$, $O^2$, $H^2$ and $H^1$ represented with yellow, black, green, blue and red lines, respectively. The MM region up to 8 Å with respect to any QM atom is allowed to move and the rest is frozen.
C) Representative conformation of the reaction site.

# 4 Concluding remarks

The present work presents a computationally economical scheme for estimating the quality of a reaction coordinate before using it in more expensive free energy calculations. The simple IoF scheme allows decomposing and analyzing the potential energy profiles in chemical reactions along a reaction path.

Although the analysis of the trajectories using potential energy profiles does not contemplate the thermal and entropic fluctuations necessary to obtain detailed kinetic and thermodynamic information, this analysis allows generating a quick and simple analysis of the degree of freedom used to describe the reactive process. This kind of analysis is useful in narrow reaction tube processes, for which the PES is similar to the free energy surface and is much more computationally efficient than schemes based on molecular dynamics.

In the case of AM calculations, it is possible to analyze the energetic change associated with the corresponding degree of freedom used to describe the reaction progress in order to evaluate the quality of the calculation. If the major energy change occurs in the selected degree of freedom, the path obtained will be reasonably good in describing the reactive process. However, if there are other degrees of freedom with significant contributions to the energy changes, the description of the reaction path may be a flawed representation of the process.

In the case of the energetic profiles obtained by NEB calculations, the decomposition of the energy in atomic contributions allows evaluating the most important component to be considered as a reaction coordinate. This method has the advantage of being far less expensive than free energy analysis based in MD so it could be used as a first estimator of reaction coordinate adequacy.

An important feature of the IoF scheme is that it allows decomposing the energy changes into different contributions (atomic or associated to a degree of freedom), which are useful for obtaining microscopic information from the calculation in a subsequent analysis of the

trajectory obtained. The proposed procedure is able to identify which atoms are contributing mostly to the chemical reaction and, therefore, which atoms should be involved in the selected degree of freedom. This is helpful for refining a suboptimal reaction coordinate.

The present methodology can be extended to analyze multiple trial $\xi$ variables at no additional cost by a unique NEB trajectory making the correct variable transformations under translational-rotational invariance in order to compute the work associated with $\xi$. Furthermore, the degree of freedom selected to analyze the transition process could be extended to non-spatial variables such as the charge of an atom or a subsystem to study charge transfer of the electronic fundamental state during a reaction.

Finally, the proposed methodology may also be extended for understanding chemical processes at a molecular level by using a similar approach of integrating the free energy gradient along the minimum free energy path.

# Acknowledgement

This research was supported by grants of the Universidad de Buenos Aires, UBACYT 20020170100043BA and Agencia Nacional de Promoción Científica y Tecnológica, PICT 2015-2761, PICT 2014-1022 and CONICET (Grant 11220150100303CO). The calculations have been performed in CECAR (Centro de Computación de Alto Rendimiento), Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires. Authors acknowledge Dr. Damián Scherlis and the reviewers for valuable and constructive suggestions of this research work. Nicolás O. Foglia gratefully acknowledges CONICET for a fellowship, and Dr. Daisaku Ikeda and SGIAR for their support in his academic formation.

# References


(1) Li, W.; Ma, A. Recent developments in methods for identifying reaction coordinates. *Mol. Simulat.* **2014**, *40*, 784–793.

(2) Rosta, E.; Woodcock, H. L.; Brooks, B. R.; Hummer, G. Artificial reaction coordinate "tunneling" in free-energy calculations: The catalytic reaction of RNase H. *J. Comput. Chem.* **2009**, *30*, 1634–1641.

(3) Peters, B. Reaction Coordinates and Mechanistic Hypothesis Tests. *Annu. Rev. Phys.* **2016**, *67*, 669–690.

(4) Bowman, G.; Pande, V.; Noé, F. *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*; Advances in Experimental Medicine and Biology; Springer Netherlands, 2013; Vol. 797.

(5) Buchete, N.-V.; Hummer, G. Coarse Master Equations for Peptide Folding Dynamics. *J. Phys. Chem. B.* **2008**, *112*, 6057–6069.

(6) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **2011**, *134*, 174105.

(7) Prinz, J.-H.; Chodera, J. D.; Noé, F. Spectral Rate Theory for Two-State Kinetics. *Phys. Rev. X* **2014**, *4*, 011020.

(8) Coifman, R. R.; Kevrekidis, I. G.; Lafon, S.; Maggioni, M.; Nadler, B. Diffusion Maps, Reduction Coordinates, and Low Dimensional Representation of Stochastic Systems. *Multiscale Model. Sim.* **2008**, *7*, 842–864.

(9) Zheng, W.; Qi, B.; Rohrdanz, M. A.; Caflisch, A.; Dinner, A. R.; Clementi, C. Delineation of Folding Pathways of a $\beta$-Sheet Miniprotein. *J. Phys. Chem. B.* **2011**, *115*, 13065–13074.

(10) Ferguson, A.; Panagiotopoulos, A.; Kevrekidis, I.; Debenedetti, P. Nonlinear dimensionality reduction in molecular simulation: The diffusion map approach. *Chem. Phys. Lett.* **2011**, *509*, 1–11.

(11) Ledbetter, P. J.; Clementi, C. A new perspective on transition states: $\Xi_1$ separatrix. *J. Chem. Phys.* **2011**, *135*, 044116.

(12) Geissler, P. L.; Dellago, C.; Chandler, D. Kinetic Pathways of Ion Pair Dissociation in Water. *J. Phys. Chem. B.* **1999**, *103*, 3706–3710.

(13) E, W.; Ren, W.; Vanden Eijnden, E. Transition pathways in complex systems: Reaction coordinates, isocommittor surfaces, and transition tubes. *Chem. Phys. Lett.* **2005**, *413*, 242–247.

(14) Du, R.; Pande, V. S.; Grosberg, A. Y.; Tanaka, T.; Shakhnovich, E. S. On the transition coordinate for protein folding. *J. Chem. Phys.* **1998**, *108*, 334–350.

(15) Zinovjev, K.; Tuñón, I. Reaction coordinates and transition states in enzymatic catalysis. *WIREs Computat. Mol. Sci.* **2018**, *8*, e1329.

(16) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. String method in collective variables: Minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.* **2006**, *125*, 024106.

(17) Torrie, G.; Valleau, J. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187 – 199.

(18) Comer, J.; Gumbart, J. C.; Hénin, J.; Lelièvre, T.; Pohorille, A.; Chipot, C. The Adaptive Biasing Force Method: Everything You Always Wanted To Know but Were Afraid To Ask. *J. Phys. Chem. B.* **2015**, *119*, 1129–1151.

(19) Laio, A.; Gervasio, F. L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.* **2008**, *71*, 126601.

(20) Klähn, M.; Braun-Sand, S.; Rosta, E.; Warshel, A. On Possible Pitfalls in ab Initio Quantum Mechanics/Molecular Mechanics Minimization Approaches for Studies of Enzymatic Reactions. *J. Phys. Chem. B.* **2005**, *109*, 15645–15650.

(21) Theodor, A. *C. L. Brooks III, M. Karplus, B. M. Pettitt. Proteins: A Theoretical Perspective of Dynamics, Structure and Thermodynamics, Volume LXXI, in: Advances in Chemical Physics*, 2nd ed.; John Wiley & Sons: England, 1988.

(22) McGibbon, R. T.; Husic, B. E.; Pande, V. S. Identification of simple reaction coordinates from complex dynamics. *J. Chem. Phys.* **2017**, *146*, 044109.

(23) Quapp, W. Chemical reaction paths and calculus of variations. *Theor. Chem. Account* **2008**, *121*, 227–237.

(24) Jiang, Y.; Xue, Y.; Zeng, Y. Microsolvated Model for the Kinetics and Thermodynamics of Glycosidic Bond Dissociative Cleavage of Nucleoside D4G. *J. Phys. Chem. B.* **2018**, *122*, 1816–1825.

(25) Eurenius, K. P.; Chatfield, D. C.; Brooks, B. R.; Hodoscek, M. Enzyme mechanisms with hybrid quantum and molecular mechanical potentials. I. Theoretical considerations. *Int. J. Quantum Chem.* **1996**, *60*, 1189–1200.

(26) Jónsson, H.; Mills, G.; Jacobsen, K. W. *Classical and Quantum Dynamics in Condensed Phase Simulations*; World Scientific: New Jersey/London, 1998; pp 385–404.

(27) Henkelman, G.; Jóhannesson, G.; Jónsson, H. *Theoretical methods in condensed phase chemistry*; Springer: Dordrecht, Vol. 5; pp 269–302.

(28) Bitzek, E.; Koskinen, P.; Gähler, F.; Moseler, M.; Gumbsch, P. Structural Relaxation Made Simple. *Phys. Rev. Lett.* **2006**, *97*, 170201.

(29) Crespo, A.; Scherlis, D. A.; Martí, M. A.; Ordejón, P.; Roitberg, A. E.; Estrin, D. A. A DFT-Based QM-MM Approach Designed for the Treatment of Large Molecular Systems: Application to Chorismate Mutase. *J. Phys. Chem. B.* **2003**, *107*, 13728–13736.

(30) Nitsche, M. A.; Ferreria, M.; Mocskos, E. E.; Lebrero, M. C. G. GPU Accelerated Implementation of Density Functional Theory for Hybrid QM/MM Simulations. *J. Chem. Theory Comput.* **2014**, *10*, 959–967.

(31) Marcolongo, J. P.; Zeida, A.; Semelak, J. A.; Foglia, N. O.; Morzan, U. N.; Estrin, D. A.; González Lebrero, M. C.; Scherlis, D. A. Chemical Reactivity and Spectroscopy Explored From QM/MM Molecular Dynamics Simulations Using the LIO Code. *Front. Chem.* **2018**, *6*, 70.

(32) Morzan, U. N.; Ramírez, F. F.; Oviedo, M. B.; Sánchez, C. G.; Scherlis, D. A.; Lebrero, M. C. G. Electron dynamics in complex environments with real–time time dependent density functional theory in a QM–MM framework. *J. Chem. Phys.* **2014**, *140*, 164105.

(33) Cuevasanta, E.; Zeida, A.; Carballal, S.; Wedmann, R.; Morzan, U. N.; Trujillo, M.; Radi, R.; Estrin, D. A.; Filipovic, M. R.; Alvarez, B. Insights into the mechanism of the reaction between hydrogen sulfide and peroxynitrite. *Free Radic. Biol. Med.* **2015**, *80*, 93 – 100.

(34) Ramírez, C. L.; Zeida, A.; Jara, G. E.; Roitberg, A. E.; Martí, M. A. Improving Efficiency in SMD Simulations Through a Hybrid Differential Relaxation Algorithm. *J. Chem. Theory Comput.* **2014**, *10*, 4609–4617.

(35) Zeida, A.; Reyes, A. M.; Lichtig, P.; Hugo, M.; Vazquez, D. S.; Santos, J.; González Flecha, F. L.; Radi, R.; Estrin, D. A.; Trujillo, M. Molecular Basis of Hy-

droperoxide Specificity in Peroxiredoxins: The Case of AhpE from Mycobacterium tuberculosis. *Biochemistry* **2015**, *54*, 7237–7247.

(36) Foglia, N. O.; Morzan, U. N.; Estrin, D. A.; Scherlis, D. A.; Gonzalez Lebrero, M. C. Role of Core Electrons in Quantum Dynamics Using TDDFT. *J. Chem. Theory Comput.* **2017**, *13*, 77–85.

(37) Boubeta, F. M.; Bari, S. E.; Estrin, D. A.; Boechi, L. Access and Binding of H2S to Hemeproteins: The Case of HbI of Lucina pectinata. *J. Phys. Chem. B.* **2016**, *120*, 9642–9653.

(38) Bringas, M.; Semelak, J.; Zeida, A.; Estrin, D. A. Theoretical investigation of the mechanism of nitroxyl decomposition in aqueous solution. *J. Inorg. Biochem.* **2016**, *162*, 102 – 108.

(39) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049–1074.

(40) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple Erratum Phys. Rev. Lett. 77, 3865 (1996). *Phys. Rev. Lett.* **1997**, *78*, 1396–1396.

(41) Godbout, N.; Salahub, D. R.; Andzelm, J.; Wimmer, E. Optimization of Gaussian–type basis sets for local spin density functional calculations. Part I. Boron through neon, optimization technique and validation. *Can. J. Chemistry* **1992**, *70*, 560–571.

(42) EMSL Basis Set Exchange. `https://bse.pnl.gov`, Accessed: 2016-01-30.

(43) Morokuma, K. Potential energy surface of the SN2 reaction in hydrated clusters. *J. Am. Chem. Soc.* **1982**, *104*, 3732–3733.

(44) Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. Theoretical examination of the SN2 reaction involving chloride ion and methyl chloride in the gas phase and aqueous solution. *J. Am. Chem. Soc.* **1985**, *107*, 154–163.

(45) Chandrasekhar, J.; Jorgensen, W. L. Energy profile for a nonconcerted SN2 reaction in solution. *J. Am. Chem. Soc.* **1985**, *107*, 2974–2975.

(46) Brickel, S.; Meuwly, M. Molecular Determinants for Rate Acceleration in the Claisen Rearrangement Reaction. *J. Phys. Chem. B* **2019**, *123*, 448–456.

(47) Galopin, C. C.; Zhang, S.; Wilson, D. B.; Ganem, B. On the mechanism of chorismate mutases: Clues from wild-type E. coli enzyme and a site-directed mutant related to yeast chorismate mutase. *Tetrahedron Lett.* **1996**, *37*, 8675 – 8678.

(48) Wiest, O.; Houk, K. N. Stabilization of the Transition State of the Chorismate-Prephenate Rearrangement: An ab Initio Study of Enzyme and Antibody Catalysis. *J. Am. Chem. Soc.* **1995**, *117*, 11628–11639.

(49) Martí, S.; Andrés, J.; Moliner, V.; Silla, E.; Tuñón, I.; Bertrán, J. A QM/MM Study of the Conformational Equilibria in the Chorismate Mutase Active Site. The Role of the Enzymatic Deformation Energy Contribution. *J. Phys. Chem. B.* **2000**, *104*, 11308–11315.

(50) Zeida, A.; Babbush, R.; González Lebrero, M. C.; Trujillo, M.; Radi, R.; Estrin, D. A. Molecular Basis of the Mechanism of Thiol Oxidation by Hydrogen Peroxide in Aqueous Solution: Challenging the SN2 Paradigm. *Chem. Res. Toxicol.* **2012**, *25*, 741–746.

(51) Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.

(52) Ranaghan, K. E.; Mulholland, A. J. Investigations of enzyme-catalysed reactions with

combined quantum mechanics/molecular mechanics (QM/MM) methods. *Int. Rev. Phys. Chem.* **2010**, *29*, 65–133.

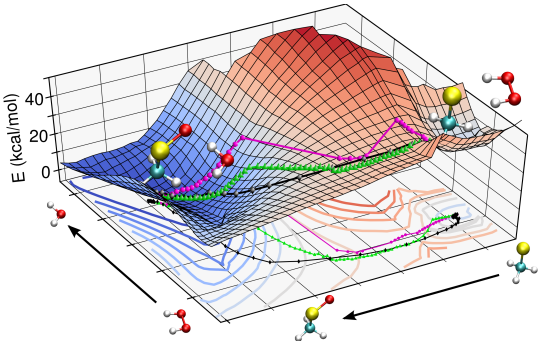
(53) Ranaghan, K.; Shchepanovska, D.; Bennie, S.; Lawan, N.; Macrae, S.; Zurek, J.; Manby, F.; Mulholland, A. Projector-Based Embedding Eliminates Density Functional Dependence for QM/MM Calculations of Reactions in Enzymes and Solution. *J. Chem. Inf. Model.* **2019**, *59*, 2063–2078.

(54) Chook, Y. M.; Gray, J. V.; Ke, H.; Lipscomb, W. N. The Monofunctional Chorismate Mutase from Bacillus subtilis: Structure Determination of Chorismate Mutase and Its Complexes with a Transition State Analog and Prephenate, and Implications for the Mechanism of the Enzymatic Reaction. *J. Mol. Biol.* **1994**, *240*, 476 – 500.

(55) Crespo, A.; Martí, M. A.; Estrin, D. A.; Roitberg, A. E. Multiple-Steering QM-MM Calculation of the Free Energy Profile in Chorismate Mutase. *J. Am. Chem. Soc.* **2005**, *127*, 6940–6941.

(56) Li, S.; Peterson, N. A.; Kim, M.-Y.; Kim, C.-Y.; Hung, L.-W.; Yu, M.; Lekin, T.; Segelke, B. W.; Lott, J. S.; Baker, E. N. Crystal Structure of AhpE from Mycobacterium tuberculosis, a 1-Cys Peroxiredoxin. *J. Mol. Biol.* **2005**, *346*, 1035 – 1046.

**For table of contents use only**




**Title**: Reaction path analysis from potential energy contributions using forces: an accessible estimator of reaction coordinate adequacy

**Authors**: Nicolás O. Foglia, Mariano C. Gonzalez Lebrero, Rodolfo R. Biekofsky, Dario A. Estrin