Mapping Enzymatic Catalysis Using the Effective Fragment Molecular Orbital Method: Towards all ab initio Biochemistry

Casper Steinmann¹, Dmitri G. Fedorov², Jan H. Jensen¹*

¹ Department of Chemistry, University of Copenhagen, Copenhagen, Denmark, ² Nanosystems Research Institute (NRI), National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Ibaraki, Japan

Abstract

We extend the Effective Fragment Molecular Orbital (EFMO) method to the frozen domain approach where only the geometry of an active part is optimized, while the many-body polarization effects are considered for the whole system. The new approach efficiently mapped out the entire reaction path of chorismate mutase in less than four days using 80 cores on 20 nodes, where the whole system containing 2398 atoms is treated in the ab initio fashion without using any force fields. The reaction path is constructed automatically with the only assumption of defining the reaction coordinate a priori. We determine the reaction barrier of chorismate mutase to be $18.3 \pm 3.5$ kcal mol$^{-1}$ for MP2/cc-pVDZ and $19.3 \pm 3.6$ for MP2/cc-pVTZ in an ONIOM approach using EFMO-RHF/6-31G(d) for the high and low layers, respectively.

Introduction

Fragment-based quantum mechanical methods [1–12] are becoming increasingly popular [13], and have been used to describe a very diverse set of molecular properties for large systems. Although these methods have been applied to refine the energetics of some enzymatic reactions [14,15] they are usually not efficient enough to allow for many hundreds of single point calculations needed to map out a reaction path for a system containing thousands of atoms, although geometry optimizations of large systems can be performed for systems consisting of several hundreds of atoms [8,9,11,16–18]. In fact, typical applications of fragment-based methods to biochemical systems, for example, to protein-ligand binding [19], are based on performing a few single point calculations for structures obtained at a lower level of theory (such as with force fields). Although many force fields are well tuned to treat typical proteins, for ligands they can be problematic.

In this work we extend the effective fragment molecular orbital (EFMO) method [20,21] into the frozen domain (FD) formalism [18], originally developed for the fragment molecular orbital (FMO) method [22–25]. For FMO, there is also the partial energy gradient method [26].

EFMO is based on dividing a large molecular system into fragments and performing ab initio calculations of fragments and their pairs, and combining their energies in the energy of the whole system (see more below). In the FD approach we employ here, one defines an active region associated with the active site, and the cost of a geometry optimization is then essentially given by the cost associated with the active region.

However, unlike the quantum-mechanical/molecular mechanical (QM/MM) method [27] with non-polarizable force fields, the polarization of the whole system is accounted for in the FMO and EFMO methods: in the former via the explicit polarizing potential and in the latter via fragment polarizabilities. Another important difference between EFMO and QM/MM is that the former does not involve force fields, and the need to elaborately determine parameters for ligands does not exist in EFMO. Also, in EFMO all fragments are treated with quantum mechanics, and the problem of the active site size [28] does not arise.

The paper is organized as follows: First, we derive the EFMO energy and gradient expressions for the frozen domain approach, when some part of the system is frozen during the geometry optimization. Secondly, we predict the reaction barrier of barrier of the conversion of chorismate to prephenate (Figure 1) in chorismate mutase. The reaction has been studied previously using conventional QM/MM techniques [29–40]. The EFMO method is similar in spirit to QM/MM in using a cheap model for the less important part of the system and the mapping is accomplished with a reasonable amount of computational resources (four days per reaction path using 80 CPU cores). Finally we summarize our results and discuss future directions.

Competing Interests: The authors have declared that no competing interests exist.
* E-mail: jhjensen@chem.ku.dk
Background and Theory

The EFMO energy of a system of \( N \) fragments (monomers) is given by

\[
E^{\text{EFMO}} = \sum_{I} E_{I}^{0} + \sum_{IJ} \left( \Delta E_{IJ}^{0} - E_{IJ}^{\text{POL}} \right) + \sum_{R_{IJ} > R_{\text{resdim}}} E_{IJ}^{\text{ES}} + E_{\text{tot}}^{\text{POL}},
\]

(1)

where \( E_{I}^{0} \) is the gas phase energy of monomer \( I \). The second sum in equation 1 is the pairwise correction to the monomer energy and only applies for pairs of fragments (dimers) separated by an interfragment distance \( R_{IJ} \) (defined previously [20]) less than a threshold \( R_{\text{resdim}} \). The correction for dimer \( IJ \) is

\[
\Delta E_{IJ}^{0} = E_{IJ}^{0} - E_{I}^{0} - E_{J}^{0}.
\]

(2)

\( E_{IJ}^{\text{POL}} \) and \( E_{IJ}^{\text{ES}} \) are the classical pair polarization energy of dimer \( IJ \) and the classical total polarization energy, respectively. Both energies are evaluated using the induced dipole model [41,42] based on distributed polarizabilities [43]. The final sum over \( E_{IJ}^{\text{ES}} \) is the classical electrostatic interaction energy and applies only to dimers separated by a distance greater than \( R_{\text{resdim}} \). These energies are evaluated using atom-centered multipole moments through quadrupoles [44]. The multipole moments and distributed polarizabilities are computed on the fly for each fragment [20,21].

In cases where only part of a molecular system is to be optimized by minimizing the energy, equation 1 can be rewritten, resulting in a method conceptually overlapping with QM/MM in using a cheap model for the less important part of the system. Consider a system \( S \) (illustrated on Figure 2) where we wish to optimize the positions of atoms in region \( A \), while keeping the atoms in region \( b \) and \( F \) frozen (the difference between \( b \) and \( F \) will be discussed below). With this definition, we rewrite the EFMO energy as

\[
E^{\text{EFMO}} = E_{F}^{0} + E_{b}^{0} + E_{A}^{0} + E_{F/b}^{0} + E_{F/A}^{0} + E_{A/b}^{0} + E_{\text{tot}}^{\text{POL}},
\]

(3)

where \( E_{A}^{0} \) is the internal energy of region \( A \). Region \( A \) is made of fragments containing atoms whose position is optimized, and \( A \) can also have some frozen atoms.

\[
E_{A}^{0} = \sum_{I \in A} E_{I}^{0} + \sum_{J \in A} \left( \Delta E_{IJ}^{0} - E_{IJ}^{\text{POL}} \right) + \sum_{R_{IJ} > R_{\text{resdim}}} E_{IJ}^{\text{ES}}.
\]

(4)

Similarly, \( E_{b}^{0} \) is the internal energy of \( b \)

\[
E_{b}^{0} = \sum_{I \in b} E_{I}^{0} + \sum_{J \in b} \left( \Delta E_{IJ}^{0} - E_{IJ}^{\text{POL}} \right) + \sum_{R_{IJ} > R_{\text{resdim}}} E_{IJ}^{\text{ES}}.
\]

(5)

Region \( A \) is surrounded by a buffer \( b \) because fragment pairs computed with QM containing one fragment outside of \( A \) (i.e., in \( b \)) can still contribute to the total energy gradient (see below). On the other hand, fragment pairs with one fragment in \( F \) can also contribute to the total gradient, but they are computed using a simple classical expression rather than with QM. Note that the relation between the notation used in FMO/FD and that we use here is as follows: \( A,F \) and \( S \) are the same. The buffer region \( B \)
includes A, but b does not, i.e., A and b share no atoms. Formally, A and b are always treated at the same level of theory by assigning fragments to the same layer.

In the EFMO method, covalent bonds between fragments are not cut. Instead, electrons from a bond connecting two fragments are placed entirely to one of the fragments. The electrons of the fragments are kept in place by using frozen orbitals across the bond. [21,45,46] Fragments connected by a covalent bond share atoms (Figure 3) through the bonding region so it is possible that one side changes the wave function of the bonding region [21]. It is therefore necessary to re-evaluate the internal ab initio energy of region b for each new geometry step.

Figure 3. Cross region fragmentation. The fragmentation procedure shares an atom (here C\textsuperscript{1} and C\textsuperscript{5} is the shared atom) between two neighboring and covalently bonded fragments. Even though these fragments are in separate regions, they still share an atom across that region as illustrated.

doi:10.1371/journal.pone.0060602.g003

Figure 4. EFMO:S model of chorismate mutase used in this study. The entire model contains 2398 atoms. There are 1341 atoms in green belonging to the frozen region (F), 928 atoms in blue belonging to the buffer region (b) and 129 atoms in red belonging to the active region (A).

doi:10.1371/journal.pone.0060602.g004
The internal geometries of fragments in region $F$ are completely frozen so the internal energy is constant and is therefore neglected

\[ E_F^0 = 0. \]  

(6)

However, it is still necessary to compute the multipole moments and polarizability tensors (and therefore the wave function) of the fragments in $F$ once at the beginning of a geometry-optimization to evaluate $E_{\text{POL}}^0$ in equation 3 as well as some inter-region interaction energies defined as

\[ E_{b/A}^0 = \sum_{I \in b, J \in A} \left( \Delta E_{IJ}^0 - E_{\text{POL}}^0 \right) + \sum_{I \in b, J \in A} E_{IJ}^{\text{ES}}, \]  

(7)

\[ E_{F/A}^0 = \sum_{I \in A, J \in F} E_{IJ}^{\text{ES}}. \]  

(8)

Equation 8 assumes that $b$ is chosen so that fragments in $A$ and $F$ are sufficiently separated (i.e., $r_{IJ} > R_{\text{resdim}}$) so the interaction is evaluated classically. If all atoms in region $b$ are frozen, then $E_{F/b}^0$ is constant and can be neglected. However, this assumes that the positions of all atoms at both sides of the bonds connecting fragments are frozen.

The final expression for the EFMO frozen domain (EFMO/FD) energy is

\[ E_{\text{EFMO}} = E_b^0 + E_A^0 + E_{b/A}^0 + \sum_{I \in A, J \in F} E_{IJ}^{\text{ES}} + E_{\text{POL}}^0. \]  

(10)

---

**Figure 5.** EFMO:L model of chorismate mutase used in this study. The entire model contains 2398 atoms. There are 1006 atoms in green belonging to the frozen region ($F$), 1151 atoms in blue belonging to the buffer region ($b$) and 241 atoms in red belonging to the active region ($A$). doi:10.1371/journal.pone.0060602.g005
Finally, we note that due to the frozen geometry of \( b \) we can further gain a speedup by not evaluating dimers in \( b \) (cross terms between \( A \) and \( b \) are handled explicitly according to equation 7) since they do not contribute to the energy or gradient of \( A \). This corresponds to the frozen domain with dimers (EFMO/FDD), and equation 5 becomes

\[
E_h^b = \sum_{i \neq b} E_j^b.
\]  

Equation 13 does not apply to non-frozen atoms shared with region \( A \).

The frozen domain formulation of EFMO was implemented in GAMESS [47] and parallelized using the generalized distributed data interface [48,49].

\[
\frac{\partial E_{\text{EFMO}}}{\partial x_A} = \frac{\partial E^0_F}{\partial x_A} + \frac{\partial E^0_{EFMO}}{\partial x_A} + \frac{\partial E^0_{POL}}{\partial x_A},
\]

\[
\frac{\partial E_{\text{EFMO}}}{\partial x_b} = 0,
\]

\[
\frac{\partial E_{\text{EFMO}}}{\partial x_f} = 0,
\]

and the details of their evaluation has been discussed previously [20,21]. Equation 13 does not apply to non-frozen atoms shared with region \( A \).

The frozen domain formulation of EFMO was implemented in GAMESS [47] and parallelized using the generalized distributed data interface [48,49].

**Methods**

**Preparation of the Enzyme Model**

We followed the strategy by Claeyssens et al. [40] The structure of chorismate mutase (PDB: 2CHT) solved by Choik et al. [50] was used as a starting point. Chains A, B and C were extracted using PyMOL [51] and subsequently protonated with PDB2PQR [52,53] and PROPKA [54] at pH = 7. The protonation state of all residues can be found in Table S1. The inhibitor between chain A and C was replaced with chorismate in the reactant state (1, Figure 1) modeled in Avogadro [55,56].

The entire complex (chorismate mutase and chorismate) was solvated in water [TIP3P [57]] using GROMACS. [58,59] To neutralize the system 11 Na\(^{+}\) counter ions were added. The protein and counter ions were treated with the CHARMM27 [60,61] force field in GROMACS. Force-field parameters for chorismate were generated using the SwissParam [62] tool. To equilibrate the complex a 100 ps NVT run at \( T = 300 \) K was followed by a 100ps NPT run at \( P = 1 \) bar and \( T = 300 \) K. The production run was an isothermal-isobaric trajectory run for 10 ns. A single conformation was randomly selected from the last half of the simulation and energy minimized in GROMACS to a force threshold of \( F_{\text{max}} = 300 \) KJmol\(^{-1}\) nm\(^{-1}\). During equilibration and final molecular dynamics (MD) simulation, the C3 and C4 atoms of chorismate (see Figure 1) were harmonically constrained to a distance of 3.3 Å to keep it in the reactant state. Finally, a sphere of 16 Å around the C1 atom of chorismate was extracted in PyMOL and hydrogens were added to correct the valency where the backbone was cut. The final model contains a total of 2398 atoms.

**Mapping the Reaction Path**

To map out the reaction path, we define the reaction coordinate similarly to Claeyssens et al. [40] as the difference in bond length between the breaking O2-C1 bond and the forming C4-C3 bond in chorismate (see also Figure 1), i.e.

\[
R = R_{21} - R_{43}.
\]

The conversion of chorismate \((R = 2.0 \text{ Å}, R_{21} = 1.4 \text{ Å}, R_{43} = 3.4 \text{ Å})\) to prephenate \((R = 1.9 \text{ Å}, R_{21} = 3.3 \text{ Å}, R_{43} = 1.4 \text{ Å})\) in the enzyme was mapped by constraining the two bond lengths in equation 15 with a harmonic force constant of 500 kcal mol\(^{-1}\)
Table 1. EFMO-RHF and FMO2-RHF results for chorismate mutase.

<table>
<thead>
<tr>
<th>Model</th>
<th>$r_{\text{resdim}}$</th>
<th>modfd</th>
<th>$R_F$</th>
<th>$R_T$</th>
<th>$R_{WP}$</th>
<th>$E_{TS-R}$</th>
<th>$E_{P-R}$</th>
<th>$T_{\text{rel}}$</th>
<th>$T_{\text{full}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>EFMO:S</td>
<td>1.5</td>
<td>3</td>
<td>−1.95</td>
<td>−0.36</td>
<td>1.42</td>
<td>46.25</td>
<td>−1.32</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>EFMO:S</td>
<td>1.5</td>
<td>1</td>
<td>−1.96</td>
<td>−0.36</td>
<td>1.42</td>
<td>46.49</td>
<td>−1.34</td>
<td>2.0</td>
<td>1.7</td>
</tr>
<tr>
<td>EFMO:S</td>
<td>2.0</td>
<td>3</td>
<td>−1.96</td>
<td>−0.12</td>
<td>1.56</td>
<td>46.46</td>
<td>−0.91</td>
<td>1.3</td>
<td>1.2</td>
</tr>
<tr>
<td>EFMO:L</td>
<td>1.5</td>
<td>3</td>
<td>−1.97</td>
<td>−0.35</td>
<td>1.57</td>
<td>46.42</td>
<td>−3.61</td>
<td>2.1</td>
<td>1.8</td>
</tr>
<tr>
<td>FMO2:S</td>
<td>1.5</td>
<td>3</td>
<td>−1.93</td>
<td>−0.33</td>
<td>1.41</td>
<td>47.21</td>
<td>−1.40</td>
<td>6.7</td>
<td>7.5</td>
</tr>
</tbody>
</table>

Reaction barriers of chorismate mutase calculated with different levels of theory. $r_{\text{resdim}}$ is unitless. The reaction coordinates for the reactant, transition state and product are $R_F$, $R_T$ and $R_{WP}$ respectively and given in Å, barrier height of the transition state $E_{TS-R}$ and overall reaction energy $E_{P-R}$ in kcal/mol. $T_{\text{rel}}$ are relative timings to EFMO-RHF/6-31G(d) using the EFMO:S model with the fully minimized reaction coordinate on the trajectory subject to harmonic constraints. $T_{\text{full}}$ are for the entire path. doi:10.1371/journal.pone.0060602.t001

In order to refine the energetics, for each minimized step on the reaction path we performed two-layer ONIOM [64,65] calculations

$$E_{\text{high}} = E_{\text{low}} + E_{\text{model}} - E_{\text{model}^*}$$

where $E_{\text{low}} = E_{\text{EFMO}}$ according to equation 3. This can be considered a special case of the more general multilayer ONIOM based on FMO [66], using EFMO instead of FMO. The high level model system is chorismate in the gas-phase calculated using B3LYP [67–69] (DFTTYP = B3LYP in $\text{CONTRL}$) or MP2 (MPLEVL = 2 in $\text{CONTRL}$) with either 6-31G(d) or the cc-pVDMZ, cc-pVTZ and cc-pVQZ basis sets by Dunning [70].

We also carried out multilayer EFMO and FMO [71] single-point calculations where region $F$ is described by RHF/6-31G(d) and $b$ and $A$ (for EFMO) or $B$ ($B = A$) for FMO [18]) is calculated using MP2/6-31G(d).

The FDD approximation in equation 11 is enabled by specifying MODFD = 3 in $\text{FMO}$, similarly to the frozen domain approximation for ONIOM calculated using various levels of theory for the FS geometries. The red curve is the EFMO-RHF/6-31G(d) result also presented in Figure 8. Blue (B3LYP) and green (MP2) curves are ONIOM results with chorismate calculated in the gas-phase using the 6-31G(d) (solid lines), cc-pVDMZ (dashed lines) or cc-pVQZ (dotted line) basis set. doi:10.1371/journal.pone.0060602.g009

Figure 8. EFMO-RHF/6-31G(d) barrier for chorismate mutase. S15FD3 and S15FD3_FMO are EFMO:S and FMO:S, respectively, both with $r_{\text{resdim}} = 1.5$, and the dimer approximation in region $b$ (Equation 11). S15FD1 is similar to S15FD3 but without the dimer approximation in region $b$. S20FD3 is also similar to S15FD3 but with $r_{\text{resdim}} = 2.0$, instead. Finally, L15FD3 is EFMO:L with $r_{\text{resdim}} = 1.5$, and the dimer approximation (FDD) in region $b$. doi:10.1371/journal.pone.0060602.g008

Figure 9. ONIOM results calculated with various levels of theory for EFMO:S geometries. The red curve is the EFMO-RHF/6-31G(d) result also presented in Figure 8. Blue (B3LYP) and green (MP2) curves are ONIOM results with chorismate calculated in the gas-phase using the 6-31G(d) (solid lines), cc-pVDMZ (dashed lines) or cc-pVQZ (dotted line) basis set. doi:10.1371/journal.pone.0060602.g009

\[ \text{Mapping Enzymatic Catalysis Using the EFMO Method} \]
The reaction coordinates in Å for the reactant, transition state and product are \( R_R, R_{TS} \) and \( R_P \), respectively. The reaction barriers for the TSs are around 46 kcal mol\(^{-1}\). The same is true when going to the product \( R_P \). Only the large model (EFMO:L) shows a difference in energy near the product \( R_P \) with a lowering of the relative energy by 4 kcal mol\(^{-1}\) compared to the other settings.

The reaction coordinates are also similar for the small systems \( R_P = 1.41 \) Å, except for \( R_{redin} = 2.0 \) which is \( R_P = 1.56 \) Å with some minor kinks on the energy surface from optimization of the structures without constraints at \( R_P \). The EFMO:L model has a different reaction coordinate for the product \( R_P = 1.57 \) Å and also a shifted reaction coordinate for the transition state \( R_{TS} = -0.12 \) Å which we can attribute to a better description of more separated pairs in the active region but more importantly that around the TS, the energy surface is very flat. Interestingly, using FMO2 shows no significant change in either reaction barriers or reaction coordinates for the reactant, transition state or product which differ from EFMO:S by 0.02 Å, 0.03 Å and 0.01 Å respectively. Timings are discussed below.

Previous work by Ranaghan et al. [36,37] obtained an RHF barrier of 36.6 kcal mol\(^{-1}\) which is 10 kcal/mol lower than what we obtained. Also, they observed that the transition state happened earlier at \( R_{TS} = -0.3 \) Å. The difference in reaction barrier from our findings is attributed to a poorer enzyme structure and other snapshots do yield similar or better reaction barriers [see below]. Furthermore, the same study by Ranaghan et al. found that the reaction is indeed exothermic with a reaction energy of around \(-30 \) kcal mol\(^{-1}\) at the RHF/6-31G(d) level of theory. We expect this difference from our results to arise from the fact the study by Ranaghan et al. used a fully flexible model for both the substrate and the enzyme where the entire protein is free to adjust contrary to our model where we have chosen active fragments and atoms in a sphere around a central fragment. This is perhaps not the best solution if one includes too few fragments in the active region because fragments in the buffer region are unable to move and can cause steric clashes. The lowering of the energy for EFMO:L suggests this.

**Refined Reaction Energetics**

For the smallest EFMO:S system ONIOM results are presented on Figure 9 and in Table 2 for various levels of theory. By calculating the MP2/cc-pVDZ.EFMO-RHF/6-31G(d) energy using ONIOM we obtain a 19.8 kcal mol\(^{-1}\) potential energy

---

**Table 2.** Reaction barriers of chorismate mutase calculated using ONIOM.

<table>
<thead>
<tr>
<th>Methods</th>
<th>( R_R )</th>
<th>( R_{TS} )</th>
<th>( R_P )</th>
<th>( E_{TS-R} )</th>
<th>( E_{P-R} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>MP2/6-31G(d)</td>
<td>-1.83</td>
<td>0.13</td>
<td>1.56</td>
<td>22.24</td>
<td>-3.20</td>
</tr>
<tr>
<td>MP2/cc-pVDZ</td>
<td>-1.83</td>
<td>-0.36</td>
<td>1.56</td>
<td>19.75</td>
<td>-5.48</td>
</tr>
<tr>
<td>MP2/cc-pVTZ</td>
<td>-1.83</td>
<td>0.13</td>
<td>1.56</td>
<td>21.79</td>
<td>-1.14</td>
</tr>
<tr>
<td>MP2/cc-pVQZ</td>
<td>-1.83</td>
<td>0.13</td>
<td>1.56</td>
<td>21.68</td>
<td>-0.82</td>
</tr>
<tr>
<td>B3LYP/6-31G(d)</td>
<td>-1.83</td>
<td>0.13</td>
<td>1.39</td>
<td>25.19</td>
<td>3.81</td>
</tr>
<tr>
<td>B3LYP/cc-pVDZ</td>
<td>-1.83</td>
<td>0.13</td>
<td>1.39</td>
<td>23.81</td>
<td>2.58</td>
</tr>
<tr>
<td>B3LYP/cc-pVTZ</td>
<td>-1.83</td>
<td>0.13</td>
<td>1.56</td>
<td>24.62</td>
<td>4.36</td>
</tr>
<tr>
<td>B3LYP/cc-pVQZ</td>
<td>-1.83</td>
<td>0.13</td>
<td>1.56</td>
<td>24.66</td>
<td>4.16</td>
</tr>
</tbody>
</table>

The reaction coordinates in Å for the reactant, transition state and product are \( R_R, R_{TS} \) and \( R_P \), respectively. The barrier height of the transition state \( E_{TS-R} \) and the overall reaction energy \( E_{P-R} \) are in kcal/mol.

**Table 3.** Reaction barriers of chorismate mutase calculated using multilayer EFMO and FMO2 calculations.

<table>
<thead>
<tr>
<th>Methods</th>
<th>( R_R )</th>
<th>( R_{TS} )</th>
<th>( R_P )</th>
<th>( E_{TS-R} )</th>
<th>( E_{P-R} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>EFMO-RHF:MP2/6-31G(d)</td>
<td>-1.64</td>
<td>-0.11</td>
<td>1.39</td>
<td>27.64</td>
<td>-4.70</td>
</tr>
<tr>
<td>FMO2-RHF:MP2/6-31G(d)</td>
<td>-1.64</td>
<td>-0.11</td>
<td>1.88</td>
<td>29.22</td>
<td>-6.41</td>
</tr>
</tbody>
</table>

The reaction coordinates in Å for the reactant, transition state and product are \( R_R, R_{TS} \) and \( R_P \), respectively. The barrier height of the transition state \( E_{TS-R} \) and the overall reaction energy \( E_{P-R} \) are in kcal/mol.

---

**Effects of Methodology, Region Sizes and Approximations**

The reaction barriers obtained in the enzyme using harmonic constraints are plotted on Figure 8 and listed in Table 1 for different settings of region sizes and approximations. All calculated reaction barriers are within 0.5 kcal mol\(^{-1}\) from each other when going from the reactant \( (R_R) \) to the proposed transition \( (R_{TS}) \) state where the reaction barriers for the TSs are around 46 kcal mol\(^{-1}\). The same is true when going to the product \( R_P \). Only the large model (EFMO:L) shows a difference in energy near the product \( R_P \) with a lowering of the relative energy by 4 kcal mol\(^{-1}\) compared to the other settings.

The reaction coordinates are also similar for the small systems \( R_P = 1.41 \) Å, except for \( R_{redin} = 2.0 \) which is \( R_P = 1.56 \) Å with some minor kinks on the energy surface from optimization of the structures without constraints at \( R_P \). The EFMO:L model has a different reaction coordinate for the product \( R_P = 1.57 \) Å and also a shifted reaction coordinate for the transition state \( R_{TS} = -0.12 \) Å which we can attribute to a better description of more separated pairs in the active region but more importantly that around the TS, the energy surface is very flat. Interestingly, using FMO2 shows no significant change in either reaction barriers or reaction coordinates for the reactant, transition state or product which differ from EFMO:S by 0.02 Å, 0.03 Å and 0.01 Å respectively. Timings are discussed below.

Previous work by Ranaghan et al. [36,37] obtained an RHF barrier of 36.6 kcal mol\(^{-1}\) which is 10 kcal/mol lower than what we obtained. Also, they observed that the transition state happened earlier at \( R_{TS} = -0.3 \) Å. The difference in reaction barrier from our findings is attributed to a poorer enzyme structure and other snapshots do yield similar or better reaction barriers [see below]. Furthermore, the same study by Ranaghan et al. found that the reaction is indeed exothermic with a reaction energy of around \(-30 \) kcal mol\(^{-1}\) at the RHF/6-31G(d) level of theory. We expect this difference from our results to arise from the fact the study by Ranaghan et al. used a fully flexible model for both the substrate and the enzyme where the entire protein is free to adjust contrary to our model where we have chosen active fragments and atoms in a sphere around a central fragment. This is perhaps not the best solution if one includes too few fragments in the active region because fragments in the buffer region are unable to move and can cause steric clashes. The lowering of the energy for EFMO:L suggests this.

---

**Results and Discussion**

The activation enthalpy is obtained in two different ways by calculating averages of \( M \) adiabatic reaction paths. The starting points of the \( M \) pathways were randomly extracted from the MD simulation, followed by the reaction path mapping procedure described above for each pathway individually. One way to obtain the activation enthalpy averages the barriers from each individual adiabatic reaction path [72].

\[
\Delta H^1 = \frac{1}{M} \sum_{i=1}^{M} (E_{TS,i} - E_{R,i}) - 1.6 \text{ kcal mol}^{-1}. \tag{17}
\]

Here \( M \) is the number of reaction paths (\( M = 7 \), Figure 6 and Figure 7) \( E_{TS,i} \) is the highest energy on the adiabatic reaction path while \( E_{R,i} \) is the lowest energy with a negative reaction coordinate. 1.6 kcal mol\(^{-1}\) corrects for the change in zero point energy and thermal contributions [72].

The other way of estimating the activation enthalpy is [37]:

\[
\Delta H^2 = \langle E_{TS} \rangle - \langle E_R \rangle - 1.6 \text{ kcal mol}^{-1}. \tag{18}
\]

Here \( \langle E_{TS} \rangle \) and \( \langle E_R \rangle \) are, respectively, the highest energy and lowest energy with a negative reaction coordinate on the averaged adiabatic path (bold line in Figure 6 and Figure 7). The brackets here mean averaging over 7 reaction paths; and the difference of Eqs 17 and 18 arises because of the non-commutativity of the sum and the min/max operation over coordinates: in Eq 17 we found a minimum and a maximum for each curve separately, and averaged the results, but in Eq 18 we first averaged and then found the extrema. As discussed below, the two reaction enthalpies are within 0.2 kcal/mol, which indicates that the TS occurs at roughly the same value of the reaction coordinate for most paths.
barrier. Furthermore, the reaction energy is lowered from $-1.3 \text{ kcal mol}^{-1}$ to $-3.5 \text{ kcal mol}^{-1}$. Increasing the basis set size through cc-pVTZ and cc-pVQZ reduces the barrier to 21.8 kcal mol$^{-1}$ and 21.7 kcal mol$^{-1}$, respectively, and the reaction energy is $-1.1 \text{ kcal mol}^{-1}$ and 0.8 kcal mol$^{-1}$. Using the smaller 6-31G(d) basis set with MP2, the reaction barrier is 22.2 kcal mol$^{-1}$ and reaction energy is $-3.2 \text{ kcal mol}^{-1}$. The B3LYP results are improvements for the TS only reducing the barrier to 23.0 kcal mol$^{-1}$ for B3LYP/cc-pVDZ:EFMO-RHF/6-31G(d). The same is not true for the product where the energy is increased by about 3 kcal mol$^{-1}$. For the other systems treated using EFMO-RHF/6-31G(d) discussed in the previous section ONIOM corrected results at the MP2 or B3LYP level of theory using a cc-pVDZ basis set are listed in tables S2 to S5 and show differences from the above by less than 1 kcal mol$^{-1}$, again the reaction coordinates change slightly between the tested options. The effect of including correlation effects by means of MP2 and systematically larger basis sets is that the potential energy barrier for the reaction rises less than 1 kcal mol$^{-1}$.

The resuls presented here for MP2 are in line with what has been observed previously by Ranaghan et al. [37] and Claeyssens et al. [40]. Overall, the reaction barrier is reduced to roughly half of the RHF barrier and the observed coordinates for the reaction shift slightly. We do note that this study and the study by Ranaghan et al. use ONIOM style energy corrections for the correlation and not geometry optimizations done at a correlated level. Overall, we observe that the predicted reaction coordinate for the approximate transition state in the conversion of chorismate to prephenate happens around 0.2 Å later than in those studies.

The results for the multilayer single points along the energy surface are presented in Table 3. The barrier calculated at the EFMO-RHF:MP2/6-31G(d) level of theory is predicted to be 27.6 kcal mol$^{-1}$ which is 5.4 kcal mol$^{-1}$ higher than the ONIOM barrier and the reaction coordinates are shifted for both the reactant and the TS from $R_R = -1.95 \text{ Å}$ to $R_R = -1.64 \text{ Å}$ and $R_T = -0.36 \text{ Å}$ to $R_T = -0.11 \text{ Å}$. Similar results are obtained at the FMO2-RHF:MP2/6-31G(d) level of theory. The difference from the ONIOM corrected values in table 3 is likely due to the inclusion of dispersion effects between the chorismate and the enzyme which is apparently weaker at the transition state compared to the reactant state.

**Ensemble Averaging**

In Figure 6 and Figure 7 we show 7 adiabatic reaction paths mapped with EFMO-RHF/6-31G(d) starting from 7 snapshots taken from the MD simulation; the energetics were refined with ONIOM at the MP2/cc-pVDZ and MP2/cc-pVTZ level. In EFMO, we used a small active region (EFMO:S) and $R_{\text{resdim}} = 1.5$ and no dimer calculations in region b (S13F9 in Figure 6). Out of the 7 trajectories one is described in detail in the previous sub-section.

For MP2/cc-pVDZ:EFMO-RHF/6-31G(d) the reaction enthalpies are $\Delta H^T = 18.3 \pm 3.5 \text{ kcal mol}^{-1}$ and $\Delta H^R = 18.2 \text{ kcal mol}^{-1}$ [cf. Equations (17) and (19)], the latter having an uncertainty of the mean of 6.9 kcal mol$^{-1}$. For MP2/cc-pVTZ:EFMO-RHF/6-31G(d) the reaction enthalpies are $\Delta H^T = 19.3 \pm 3.7 \text{ kcal mol}^{-1}$ and $\Delta H^R = 18.8 \text{ kcal mol}^{-1}$ with an uncertainty of the mean of 7.1 kcal mol$^{-1}$. These barriers are ca 5.5 (6.3) kcal mol$^{-1}$ higher than the experimental value of 12.7 (4.4) kcal mol$^{-1}$ for MP2/cc-pVDZ (MP2/cc-pVTZ). For comparison, the activation enthalpy obtained by Claeyssens et al. [40,72] (9.7 ± 1.8 kcal mol$^{-1}$) is underestimated by 3.0 kcal mol$^{-1}$.

There are several differences between our study and that of Claeyssens et al. that could lead to an overestimation of the barrier height: biasing the MD towards the TS rather than the reactant, a larger enzyme model (7218 vs 2390 atoms), and more conformational freedom when computing the potential energy profile.

With regard to the latter point, while Figure 8 shows that increasing the active region has a relatively small effect on the barrier this may not be the case for all snapshots. We did identify one trajectory that failed to produce a meaningful reaction path and is presented in Figure S1. Here, the energy of the barrier becomes unrealistically high because of very little flexibility in the active site and unfortunate placement of Phe57 (located in the buffer region, Figure S2), which hinders the conformational change needed for the successful conversion to prephenate yielding an overall reaction energy of around +11 kcal mol$^{-1}$.

As noted above, the EFMO:L settings is a possible solution to this as more of the protein in available to move, but as seen from Table 1 the computational cost doubles.

**Timings**

Using the computationally most efficient method tested here (EFMO:S), $R_{\text{resdim}} = 1.5$, and skipping dimers in the buffer region b, an adiabatic reaction path, which requires a total of 467 gradient evaluations, can be computed in four days using 80 CPU cores (20 nodes with 4 cores each) at the RHF/6-31G(d) level of theory. As shown in Table 1, the same calculation using FMO2 requires taking roughly $T_{\text{rel}} = 7.5$ times longer.

Increasing $R_{\text{resdim}}$ from 1.5 to 2 has a relatively minor effect of the CPU time (a factor of 1.2), while performing the dimer calculations in the buffer region nearly doubles (1.7) the CPU time. Increasing the size of active region from 2.0 Å to 3.0 Å around chorismate nearly doubles (1.8) the CPU time. This is mostly due to the fact that more dimer calculations must be computed, but the optimizations also require more steps (513 gradient evaluations) to converge due to the larger number of degrees of freedom that must be optimized.

Looking at a single minimization for a specific reaction coordinate $R = -1.79 \text{ Å}$, the most efficient method takes 4.5 hours. Here, the relative timings $T_{\text{rel}}$ are all larger than for the full run ($T_{\text{full}}$ due to a slight increase in the number of geometry steps (around 25) taken for all but FMO2 which is identical to the reference (22 steps). Thus, the overall cost of performing the FMO2 minimization is 6.7 times as expensive as EFMO.

**Conclusions**

In this paper we have shown that the effective fragment molecular orbital (EFMO) method [20,21] can be used to efficiently map out enzymatic reaction paths provided the geometry of a large part of the enzyme and solvent is frozen. In EFMO one defines an active region associated with the active site, and the cost of a geometry optimization is then essentially the cost of running quantum-mechanical calculations of the active domain. This is similar to the cost of QM/MM, if the QM region is the same; the difference is that in EFMO we freeze the coordinates of the rest of the system, whereas in QM/MM they are usually fully relaxed. On the other hand, EFMO does not require parameters and can be better considered an approximation to a full QM calculation rather than a QM/MM approach.

In this work we used the mapping technique based on running a classical MD simulation, selecting some trajectories, freezing the coordinates of the outside region, and doing constrained geometry
optimizations along a chosen reaction coordinate. An alternative to this approach is to run full MD simulation of a chemical reaction using EFMO. This has already been done for many chemical reactions using FMO-MD [73–75] and can be done in future with EFMO.

A potential energy profile for the chorismate to prephenate reaction in chorismase has been computed in 4 days using 80 CPU cores for an RHF/6-31G(d) description of a truncated model of the enzyme containing 2398 atoms. For comparison, a corresponding FMO2 calculation takes about 7.5 times more. The cost of EFMO calculations is mainly determined by the size of the buffer- and active region. Comparing to a QM/MM calculation with a QM region of the same size, EFMO as a nearly linear scaling method, becomes faster than QM if the system size is sufficiently large; especially for correlated methods like MP2 where this cross-over should happen with relatively small sizes.

Our computed conformationally-averaged activation enthalpy is in reasonable agreement to the experimental value, although overestimated by 5.5 kcal/mol.

The energetics of this reaction depends on the level of calculation. We have shown that by using a level better than RHF, for instance, MP2 or DFT, considerably improves the energetics and by using such an appropriate level to also determine the reaction path following the formalism in this work can be used to provide a general and reliable way in future.

EFMO, as one of the fragment-based methods [13], can be expected to be useful in various biochemical studies, such as in enzymatic catalysis and protein-ligand binding. It should be noted that in addition to its parameter-free ab initio based nature, EFMO and FMO also offer chemical insight on the processes by providing subsystem information, such as the properties of individual fragments (e.g., the polarization energy) as well as the pair interaction energies between fragments [76,77]. This can be of considerable use in fragment-based drug discovery [78,79].

Supporting Information
Figure S1 Reaction barrier calculated at the MP2/cc-pVDZ:EFMO-RHF/6-31G(d) level of theory for EFMO:S using $R_{\text{resdim}} = 1.5$ and FDD (modfd = 3). This snapshot shows the effect of not having enough flexibility in the active region around the substrate.

Table S1 Complete listing of all residues in the protein model (PDB: 2CHT) along with their protonation state after being protonated using the PDB2PQR tool.

Table S2 Reaction barriers of chorismate mutase calculated using ONIOM for EFMO:S using $R_{\text{resdim}} = 1.5$ and FD (modfd = 1).

Table S3 Reaction barriers of chorismate mutase calculated using ONIOM for EFMO:S using $R_{\text{resdim}} = 2.0$ and FDD (modfd = 3).

Table S4 Reaction barriers of chorismate mutase calculated using ONIOM for EFMO:L using $R_{\text{resdim}} = 1.5$ and FDD (modfd = 3).

Table S5 Reaction barriers of chorismate mutase calculated using ONIOM for FMO2:S using $R_{\text{resdim}} = 1.5$ and FDD (modfd = 3).

Acknowledgments
CS and JHJ thank the Danish Center for Scientific Computing at the University of Copenhagen for providing computational resources.

Author Contributions
Conceived and designed the experiments: CS. Performed the experiments: CS. Analyzed the data: CS DF JJ. Contributed reagents/materials/analysis tools: CS. Wrote the paper: CS DF JJ.

References


Mapping Enzymatic Catalysis Using the EFMO Method


