

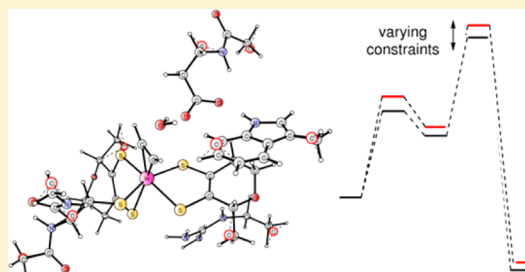
# On the Effect of Varying Constraints in the Quantum Mechanics Only Modeling of Enzymatic Reactions: The Case of Acetylene Hydratase

Rong-Zhen Liao and Walter Thiel\*

Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, D-45470, Mülheim an der Ruhr, Germany

## S Supporting Information

**ABSTRACT:** Quantum mechanics only (QM-only) studies of enzymatic reactions employ a coordinate-locking scheme, in which certain key atoms at the periphery of the chosen cluster model are fixed to their crystal structure positions. We report a case study on acetylene hydratase to assess the uncertainties introduced by this scheme. Random displacements of 0.1, 0.15, and 0.2 Å were applied at the ten terminal atoms fixed in the chosen 116-atom cluster model to generate sets of ten distorted structures for each given displacement. The relevant stationary points were reoptimized under these modified constraints to determine the variations of the computed energies and geometries induced by the displacements of the fixed atoms. Displacements of 0.1 Å cause a relatively minor perturbation that can be accommodated during geometry optimization, resulting in rather small changes in key bond distances and relative energies (typically of the order of 0.01 Å and 1 kcal/mol), whereas displacements of 0.2 Å lead to larger fluctuations (typically twice as high) and may sometimes even cause convergence to different local minima during geometry optimization. A literature survey indicates that protein crystal structures with a resolution higher than 2.0 Å are normally associated with a coordinate error of less than 0.1 Å for the backbone atoms. Judging from the present results for acetylene hydratase, such uncertainties seem tolerable in the design of QM-only models with more than 100 atoms, which are flexible enough to adapt during geometry optimization and thus keep the associated uncertainties in the computed energies and bond distances at tolerable levels (around 1 kcal/mol and 0.01 Å, respectively). On the other hand, crystal structures with significantly lower resolution should be used with great caution when setting up QM-only models because the resulting uncertainties in the computational results may become larger than acceptable. The present conclusions are mostly based on systematic DFT(B3LYP) calculations with a medium-size basis set. Test calculations on selected structures confirm that similar results are obtained for larger basis sets, different functionals ( $\omega$ B97X, BMK, M06), and upon including solvation and zero-point corrections, even though the fluctuations in the computed relative energies become somewhat larger in some cases.



## I. INTRODUCTION

The quantum mechanics only (QM-only) approach,<sup>1–6</sup> alternatively known as the quantum chemical cluster approach, has been successfully applied to mechanistic investigations of a wide spectrum of different enzymes. The basic idea of this approach is to design a cluster model of the active site on the basis of available crystal structures and treat it quantum mechanically as accurately as possible. In most cases, density functional theory (DFT) methods, in particular the B3LYP<sup>7</sup> functional, have been used. To account for the polarization and steric effects from the protein surrounding, two simple procedures are normally adopted. The polarization effect is modeled by embedding the cluster into a dielectric continuum with a dielectric constant of 4. The particular choice of the dielectric constant becomes irrelevant when the size of the QM-only model reaches 150–200 atoms, as shown by recent studies on four different types of enzymes.<sup>8–11</sup> The steric effect introduced by the protein environment is taken into account by a coordinate-locking scheme, in which certain key atoms, typically those where the truncation is made, are fixed to their crystal structure positions during the geometry optimizations.

The coordinate-locking scheme keeps the optimized structures reasonably close to the X-ray structure and prevents artificial movements and rotations of various groups in the model. A proper application of this technique can ensure the structural integrity of the model and at the same time permit enough flexibility during geometry optimization. In a mechanistic study of the zinc-dependent enzyme thermolysin, Siegbahn and co-workers found that the geometries and relative energies of all stationary points were quite similar in cluster calculations with and without constraints.<sup>12</sup> For geometries, the mean absolute deviation of the zinc–ligand distances was 0.03 Å, with a maximum deviation of 0.23 Å. For energies, the average and maximum deviations were 1.3 and 2.6 kcal/mol, respectively. In an investigation of another zinc-containing enzyme, stromelysin, they used an alternative procedure to mimic the steric effects by including backbones to connect all four important residues in the model.<sup>13</sup> The difference in the barriers calculated with and without backbones was 1.6 kcal/

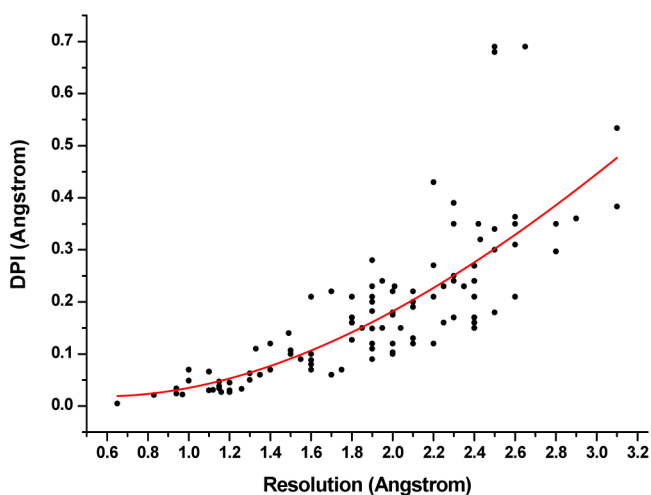
Received: November 28, 2012

Revised: March 13, 2013

Published: March 21, 2013

mol, while the reaction energy deviated somewhat more (by 5.1 kcal/mol). Siegbahn also tested the use of two different crystal structures of the dinuclear copper enzyme catechol oxidase, with Cu–Cu distances of 2.9 and 4.2 Å, to check the effect of backbone constraints at different positions.<sup>14</sup> After geometry optimization, the total energies of the  $\mu$ -oxo complexes derived from these two different structures deviated by 2.3 kcal/mol. In the case of photosystem II,<sup>15</sup> cluster models were generated from crystal structures of different resolution (3.5 and 1.9 Å); both models gave similar transition states and barriers (5.7 vs 6.0 kcal/mol) for the critical O–O bond formation step, which may be due to the fact that the positions of the backbone atoms are quite similar in the X-ray structures, with differences of a few tenths of an Ångström, except for Asp170 (2.7 Å) and Glu189 (2.0 Å). Himo and co-workers compared constrained and completely unconstrained models for the dinuclear zinc enzyme phosphotriesterase (82 atoms) and found energy differences of 2–5 kcal/mol for various stationary points.<sup>16</sup> As a general conclusion, the reliance on the crystal structure is not considered critical in QM-only studies, if its resolution is reasonable and if the geometry, particularly around the active site, does not change much during the reaction.<sup>15</sup> If the protein structure undergoes large displacements during the reaction, which is not common, the standard QM-only approach may fail because of the locking procedure used.

The robustness and success of the QM-only model depend on the quality of the starting crystal structure and on the flexibility of the chosen model, both of which should resemble the natural enzyme as much as possible. In principle, if more than one crystal structure is available, one should use the one with the highest resolution. The accuracy of crystal structures can be assessed from the atomic coordinate error, which can be estimated by three different methods, namely, the Luzzati plot,<sup>17</sup> the  $\sigma_A$  plot,<sup>18</sup> and the Cruickshank diffraction-component precision index (DPI).<sup>19</sup> The DPI indexes for 109 protein structures at different resolution (for data see Supporting Information, Table S1) are plotted in Figure 1. Evidently, the coordinate error is generally less than 0.1 Å at high resolution (0.8 to 1.5 Å) and in the range between 0.1 and 0.3 Å at medium resolution (1.5 to 2.5 Å), apart from a few exceptions. Since the C $\alpha$  atoms of the protein backbone are



**Figure 1.** Plots of DPI (in Å) vs resolution (in Å) for 109 protein structures obtained from X-ray crystallography. A polynomial curve is used to fit all points.

often fixed in the QM-only model, it is of interest to estimate the coordinate error also specifically for the backbone atoms. Corresponding comparisons for eight crystal structures<sup>20–25</sup> (for data see Supporting Information, Table S2) reveal that the error for the backbone atoms is approximately half of that for all atoms, due to the lower flexibility of the backbone relative to the side chains. Therefore, the coordinate error for the backbone atoms should normally be less than 0.1 Å for structures with a resolution higher than 2.0 Å, while it could be 0.2 Å or even larger for structures with lower resolution.

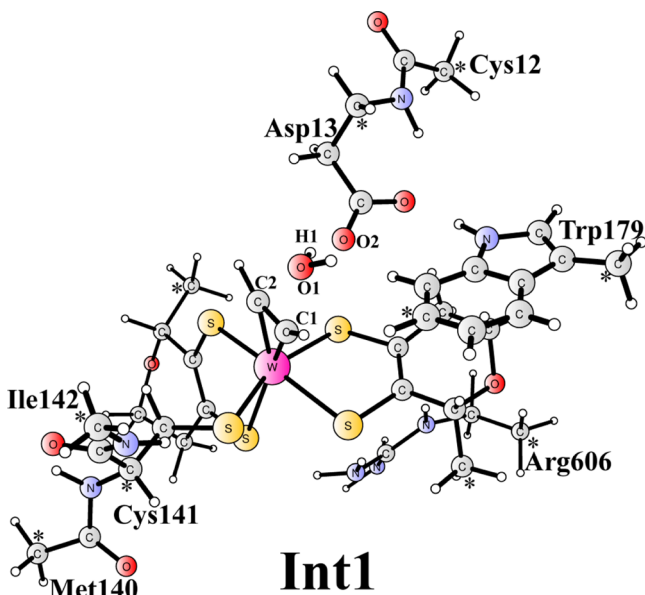
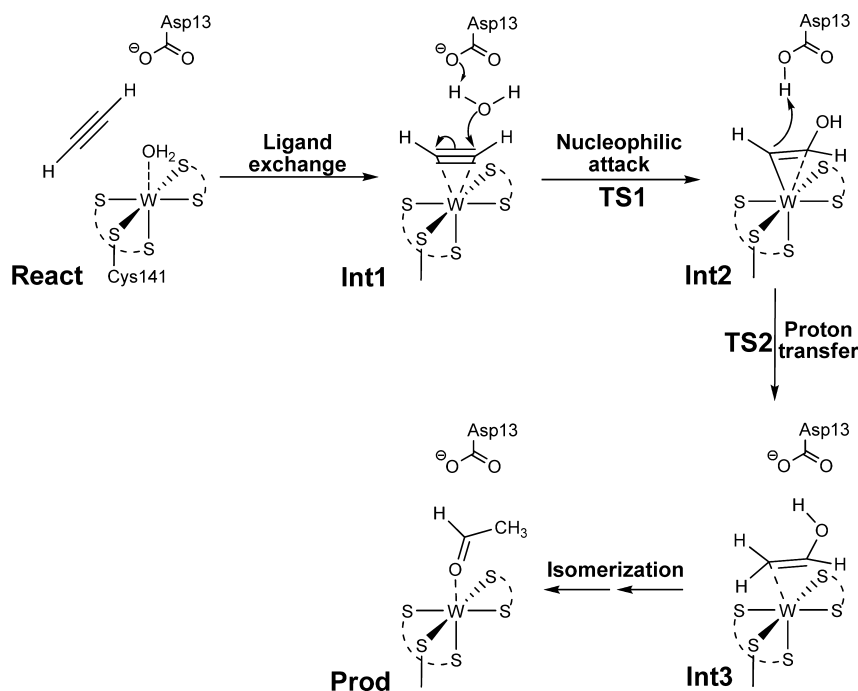
The uncertainties in X-ray coordinates may affect the accuracy of the structures and energetics obtained from QM-only modeling of enzymatic reactions (because of the commonly applied coordinate-locking procedure). In this article, we address this issue using the tungsten-dependent acetylene hydratase as an example. The reaction mechanism of this enzyme (Scheme 1) has been established using the QM-only approach,<sup>26</sup> and the QM-only results have been compared thoroughly with those from QM/MM calculations.<sup>27</sup> The mechanism consists of five steps. First, the acetylene molecule displaces the tungsten-bound water molecule and becomes coordinated to the metal in a  $\eta^2$  fashion. In the second step, the liberated water molecule performs a nucleophilic attack on the acetylene carbon, activated by the anionic Asp13 residue. This is followed by proton transfer from Asp13 to the substrate to afford vinyl alcohol. The subsequent two isomerization steps take place with the assistance of the tungsten ion and Asp13. This proposed mechanism has been used to rationalize the factors that govern the chemoselectivity of this enzyme, which does not promote the hydration of propyne, ethylene, and acetonitrile.<sup>28</sup> A biomimetic tungsten complex also employs a similar mechanism to catalyze the acetylene hydration.<sup>29</sup>

In the present paper, we are only interested in methodological issues, and we will thus focus on the vinyl alcohol formation process (from **Int1** to **Int3**). Random displacements at **Int1** (Figure 2) of the QM-only model<sup>26</sup> were performed for the ten atoms (labeled by asterisks) that were fixed during the geometry optimizations. Three distances of 0.1, 0.15, and 0.2 Å were selected for the random displacements, and ten structures were generated for each one, resulting in a total of 30 structures. The value of 0.1 Å roughly corresponds to the upper limit of the coordinate error of backbone atoms for protein structures with a resolution of better than 2.0 Å (see above). In contrast to earlier tests by Siegbahn,<sup>15</sup> one of the fixed atoms belongs to a nearby amino acid (Asp13) that is directly involved in the reaction.

## II. COMPUTATIONAL DETAILS

The QM-only model is exactly the same as in the previous study<sup>26</sup> and will thus not be described in detail here. It consists of 116 atoms (Figure 2) with a total charge of  $-1$ . The present calculations were carried out using exactly the same methods as in the previous QM-only investigation of the mechanism of acetylene hydratase.<sup>26</sup> They employed the hybrid functional B3LYP,<sup>7</sup> as implemented in the Gaussian09 code.<sup>30</sup> For geometry optimizations, we used the 6-31G(d) basis set for C, N, O, and H, the 6-311+G(d) basis for S, and the LANL2TZ(f)<sup>31</sup> pseudopotential and basis for tungsten (labeled as BS1).

For one structure with 0.1 Å random displacements, single-point calculations were done with a larger basis set, i.e., 6-311+G(2d,2p) for all elements except for W, which was described by the LANL2TZ(f) basis set and pseudopotential

Scheme 1. Reaction Mechanism of Acetylene Hydratase Suggested in a Previous QM-Only Study<sup>26</sup>

**Figure 2.** Optimized structure of **Int1** of the QM-only model of acetylene hydratase. Atoms marked with asterisks were fixed at their crystal structure positions during the geometry optimizations.

(labeled as **BS2**). In addition, the def2-TZVPPD<sup>32</sup> basis set (with the corresponding pseudopotential for W) was used in single-point calculations with the B3LYP functional. Further single-point calculations were performed using the  $\omega$ B97X,<sup>33</sup> M06,<sup>34</sup> and BMK<sup>35</sup> functionals with the **BS2** basis set. Solvation effects were evaluated by single-point calculations at the optimized gas-phase structures using the conductor-like polarizable continuum model (CPCM)<sup>36,37</sup> with  $\epsilon = 4$ , applying the same level of theory as in the geometry optimizations. Analytic frequency calculations were also carried out at the same level of theory as the geometry optimizations to obtain zero-point energies and to confirm the nature of the various

stationary points. In the following, unless stated otherwise, we report energies obtained at the B3LYP/BS1 level, to illustrate the effect of varying constraints.

### III. RESULTS AND DISCUSSION

**III.1. Random Displacements of 0.1 Å.** Ten structures were generated with random displacements of 0.1 Å for the fixed atoms (marked by asterisks in Figure 2). One typical structure (*Structure1*) was first chosen for comparisons with the original model (labeled as *Original*). The QM relative energies for *Original* and *Structure1* at various levels are listed in Table 1.

The relative energies from B3LYP calculations change significantly upon basis set extension from **BS1** to **BS2**, both for *Original* and *Structure1*; for example, the energies of **TS1** and **TS2** increase by 6–7 kcal/mol. On the other hand, changes are small for all stationary points when going from **BS2** to def2-TZVPPD, indicating convergence with regard to the basis set. The differences between the relative energies calculated at the B3LYP level for *Original* and *Structure1* are not affected much by basis set extension; for example, when going from *Original* to *Structure1*, the relative energy of **TS2** rises by 1.9, 2.0, and 1.3 kcal/mol for **BS1**, **BS2**, and def2-TZVPPD, respectively. The corresponding differences for the other three stationary points vary even less (within 0.5 kcal/mol). These results suggest that medium-size basis sets, in this case **BS1**, are capable of capturing the energy differences introduced by the random displacements, thus obviating the need for using computationally demanding large basis sets for this purpose.

The adequacy of the B3LYP functional for this enzyme was previously examined by performing single-point calculations (**BS2** basis set) with the B3LYP\* (15% HF exchange),<sup>38</sup> TPSS,<sup>39</sup> and BB1K<sup>40</sup> functionals.<sup>26</sup> The calculated barriers for the rate-limiting step were 2–5 kcal/mol lower than the B3LYP barrier.<sup>26</sup> Here, we report single-point calculations with three more functionals ( $\omega$ B97X, BMK, and M06; **BS2** basis set) to further assess the sensitivity of the energetics to the choice of functional. As shown in Table 1, all three functionals give

**Table 1. Relative Energies in kcal/mol of Various Stationary Points at Different Levels for Original and Structure1 with 0.1 Å Random Displacements at the Fixed Atoms**

		Int1	TS1	Int2	TS2	Int3
B3LYP/BS1	<i>Original</i>	0	13.7	8.0	17.4	-18.4
	<i>Structure1</i>	0	15.1	9.2	19.3	-17.8
	Deviation	0	+1.4	+1.2	+1.9	+0.6
B3LYP/BS2	<i>Original</i>	0	19.4	16.4	24.8	-12.0
	<i>Structure1</i>	0	21.0	17.8	26.8	-11.2
	Deviation	0	+1.6	+1.4	+2.0	+0.8
B3LYP/def2-TZVPPD	<i>Original</i>	0	19.7	16.2	24.0	-13.3
	<i>Structure1</i>	0	20.6	16.8	25.3	-13.2
	Deviation	0	+0.9	+0.8	+1.3	+0.1
$\omega$ B97X/BS2	<i>Original</i>	0	19.0	12.6	25.7	-14.0
	<i>Structure1</i>	0	20.1	13.2	27.2	-14.0
	Deviation	0	+1.1	+0.6	+1.5	0
BMK/BS2	<i>Original</i>	0	18.6	12.3	23.9	-16.2
	<i>Structure1</i>	0	19.8	13.0	25.3	-16.3
	Deviation	0	+1.2	+0.7	+1.4	-0.1
M06/BS2	<i>Original</i>	0	16.9	11.7	22.0	-12.8
	<i>Structure1</i>	0	20.2	14.4	25.6	-10.6
	Deviation	0	+3.3	+2.7	+3.6	+1.8
B3LYP final <sup>a</sup>	<i>Original</i>	0	20.6	18.2	22.2	-8.3
	<i>Structure1</i>	0	23.7	20.8	25.7	-5.7
	Deviation	0	+3.1	+2.6	+3.5	+2.6

<sup>a</sup>With solvation, basis set, and ZPE corrections.

barriers and reaction energies that are quite similar to the B3LYP values. For instance, the rate-limiting barriers for *Original* are 25.7, 23.9, 22.0, and 24.8 kcal/mol for  $\omega$ B97X, BMK, M06, and B3LYP, respectively. The differences between the relative energies calculated for *Original* and *Structure1* are smallest for  $\omega$ B97X and BMK (maximum of 1.5 and 1.4 kcal/mol, respectively), somewhat larger for B3LYP (maximum of 2.0 kcal/mol), and largest for M06 (maximum of 3.6 kcal/mol).

In QM-only studies of enzymes, the final energies reported usually include basis set, solvation, and ZPE corrections. Table 1 presents these final energies from B3LYP calculations to assess the cumulative effect of these corrections (for details see Supporting Information, Table S3). The differences between *Original* and *Structure1* become somewhat larger for all stationary points when considering these final energies. The largest difference is found for TS2, which amounts to +1.9 kcal/mol at the B3LYP/BS1 level and increases to +3.5 kcal/mol for the final energies.

The relative energies of various stationary points at the B3LYP/BS1 level for the ten structures generated by different 0.1 Å random displacements are listed in Table 2. The average relative energies of these ten structures are 14.3, 8.5, 18.0, and -17.5 kcal/mol, for TS1, Int2, TS2, and Int3, respectively. The associated root-mean-square deviations (RMSDs) are all less than 2 kcal/mol, with the maximum absolute deviation (MAD) reaching 4.0 kcal/mol for Int3. Moreover, the average relative energies differ from the values obtained for *Original* by less than 1 kcal/mol. For example, the average overall barrier is 18.0 kcal/mol, and thus is only slightly higher than the *Original* value of 17.4 kcal/mol. The average reaction energy of -17.5 kcal/mol is also quite close to the *Original* value of -18.4 kcal/mol.

The results for *Structure1* (see above) suggest that the accumulated basis set, solvation, and ZPE corrections could increase the differences between the relative energies of *Original* and randomly displaced structures by more than 1 kcal/mol. We thus performed some computations on all ten

**Table 2. Relative Energies in kcal/mol (B3LYP/BS1) of Various Stationary Points for the Original Structure and 10 Structures with 0.1 Å Random Displacements at the Fixed Atoms<sup>a</sup>**

	Int1	TS1	Int2	TS2	Int3
<i>Original</i>	0	13.7	8.0	17.4	-18.4
<i>Structure1</i>	0	15.1	9.2	19.3	-17.8
<i>Structure2</i>	0	13.8	8.2	17.9	-18.4
<i>Structure3</i>	0	15.1	9.2	19.1	-17.2
<i>Structure4</i>	0	14.6	8.7	18.6	-18.5
<i>Structure5</i>	0	13.3	7.3	16.2	-18.7
<i>Structure6</i>	0	14.4	8.9	18.2	-14.4
<i>Structure7</i>	0	14.0	8.6	17.3	-15.8
<i>Structure8</i>	0	13.2	7.1	17.6	-18.5
<i>Structure9</i>	0	15.4	9.7	19.0	-16.4
<i>Structure10</i>	0	13.6	8.0	17.2	-18.8
average	0	14.3	8.5	18.0	-17.5
RMSD	0	0.9	0.9	1.1	1.7
MAD	0	1.7	1.7	1.9	4.0

<sup>a</sup>RMSD: root mean square deviation. MAD: maximum absolute deviation.

structures to get more reliable statistics. For the sake of efficiency, this was done only for Int1 and Int2 (for results see Table S4, Supporting Information). The average energy of Int2 is higher than that of *Original*, by 0.5 kcal/mol at the B3LYP/BS1 level and by 0.7 kcal/mol for the final energies. These results confirm that the use of B3LYP/BS1 energies is sufficient to analyze the energetic effects of varying constraints.

To summarize, we find rather small differences (RMSD less than 2 kcal/mol) between the relative energies calculated for *Original* and for the ten structures generated by varying the constraints on the fixed atoms (0.1 Å random displacements). Hence, for proteins with a coordinate error of the backbone atoms of less than 0.1 Å (i.e., typically a resolution of less than



2.0 Å), the uncertainty in the coordinates for the fixed atoms in the QM-only approach should have a rather small effect on the computed reaction energetics.

**III.2. Random Displacements of 0.15 and 0.2 Å.** Since crystal structures with resolution lower than 2.0 Å may have large coordinate errors, up to 0.2 Å as discussed in the Introduction, we decided to quantify the resulting uncertainties in the calculated energies as well. For this purpose, we generate two further sets of structures (ten in each set), with random displacements of 0.15 and 0.2 Å from the *Original* reference structure. Being only interested in the relative energy differences caused by the coordinate displacements of the fixed atoms, we focus on one representative example, namely, the energy of **Int2** relative to **Int1**, to analyze the energetic consequences of these displacements.

Regardless of the chosen displacement, the average relative energy of **Int2** lies in the range of 8–9 kcal/mol, with deviations of up to 1 kcal/mol from the *Original* reference value (Table 3). However, when increasing the displacements from

**Table 3. Energies of Int2 Relative to Int1 in kcal/mol (B3LYP/BS1) for the Original Structure and 10 Structures with 0.1, 0.15, and 0.2 Å Random Displacements at the Fixed Atoms<sup>a</sup>**

	0.1 Å	0.15 Å	0.2 Å
<i>Original</i>	8.0	8.0	8.0
<i>Structure1</i>	9.2	7.1	7.9
<i>Structure2</i>	8.2	9.6	12.0
<i>Structure3</i>	9.2	9.3	8.6
<i>Structure4</i>	8.7	8.7	7.7
<i>Structure5</i>	7.3	6.4	10.2
<i>Structure6</i>	8.9	8.4	11.1 (4.4)
<i>Structure7</i>	8.6	8.2	7.2 (2.6)
<i>Structure8</i>	7.1	9.2	11.1
<i>Structure9</i>	9.7	8.0	7.0
<i>Structure10</i>	8.0	11.9	8.4
average	8.5	8.7	9.1 (8.0)
RMSD	0.9	1.6	2.1 (2.7)
MAD	1.7	3.9	4.0 (5.4)

<sup>a</sup>See text.

0.1 to 0.15 and 0.2 Å, the fluctuations in the computed energies become more pronounced: the RMSD and MAD values rise from 0.9 to 1.6 and 2.1 kcal/mol and from 1.7 to 3.9 and 4.0 kcal/mol, respectively. The large fluctuations for the 0.2 Å displacements suggest that low-resolution crystal structures should be used with great caution when designing cluster models. In such cases, it would seem advisable to choose large QM-only models to reduce the potential error introduced by the coordinate uncertainties at the fixed outer atoms.

It should be pointed out that the geometry optimizations for structures 6 and 7 with 0.2 Å displacements encountered problems with multiple local minima. Our standard procedure of reoptimization of the previously published structures<sup>26</sup> (with modified positions of the fixed atoms) led to conformations of **Int1** and **Int2** with dihedral angles C–C<sub>α</sub>–C<sub>β</sub>–S of Cys141 between 162° and 167° (see Supporting Information, Tables S8–S9) that are much larger than the reference values (*Original*) of 91° and 110° for structures 6 and 7, respectively. Upon re-examination of the potential energy surface, we found alternative local minima of **Int1** and **Int2** with dihedral angles C–C<sub>α</sub>–C<sub>β</sub>–S of Cys141 between 76° and 99° (see Supporting

Information, Tables S8–S9) that are much closer to the reference values (see above). The computed relative energies (**Int2** vs **Int1**) for structures 6 and 7 are quite low for the conformers with large dihedral angles (4.4 and 2.6 kcal/mol) and are in the usual range for the conformers with the more standard dihedral angles (11.1 and 7.2 kcal/mol). Since we have no clear-cut technical preference for either of the two types of local minima, we report relative energies and statistical data for both in Table 3 (values derived from the conformers with large dihedral angles in parentheses). Further potential energy surface scans were performed for Structure2 with 0.2 Å displacements and Structure10 with 0.15 Å displacements, the energies of which deviate quite strongly from the original reference value. These calculations confirmed that the obtained structures belong to the same local minimum.

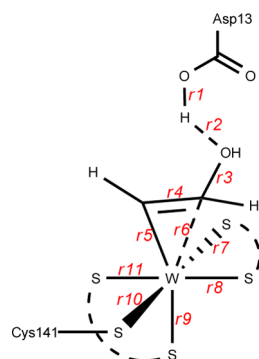
**III.3. Effects on the Optimized Geometries.** The structure of the key intermediate **Int2** was used to analyze the geometric changes due to varying constraints. Eleven important bond distances were selected for comparison, including seven W-related coordination distances. They are listed in Table 4 (for detailed data for each structure see Supporting Information, Tables S5–S7). The trends for the geometric changes are consistent with those for the energy changes (Table 3). For the 0.1 Å displacement, the RMSD values for the differences in the computed distances (displaced structures vs *Original*) are all below 0.01 Å, except for *r2*, the hydrogen-bond distance between Asp13 and the hydroxyl group of the vinyl anion (RMSD 0.015 Å, MAD 0.029 Å). When the displacements are extended, the differences become larger, of course. For the 0.2 Å displacement, there are four (six) bonds with RMSD greater than 0.01 Å, and the MAD value amounts to 0.05 Å for *r2* (0.10 Å for *r10*, data in parentheses refer to the conformers of structures 6 and 7 with large dihedral angle, see above). However, these MAD values are still much smaller than the one reported for thermolysin (0.23 Å).<sup>12</sup> Overall, the changes of the bond distances upon varying the constraints on the fixed atoms thus remain fairly small in **Int2**.

As discussed in Section III.2, the most important qualitative variation in the optimized geometries concerns the dihedral angle C–C<sub>α</sub>–C<sub>β</sub>–S of Cys141, which in 2 out of the 30 cases considered can adopt a much larger value than usual that corresponds to a different type of local minimum (see Table S9, Supporting Information, for a complete list of numerical values for **Int2**). The trends in the fluctuations of this dihedral angle are analogous to those for the relative energy and the bond lengths. When increasing the displacements from 0.1 to 0.15 and 0.2 Å, the RMSD value rises from 6.2° to 10.2° and 13.5° (37.4°). The very large RMSD value of 37.4° reflects the multiple minimum problem encountered for the 0.2 Å displacement (see Section III.2).

## IV. CONCLUSIONS

QM-only studies of enzymatic reactions normally employ a coordinate-locking scheme, in which certain key atoms at the periphery of the chosen model system are fixed to their crystal structure positions. We have assessed the uncertainties introduced by this scheme, using acetylene hydratase as a representative example. For this purpose, random displacements of 0.1, 0.15, and 0.2 Å were applied at the ten terminal atoms fixed in the chosen cluster model to generate sets of ten distorted structures for each displacement. The relevant stationary points were reoptimized under these modified

**Table 4. Important Distances (in Å) of Int2 for the Original Structure and 10 Structures with 0.1, 0.15, and 0.2 Å Random Displacements at the Fixed Atoms<sup>a</sup>**



		r1	r2	r3	r4	r5	r6	r7	r8	r9	r10	r11
	original	1.059	1.395	1.479	1.403	1.933	2.167	2.517	2.568	2.562	2.510	2.387
0.1 Å	average	1.056	1.402	1.478	1.402	1.933	2.169	2.517	2.571	2.562	2.509	2.386
	RSMD	0.004	0.015	0.002	0.001	0.001	0.005	0.004	0.008	0.005	0.007	0.005
	MAD	0.008	0.029	0.003	0.002	0.002	0.007	0.008	0.020	0.012	0.018	0.009
0.15 Å	average	1.057	1.405	1.479	1.402	1.933	2.166	2.519	2.579	2.564	2.508	2.385
	RSMD	0.006	0.020	0.002	0.001	0.001	0.005	0.004	0.017	0.007	0.008	0.007
	MAD	0.009	0.038	0.004	0.002	0.002	0.007	0.006	0.035	0.011	0.014	0.014
0.2 Å	average	1.057	1.409	1.480	1.403	1.933	2.166	2.515	2.569	2.566	2.507	2.386
	RSMD	0.007	0.027	0.003	0.002	0.001	0.009	0.010	0.017	0.014	0.019	0.010
	MAD	0.013	0.050	0.006	0.003	0.002	0.018	0.022	0.037	0.025	0.042	0.020

<sup>a</sup>For the 0.2 Å displacements, the statistical data were evaluated using structures 6 and 7 with standard dihedral angles.

constraints to determine the variations of the computed energies and geometries caused by displacements of the fixed atoms.

For random displacements of 0.1 Å, the relative energies of the stationary points show rather small fluctuations around the reference values at the *Original* geometries, as indicated by RMSD (MAD) values of about 1 (2) kcal/mol for **TS1**, **Int2**, and **TS2** and 1.7 (4.0) kcal/mol for **Int3** (Table 2). As expected, the fluctuations increase for larger displacements. This has been studied mainly for **Int2**. For each of the three chosen displacements, the average energy (10 structures) of **Int2** relative to **Int1** remains close to the *Original* one, with deviations of less than 1 kcal/mol, but the RMSD (MAD) values roughly double from 0.9 (1.7) to 2.1 (4.0) kcal/mol when increasing the random displacements from 0.1 to 0.2 Å (Table 3). Likewise, the fluctuations in the computed bond distances become much more pronounced, with the largest RMSD (MAD) values rising from 0.015 (0.019) to 0.027 (0.050) Å (Table 4). For the 0.2 Å displacements, we find convergence to different types of local minima for two out of the ten distorted structures. The statistical data given above refer to a consistent set of minimum structures with similar shape; the RMSD (MAD) values increase to 2.7 (5.4) kcal/mol and 0.048 (0.102) Å when evaluated with the data from the other two local minima (i.e., they roughly triple compared with the 0.1 Å displacements).

In the chosen 116-atom model of acetylene hydratase,<sup>26</sup> the ten fixed atoms are not directly connected to the reaction center but separated by at least two covalent bonds. In this case, displacements of 0.1 Å for the fixed atoms seem to cause a relatively minor perturbation that can be accommodated during geometry optimization, resulting in rather small changes in key bond distances and relative energies. On the other hand, displacements of 0.2 Å for the fixed atoms lead to significantly greater variations that would commonly be considered too large

to be acceptable and may even cause convergence to different local minima. One major reason for these rather large variations may be that one of the fixed *C $\alpha$*  atoms belongs to the Asp13 residue that is directly involved in both proton transfer reactions.

Generalizing our findings from the present case study on acetylene hydratase, we note that crystal structures with a resolution higher than 2.0 Å are normally associated with a coordinate error less than 0.1 Å for the backbone atoms (see Introduction). These uncertainties should be tolerable in the design of QM-only models with more than 100 atoms, which are flexible enough to adapt during geometry optimization, so that the associated uncertainties in the computed energies and bond distances are expected to be of the order of 1 kcal/mol and 0.01 Å. Crystal structures with significantly lower resolution should be used with great caution when setting up QM-only models—in such cases, it is advisable to use very large cluster models, with the fixed atoms further apart from the reaction center to minimize their influence.

On the technical side, the comparison of using different basis sets suggests that a medium basis set (**BS1**) is sufficient to capture the energy fluctuations due to varying constraints. However, it should be pointed out that the computation of more reliable energies requires the use of larger basis sets, like **BS2** or def2-TZVPPD, which are affordable for the model size used here (116 atoms). Single-point calculations with the  $\omega$ B97X, BMK, and M06 functionals give similar barriers for both reaction steps, with deviations from the B3LYP values of less than 3 kcal/mol (Table 1). Furthermore, the displacement-induced energy fluctuations are also similar for  $\omega$ B97X and BMK and somewhat more pronounced for M06.

## ■ ASSOCIATED CONTENT

### ■ Supporting Information

Individual data on crystal structures and additional results on relative energies and geometries (as described in the text), procedure for the preparation of structures with displacements, and Cartesian coordinates for selected structures. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## ■ AUTHOR INFORMATION

### Corresponding Author

\*E-mail: [thiel@kofo.mpg.de](mailto:thiel@kofo.mpg.de). Tel.: +49(0)2083062150.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was supported by the Max Planck Society and the Volkswagenstiftung.

## ■ REFERENCES

- (1) Siegbahn, P. E. M.; Blomberg, M. R. A. Transition-Metal Systems in Biochemistry Studied by High-Accuracy Quantum Chemical Methods. *Chem. Rev.* **2000**, *100*, 421–438.
- (2) Siegbahn, P. E. M.; Borowski, T. Modeling Enzymatic Reactions Involving Transition Metals. *Acc. Chem. Res.* **2006**, *39*, 729–738.
- (3) Himo, F. Quantum Chemical Modeling of Enzyme Active Sites and Reaction Mechanisms. *Theor. Chem. Acc.* **2006**, *116*, 232–240.
- (4) Ramos, M. J.; Fernandes, P. A. Computational Enzymatic Catalysis. *Acc. Chem. Res.* **2008**, *41*, 689–698.
- (5) Himo, F.; Siegbahn, P. E. M. Recent Developments of the Quantum Chemical Cluster Approach for Modeling Enzyme Reactions. *J. Biol. Inorg. Chem.* **2009**, *14*, 643–651.
- (6) Siegbahn, P. E. M.; Himo, F. The Quantum Chemical Cluster Approach for Modeling Enzyme Reactions. *WIREs Comput. Mol. Sci.* **2011**, *1*, 323–336.
- (7) Becke, A. D. Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (8) Sevastik, R.; Himo, F. Quantum Chemical Modeling of Enzymatic Reactions: The Case of 4-Oxalocrotonate Tautomerase. *Bioorg. Chem.* **2007**, *35*, 444–457.
- (9) Hopmann, K. H.; Himo, F. Quantum Chemical Modeling of the Dehalogenation Reaction of Haloalcohol Dehalogenase. *J. Chem. Theory Comput.* **2008**, *4*, 1129–1137.
- (10) Georgieva, P.; Himo, F. Quantum Chemical Modeling of Enzymatic Reactions: The Case of Histone Lysine Methyltransferase. *J. Comput. Chem.* **2010**, *31*, 1707–1714.
- (11) Liao, R.-Z.; Yu, J.-G.; Himo, F. Quantum Chemical Modeling of Enzymatic Reactions: The Case of Decarboxylation. *J. Chem. Theory Comput.* **2011**, *7*, 1494–1501.
- (12) Pelmenchikov, V.; Blomberg, M. R. A.; Siegbahn, P. E. M. A Theoretical Study of the Mechanism for Peptide Hydrolysis by Thermolysin. *J. Biol. Inorg. Chem.* **2002**, *7*, 284–298.
- (13) Pelmenchikov, V.; Siegbahn, P. E. M. Catalytic Mechanism of Matrix Metalloproteinases: Two-Layered ONIOM Study. *Inorg. Chem.* **2002**, *41*, 5659–5666.
- (14) Siegbahn, P. E. M. The Catalytic Cycle of Catechol Oxidase. *J. Biol. Inorg. Chem.* **2004**, *9*, 577–590.
- (15) Siegbahn, P. E. M. The Effect of Backbone Constraints: The Case of Water Oxidation by the Oxygen-Evolving Complex in PSII. *ChemPhysChem* **2011**, *12*, 3274–3280.
- (16) Chen, S.-L.; Fang, W.-H.; Himo, F. Technical Aspects of Quantum Chemical Modeling of Enzymatic Reactions: The Case of Phosphotriesterase. *Theor. Chem. Acc.* **2008**, *120*, 515–522.
- (17) Luzzati, P. V. Traitement Statistique des Erreurs Dans la Détermination des Structures Cristallines. *Acta Crystallogr.* **1952**, *5*, 802–810.
- (18) Read, R. J. Improved Fourier Coefficients for Maps Using Phases from Partial Structures with Errors. *Acta Crystallogr.* **1986**, *A42*, 140–149.
- (19) Cruickshank, D. W. J. Remarks about Protein Structure Precision. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1999**, *D55*, 583–601.
- (20) Stec, B.; Zhou, R.; Teeter, M. M. Full-Matrix Refinement of the Protein Crambin at 0.83 Å and 130 K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1995**, *D51*, 663–681.
- (21) Tickle, I. J.; Laskowski, R. A.; Moss, D. S. Error Estimates of Protein Structure Coordinates and Deviations from Standard Geometry by Full-Matrix Refinement of  $\gamma$ B- and  $\beta$ B2-Crystallin. *Acta Crystallogr.* **1998**, *D54*, 243–252.
- (22) Sandalova, T.; Schneider, G.; Käck, H.; Lindqvist, Y. Structure of Dethiobiotin Synthetase at 0.97 Å Resolution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1999**, *D55*, 610–624.
- (23) Harata, K.; Abe, Y.; Muraki, M. Full-Matrix Least-Squares Refinement of Lysozymes and Analysis of Anisotropic Thermal Motion. *Proteins: Struct., Funct., Genet.* **1998**, *30*, 232–243.
- (24) Sauter, C.; Otálora, F.; Gavira, J.-A.; Vidal, O.; Giegé, R.; Garcia-Ruiz, J. M. Structure of Tetragonal Hen Egg-white Lysozyme at 0.94 Å from Crystals Grown by the Counter-diffusion Method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2001**, *D57*, 1119–1126.
- (25) Harata, K.; Abe, Y.; Muraki, M. Crystallographic Evaluation of Internal Motion of Human Alpha-Lactalbumin Refined by Full-Matrix Least-Squares Method. *J. Mol. Biol.* **1999**, *287*, 347–358.
- (26) Liao, R. Z.; Yu, J. G.; Himo, F. Mechanism of Tungsten-Dependent Acetylene Hydratase from Quantum Chemical Calculations. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 22523–22527.
- (27) Liao, R. Z.; Thiel, W. Comparison of QM-Only and QM/MM Models for the Mechanism of Tungsten-Dependent Acetylene Hydratase. *J. Chem. Theory Comput.* **2012**, *8*, 3793–3803.
- (28) Liao, R. Z.; Himo, F. Theoretical Study of the Chemoselectivity of Tungsten-Dependent Acetylene Hydratase. *ACS Catal.* **2011**, *1*, 937–944.
- (29) Liu, Y.-F.; Liao, R.-Z.; Ding, W.-J.; Yu, J.-G.; Liu, R.-Z. Theoretical Investigation of the First-Shell Mechanism of Acetylene Hydration Catalyzed by a Biomimetic Tungsten Complex. *J. Biol. Inorg. Chem.* **2011**, *16*, 745–752.
- (30) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A. et al. *Gaussian 09*, revision B.01; Gaussian, Inc.: Wallingford CT, 2009.
- (31) Roy, L. E.; Hay, P. J.; Martin, R. L. Revised Basis Sets for the LANL Effective Core Potentials. *J. Chem. Theory Comput.* **2008**, *4*, 1029–1031.
- (32) Rappoport, D.; Furche, F. Property-Optimized Gaussian Basis Sets for Molecular Response Calculations. *J. Chem. Phys.* **2010**, *133*, 134105.
- (33) Chai, J.-D.; Head-Gordon, M. Systematic Optimization of Long-Range Corrected Hybrid Density Functionals. *J. Chem. Phys.* **2008**, *128*, 084106.
- (34) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Non-covalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Functionals. *Theor. Chem. Acc.* **2008**, *120*, 215–241.
- (35) Boese, A. D.; Martin, J. M. L. Development of Density Functionals for Thermochemical Kinetics. *J. Chem. Phys.* **2004**, *121*, 3405–3416.
- (36) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, Structures, and Electronic Properties of Molecules in Solution with the C-PCM Solvation Model. *J. Comput. Chem.* **2003**, *24*, 669–681.
- (37) Gaussian03 and Gaussian09 differ in the CPCM implementation; therefore, the solvation effects reported here are slightly different from those in ref 26.

(38) Reiher, M.; Salomon, O.; Hess, B. A. Reparametrization of Hybrid Functionals Based on Energy Differences of States of Different Multiplicity. *Theor. Chem. Acc.* **2001**, *107*, 48–55.

(39) Tao, J. M.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. Climbing the Density Functional Ladder: Nonempirical Meta-Generalized Gradient Approximation Designed for Molecules and Solids. *Phys. Rev. Lett.* **2003**, *91*, 146401.

(40) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. Development and Assessment of a New Hybrid Density Functional Model for Thermochemical Kinetics. *J. Phys. Chem. A* **2004**, *108*, 2715–2719.