Towards accurate QM/MM reaction barriers with large QM regions using Domain Based Pair Natural Orbital Coupled Cluster Theory

Giovanni Bistoni,*[1] Iakov Polyak,[1,2] Manuel Sparta,[1,3] Walter Thiel,[1] Frank Neese*[1]

[1]Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz, D-45470 Mülheim an der Ruhr, Germany. Email: giovanni.bistoni@kofo.mpg.de, frank.neese@kofo.mpg.de

ABSTRACT. The hydroxylation reaction catalyzed by p-hydroxybenzoate hydroxylase and the Baeyer-Villiger reaction catalyzed by cyclohexanone monooxygenase are investigated by means of quantum mechanical/molecular mechanical (QM/MM) calculations at different levels of QM theory. The geometries of the stationary points along the reaction profile are obtained from QM/MM geometry optimizations, in which the QM region is treated by density functional theory (DFT). Relative energies are determined from single-point QM/MM calculations using the domain-based local pair natural orbital coupled cluster DLPNO-CCSD(T) method as QM component. The results are compared with single-point DFT/MM energies obtained using popular density functionals and with available experimental and computational data. It is found that the choice of the QM method strongly affects the computed energy profiles for these reactions. Different density functionals provide qualitatively different energy barriers (variations of the order of 10 kcal/mol in both reactions), thus limiting the confidence in DFT/MM computational predictions of energy profiles. On the other hand, the use of the DLPNO-CCSD(T) method in conjunction with large QM regions and basis sets makes it

[2]Current Address: School of Chemistry, Cardiff University, Cardiff, CF10 3AT, UK

[3]Current Address: Teknova AS, Tordenskjolds gate 9, NO-4612 Kristiansand, NO

possible to achieve high accuracy. A critical discussion of all the technical aspects of the calculations is given with the aim of aiding computational chemists in the application of the DLPNO-CCSD(T) methodology in QM/MM calculations.

1. INTRODUCTION

Enzymes are natural catalysts with outstanding efficiency and selectivity. A thorough understanding of the mechanism of enzymatic reactions can open the route to the development of artificial catalysts with improved performance. This is of great interest in pharmaceutical, chemical, and biotechnology industries.

For the modeling of enzymatic reactions, hybrid quantum mechanical/molecular mechanical (QM/MM) approaches have become the method of choice. [1-2] In these approaches, an accurate QM method is used for describing the region of the system in which the chemical transformation occurs and the remainder is treated using an established force field. Hence, the accuracy of QM/MM calculations is limited by: (i) the level of theory employed for the QM region; (ii) the level of theory used in the description of the MM part; (iii) the QM/MM coupling scheme adopted.

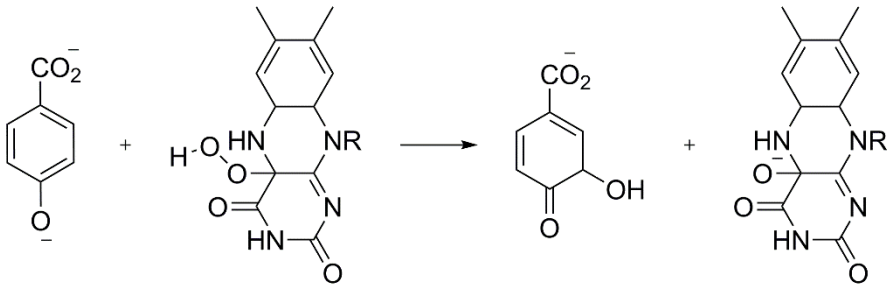
The first issue was addressed already about 10 years ago.[3-4] The accuracy of computed enzymatic reaction barriers was found to benefit strongly from the use of highly correlated techniques as QM components,[3-4] in particular of a local variant of the singles and doubles coupled cluster method with perturbative treatment of triple excitations, CCSD(T).[5] Using a QM region of 49 atoms and a triple-ζ quality basis set, the reaction barrier for the hydroxylation reaction in the active site of *p*-hydroxybenzoate hydroxylase (PHBH) was calculated in good agreement with available experimental data.[3-4, 6-10]

Over the last decade, local correlation methods have made great progress in exploiting the short range nature of electron correlation,[5, 11-19] allowing the inherent steep scaling of correlated wavefunction based methods to be reduced to linear or near-linear scaling. Our group has

contributed with the development of the domain-based local pair natural orbital coupled cluster method with single, double, and perturbatively included triple excitations, DLPNO-CCSD(T). [20-26] This approach allows one to perform coupled cluster calculations for systems with hundreds of atoms while retaining about 99.9% of the canonical CCSD(T) correlation energy.[27] Single-point energy calculations for entire proteins using more than 12000 basis functions were shown to be possible.[26] Hence, the use of this technique in QM/MM studies of large proteins and enzymes can potentially open the way to a more reliable description of biochemical processes.

Herein, QM/MM energy barriers in enzymatic reactions are computed using the DLPNO-CCSD(T) method as the QM component. Results are compared with available experimental data and with those obtained using Density Functional Theory (DFT) as the QM component. The convergence of the results with respect to the basis set size, the size of the QM region, and other technical aspects of the calculations is also discussed.

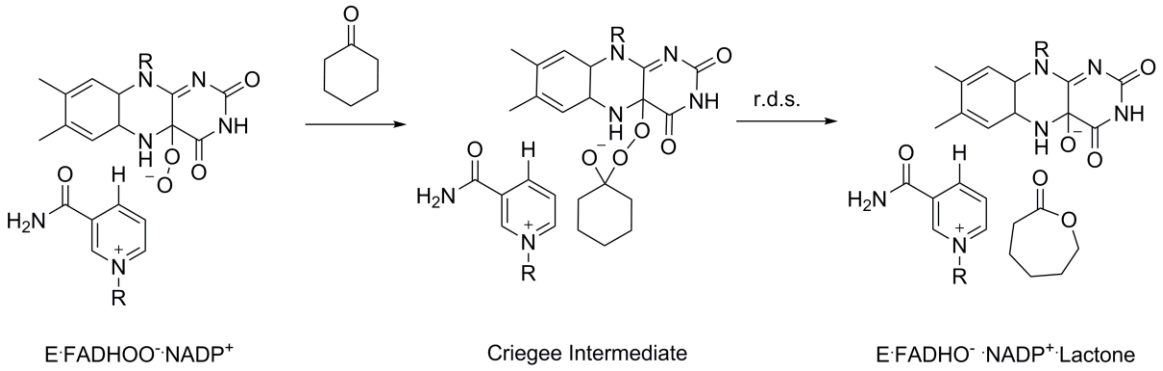
For this study two enzymes were selected, i.e. the already mentioned PHBH and cyclohexanone monooxygenase (CHMO).[28] PHBH is a flavoprotein monooxygenase that catalyzes the hydroxylation of p-hydroxybenzoate by the cofactor flavin hydroperoxide



Scheme 1: Hydroxylation reaction in the active site of PHBH.

(FADHOOH), as shown in Scheme 1. This reaction has already been the subject of a wide range of experimental[7, 9-10] and computational[3-4, 29-33] investigations. Experimental values for barrier of the rate-determining step are available.[3-4, 6-10] The present DLPNO-CCSD(T) predictions will be compared with available experimental data and previously reported theoretical results.[3-4]

The second reaction involves CHMO,[28] which is a prominent member of the Baeyer-Villiger monooxygenase enzyme family,[34, 35] responsible for conversion of ketones into esters and lactones. Due to its high stereoselectivity towards a rather wide range of valuable reactants, CHMO has been intensely studied experimentally[36-37] and continues to be subjected to directed evolution techniques to further improve or change its catalytic properties.[38] In particular, it catalyzes the O_2-driven oxidation of cyclohexanone into lactone (see Scheme 2).



Scheme 2: Cyclohexanone monooxygenation in the active site of CHMO. The second reaction is the rate-determining step (r.d.s.).

Previously published QM/MM calculations[39] suggested that this reaction proceeds via formation of a Criegee intermediate, with the following migration step being rate-determining; in this earlier QM/MM study, energy profiles were computed using DFT as the QM component and a QM region of 99 atoms. Here we present coupled-cluster based QM/MM results for this reaction that can be used as a valuable benchmark reference for lower levels of theory. We also address the convergence of the results with the size of the QM region and with other technical aspects of the calculations.

The paper is organized as follows. Computational details are given in Section 2. In Section 3 we discuss the present QM/MM results for the PHBH- and CHMO-catalyzed reactions. Conclusions are offered in Section 4.

2. COMPUTATIONAL DETAILS

The general QM/MM setup and procedures were the same as in previous QM/MM studies of these enzymes (see Refs. 3, 4, 29 for PHBH and Ref. 39 for CHMO). Single-point QM/MM energies were evaluated in an additive scheme as the sum of the QM energy of the QM region in the point-charge field of the MM part (electrostatic embedding) and of the MM term comprised of the MM energy of the MM region and the non-electrostatic QM/MM interaction energy. All single-point QM energy calculations were carried out with a development version of the ORCA suite of programs.[40-41] The MM term was either available from previous work[4,39] or evaluated using the ChemShell software.[42-44] The QM/MM methodology applied in the geometry optimizations has been described previously.[3-4, 39] **2.1 Hydroxylation of p-hydroxybenzoate**

Geometries for the reactant and transition states (see Scheme 1) optimized at the DFT(B3LYP/TZVP)/GROMOS level were taken from Ref. 4. A total of 10 snapshots were used for both reactant and transition states. The snapshots are labeled by 1-10, consistent with the previous work.[4] Single-point QM/MM calculations using the DLPNO-CCSD/TightPNO [26-27] method as QM component were carried out at the DFT(B3LYP/TZVP)/GROMOS geometries. The original QM region, [3-4] including all the atoms shown in Scheme 1, was used (49 atoms). The MM point charges from the remaining protein environment were included in the calculations (about 20000 point charges).

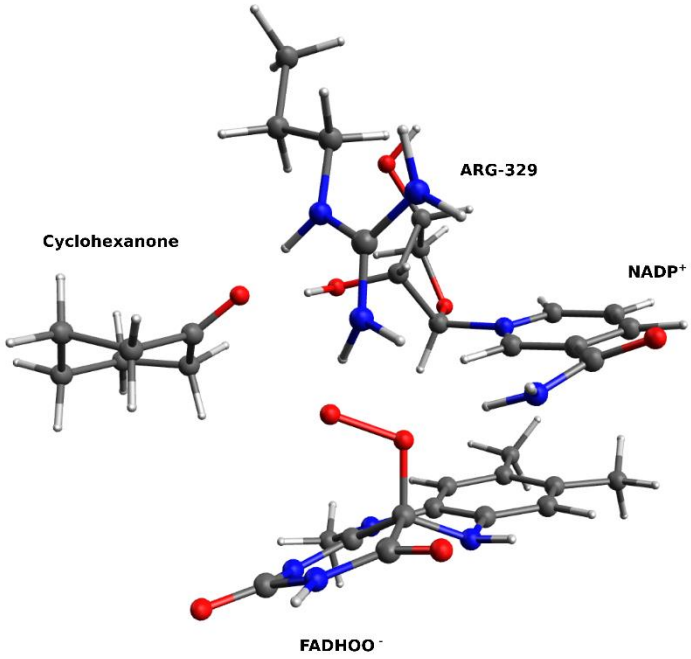
The perturbative triples corrections were computed using two different approximations. The (T0) correction relies on the so-called semi-canonical approximation, which neglects the couplings between different triples by the off-diagonal Fock matrix elements.[25] The (T) correction refers to the recently implemented iterative algorithm.[45] In the following, the DLPNO-CCSD method using the (T0) correction is denoted as DLPNO-CCSD(T0) whereas the DLPNO-CCSD method with the iterative triples correction is abbreviated as DLPNO-CCSD(T). In all cases, the resolution of identity (RI) approximation[46] for both Coulomb- and

exchange-type integrals (RIJK) was used in the Hartree-Fock (HF) part. Augmented correlation-consistent polarized valence basis sets of triple-ζ quality, namely aug-cc-pVTZ, [47-48] were used in conjunction with matching auxiliary basis sets (aug-cc-pVTZ/C and aug-cc-pVTZ/JK). See Ref. 4 for a discussion regarding the basis set convergence of the coupled cluster energies for the same reaction.

Single-point DFT/MM calculations were carried out using the BLYP, B3LYP, and M062X exchange-correlation functionals.[49] For BLYP and B3LYP the RI approximation was employed and the –D3 correction of Grimme was included[50-51] with Becke-Johnson damping.[52] The Ahlrichs def2-QZVP basis set [53] in conjunction with matching auxiliary basis sets was used in all cases.

2.2 Cyclohexanone monooxygenation

Starting from the X-ray structural data,[54] previously published DFT/MM calculations elucidated the mechanism of cyclohexanone oxygenation in the active site of CHMO.[39, 55] Four relevant stationary points for the reaction of wild-type (WT) CHMO with cyclohexanone were located:[38] the reactant state, the Criegee intermediate, the transition state for the rate-determining migration step, and the product state (see Scheme 2). In this work, the energy of these stationary points was calculated by means of QM/MM calculations at different levels of QM theory. The original[38] QM region (99 atoms, labeled as R1 in the following) as well as two larger QM regions of approximately double (R2) and triple (R3) size were employed.



Scheme 3: Reactant state structure (QM region R1 only) optimized at the QM(B3LYP/TZVP)/CHARMM level.

QM region R1 comprised the cyclohexanone substrate, all atoms from the isoalloxazine ring of the peroxyflavin, the side chain of Arg329, the nicotinamide ring, and the adjacent ribose of the $NADP^+$ cofactor (See Scheme 3). QM region R2 comprised all atoms included in R1, plus the side chains and parts of the backbones of the amino acid residues positioned closest to cyclohexanone in the active region (Phe434, Thr435, Asn436, Leu437, Leu145, Leu146, and Trp488), with 206 atoms in total (including link atoms). The overall charge of R1 and R2 was +1.[38] QM region R3 comprised three more residues surrounding the substrate (Phe248, Phe279, and Phe507), two residues interacting with Arg329 - Asp59 and Pro330 as well as a water molecule connecting Asp59 and Pro330 via two H-bonds. The overall number of atoms in R3 is 307 (including link atoms), and the overall charge of the QM region was reduced to 0 due to incorporation of the negatively charged Asp59 residue. Note that the definition of QM regions R2 and R3 involved cuts in the backbone through the polar peptide bonds (one in R2 and two in R3), which is usually avoided in QM/MM calculations. However, the rather large size of

these QM regions and the use of the charge shift scheme for the link atoms diminished possible overpolarization effects.

Geometry optimizations were carried out at the DFT(B3LYP/TZVP)/CHARMM level, as described in Ref. 39. QM/MM geometries for QM region R1 were taken from Ref. 39 whilst QM/MM geometries for QM region R2 were optimized herein at the same level of theory. QM/MM energies for QM region R3 were obtained at the geometries optimized with QM region R2.

For each definition of the QM region, single-point energies were computed at the corresponding DFT(B3LYP/TZVP)/CHARMM geometries employing DLPNO-CCSD(T) and DLPNO-CCSD(T0) as QM components. In all cases, TightPNO settings were used. The RIJK approximation was applied in the HF part. The Ahlrichs def2-TZVPP and def2-QZVPP basis sets were used in conjunction with matching auxiliary basis sets. The MM point charges from the remaining protein environment were included in the calculations (about 40000 point charges, depending on the size of the QM region).

Single-point DFT/MM calculations were carried out at the same geometries using the BLYP, B3LYP, and M062X exchange-correlation functionals. For BLYP and B3LYP the RI approximation was employed and the –D3 correction of Grimme was included (BLYP-D3 and B3LYP-D3) with Becke-Johnson damping.[52] The Ahlrichs def2-TZVPP basis set in conjunction with matching auxiliary basis sets was used in all cases.

RESULTS AND DISCUSSION

3.1 PHBH

For the hydroxylation reaction in the active site of PHBH we considered a total of 10 snapshots optimized at the DFT(B3LYP/TZVP)/GROMOS level for both reactant and transition states (see Computational Details). The energy of each of these snapshots was computed by means of single-point QM/MM calculations using different QM methods. The resulting energy barriers are reported in Table 1. The DLPNO-CCSD(T) results employing the

iterative algorithm for the triples correction are the most accurate ones and can be taken as reference.

Consistent with previously published results,[4] dynamic electron correlation was found to play a fundamental role in this system. On average, DLPNO-CCSD(T) barriers are smaller than their HF counterparts by about 23 kcal/mol. The perturbative triples correction contributes significantly to this energy difference, with DLPNO-CCSD(T) barriers being on average 8.1 kcal/mol lower in energy than the DLPNO-CCSD barriers.

In the large majority of cases studied up to now, the gain in accuracy of DLPNO-CCSD(T) with respect to DLPNO-CCSD(T0) for relative energies is modest.[45] However, for the reaction under study, the large contribution of the triples correction to the computed energy barriers makes the results very sensitive to approximations in the treatment of the triples correction. In fact, the DLPNO-CCSD(T) method provides notably smaller energy barriers with respect to DLPNO-CCSD(T0). The average difference is 1.1 kcal/mol, demonstrating that the use of the iterative algorithm provides a non-negligible gain in accuracy in this case.

The previously published LCCSD and LCCSD(T0) results are given in parentheses for comparison.[4] DLPNO-CCSD(T0) and LCCSD(T0) results are very close to each other, with an average deviation of only 0.4 kcal/mol. Considering the different approximations involved in the two approaches, which rely on completely different implementations, these results demonstrate the robustness of different local correlation schemes for the system under study.

It is interesting to compare these results with the ones obtained with the dominant computational methodology used as QM component in QM/MM calculations, i.e. with DFT. Three functionals were chosen as representative examples of the ones commonly adopted in QM/MM simulations, i.e. a pure GGA functional (BLYP), a popular hybrid functional (B3LYP), and the meta-hybrid M062X. BLYP and B3LYP energies were corrected by using Grimme's –D3 correction.

The reaction barriers computed at the DFT/MM level are reported in Table 2. A graphical representation of these results is given in Figure 1, which also includes the coupled cluster results for comparison. The horizontal red line in the figure marks the experimental value of 13.3 kcal/mol. This value has been obtained from the experimental activation enthalpy of 12 kcal/mol by subtracting the computed ZPE correction of -1.3 kcal/mol, as detailed in Ref. 3.

Table 1. QM/MM activation barriers (kcal/mol) calculated at different levels of QM theory using the aug-cc-pVTZ basis set. Results do not include the ZPE correction. All reported values include the effect of the MM environment (taken from Ref. 4). The corresponding values obtained from Ref. 4 with a different local coupled cluster scheme are reported in parentheses.

Entry	HF	DLPNO-CCSD	DLPNO-CCSD(T0)	DLPNO-CCSD(T)
1	36.0	21.1 (20.4)	14.5 (13.8)	13.3
2	41.3	24.4 (24.0)	16.9 (16.1)	15.6
3	38.8	23.6 (23.2)	16.4 (16.2)	15.2
4	39.3	23.0 (22.7)	15.8 (15.9)	14.5
5	35.9	20.5 (19.9)	13.8 (13.4)	12.8
6	38.8	22.1 (21.5)	14.9 (14.4)	13.8
7	32.1	17.6 (16.8)	11.6 (10.9)	10.7
8	37.8	21.4 (20.8)	14.3 (14.4)	13.2
9	39.8	22.9 (22.3)	15.5 (14.5)	14.3
10	41.4	24.3 (23.6)	16.6 (16.4)	15.3
average	38.1	22.1 (21.5)	15.0 (14.6)	13.9

Table 2. QM/MM activation barriers (kcal/mol) calculated using different density functionals as QM component. The def2-QZVP basis set was used. Results do not include the ZPE correction. All reported values include the effect of the MM environment (taken from Ref. 4). For completeness, the corresponding values obtained without the -D3 correction are reported in parentheses.

Entry	BLYP-D3	B3LYP-D3	M06-2X
1	-1.7 (-0.6)	7.7 (8.6)	21.5
2	-0.2 (1.6)	9.9 (11.4)	24.4
3	-0.4 (0.7)	9.4 (10.2)	23.8
4	-1.2 (0.5)	8.7 (10.1)	23.1
5	-2.4 (-0.8)	7.1 (8.4)	20.8
6	-1.3 (0.5)	8.4 (9.9)	22.3
7	-3.2 (-1.7)	5.5 (6.7)	18.0
8	-2.3 (-0.5)	7.5 (9.0)	21.5
9	-0.8 (1.1)	9.0 (10.5)	23.0
10	-0.5 (1.3)	9.7 (11.1)	24.2
average	**-1.4 (0.2)**	**8.3 (9.6)**	**22.3**

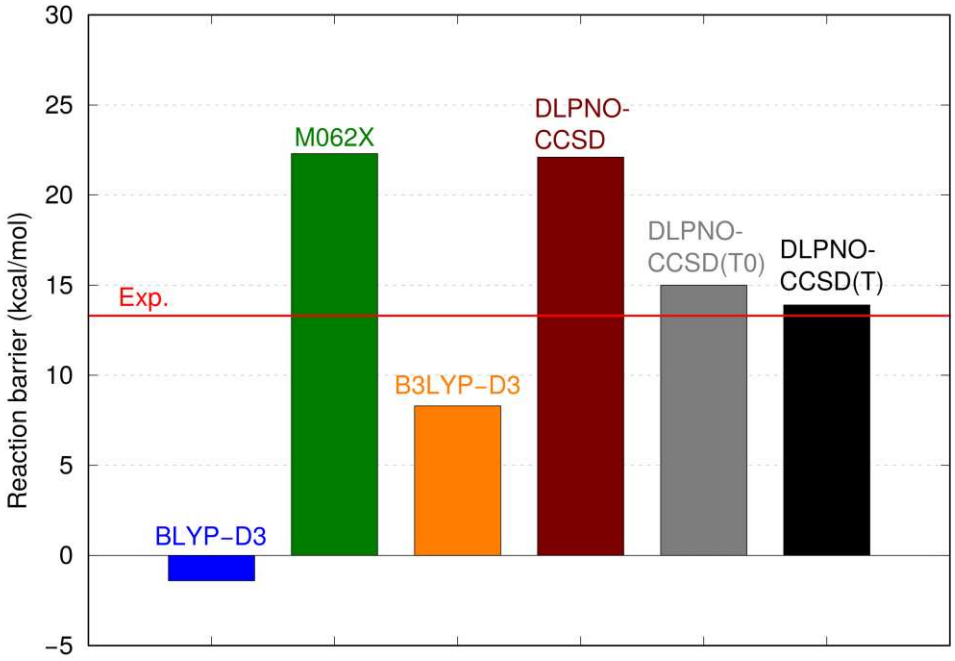


Figure 1 QM/MM reaction barriers for the hydroxylation reaction in the active site of PHBH at various level of QM theory. The results are obtained by averaging the values obtained from 10 different snapshots (see Tables 1 and 2). The horizontal red line marks the available experimental value.

The first eye-catching feature of Figure 1 is the wide range obtained in the computed energy barriers when different density functionals are employed. BLYP-D3 predicts the wrong energy ordering between transition states and intermediates, resulting in a negative "energy barrier" of -1.3 kcal/mol. B3LYP-D3 performs comparatively much better, providing an energy barrier of 8.3 kcal/mol, which is 5.6 kcal/mol smaller than the DLPNO-CCSD(T) value. The popular M062X functional significantly overestimates the reaction barrier, thus performing similarly to DLPNO-CCSD. The large differences between the various DFT/MM computational predictions are probably a consequence of the fundamental role that dynamic electron correlation plays in this system. As different functionals describe the dynamic electron correlation differently, the computed reaction barriers are inherently very sensitive to the nature of the exchange-correlation functional employed in the QM treatment.

Note that the high sensitivity of QM/MM energy profiles to the density functional was recently observed for different enzymatic reactions.[56, 57] It was found that accurate energies can be obtained by means of a projector-based embedding scheme in which a small subset of atoms in the QM region is treated at the CCSD(T) level of theory. Similar results concerning the sensitivity of computational predictions to the density functional were observed for the alcohol-mediated Morita Baylis-Hillman reaction.[58] In this case, Harvey et al have shown that reliable results can be obtained by employing the DLPNO-CCSD(T) method in conjunction with an appropriate treatment of environmental effects.[59]

As a final remark, it is worth emphasizing that the computed barriers just discussed were obtained by averaging the values from a limited number of snapshots, which do not constitute a statistically meaningful Boltzmann ensemble. Moreover, the accuracy of any QM/MM calculation is limited by the accuracy of the MM force field employed and by the method used in the geometry optimizations. Nevertheless, and with all due caution, the excellent agreement between the DLPNO-CCSD(T)/MM computational prediction and the experimental value is certainly remarkable.

3.2 CHMO

The same QM methods as in the previous section were also tested for the cyclohexanone monooxygenation reaction in the active site of CHMO (Scheme 2). For this reaction, experimental data for the barriers are not available for comparison and previous computational investigations were only carried out at the DFT/MM level of theory using a QM region of 99 atoms.

Hence, we first tested the convergence of the computational results with the size of the QM region. Three different QM regions of increasing size, namely R1 (the original QM region[38] with 99 atoms), R2 and R3 (206 and 307 atoms, respectively) were defined. For each QM region, single-point DLPNO-CCSD(T0)(TightPNO/def2-TZVPP)/MM calculations were

carried out for the stationary points along the reaction profile (the reactant state, the Criegee intermediate, the transition state, and the product state). The chosen DLPNO methodology typically provides a good balance between accuracy and computational cost.[27]

Table 3. DLPNO-CCSD(T0)(TightPNO/def2-TZVPP)/MM relative energies (kcal/mol) of the stationary points for the cyclohexanone monooxygenation reaction using QM regions of different size. The energy of the reactant state is used as reference. Results do not include the ZPE correction. All reported values include the effect of the MM environment (reported in parenthesis). $\Delta^{\ddagger}E$ is the energy difference between the transition state and the Criegee intermediate.

QM region	reactant	Criegee	TS	product	$\Delta^{\ddagger}E$
R1	0.0 (0.0)	-4.5 (4.1)	8.1 (3.0)	-68.6 (2.4)	12.6[a]
R2	0.0 (0.0)	-2.0 (2.2)	10.5 (-0.3)	-71.1 (0.6)	12.5
R3	0.0 (0.0)	-1.5 (0.6)	9.9 (-0.4)	-71.9 (-0.2)	11.3

[a]The corresponding value calculated using DLPNO-CCSD(T0)/def2-QZVPP basis set is 12.7 kcal/mol.

The results of this study are shown in Table 3, which reports the relative energies of the stationary points with respect to the reactant state. In all cases, the Criege intermediate is predicted to be the lowest energy point of the reaction profile before the transition state. These results differ from the ones obtained previously by means of DFT/MM calculations, which predicted the reactant state to be lowest in energy.[39,55] We will come back to this point later when discussing the quality of DFT/MM predictions in this context.

The size of the QM region affects the energy profile. In particular, the relative energy of the Criegee intermediate with respect to the reactant state was found to be -4.5, -2.0, and -1.5 kcal/mol for R1, R2, and R3, respectively. The energy differences between the transition state and the Criegee intermediate ($\Delta^{\ddagger}E$) are 12.6, 12.5, and 11.3 kcal/mol for R1, R2, and R3, respectively. These results indicate that very large QM regions are needed if one aims at quantitative accuracy. On the other hand, the reaction profiles are qualitatively similar for all three QM regions, and the relative energies of the stationary points change by less than 1 kcal/mol when going from R2 (206 atoms) to R3 (307 atoms).

In order to test the dependence of our results with respect to the basis set size, $\Delta^{\ddagger}E$ values were also calculated using the same settings but with the def2-QZVPP basis set. This calculation was carried out only for the smallest QM region (R1, 99 atoms). It was found that the increase of the basis set size does not significantly affect the $\Delta^{\ddagger}E$ values (see Table 3), indicating that triple-ζ quality basis sets already provide reliable coupled cluster energies for this reaction.

Having established the dependence of our results on the size of the QM region and on the basis set, DLPNO-CCSD(T) results were used as a benchmark for different QM methods in QM/MM calculations. With this aim, the same methods as in the previous section were applied to calculate the cyclohexanone monooxygenation energy profile while keeping the size of the QM region fixed (R2). The results are reported in Table 4. The benchmark DLPNO-CCSD(T) values are given in bold for easy reference.

Table 4. QM/MM relative energies (kcal/mol) of the stationary points for the cyclohexanone monooxygenation reaction using different QM methods. The energy of the reactant state is used as reference. Results do not include the ZPE correction. All reported values include the effect of the MM environment. $\Delta^{\ddagger}E$ is the energy difference between the transition state and the Criegee intermediate. DLPNO-CCSD(T) results are given in bold for easy reference.

QM region	Method	reactant	Criegee	TS	product	$\Delta^{\ddagger}E$
R2	HF	0.0	15.3	49.2	-77.7	33.9
R2	DLPNO-CCSD	0.0	0.3	19.4	-72.1	19.1
R2	DLPNO-CCSD(T0)	0.0	-2.0	10.5	-71.1	12.5
R2	**DLPNO-CCSD(T)**	**0.0**	**-2.2**	**9.5**	**-71.0**	**11.7**
R2	BLYP	0.0	6.3	5.6	-67.8	-0.7
R2	BLYP-D3	0.0	3.7	0.9	-67.9	-2.8
R2	B3LYP	0.0	5.7	13.9	-71.1	8.2
R2	B3LYP-D3	0.0	3.6	10.0	-71.1	6.4
R2	M062X	0.0	-3.2	15.5	-75.2	18.7

When using the HF method as QM component, the Criegee intermediate is found to be 15.3 kcal/mol higher in energy than the reactant state, which is in striking contrast to the DLPNO-CCSD(T) results just discussed. The energy of the transition state is shifted to even higher

energies, leading to an overall $\Delta^{\ddagger}E$ value of 33.9 kcal/mol at the HF/MM level. Hence, dynamic electron correlation plays a fundamental role also in this reaction.

Although the DLPNO-CCSD/MM energies are significantly better than the HF/MM ones, the proper energy ordering between the Criegee intermediate and the reactant state is only

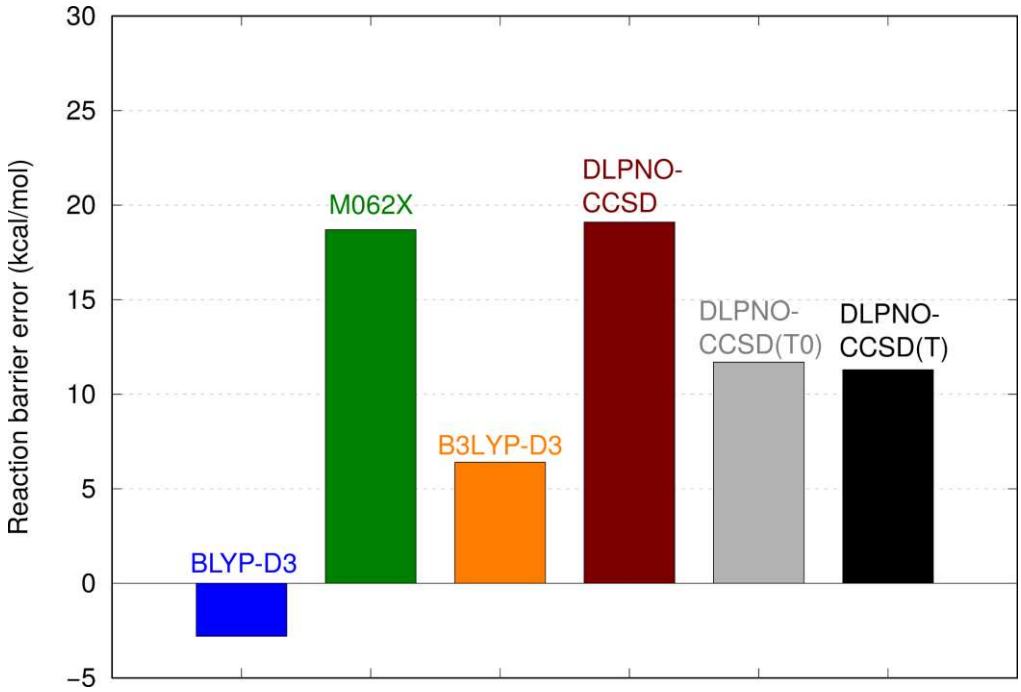


Figure 2 QM/MM reaction barriers for the cyclohexanone monooxygenation reaction in the active site of CHMO at various level of QM theory (QM region: R2).

obtained upon inclusion of the triples correction. Note that the DLPNO-CCSD(T) and DLPNO-CCSD(T0) results differ in the predicted $\Delta^{\ddagger}E$ values again by about 1 kcal/mol, consistent with the PHBH case.

It is interesting to compare the above results with those obtained at the DFT/MM level. With the only exception of M062X, all functionals predict the opposite energy ordering between the Criege intermediate and the reactant state with respect to DLPNO-CCSD(T). They also show a

very large variability in the computed $\Delta^{\ddagger}E$ values, ranging from -2.8 to 18.3 kcal/mol for BLYP-D3 and M062X, respectively. A graphical representation of these results is shown in Figure 2.

These results closely resemble our findings in the PHBH case: BLYP-D3 does not predict an energy barrier; B3LYP-D3 provides qualitatively correct results but underestimates the reaction energies by about 5 kcal/mol; M062X provides results that are similar to the DLPNO-CCSD ones, overestimating the energy barrier by 7.0 kcal/mol. Hence, the choice of the QM method is crucial for the accuracy of QM/MM predictions. In this specific case, reaction barriers are found to be much more sensitive to this choice rather than to the size of the QM region.

The present work demonstrates that the novel linear scaling DLPNO-CCSD(T) method can be effectively used to study chemical reactions in complex condensed phase systems. Single-point energy calculations at the DLPNO-CCSD(T)/MM level for fairly large QM regions can be now performed in a reasonable amount of time. As an example, the DLPNO-CCSD(T)/TightPNO/def2-TZVPP/MM single-point energy calculations for QM region R2 (206 atoms) required about one week (using 8 cores of a single node with two 8-core Intel Xeon E5-2670 CPUs and 256 GB of main memory).

3. CONCLUSIONS

The hydroxylation reaction catalyzed by PHBH and the Baeyer-Villiger reaction catalyzed by CHMO were investigated by means of QM/MM simulations employing highly correlated wavefunction-based methods as QM components. B3LYP/MM geometries were used in conjunction with single-point DLPNO-CCSD(T)/MM energy evaluations using the recently introduced iterative algorithm for the treatment of the perturbative triples correction. The results were compared with those obtained from analogous single-point QM/MM calculations in which various density functionals as well as the HF, DLPNO-CCSD, and DLPNO-CCSD(T0) methods were used as QM components.

For the hydroxylation reaction, we adopted a QM region of 49 atoms that had already been employed in previous QM/MM studies. The computed DLPNO-CCSD(T0)(TightPNO/aug-cc-pVTZ)/MM energy barriers were found to be in good agreement with previously published LCCSD(T0)/MM results and available experimental data. The use of the iterative algorithm for the treatment of the perturbative triples correction further improved the quality of the computed activation energies by about 1 kcal/mol, thus providing new benchmark data for this reaction.

For the Baeyer-Villiger reaction catalyzed by CHMO we considered three different QM regions with 99, 206, and 307 atoms. The computed DLPNO-CCSD(T)(TightPNO/aug-cc-pVTZ)/MM energy profiles were found to be only weakly affected by the size of the QM region and by the choice of basis set. Also for this reaction, DLPNO-CCSD(T) and DLPNO-CCSD(T0) activation energies were found to differ by about 1 kcal/mol.

For both reactions, HF/MM calculations severely overestimate the energy barriers obtained at the DLPNO-CCSD(T) level. Hence, dynamic electron correlation plays a fundamental role. Probably for this reason, the DFT/MM results were found to be extremely sensitive to the specific exchange-correlation functional employed, leading to a wide range of predicted energy barriers. In these and similar situations, the DLPNO-CCSD(T) method in conjunction with a triple-ζ quality basis set as QM component provides more reliable QM/MM computational predictions.

Finally, we note that DLPNO-CCSD(T)/MM calculations are now feasible for fairly large QM regions, as demonstrated in the case of CHMO (up to 307 atoms presently). This allows us to converge QM/MM relative energies both with regard to the QM methodology and the size of the QM region, and thus to approach quantitative accuracy in QM/MM studies of the energetics of enzymatic reactions.

ACKNOWLEDGMENT

We gratefully acknowledge Max Planck Society for financial support.

References:

1. Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198-1229.
2. Chung, L. W., Sameera,. W. M. C.; Ramozzi, R.; Page, A. J.; Hatanaka, M.; Petrova, G. P.; Harris, T. V.; Li, X.; Ke, Z.; Liu, F.; Li, H.-B.; Ding, L.; Morokuma, K. The ONIOM Method and Its Applications. *Chem. Rev.* **2015**, 115, 5678-5796.
3. Claeyssens, F.; Harvey, J. N.; Manby, F. R.; Mata, R. A.; Mulholland, A. J.; Ranaghan, K. E.; Schütz, M.; Thiel, S.; Thiel, W.; Werner, H.-J. High Accuracy Computation of Reaction Barriers in Enzymes. *Angew. Chem., Int. Ed.* **2006**, 45, 6856-6859.
4. Mata, R. A.; Werner, H.-J.; Thiel, S.; Thiel, W. Toward Accurate Barriers for Enzymatic Reactions: QM/MM Case Study on p-Hydroxybenzoate Hydroxylase. *J. Chem. Phys.* **2008**, 128, 025104.
5. Schütz, M.; Werner, H. J. Low-Order Scaling Local Electron Correlation Methods. IV. Linear Scaling Local Coupled-Cluster (LCCSD). *J. Chem. Phys.* **2001**, 114, 661-681.
6. Entsch, B.; Van Berkel, W. J. H. Structure and Mechanism of para-Hydroxybenzoate Hydroxylase. *FASEB J.* **1995**, 9, 476-483.
7. Van Berkel, W. J. H.; Müller, F., The Temperature and pH Dependence of Some Properties of p-Hydroxybenzoate Hydroxylase from *Pseudomonas Fluorescens*. *Eur. J. Biochem.* **1989**, 179, 307-314.
8. Schreuder, H. A.; van der Laan, J. M.; Hol, W. G. J.; Drenth, J. Crystal Structure of p-Hydroxybenzoate Hydroxylase Complexed with Its Reaction Product 3,4-Dihydroxybenzoate. *J. Mol. Biol.* **1988**, 199, 637-648.
9. Ortiz-Maldonado, M.; Cole, L. J.; Dumas, S. M.; Entsch, B.; Ballou, D. P. Increased Positive Electrostatic Potential in p-Hydroxybenzoate Hydroxylase Accelerates Hydroxylation but Slows Turnover. *Biochemistry* **2004**, 43, 1569-1579.
10. Entsch, B.; van Berkel, W. J. H. Structure and mechanism of para-hydroxybenzoate hydroxylase. *FASEB J.* **1995**, *9*, 476.
11. Saebø, S.; Pulay, P. Local Configuration-Interaction - an Efficient Approach for Larger Molecules. *Chem. Phys. Lett.* **1985**, *113*, 13-18.
12. Saebø, S.; Pulay, P. 4th-Order Møller-Plesset Perturbation-Theory in the Local Correlation Treatment.1. Method. *J. Chem. Phys.* **1987**, *86*, 914-922.
13. Saebø, S.; Pulay, P. The Local Correlation Treatment. 2. Implementation and Tests. *J. Chem. Phys.* **1988**, *88*, 1884-1890.
14. Saebø, S.; Pulay, P. Local Treatment of Electron Correlation. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213-236.
15. Hampel, C.; Werner, H.-J. Local Treatment of Electron Correlation in Coupled Cluster Theory. *J. Chem. Phys.* **1996**, *104*, 6286-6297.
16. Schütz, M.; Rauhut, G.; Werner, H. J. Local Treatment of Electron Correlation in Molecular Clusters: Structures and Stabilities of $(H_2O)_n$, n = 2-4. *J. Phys. Chem. A* **1998**, *102*, 5997-6003.
17. Schütz, M.; Hetzer, G.; Werner, H. J. Low-Order Scaling Local Electron Correlation Methods. I. Linear Scaling Local MP2. *J. Chem. Phys.* **1999**, *111*, 5691-5705.
18. Schütz, M.; Werner, H. J. Local Perturbative Triples Correction (T) with Linear Cost Scaling. *Chem. Phys. Lett.* **2000**, *318*, 370-378.
19. Schütz, M.; Yang, J.; Chan, G. K.; Manby, F. R.; Werner, H. J. The Orbital-Specific Virtual Local Triples Correction: OSV-L(T). *J. Chem. Phys.* **2013**, *138*, 054109.

20. Neese, F.; Wennmohs, F.; Hansen, A. Efficient and Accurate Local Approximations to Coupled-Electron Pair Approaches: An Attempt to Revive the Pair Natural Orbital Method. *J. Chem. Phys.* **2009**, *130*, 114108.
21. Neese, F.; Hansen, A.; Liakos, D. G. Efficient and Accurate Approximations to the Local Coupled Cluster Singles Doubles Method Using a Truncated Pair Natural Orbital Basis. *J. Chem. Phys.* **2009**, *131*, 064103.
22. Hansen, A.; Liakos, D. G.; Neese, F. Efficient and Accurate Local Single Reference Correlation Methods for High-Spin Open-Shell Molecules Using Pair Natural Orbitals. *J. Chem. Phys.* **2011**, *135*, 214102.
23. Liakos, D. G.; Hansen, A.; Neese, F. Weak Molecular Interactions Studied with Parallel Implementations of the Local Pair Natural Orbital Coupled Pair and Coupled Cluster Methods. *J. Chem. Theory Comput.* **2011**, *7*, 76-87.
24. Riplinger, C.; Neese, F. An Efficient and Near Linear Scaling Pair Natural Orbital Based Local Coupled Cluster Method. *J. Chem. Phys.* **2013**, *138*.
25. Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural Triple Excitations in Local Coupled Cluster Calculations with Pair Natural Orbitals. *J. Chem. Phys.* **2013**, *139*, 134101.
26. Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. Sparse Maps—a Systematic Infrastructure for Reduced-Scaling Electronic Structure Methods. II. Linear Scaling Domain Based Pair Natural Orbital Coupled Cluster Theory. *J. Chem. Phys.* **2016**, *144*, 024109.
27. Liakos, D. G.; Sparta, M.; Kesharwani, M. K.; Martin, J. M. L.; Neese, F. Exploring the Accuracy Limits of Local Pair Natural Orbital Coupled-Cluster Theory. *J. Chem. Theory Comput.* **2015**, *11*, 1525-1539.
28. Schwab, J. M.; Li, W.; Thomas, L. P. Cyclohexanone Oxygenase: Stereochemistry, Enantioselectivity, and Regioselectivity of an Enzyme-Catalyzed Baeyer-Villiger Reaction. *J. Am. Chem. Soc.* **1983**, *105*, 4800-4808.
29. Senn, H. M.; Thiel, S.; Thiel, W. Enzymatic Hydroxylation in p-Hydroxybenzoate Hydroxylase: A Case Study for QM/MM Molecular Dynamics. *J. Chem. Theory Comput.* **2005**, *1*, 494-505.
30. Kästner, J.; Thiel, W. Bridging the Gap between Thermodynamic Integration and Umbrella Sampling Provides a Novel Analysis Method: "Umbrella Integration". *J. Chem. Phys.* **2005**, *123*, 144104.
31. Billeter, S. R.; Hanser, C. F. W.; Mordasini, T. Z.; Scholten, M.; Thiel, W.; van Gunsteren, W. F. Molecular Dynamics Study of Oxygenation Reactions Catalysed by the Enzyme p-Hydroxybenzoate Hydroxylase. *Phys. Chem. Chem. Phys.* **2001**, *3*, 688-695.
32. Ridder, L.; Palfey Bruce, A.; Vervoort, J.; Rietjens I. M. C. M. Modelling Flavin and Substrate Substituent Effects on the Activation Barrier and Rate of Oxygen Transfer by p-Hydroxybenzoate Hydroxylase. *FEBS Letters* **2000**, *478*, 197-201.
33. Ridder, L.; Mulholland, A. J.; Rietjens, I. M. C. M.; Vervoort, J. Combined Quantum Mechanical and Molecular Mechanical Reaction Pathway Calculation for Aromatic Hydroxylation by p-Hydroxybenzoate-3-Hydroxylase. *J. Mol. Graph.* **1999**, *17*, 163-175.
34. Leisch, H.; Morley, K.; Lau, P. C. K. Mechanistic Studies on Cyclohexanone Oxygenase. *Chem. Rev.* **2011**, *111*, 4165-4222.
35. Messiha, H. L.; Ahmed, S. T.; Karuppiah, V.; Suardíaz, R.; Avalos, G. A. A.; Fey, N.; Yeates, S; Toogood, H. S.; Mulholland, A. J.; Scrutton, N. S. Biocatalytic Routes to Lactone Monomers for Polymer Productio*n. Biochemistry*, **2018**, 57, 1997–2008.
36. Ryerson, C. C.; Ballou, D. P.; Walsh, C. *Biochemistry* **1982**, *21*, 2644-2655.
37. Sheng, D.; Ballou, D. P.; Massey, V. Mechanistic Studies of Cyclohexanone Monooxygenase: Chemical Properties of Intermediates Involved in Catalysis. *Biochemistry* **2001**, *40*, 11156-11167.
38. Reetz, M. T. Laboratory Evolution of Stereoselective Enzymes: A Prolific Source of Catalysts for Asymmetric Reactions. *Angew. Chem., Int. Ed.* **2011**, *50*, 138-174.

39. Polyak, I.; Reetz, M. T.; Thiel, W. Quantum Mechanical/Molecular Mechanical Study on the Mechanism of the Enzymatic Baeyer–Villiger Reaction. *J. Am. Chem. Soc.* **2012**, *134*, 2732-2741.
40. Neese, F, The ORCA Program System. *WIREs Comput. Mol. Sci.* **2012**, *2*, 73-78.
41. Neese, F. Software Update: The ORCA Program System, Version 4.0. *WIREs Comput. Mol. Sci* **2017**, *8*, e1327.
42. Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjøvoll, M.; Fahmi, A.; Schäfer, A.; Lennartz, C. QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis. *J. Mol. Struct. (Theochem)* **2003**, *632*, 1-28.
43. Metz, S.; Kästner, J.; Sokol, A. A.; Keal, T. W.; Sherwood, P. ChemShell - a modular software package for QM/MM simulations. *WIREs Comput. Mol .Sci.* **2013***, 4,* 101-110.
44. ChemShell, a Computational Chemistry Shell, see www.chemshell.org .
45. Guo, Y.; Riplinger, C.; Becker, U.; Liakos, D. G.; Minenkov, Y.; Cavallo, L.; Neese, F. An Improved Linear Scaling Perturbative Triples Correction for the Domain Based Local Pair-Natural Orbital Based Singles and Doubles Coupled Cluster Method [DLPNO-CCSD(T)]. *J. Chem. Phys.* **2018**, *148*, 011101.
46. Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. Auxiliary Basis Sets to Approximate Coulomb Potentials. *Chem. Phys. Lett.* **1995**, *240*, 283-290.
47. Dunning, T. H. Gaussian-Basis Sets for Use in Correlated Molecular Calculations .1. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007-1023.
48. Kendall, R. A.; Dunning Jr, T. H.; Harrison, R. J. Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, *96*, 6796-6806.
49. Zhao, Y.; Truhlar, D. G. Density Functionals with Broad Applicability in Chemistry. *Acc. Chem. Res.* **2008**, *41*, 157-167.
50. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
51. Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C. Dispersion-Corrected Mean-Field Electronic Structure Methods. *Chem. Rev.* **2016**, *116*, 5105-5154.
52. Weigend, F.; Ahlrichs, R. Balanced Basis Sets of Split Valence, Triple Zeta Valence and Quadruple Zeta Valence Quality for H to Rn: Design and Assessment of Accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297-3305.
53. Becke, A. D.; Johnson, E. R. A Density-Functional Model of the Dispersion Interaction. *J. Chem. Phys.* **2005**, 123, 154101.
54. Mirza, I. A.; Yachnin, B. J.; Wang, S.; Grosse, S.; Bergeron, H.; Imura, A.; Iwaki, H.; Hasegawa, Y.; Lau, P. C. K.; Berghuis, A. M. Crystal Structures of Cyclohexanone Monooxygenase Reveal Complex Domain Movements and a Sliding Cofactor. *J. Am. Chem. Soc.* **2009**, *131*, 8848-8854.
55. Polyak, I.; Reetz, M. T.; Thiel, W. Quantum Mechanical/Molecular Mechanical Study on the Enantioselectivity of the Enzymatic Baeyer–Villiger Reaction of 4-Hydroxycyclohexanone. *J. Phys. Chem. B* **2013**, *117*, 4993-5001.
56. Bennie, S. J.; van der Kamp, M. W.; Pennifold, R. C. R.; Stella, M.; Manby, F. R.; Mulholland, A. J. Projector-Embedding Approach for Multiscale Coupled-Cluster Calculations Applied to Citrate Synthase *J. Chem. Theory Comput.* **2016**, 12, 2689−2697.
57. Zhang, X.; Bennie, S. J.; van der Kamp, M. W.; Glowacki, D. R.; Manby, F. R.; Mulholland, A. J. Multiscale analysis of enantioselectivity in enzyme-catalysed 'lethal synthesis' using projector-based embedding *R. Soc. open sci.* **2018**, 5, 171390.

58. Plata, R. E.; Singleton, D. A. A Case Study of the Mechanism of Alcohol-Mediated Morita Baylis–Hillman Reactions. The Importance of Experimental Observations *J. Am. Chem. Soc.* **2015**, 137, 3811−3826.
59. Liu, Z.; Patel, C.; Harvey, J. N.; Sunoj, R. B. The True Catalyst Revealed: The Intervention of Chiral Ca and Mg Phosphates in Brønsted Acid Promoted Asymmetric Mannich Reactions *Phys. Chem. Chem. Phys.* **2017**, 19, 30647−30657.

Graphical TOC Entry