ACS OMEGA

# New Insights into the Catalytic Mechanism of Aldose Reductase: A QM/MM Study

Marie-Pierre Dréanic,[†,‡] Colin M. Edge,[†] and Tell Tuttle*,[‡]

†Medicines Research Centre, GlaxoSmithKline, Gunnels Wood Road, Stevenage, Hertfordshire SG1 2NY, U.K.

‡Department of Pure and Applied Chemistry, WestCHEM, University of Strathclyde, 295 Cathedral Street, Glasgow G1 1XL, U.K.

S Supporting Information

ABSTRACT: Aldose reductase is the first enzyme of the polyol pathway in which glucose is converted to fructose via sorbitol. The understanding of this key enzyme is important as it has been linked to some diabetes mellitus complications. The mechanism of the enzyme was investigated using a hybrid quantum mechanics/molecular mechanics (QM/MM) method. It was found that depending on the protonation state of His110 the mechanism can be concerted or stepwise and the proton donor can be either Tyr48 or His110. These findings are different from the previous theoretical studies based on QM/MM calculations using either AM1 or HF/4-31G, in which the reduction is, respectively, a stepwise or one-step process. The QM/MM energy barriers for the reduction of D-glyceraldehyde were evaluated at a B3LYP/6-31G* level for both HIP and HIE protonation states of His110. These were, respectively, 6.5 ± 2.2 and 16.7 ± 1.0 kcal/mol, which makes only the HIE protonation state consistent with the experimental value of 14.8 kcal/mol derived from kinetics experiments and makes Tyr48 the most probable proton donor.
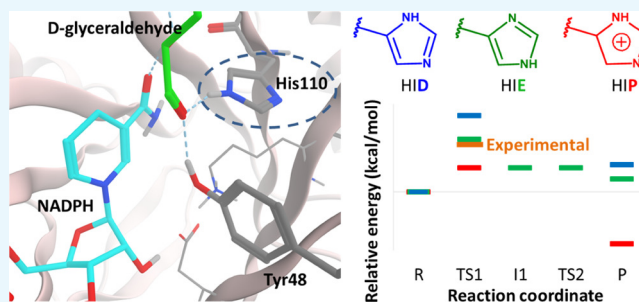
1. INTRODUCTION

Aldose reductase (AR) (EC 1.1.1.21) is a cytosolic reduced nicotinamide adenine dinucleotide phosphate (NADPH)-dependent oxidoreductase enzyme that belongs to the superfamily of aldo-keto reductases.[1,2] Aldose reductase has been identified as the first enzyme involved in the polyol pathway of glucose metabolism that converts glucose to fructose via sorbitol.[3] This is of particular interest for the pharmaceutical industry as glucose overutilization through the polyol pathway has been linked to tissue-based pathologies associated with diabetes mellitus complications.[3,4] AR has thus been widely studied to develop potent AR inhibitors to prevent or delay the onset and progression of these complications.[5] As a result, the protein data bank (PDB) accounts to date (February, 2017) for an impressive number of X-ray crystallographic structures (136) of human aldose reductase.[6]

The human AR enzyme comprises 315 amino acid residues and has a β/α barrel structure (Figure 1a,b).[7] The barrel is composed of eight parallel β-strands and eight adjacent peripheral α-helical segments that are running antiparallel to the β-sheet. The catalytic active site is located in the barrel core. The nicotinamide adenine dinucleotide phosphate (NADP) cofactor is situated at the top of the COOH-terminal end of the β/α barrel, with the nicotinamide ring projecting into the center of the barrel and the pyrophosphate part on the border of the barrel.

The reaction mechanism of aldose reductase in the direction of aldehyde reduction comprises two steps.[8] The first step is the transfer of the pro-R hydride of NADPH to the re face of the

substrate's carbonyl carbon. The second step is the donation of a proton to reduce the carbonyl to an alcohol (Scheme 1).

Despite the acceptance of this general mechanism, several key features remain unclear. On the one hand, it is not known whether the reaction occurs in a concerted or stepwise manner. On the other hand, it is not clear which of the proximal residues, Tyr48 or His110, acts as the proton donor. Indeed, both of these residues could potentially occupy this function, as crystal structures indicate that they are well positioned to be potential proton donors during catalysis; in crystal structure 1ADS, a water molecule in close proximity to the nicotinamide is hydrogen bonded to both Tyr48 and His110 and thus indicates a possible position for the substrate (Figure 1c). A comparison of the relative pK_as of the residues suggests that the lower value of histidine (pK_a = 6−7) relative to tyrosine (pK_a = 10) would make it a more likely candidate to donate a proton.[9,10] However, the proximity of the Lys77−Asp43 pair in the binding site has been proposed to lower the pK_a of Tyr48− 8.25 through hydrogen bonding.[10]

In the literature, there are several computational studies that investigate which of the two potential residues is the proton donor.[11−13] These include two quantum mechanics/molecular mechanics (QM/MM) studies, one by Lee and co-workers[13] and the other by Várnai and co-workers,[11] and one empirical valence bond (EVB) study by Várnai and Warshel.[12] From the
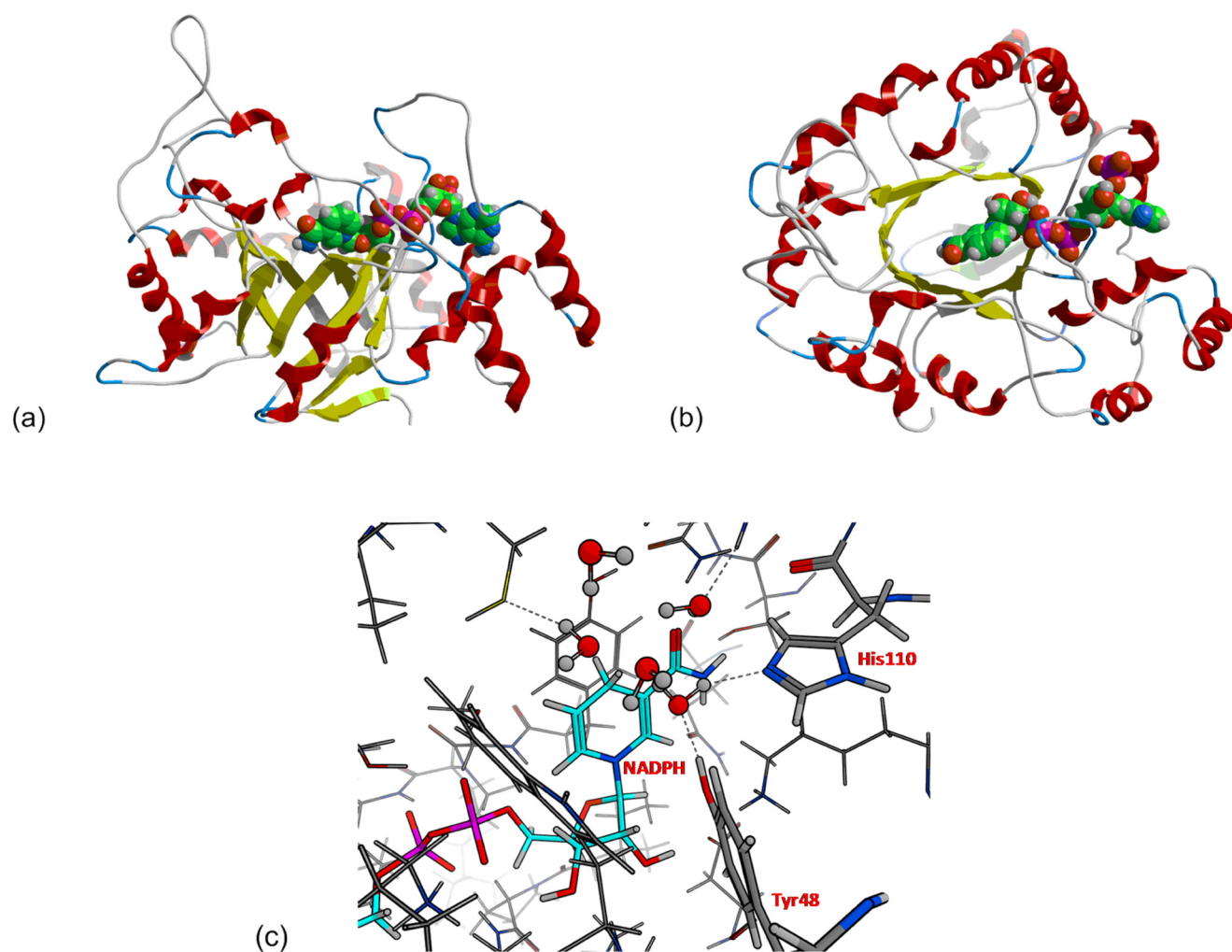
Figure 1. (a, b) View of the α-carbon backbone trace (schematic diagram) of the aldose reductase structure with bound NADPH. (a) View perpendicular to the β/α barrel with NADPH shown in green space-filling model. (b) The structure viewed down the COOH-terminal end of the β/α barrel. (c) Aldose reductase active site (PDB ID: 1ADS) with crystallographic waters shown.

Scheme 1. Schematic Representation of Aldehyde Reduction by Aldose Reductase
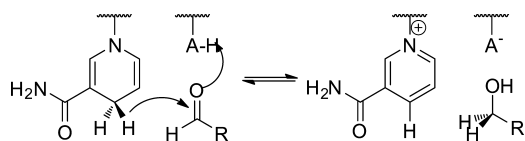
results of the two QM/MM studies, which are summarized in Table 1, it is evident that they differ in both their proposed mechanisms and calculated energetics. The results of Lee and co-workers[13] show a concerted mechanism, whereas Várnai and co-workers reported a stepwise mechanism.[11] The difference in the calculated energetics of the reactions is also significant, with a difference for the calculated relative energies of around 10 kcal/mol. The experimental activation free energy, determined from reaction rate studies, is 14.8 kcal/mol.[14] Thus, both computational studies overestimate the activation energy with a relative energy of 21.3 kcal/mol for Lee and co-workers[13] and 31.8 kcal for Várnai and co-workers.[11] Nevertheless, both studies agree that the reaction mechanism is more favorable with the His110 model than with the Tyr48 model, as the relative activation energies in both studies are smaller when employing His110 as the proton donor.

Table 1. Summary of Previous QM/MM Results for the Reduction of D-Glyceraldehyde (GLD) by Aldose Reductase[a]

study	TS1	I	TS2	P
	Proton Donor His110			
Lee et al.[b]	21.2			−12.4
Várnai et al.[c]	31.8	25.4	35.4	−5.9
	Proton Donor Tyr48			
Lee et al.[b]	24.3			−3.7
Várnai et al.[c]	41.2	33.6	34.7	10.3

[a]Electronic energies (ΔE) in kcal/mol are given relative to the reactant state for each system studied. [b]Ref 13. [c]Ref 11.

In the EVB study from Várnai and Warshel,[12] the energy profile was only evaluated for the tyrosine proton donor hypothesis as their detailed pKa studies on both Tyr48 and His110 suggested that the Tyr48 proton donor hypothesis would be the most probable mechanism. The activation free energy was calculated to be 17 kcal/mol and thus in good agreement with experimental results. The better agreement of the EVB results is not a surprise as the method comprises significant sampling and is thus able to evaluate free energies that can be directly compared to experiment results. On the

Scheme 2. Proposed Mechanisms for the Different Protonated States of His110: (a) HIP, (b) HIE, and (c) HID
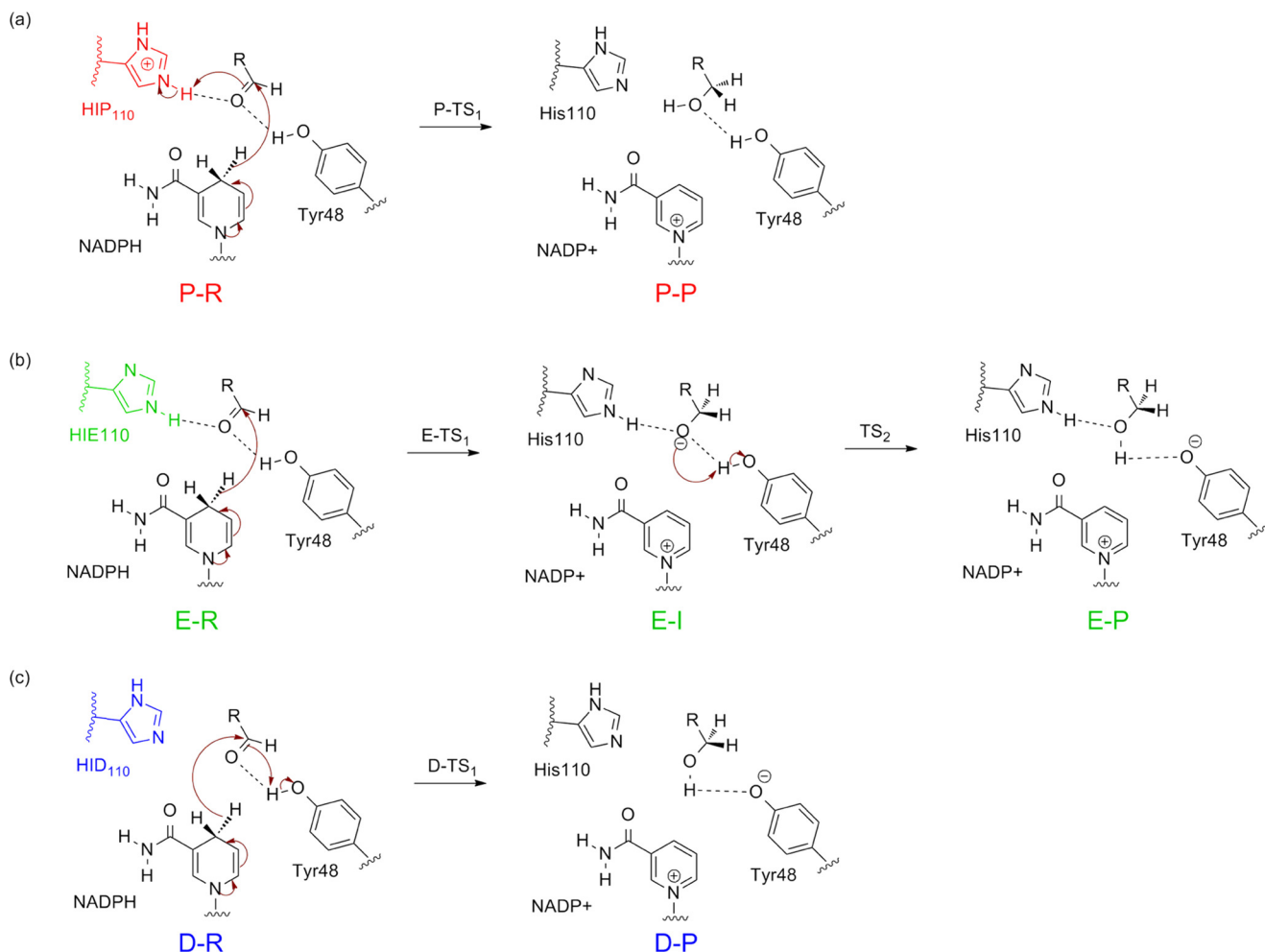
contrary, in the two QM/MM methods described above, no sampling is included and thus only potential energies are calculated. Entropic contributions are thus not included in these original calculations. Although entropic and thermal contributions can play a significant role in determining transition-state energies, previous work has shown that the entropic contributions to the activation energies for some enzyme reactions can be minimal,[15] and as such the underlying difference in the quality of the results between the EVB and QM/MM calculations is not necessarily due entirely to the exclusion of entropic effects.

Overall, the opposing nature of the conclusions from these two QM/MM studies, combined with the low chemical accuracy of the calculated activation energies, indicates that a more detailed study into this important mechanism is warranted. In the present work, we have examined the catalytic mechanism of aldose reductase with a QM/MM approach employing this time density functional theory (DFT) as the QM methodology. The structures of transition states (TSs) and intermediates involved in the reaction, the energy profiles, and the roles of key residues are presented herein. The detailed interpretation of the catalytic mechanism that results from this work is helpful for the design of mechanism-based inhibitors like transition-state analogue or covalent inhibitors.[16] Finally, one of the main objectives of this work is to determine how the methodological choices in a QM/MM calculation can have significant effects on both the calculated energetics and the

resulting interpretation of the preferred mechanism. Therefore, the extent to which using a modern density functional and a larger QM region can affect previous results, both quantitatively and qualitatively, is also discussed.

2. RESULTS AND DISCUSSION

The proposed mechanism for the different protonation states is represented in Scheme 2, and the associated relative energies are shown in Figure 2. To clearly differentiate the intermediates from the different models, a labeling scheme is introduced, where the abbreviated name of the model (P for HIP, E for HIE, and D for HID) precedes the abbreviated name of the intermediate (R for reactant, TS for transition state, I for intermediate, and P for product). Thus, we have for example P−R that stands for reactant of the HIP model. In the following sections, a detailed description of the three different reaction mechanisms is given.

2.1. Mechanism with HIP110. The results for the mechanism of GLD reduction by AR in the case of a protonated histidine show a single-step mechanism with associated activation energy of 8.1 kcal/mol (Scheme 2a and Figure 2). A schematic representation of the starting enzyme−substrate (ES) complex (P−R), the transition state (P−TS1), and the final enzyme−product (EP) complex (P−P), including only the closest atoms around substrate, is given in Figure 3a−c, respectively.
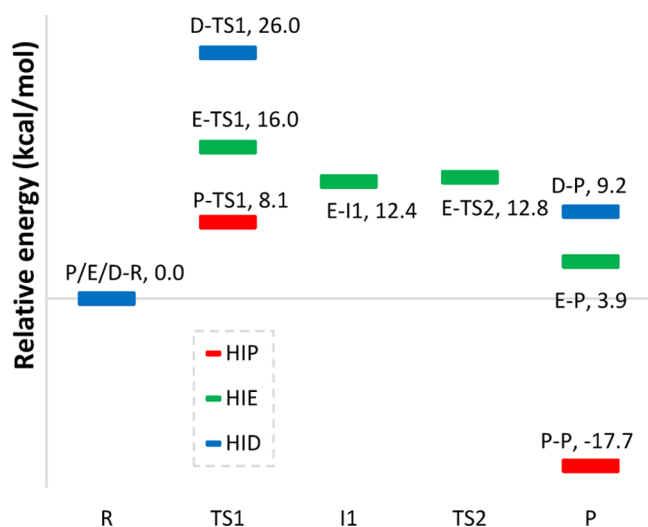
Figure 2. Comparison of relative energies for the three protonation models: HIP, HIE, and HID. Electronic energies (ΔE) in kcal/mol, calculated at the B3LYP/6-31G* level, are given relative to the reactant state for each system studied.

Significant interactions help to maintain atoms in the ES complex, P−R, in a suitable position for reactions, namely, H-bonds between both His110 and Tyr48 hydroxyl group and the carbonyl group of GLD (H5(His110)···O4(GLD) = 1.83 Å and H6(Tyr48)···O4(GLD) = 2.02 Å), the H-bond between the NADPH amide group and the 2-hydroxy of GLD (O8-(NADPH)···H9(GLD) = 1.80 Å), and finally the H-bond between the 3-hydroxy of GLD with a water molecule (H(H_2O)···O13(GLD) = 2.15 Å). In the transition state (characterized by an imaginary frequency of −667 cm^{-1}), the NADPH hydride is approximately halfway between C1 and C3, the C1···H2 and C3···H2 distances being 1.45 and 1.26 Å, respectively (Figure 3b). In addition, the comparison of the enzyme−substrate complex (Figure 3a) and TS (Figure 3b) geometries shows the beginning of transition from a planar sp^2 to a tetrahedral sp^3 for the GLD carboxyl. In the same way, the donation of the hydride by NADPH makes the nicotinamide ring become more planar. The TS structure, P−TS1, does not clearly show whether His110 or Tyr48 is the proton donor: both H5 from His110 and H6 from Tyr48 are now closer to O4 (GLD) (H5(His110)···O4(GLD) distance is 1.61 Å and H6(Tyr48)···O4(GLD) distance is 1.83 Å), and these may contribute to the stabilization of the TS. However, from the product complex (Figure 3c), it is clear that the proton donor is

Figure 3. (a−c) Reaction intermediates of the GLD reduction by AR with HIP110 as studied by the QM/MM model. (a) Enzyme−substrate complex, (b) transition state, (c) enzyme−product complex (distances shown in green, atom numbers in brown, and residue names in red), and (d) superposition of the three intermediates of the reaction (P−R in pink, P−TS1 in green, P−P in cyan, and hydride in yellow).
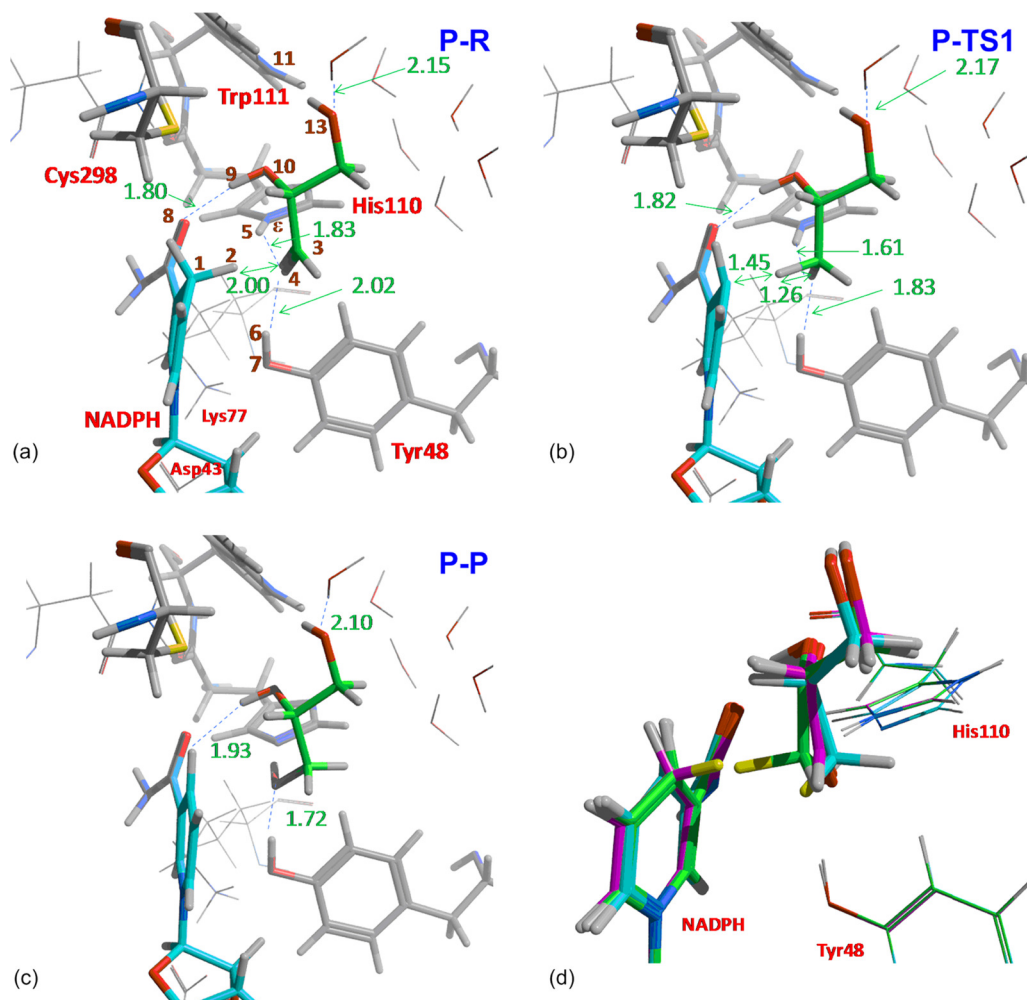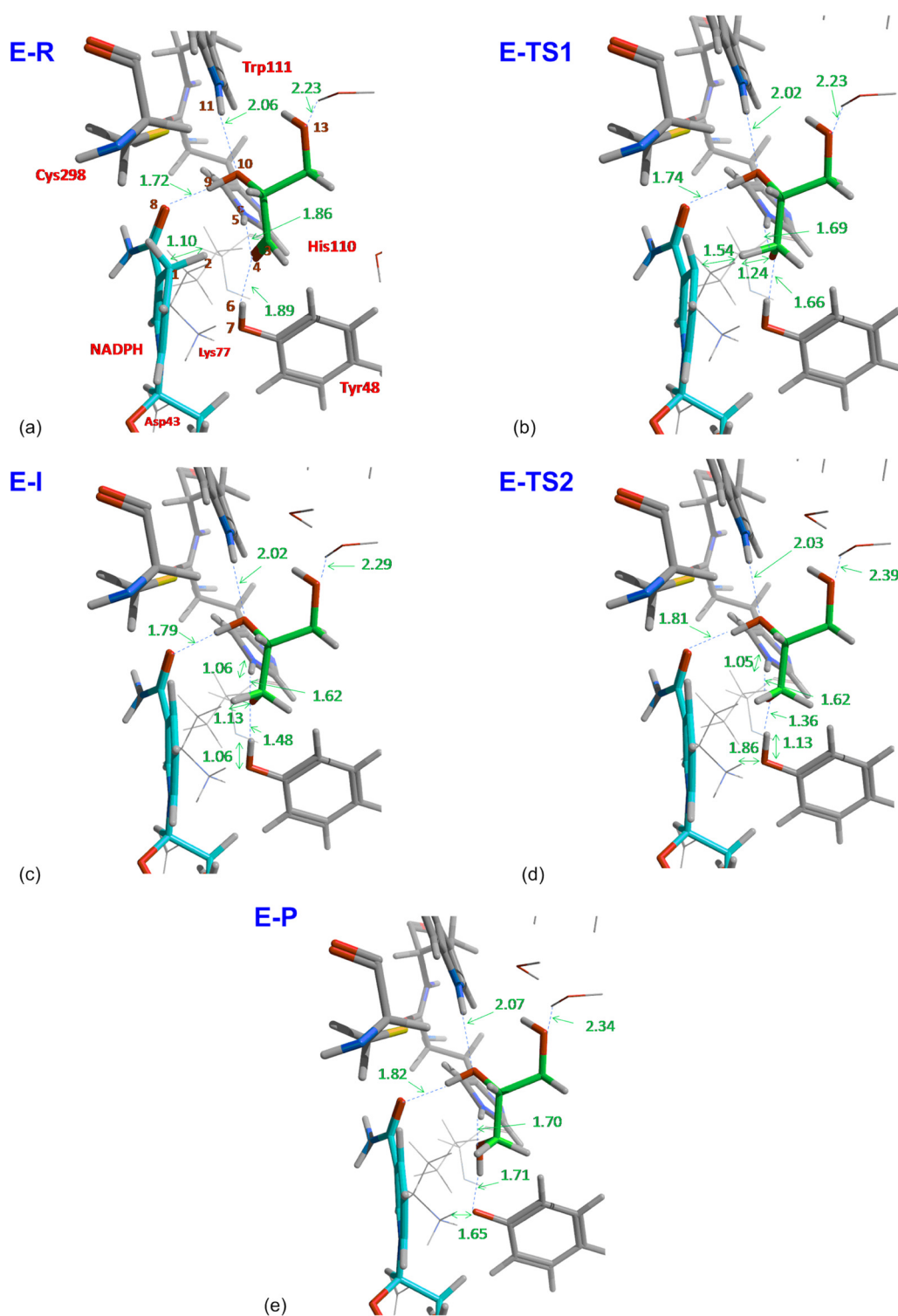
Figure 4. Reaction intermediates of the GLD reduction by AR with HIE110 as studied by the QM/MM model. (a) Enzyme−substrate complex, (b) transition state 1, (c) enzyme−intermediate complex, (d) transition state 2, and (e) enzyme−product complex (distances shown in green, atom numbers in brown, and residue names in red).

His110; thus, the role of Tyr48 is stabilizing the incipient negative charge on the aldehyde group of GLD. From these results, it can be concluded that the mechanism for the protonated histidine system is concerted and asynchronous, where the approach of the NADPH hydride to the carbonyl carbon of GLD triggers the proton transfer. In the final product complex (EP), the hydride is definitively bonded to the D-

glycerol carbon C3, and the proton H4 from His110 has been completely transferred to D-glycerol oxygen O4. A strong hydrogen bond involving H6 of Tyr48 persists, the O4···H6 distance being 1.72 Å.

We examined the possibility of an alternative mechanism where the proton transfer (from H5(His110) to O4(GLD)) and the attack of the hydride on the C3 group occur in two

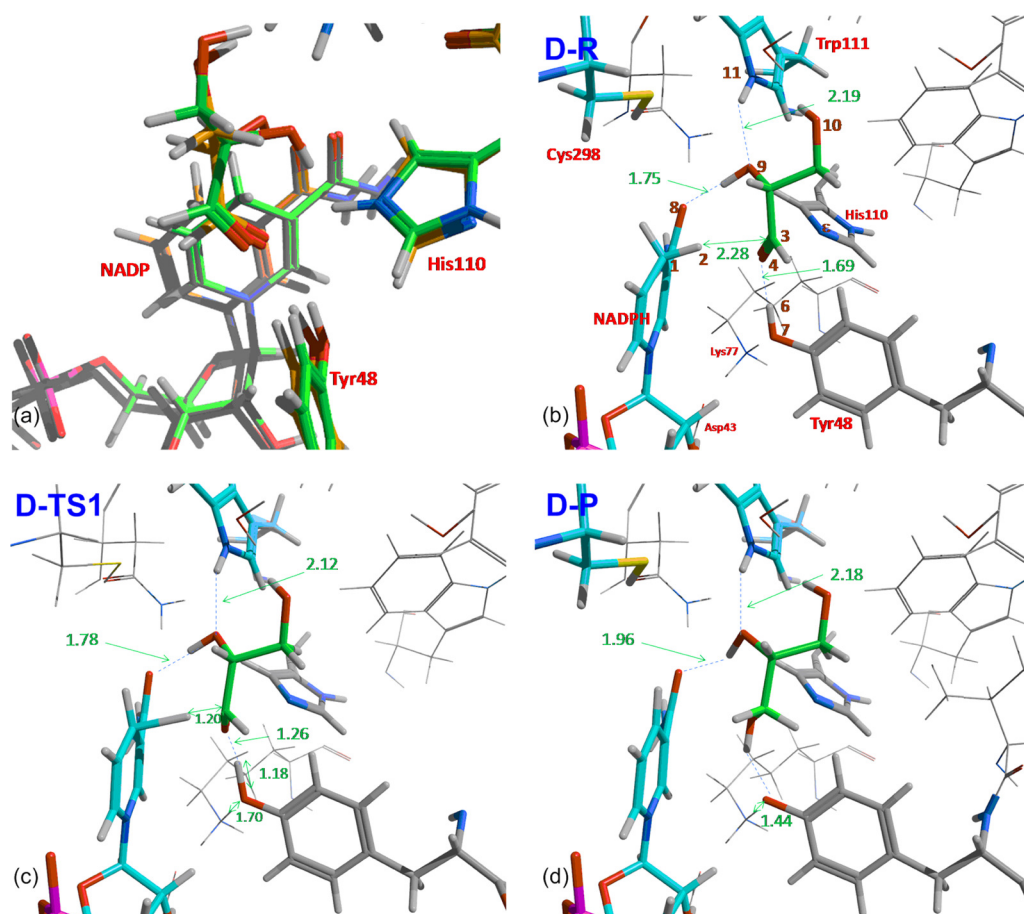**Figure 5.** (a) Comparison of NADPH position between HID (green) and HIE (orange) after a 1 ns molecular dynamics (MD) simulation. (b−d) Reaction intermediates of the D-glyceraldehyde reduction by AR with HID110 as studied by the QM/MM model: (b) enzyme−substrate complex, (c) transition state, (d) enzyme−product complex (distances shown in green, atom numbers in brown, and residue names in red).

separate steps. In spite of an extensive search, it was not possible to locate any intermediate corresponding to the alkoxide and thus the possibility of a two-step nonconcerted mechanism was discounted. Finally, the mechanism with a proton transfer from Tyr48 was also intensively investigated, but no transition state that could lead to a proton transfer from Tyr48 could be identified.

It is also informative to superpose the three stationary points along the potential energy surface (PES; Figure 3d). The aldehyde hydrogen of the reactant maintains its location in the TS but is replaced by the NADPH hydride hydrogen at the product stage. This suggests that the active site is set up to stabilize a hydrogen at this point and is ideally arranged for this transformation.

2.2. Mechanism with HIE110. For HIE, the mechanism is constituted of two steps with an activation energy of 16.0 kcal/mol (Scheme 2b and Figure 2). A schematic representation of the starting enzyme−substrate complex (E−R), the two transition states (E−TS1 and E−TS2), the intermediate (E−I), and the final enzyme−product complex (E−P), including only the closest atoms around substrate, is given in Figure 4a−e.

The same interactions that help to maintain atoms in the ES complex of P−R can be found in E−R. These are hydrogen-bonding interactions between both His110 and Tyr48 residues and the carbonyl oxygen of GLD (H5(His110)···O4(GLD) = 1.86 Å and H6(Tyr48)···O4(GLD) = 1.89 Å) and the

interaction between the amide group of NADPH and the middle hydroxyl group of GLD (O8(NADPH)···H9 = 1.72 Å) (Figure 4a). A water molecule also stabilizes the 3-hydroxy of GLD (O13(GLD)···H_2O = 2.23 Å). Compared to P−R, E−R is further stabilized by a supplementary interaction with Trp111 (H11(Trp111)···O10(GLD) = 2.06 Å) that is not always present in the HIP simulation.

In the transition state, E−TS1 (characterized by an imaginary frequency of −562 cm^{-1}), the NADPH hydride transfer from C1 to C3 is nearly completed, the C1···H2 and C3···H2 distances being 1.54 and 1.24 Å, respectively (Figure 4b). Hydrogens from Tyr48 and His110 are both almost at the same distance to the carboxyl oxygen of GLD O4 and closer compared to E−R—the H5(His110)···O4(GLD) distance was 1.86 Å in E−R but is 1.69 Å in E−TS1, and the H6(Tyr48)···O4(GLD) distance was 1.89 Å in E−RE−R but is 1.66 Å in E−TS1. The H11(Trp111) to O10(GLD) distance is nearly unchanged from the E−R (2.06 Å) to E−TS1 (2.02 Å). The interaction between the amide group of NADPH and the 2-hydroxy of GLD (O8(NADPH)···H9(GLD) = 1.74 Å) is relatively unchanged at E−TS1 compared to E−R, suggesting that the function of this interaction is to maintain the position of the substrate through a consistently strong stabilizing interaction.

The transition from E−I to E−TS is almost barrier-less with a difference of 0.4 kcal/mol. In the intermediate, E−I, the NADPH hydride is now completely transferred from C1 to C3

Table 2. Comparison of Previous QM/MM Results from Lee and Co-Workers'[a] Model (Lee) and Várnai and Co-Workers'[b] Model (Var.) to HIP and HIE Models[c]

TS1	ΔE	I	ΔE	TS2	ΔE	P	ΔE
E−TS1	16.0	E−I	12.4	E−TS2	12.8	E−P	3.9
Var.-TS1	41.2	Var.-I	33.6	Var.-TS2	34.7	Var.-P	10.3
P−TS1	8.1					P−P	−17.7
Lee-TS1	21.2					Lee-P	−12.4

[a]Ref 13. [b]Ref 11. [c]Electronic energies (ΔE) in kcal/mol are given relative to the reactant state for each system studied.

as C3···H2 is 1.13 Å, the distance of a C−H bond. The difference between His110 and Tyr48 is clear in the intermediate structure as Tyr48 H6 is closer to the GLD O4 (1.48 Å) than His110 H5 is to O4 (1.62 Å).

In the second transition state, E−TS2 (characterized by an imaginary frequency of −305 cm^{-1}), the hydrogen bonding from Tyr48 and His110 to the GLD carbonyl oxygen O4 is further differentiated. This proton is partially transferred from Tyr48 (H6(Tyr48)···O4(GLD) =1.36 Å) compared to His110, in which the hydrogen bonding remains consistent relative to I−E (1.62 Å).

In the final EP complex, E−P, the proton H6 from Tyr48 is bonded to O4 from GLD. At this stage, a strong interaction is formed between the formed Tyr48 phenolate and Lys77, going from 1.86 Å in E−TS2 to 1.65 Å in E−P.

2.3. Mechanism with HID110. During the HID110 1 ns MD simulation, a displacement of NADPH occurred (Figure 5a), giving an unproductive complex and perhaps suggesting that this electronic state is quite unreactive. To address this issue, the energy profile for HID was determined with a new 1 ns MD but with restraints on the cofactor position.

For HID, the eventually identified mechanism is a concerted one, using Tyr48 as proton donor with an activation energy of 26 kcal/mol (Scheme 2c and Figure 2). A schematic representation of the starting enzyme−substrate complex (D−R), the transition state (D−TS1), and the final enzyme−product complex (D−P), including only the closest atoms around substrate, is given in Figure 5b−d.

In the case of HID, there are fewer interactions that help to maintain atoms in the ES complex, D−R, in a suitable position for reactions. As Nε of His110 is deprotonated, no stabilization is possible. In contrast to the other simulations, no water was observed interacting with the 3-hydroxy of GLD. As a consequence, the interaction between the Tyr48 hydroxyl group and O4 of GLD carbonyl group is strong (H6(Tyr48)···O4(GLD) = 1.69 Å), whereas in P−R and E−R, the hydrogen bond was longer, 2.02 and 1.89 Å, respectively (cf. Figure 5b, 3a, and 4a). The NADPH hydride is almost at the same distance from the GLD carboxyl carbon (H2(NADPH)···O4(GLD) = 2.28 Å) compared to E−R (2.25 Å) but slightly further compared to P−R (2.00 Å). The hydrogen bond between the GLD 2-hydroxy and Trp111 is fairly consistent between both D−R (2.19 Å) and E−R (2.20 Å).

In the transition state, D−TS1 (characterized by an imaginary frequency of −875 cm^{-1}), the NADPH hydride is moving from C1(NADPH) to C3(GLD). The hydride is very close to completely transferred as the distance to C3 is only 1.20 Å. The proton from Tyr48 is approximately halfway between O7(Tyr48) and O4(GLD) (H6···O7(Tyr48) distance is 1.18 Å and H6(Tyr48)···O4(GLD) distance is 1.26 Å.) Thus, the mechanism is concerted, indeed almost simultaneous, between the hydride transfer and the proton transfer.

In the final EP complex, D−P, the hydride is definitively bonded to D-glycerol carbon C3, and the proton H6 from Tyr48 has been completely transferred to GLD O4. At this stage, a strong H-bonding interaction is formed between the phenolate of Tyr48 and Lys77 (1.44 Å).

2.4. Comparison between the Three Reaction Models. The calculated activation barriers for HIP and HIE are of 8.1 and 16.0 kcal/mol, respectively. Thus, from an energetic point of view, both mechanisms are different. This is all the more true when we average the results obtained from QM/MM studies on other frames (one additional for HIP and two for HIE) that we also studied (details of the structures are provided in the Supporting Information). These gave an average of 6.5 ± 2.2 kcal/mol for HIP and 16.7 ± 1.0 kcal/mol for HIE. Also, it should be pointed out that we are comparing ΔE with ΔG; nevertheless, in these types of reaction, the contribution from thermal and entropic effects is expected to be small.[17] It could be concluded from these results that the mechanism with the lower activation energy is the more probable one. Nevertheless, the experimental activation free energy calculated from kinetics constants is 14.8 kcal/mol.[14] Thus, although the activation energy with the HIP model is lower, the results from the HIE model are closer to the experimental value. Given the significant difference (~10 kcal/mol) between the calculated activation energies that arises from considering the protonation state of the histidine, the clear agreement of the HIE model with the experimental data indicates that the experimental system involves an unprotonated histidine in the binding site with Tyr48 acting as the proton donor. These conclusions based on the difference in the calculated activation barriers are in agreement with the calculation of the pK_a of the residues, which has been done by Várnai and Warshel[12] that yielded an estimated pK_a of 8.5 for Tyr48 and a remarkably low value of 0.9 for His110. The study of the HID model gave a much higher energy barrier of 26.5 kcal/mol. The difference in the results of HIE and HID models demonstrates that the presence of a proton on Nε of His110 is required for the correct positioning of GLD.

2.5. Effect of Basis Set Size and QM Region Size. The goal of this work was to obtain an updated QM/MM model for the reduction of GLD by AR to determine both the mechanism of reaction and the effect that a different QM/MM methodology can have on the outcome of results. To reach that goal, we have used a more accurate QM treatment and a larger QM zone. Thus, in the following, the current results are compared to those from previous studies.

Our HIP model can be compared to the results obtained by Lee and co-workers,[13] and our HIE model, to the results obtained by Várnai and co-workers.[11] To help the comparison, energies from both studies and this work are summarized in Table 2. The structural characteristics were very similar to previous studies and are thus described in the Supporting Information.

From an energetic point of view, the results differ significantly. For the HIE model, there is notable difference in the activation energy between the work of Várnai and co-workers[11] (41.2 kcal/mol) and this work (16.0 kcal/mol), a difference of 25.2 kcal/mol. From the experimental activation free energy calculated from kinetics constants of 14.8 kcal/mol, we know that our model is in better agreement.[14]

For the HIP model, the individual influence of the QM treatment and the QM size is summarized in Table 3.

Table 3. Effect of QM Treatment and QM Size on the Activation Energy for the HIP Model[c]

QM treatment	HF/4-31G	B3LYP/6-31G*	B3LYP/6-31G*
QM/MM partitioning[a]	1	1	2
E_a	21.2[b]	7.5	8.1
ΔE	−12.4[b]	−6.0	−17.7

[a]As defined in Figure 2. [b]Ref 13. [c]Electronic activation energies (E_a) and reaction energies (ΔE) in kcal/mol are given relative to the reactant state for each system studied.

In the study by Lee and co-workers, the relative energy to the reactant of the TS for the His110 proton donor model, obtained using 4-31G, was 21.2 kcal/mol.[13] From our results, we can see that the use of the more accurate B3LYP/6-31G* method has significantly changed the calculated relative energies as we obtained an activation energy of 7.5 kcal/mol, 13.7 kcal/mol smaller than that obtained by the HF/4-31G method. The combination of B3LYP/6-31G* and a larger QM region did not significantly alter the activation energy (7.5 kcal/mol for the smaller region and 8.1 kcal/mol for the larger region). However, the effect on the relative energy of the product to the reactant (−6.0 kcal/mol for the smaller region and −17.7 for the larger region) was more substantial. Overall, the results show that a meaningful gain in accuracy for the comparison of the two potential reaction mechanisms is due mostly to the developments in accuracy and efficiency of QM methods.

3. CONCLUSIONS

Since 1992 and the first suggestion of His110 and Tyr48 as potential proton donors, there has been a long history of debate on the catalytic mechanism of AR.[7,18] Nevertheless, the common opinion seemed to favor the Tyr48 proton donor mainly because of crystallographic and mutagenesis data.[10,19−22]

Nevertheless, two previous QM/MM methodologies (using CHARMM22[23]/HF 4-31G and CHARMM22[23]/AM1[24]) have failed to validate the Tyr48 hypothesis. Furthermore, they have also given different results between them for the proposed mechanism: one predicted a concerted mechanism, whereas the other predicted a stepwise mechanism. Using a different force field and QM method (OPLS2005 and B3LYP[25,26]/6-31G*[27]) and a bigger QM region, the mechanism was reevaluated. For the first time, a different mechanism is suggested depending on the protonation state of His110. With HIP as protonation state for His110, the results show an average activation energy of 6.5 ± 2.2 kcal/mol and evidence for a highly asynchronous concerted mechanism with His110 as proton donor. With HIE, the mechanism is different, as results show an average activation energy of 16.7 ± 1.0 kcal/mol and evidence for a stepwise mechanism using Tyr48 as proton donor. Preliminary MD simulation on HID indicates that this protonation state is

unreactive and shows the importance of a proton on Nε of His110 for the reaction to occur as this residue is implicated in the positioning of the substrate prior to the reaction. Our results demonstrate that the HIP and HIE model mechanisms are significantly different in energy and that only the HIE model is in good agreement with experimental data—confirming that Tyr48 is the most probable proton donor. Finally, the effect of using modern DFT methods for the QM/MM calculation was evaluated by comparing our results to those of previous studies. We found that the changes in energetics can be substantially affected by the choice of methods and, importantly, the size of the QM site (particularly for the relative energy of the reactants and products).

4. COMPUTATIONAL METHODS

4.1. Model Systems. Theoretical studies were performed starting with an X-ray crystal structure (PDB ID: 1ADS) of the aldose reductase enzyme that has a resolution of 1.65 Å.[7] Although there are many other available crystal structures, this provided a convenient point of reference, as the structure was used in the previous QM/MM studies by Lee and co-workers.[13] The structure includes the cofactor NADP, so this was transformed into the reacting form NADPH. The "structure preparation" module of molecular operating environment[28] was used to prepare the structure: missing hydrogens were added, orientated, and the protonation states of residues were optimized.

Particular attention was paid to His110 because, at physiological pH, histidine can exhibit three different protonation states: HIP (protonated Nε and Nδ), HIE (protonated Nε), and HID (protonated Nδ) (Scheme 3).[29,30]

Scheme 3. Different Protonation States of Histidine

HID   HIE   HIP
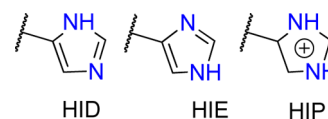
Within a protein, standard pK_a values of residues can be more or less influenced by the environment and that makes the prediction of the residues' protonation state less straightforward. Different methods exist to predict the pK_a of residues, but results from these prediction are not always reliable.[31] In this case, we have not attempted to do a QM/MM pK_a prediction, rather the initial calculation of the protonation states was carried out with the empirical modeling program PROPKA.[32] However, the calculated protonation states from this program were found to be unreliable for the system under study. As such, all possible protonation states for the histidine residue involved in the mechanism were evaluated.

D-Glyceraldehyde (GLD) was chosen as the ligand, to be consistent with the reference studies.[11,13] The accuracy of the MM force-field parameters associated to GLD was tested by performing various conformational searches and minimizations with MacroModel.[33] The consistency of the bonds and angles of the resulting structures was checked using Mogul.[34] As the crystal structure did not contain any ligand, GLD was added manually. To ensure that the re face of the carbonyl of GLD would be able to receive the hydride from the NADPH, the carbonyl oxygen of GLD was positioned within the range of hydrogen-bonding interactions with the Nε hydrogen of His110 and the hydroxyl of Tyr48. To get an adequate pose,
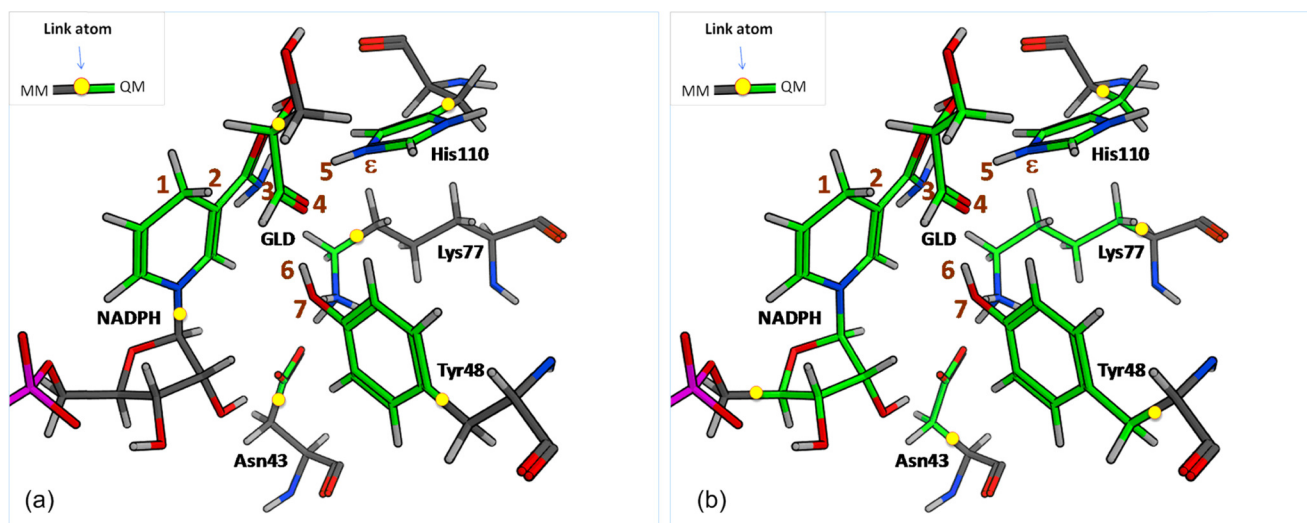
Figure 6. QM/MM partitioning (a) 1 and (b) 2. The QM region is shown in green, and the yellow dots represent link atom positions.

an optimization with constraints on the distances of these hydrogens bonds was run to reproduce the distances from the Michaelis complex (MC) described by Lee and co-workers.[13] After having deleted all crystal waters, the system was solvated in a 10 Å orthorhombic box of water using the system builder panel of Desmond.[35,36] TIP4P[37] was chosen as the model for the waters. Finally, to neutralize the system, three sodium atoms were added randomly for the HIP model and four for both HIE and HID models.

The three model systems were gradually relaxed using a standard protocol implemented in Desmond.[35,36] This protocol comprises a series of minimizations and molecular dynamics, starting from a system where only hydrogens and the solvent are free to move as the system is gradually relaxed. A molecular dynamics (MD) simulation was performed for 1 ns using Desmond[35,36] to relax further the system and to obtain a variety of snapshots for the QM/MM calculations. The NPT ensemble was used for the simulation with the Martyna, Tuckerman, and Klein method.[38] The temperature was kept fixed at 300 K. The cutoff radius for the nonbonded interactions (Coulombic and van der Waals) was fixed at 9.0 Å without any special treatment of interactions at the cutoff. The long-range Coulombic interactions in the simulation were treated by a smooth particle mesh Ewald method with a tolerance value of 1 nm.[39] At the end of the MD simulation, the average value, the standard deviation, and the slope of different properties (potential energy, pressure, temperature, and volume) were calculated and analyzed to confirm that the MD simulation was at equilibrium.

To obtain an averaged energy barrier, a minimum of two snapshots per model, with suitable hydrogen bonding between GLD and both His110 and Tyr48, were extracted from the MD output and prepared for QM/MM calculations. The selected snapshots were then MM minimized to return the system to 0 K using the truncated Newton method[40] implemented in Impact.[41] A second MM minimization using the Polak–Ribiere conjugate gradient[42] method implemented in MacroModel[33] was used to reproduce the distances from the Michaelis complex described by Lee and co-workers.[13] All of the minimizations were done with waters' oxygen atoms kept constrained. These postequilibration minimized structures, which represent the enzyme–substrate (ES) complex, were used as starting structures for the QM/MM calculations.

4.2. QM/MM Methodology. In previous QM/MM studies by Lee and co-workers[13] and Várnai and co-workers,[11] the choice of the QM region for both studies includes all hypothetical reacting species (D-glyceraldehyde, NADPH, His110, and Tyr48) and influential residues (Asp43 and Lys77).[18] An increase in accuracy can be expected if the size of the QM region is extended;[43,44] currently, QM/MM calculations can readily account for up to 100 atoms in the QM region,[45] so performing calculations on the upper side of this range could thus be considered. The QM treatment was done at an ab initio level (HF/4-31G) by Lee and co-workers[13] and at a semiempirical level (AM1)[24] by Várnai and co-workers.[11] Although issues such as boundary effects, the classical potential, and optimization strategies may all affect calculated results, in this comparison, the difference between the QM treatment of the system could be the main reason for lack of accuracy in the previous results. As such, this hypothesis will be tested in the current work. Ideally, one would perform all QM calculations with the most accurate ab initio method, together with the largest available basis set.[46] The most generally reliable and routinely used QM treatment in current QM/MM studies is DFT, particularly with the B3LYP functional.[25,45]

To study the reaction, a hybrid QM/MM Hamiltonian was employed using QSite.[47−49] For the classical region (MM), the OPLS_2005[50] force field was used to describe the protein. The QM region was modeled at the B3[25]LYP[26,51]/6-31G*[27] level of theory. The effect of the size of the QM region was examined using two different QM regions on the HIP model. The first QM region was defined as in the study by Lee and co-workers.[13] This QM region with a total of 54 atoms is represented in Figure 6a. A larger QM/MM partitioning, as defined in Figure 6b, was also used in the three models (HIP, HIE, and HID). In this second partitioning, the same residues are included, but the QM/MM frontier is positioned differently. First, the entire side chains of residues were included by cutting between Cα and Cβ. Second, the frontier within NADPH was extended by adding the ribose part of NADPH, allowing a cut between two carbons rather than between a carbon and a nitrogen atom, consistent with the best practice for the positioning of link atoms.[52] The QM/MM partitioning 2 thus

had a total of 90 atoms included for the HIP model and 89 atoms for both HIE and HID models.

To be consistent with the reference study, the link atom approach was used to saturate the valence of the QM/MM frontiers.[13,53] To avoid overpolarization of QM atoms by MM atoms to the boundary, Gaussian charge distributions were used to represent the potential of the atoms within two covalent bonds of the QM/MM cut-site using the Gaussian grid method for hydrogen cap electrostatics in QSite.[54] MM point charges were employed for the rest of the MM region.

All atoms beyond 10 Å from the reactant were consistently kept constrained during the QM/MM simulations to speed up the calculations. The equilibrated ES complex was optimized with QM/MM calculations. The potential energy surface (PES) for the reaction was explored starting from this optimized structure of the reactant. Transition states (TS) were identified by means of a micro/macroiteration scheme, with all pure MM atoms being adiabatically minimized at each TS search step.[55] This TS search calculation was done in QSite[47,48] using the standard method. This takes an initial guess of the TS as input and tries to find the closest saddle point to it. The initial guess was built by small modifications of the optimized structure of the reactant, the goal being to make it resemble as much as possible the believed transition state. To do so, the reacting bond C1–H2 (the carbon–hydride bond of the NADPH) was elongated manually to position the hydride halfway between the C1 carbon, where the hydride is initially attached, and the GLD carbonyl carbon C3. The carboxyl double bond of the GLD was also elongated to mimic the transition from a carbonyl double bond to an alcohol single bond. Also, to help the TS search process, QSite allows one to indicate as an input what bonds are supposed to be made or broken. This was done by adding a connect section to the input file, where C1–H2 hydride bond was defined as the reaction coordinate. To find the reactant and product associated with this saddle point, the TS was minimized at the same level of theory. The nature of the structures was confirmed from the analysis of the Hessian.

■ ASSOCIATED CONTENT

ⓈSupporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsomega.7b00815.

> Protonated HIP110; unprotonated HIE110; effect of the basis set and the QM size on selected geometrical characteristics; comparison of key distances; energies of the minimized snapshots (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: tell.tuttle@strath.ac.uk. Tel: +44 (0) 141 548 2290.

ORCID Ⓞ

Colin M. Edge: 0000-0002-4964-7759
Tell Tuttle: 0000-0003-2300-8921

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ABBREVIATIONS

AR, aldose reductase; GLD, D-glyceraldehyde; QM/MM, quantum mechanics/molecular mechanics; MD, molecular dynamics

■ REFERENCES

(1) Penning, T. M. The Aldo-Keto Reductases (AKRs): Overview. *Chem.-Biol. Interact.* **2015**, *234*, 236−246.
(2) Ramana, K. V.; Srivastava, S. K. Aldose Reductase: A Novel Therapeutic Target for Inflammatory Pathologies. *Int. J. Biochem. Cell Biol.* **2010**, *42*, 17−20.
(3) Petrash, J. M. All in the Family: Aldose Reductase and Closely Related Aldo-Keto Reductases. *Cell. Mol. Life Sci.* **2004**, *61*, 737−749.
(4) Maccari, R.; Ottanà, R. Targeting Aldose Reductase for the Treatment of Diabetes Complications and Inflammatory Diseases: New Insights and Future Directions. *J. Med. Chem.* **2015**, *58*, 2047−2067.
(5) Chatzopoulou, M.; Alexiou, P.; Kotsampasakou, E.; Demopoulos, V. J. Novel Aldose Reductase Inhibitors: A Patent Survey (2006–Present). *Expert Opin. Ther. Pat.* **2012**, 1303.
(6) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures. *J. Mol. Biol.* **1977**, *112*, 535−542.
(7) Wilson, D. K.; Bohren, K. M.; Gabbay, K. H.; Quiocho, F. A. An Unlikely Sugar Substrate Site in the 1.65 A Structure of the Human Aldose Reductase Holoenzyme Implicated in Diabetic Complications. *Science* **1992**, *257*, 81−84.
(8) Feldman, H. B.; Szczepanik, P. A.; Havre, P.; Corrall, R. J. M.; Yu, L. C.; Rodman, H. M.; Rosner, B. A.; Klein, P. D.; Landau, B. R. Stereospecificity of the Hydrogen Transfer Catalyzed by Human Placental Aldose Reductase. *Biochim. Biophys. Acta, Enzymol.* **1977**, *480*, 14−20.
(9) Liu, S. Q.; Bhatnagar, A.; Srivastava, S. K. Bovine Lens Aldose Reductase: pH-Dependence of Steady-State Kinetic-Parameters and Nucleotide-Binding. *J. Biol. Chem.* **1993**, *268*, 25494−25499.
(10) Bohren, K. M.; Grimshaw, C. E.; Lai, C. J.; Harrison, D. H.; Ringe, D.; Petsko, G. A.; Gabbay, K. H. Tyrosine-48 Is the Proton Donor and Histidine-110 Directs Substrate Stereochemical Selectivity in the Reduction Reaction of Human Aldose Reductase: Enzyme Kinetics and Crystal Structure of the Y48H Mutant Enzyme. *Biochemistry* **1994**, *33*, 2021−2032.
(11) Várnai, P.; Richards, W. G.; Lyne, P. D. Modelling the Catalytic Reaction in Human Aldose Reductase. *Proteins: Struct., Funct., Genet.* **1999**, *37*, 218−227.
(12) Várnai, P.; Warshel, A. Computer Simulation Studies of the Catalytic Mechanism of Human Aldose Reductase. *J. Am. Chem. Soc.* **2000**, *122*, 3849−3860.
(13) Lee, Y. S.; Hodoscek, M.; Brooks, B. R.; Kador, P. F. Catalytic Mechanism of Aldose Reductase Studied by the Combined Potentials of Quantum Mechanics and Molecular Mechanics. *Biophys. Chem.* **1998**, *70*, 203−216.
(14) Grimshaw, C. E.; Bohren, K. M.; Lai, C.-J.; Gabbay, K. H. Human Aldose Reductase: Rate Constants for a Mechanism Including Interconversion of Ternary Complexes by Recombinant Wild-Type Enzyme. *Biochemistry* **1995**, *34*, 14356−14365.
(15) 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.
(16) Lodola, A.; De Vivo, M. Chapter 11 − The Increasing Role of QM/MM in Drug Discovery. *Adv. Protein Chem. Struct. Biol.* **2012**, *87*, 337−362.

(17) 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. Engl.* **2006**, *45*, 6856−6859.

(18) El-Kabbani, O.; Ruiz, F.; Darmanin, C.; Chung, R. P.-T. Aldose Reductase Structures: Implications for Mechanism and Inhibition. *Cell. Mol. Life Sci.* **2004**, *61*, 750−762.

(19) Del Corso, A.; Cappiello, M.; Mura, U. From a Dull Enzyme to Something Else: Facts and Perspectives Regarding Aldose Reductase. *Curr. Med. Chem.* **2008**, *15*, 1452−1461.

(20) Tarle, I.; Borhani, D. W.; Wilson, D. K.; Quiocho, F. A.; Petrash, J. M. Probing the Active Site of Human Aldose Reductase. Site-Directed Mutagenesis of Asp-43, Tyr-48, Lys-77, and His-110. *J. Biol. Chem.* **1993**, *268*, 25687−25693.

(21) Ruiz, F.; Hazemann, I.; Mitschler, A.; Joachimiak, A.; Schneider, T.; Karplus, M.; Podjarny, A. The Crystallographic Structure of the Aldose Reductase−IDD552 Complex Shows Direct Proton Donation from Tyrosine 48. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 1347−1354.

(22) Blakeley, M. P.; Ruiz, F.; Cachau, R.; Hazemann, I.; Meilleur, F.; Mitschler, A.; Ginell, S.; Afonine, P.; Ventura, O. N.; Cousido-Siah, A.; et al. Quantum Model of Catalysis Based on a Mobile Proton Revealed by Subatomic X-Ray and Neutron Diffraction Studies of H-Aldose Reductase. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 1844−1848.

(23) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.

(24) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and Use of Quantum Mechanical Molecular Models. 76. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.

(25) Becke, A. D. A New Mixing of Hartree−Fock and Local Density-Functional Theories. *J. Chem. Phys.* **1993**, *98*, 1372.

(26) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623−11627.

(27) Hariharan, P. C.; Pople, J. A. The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies. *Theor. Chim. Acta* **1973**, *28*, 213−222.

(28) *Molecular Operating Environment (MOE)*, 2013.08; Chemical Computing Group Inc.: Montreal, QC, Canada, 2015.

(29) Ullmann, G. M.; Knapp, E.-W. Electrostatic Models for Computing Protonation and Redox Equilibria in Proteins. *Eur. Biophys. J.* **1999**, *28*, 533−551.

(30) Uranga, J.; Mikulskis, P.; Genheden, S.; Ryde, U. Can the Protonation State of Histidine Residues Be Determined from Molecular Dynamics Simulations? *Comput. Theor. Chem.* **2012**, *1000*, 75−84.

(31) Alexov, E.; Mehler, E. L.; Baker, N.; Baptista, A. M.; Huang, Y.; Milletti, F.; Nielsen, J. E.; Farrell, D.; Carstensen, T.; Olsson, M. H. M.; et al. Progress in the Prediction of pKa Values in Proteins. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 3260−3275.

(32) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, 525.

(33) Schrödinger, LLC. *N. MacroModel*, version 9.6; Schrödinger, LLC: New York, 2008.

(34) Bruno, I. J.; Cole, J. C.; Kessler, M.; Luo, J.; Motherwell, W. D. S.; Purkis, L. H.; Smith, B. R.; Taylor, R.; Cooper, R. I.; Harris, S. E.; et al. Retrieval of Crystallographically-Derived Molecular Geometry Information. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2133−2144.

(35) Bowers, K. J.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A. et al. In *Molecular Dynamics—Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters*, Proceedings of the

2006 ACM/IEEE Conference on Supercomputing − SC '06; ACM Press: New York, NY, 2006; p 84.

(36) (a) *Schrödinger Release 2014-3: Desmond Molecular Dynamics System*, version 3.9; D. E. Shaw Research: New York, NY, 2014. (b) *Maestro-Desmond Interoperability Tools*, version 3.9; Schrödinger: New York, NY, 2014.

(37) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926.

(38) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. Explicit Reversible Integrators for Extended Systems Dynamics. *Mol. Phys.* **1996**, *87*, 1117−1157.

(39) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577.

(40) Nash, S. G. A Survey of Truncated-Newton Methods. *J. Comput. Appl. Math.* **2000**, *124*, 45−59.

(41) Schrödinger, LLC. *N. Impact*, version 5.5; Schrödinger, LLC: New York, 2005.

(42) Polak, E.; Ribiere, G. Note sur la convergence de méthodes de directions conjuguées. *ESAIM: Math. Modell. Numer. Anal. − Modél. Math. Anal. Numér.* **1969**, *3*, 35−43.

(43) Kulik, H. J.; Zhang, J.; Klinman, J. P.; Martínez, T. J. How Large Should the QM Region Be in QM/MM Calculations? The Case of Catechol *O*-Methyltransferase. *J. Phys. Chem. B* **2016**, *120*, 11381−11394.

(44) Liao, R.-Z.; Thiel, W. Convergence in the QM-Only and QM/MM Modeling of Enzymatic Reactions: A Case Study for Acetylene Hydratase. *J. Comput. Chem.* **2013**, *34*, 2389−2397.

(45) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Computational Enzymatic Catalysis − Clarifying Enzymatic Mechanisms with the Help of Computers. *Phys. Chem. Chem. Phys.* **2012**, 12431.

(46) Lonsdale, R.; Harvey, J. N.; Mulholland, A. J. A Practical Guide to Modelling Enzyme-Catalysed Reactions. *Chem. Soc. Rev.* **2012**, 3025.

(47) Schrödinger, LLC. *Small-Molecule Drug Discovery Suite 2014-3: QSite*, version 6.4; Schrödinger, LLC: New York, NY, 2014.

(48) Murphy, R. B.; Philipp, D. M.; Friesner, R. A. A Mixed Quantum Mechanics/molecular Mechanics (QM/MM) Method for Large-Scale Modeling of Chemistry in Protein Environments. *J. Comput. Chem.* **2000**, *21*, 1442−1457.

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

(50) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

(51) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, 785.

(52) Lin, H.; Truhlar, D. G. QM/MM: What Have We Learned, Where Are We, and Where Do We Go from Here? *Theor. Chem. Acc.* **2007**, *117*, 185−199.

(53) Singh, U. C.; Kollman, P. A. A Combined Ab Initio Quantum Mechanical and Molecular Mechanical Method for Carrying out Simulations on Complex Molecular Systems: Applications to the CH$_3$Cl + Cl$^-$ Exchange Reaction and Gas Phase Protonation of Polyethers. *J. Comput. Chem.* **1986**, *7*, 718−730.

(54) *QSite 6.5 User Manual*; Schrödinger Press, 2014.

(55) Prat-Resina, X.; Bofill, J. M.; González-Lafont, À.; Lluch, J. M. Geometry Optimization and Transition State Search in Enzymes: Different Options in the Microiterative Method. *Int. J. Quantum Chem* **2004**, *98*, 367−377.