QM/MM study of the catalytic reaction of aphid myrosinase

Jafari, Sonia; Ryde, Ulf; Irani, Mehdi

Published in:
International Journal of Biological Macromolecules

DOI:
10.1016/j.ijbiomac.2024.130089

2024

Document Version:
Peer reviewed version (aka post-print)

QM/MM Study of the Catalytic Reaction of Aphid Myrosinase

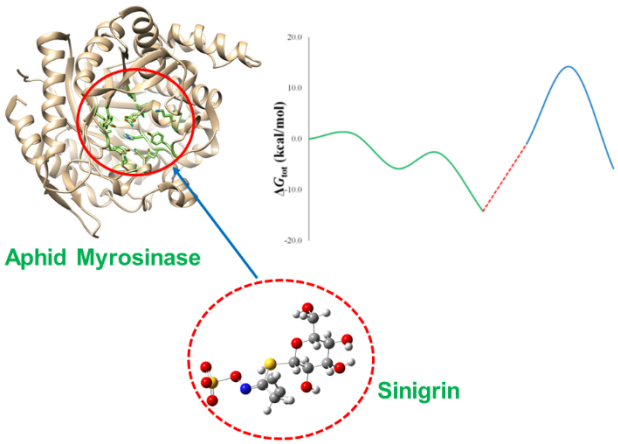
Sonia Jafari,[1] Ulf Ryde[2], Mehdi Irani,[1]*

[1]Department of Chemistry, University of Kurdistan, P. O. Box 66175-416, Sanandaj, Iran

[2]Department of Theoretical Chemistry, Lund University, P. O. Box 124, SE-221 00 Lund, Sweden

Correspondence to Mehdi Irani, m.irani@uok.ac.ir, Tel: +98 – 9128018046

2024-04-28

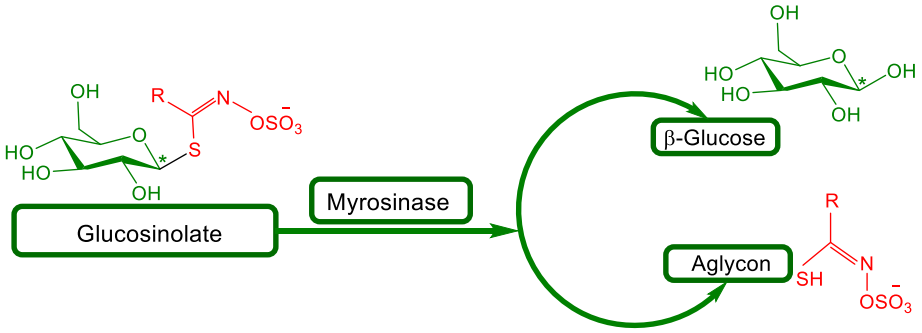


A reaction mechanism confirmed by computational methods was proposed for aphid myrosinase based on a QM/MM model.

1 Introduction

Glucosinolates are secondary metabolites derived from amino acids that are found in plants (Scheme 1). They have a common structure consisting of a β-thioglucose moiety, a sulfonated oxime moiety, and a variable side chain, which determines their classification into three main groups: aliphatic, indolic, and aromatic glucosinolates [1]. These compounds and their breakdown products are responsible for the characteristic taste and aroma of crops such as horseradish, cabbage, mustard, and broccoli, making their glucosinolate content highly valuable. Glucosinolates are well-known for their role in plant defense against pathogens and insects, but they may also have other functions in plant metabolism [2].

Myrosinase, also known as thioglucosidase or thioglucoside glucohydrolase (EC 3.2. 1.147), is an enzyme that hydrolyzes glucosinolates. This enzymatic process releases a labile aglycone that rapidly rearranges, liberating sulfur and producing a variety of toxic metabolites, such as isothiocyanates, thiocyanates, cyanoepithioalkanes, and nitriles. The specific reaction products depend on factors such as pH, the presence of ferrous ions, epithiospecifier protein, and the nature of the glucosinolate side chain.



Scheme 1. The reaction catalyzed by myrosinase, an enzyme that hydrolyzes glucosinolates. The anomeric carbon of the glucose moiety is marked with a star. R represents the variable side chain of the glucosinolate; for example, R is allyl in sinigrin.

The myrosinase–glucosinolate system in plants has been widely studied [2,3], but myrosinase-like activity has also been found in organisms, such as fungi, bacteria in the digestive tract, mammalian tissue, and aphids on cruciferous plants [2]. One of the most fascinating aspects of the co-evolution of plants and insects is how some herbivorous insects, such as cabbage aphids, have adapted to use the plant's glucosinolate–myrosinolate system as a defense against their predators, by producing their own myrosinase and sequestering glucosinolates from the plant [4].

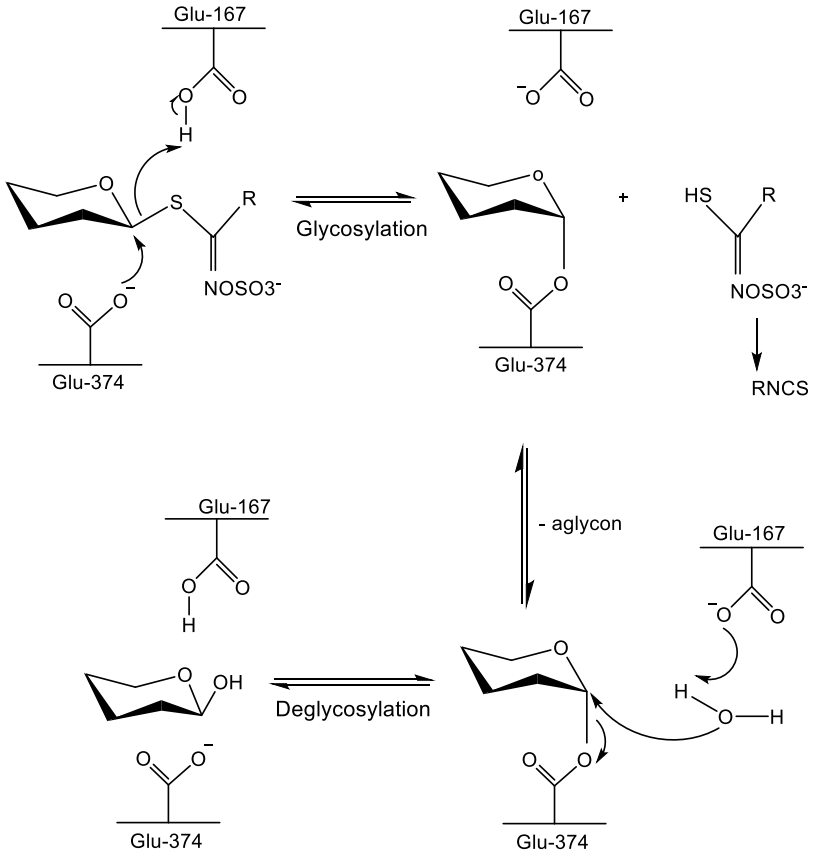
Aphid myrosinase is a globular protein with a cleft at the core of the structure, where the putative active-site residues are located. The protein has a subunit molecular mass of 57–58 kDa and an isoelectric point of 4.9. The enzyme has an optimal temperature of around 40˚C and can hydrolyze several common plant glucosinolates, such as sinigrin and glucotropaeolin [4–6]. However, unlike plant myrosinases, aphid myrosinase is inhibited by ascorbate at concentrations

that usually activate plant myrosinases [5,6].

Aphid myrosinase and plant myrosinase have very similar main structures, with superimposable α-carbon skeletons [7]. Bones et al. compared the sequences of the two enzymes and identified equivalent residues [8], shown in Table S1 in the Supporting Information. Aphid myrosinase is more similar to classic β-glycosidase than to *Sinapis alba* myrosinase in terms of structure and active-site residues. However, aphid myrosinase also has unique structural features that are not observed in either β-glycosidases or other myrosinases.

Aphid myrosinase has a different catalytic system from plant myrosinase, even though both enzymes hydrolyze glucosinolates. Aphid myrosinase uses two carboxylic amino acids, one as a nucleophile and the other as an acid/base residue (cf. Scheme 2), which is common in β-glycosidases. These amino acids are Glu-167 and Glu-374, and they donate and accept protons during the reaction. Plant myrosinase, on the other hand, has a glutamine instead of the acid/base residue. The corresponding amino acids in plant myrosinase are Gln-187 and Glu-409.

Plant and aphid myrosinases cleave glucosinolates through different mechanisms. Plant myrosinases, such as *Sinapis alba* myrosinase, require ascorbic acid to cleave glucosinolates. In the first step of the reaction, a carboxylic acid residue in the enzyme attacks the anomeric carbon of the glucosinolate molecule. This results in the formation of a glycosyl-enzyme intermediate and the release of the aglycon moiety of the glucosinolate. Next, an ascorbate molecule binds to the active site and acts as a base, deprotonating a water molecule and facilitating its nucleophilic attack on the anomeric center. Finally, glucose and ascorbate are released from the active site. A schematic representation of this mechanism can be found in Scheme S2 in the supporting information, where ascorbate's structure is depicted in green. This mechanism was proposed based on a crystallographic study by Burmeister et al. [9] and later confirmed by a quantum mechanical/molecular mechanical (QM/MM) study [10]. On the other hand, aphid myrosinases do not require ascorbic acid to cleave glucosinolates. Instead, they use an acid/base catalyst. Rossiter et al. examined the mechanism of the insect enzyme by protein homology modeling (cf. Scheme 2) [7]. They suggested that the carboxyl residue that acts as a general acid protonates the glycosidic sulfur and the other residue simultaneously performs a nucleophilic attack at the anomeric carbon, producing an enzyme-glycosyl intermediate with an inverted anomeric configuration, while the aglycon moiety leaves the active site. In the next step, the carboxylic residue that acted as the general acid in the first step activates a water molecule to attack the anomeric center of the intermediate, and the covalent bond between the enzyme and the glycosyl is broken, releasing the β-glucose molecule. Like the plant myrosinase mechanisms [9,10], this mechanism retains the anomeric carbon configuration, but it uses an acid/base catalyst rather than ascorbic acid.



Scheme 2. A generalized mechanism for aphid myrosinase based on Rossiter et al. [11]

In this work, we use computational methods to better understand the chemical processes that occur during the hydrolysis reaction of sinigrin by aphid myrosinase.We provide a pioneering exploration of the catalytic mechanism using a sophisticated QM/MM approach coupled with innovative in silico mutagenesis and molecular dynamics (MD) simulations. The distinctive aspect of our study lies in the comprehensive investigation of the structural and functional intricacies of this enzyme during the hydrolysis reaction of sinigrin. Specifically, our study of the critical Arg-312 residue and its impact on catalysis, achieved through neutralization and mutagenesis studies, unveils previously unexplored facets of the enzyme's active site. Additionally, our utilization of MD simulations to decipher the probable protonation states of key residues (a critical factor in the modeling of enzymatic reactions [10]) represents a novel strategy in elucidating the dynamic behavior of enzymatic systems. By integrating these advanced computational methodologies, our research contributes a novel perspective to the understanding of aphid myrosinase's catalytic machinery, paving the way for innovative approaches in enzyme studies and potential applications in drug design and agricultural strategies.

Our investigation holds significant potential for practical applications. Understanding the intricate details of how this enzyme facilitates the hydrolysis reaction of sinigrin provides valuable insights into the chemical defense system of aphids and their host plants. This knowledge is pivotal for agricultural research, offering opportunities for developing novel strategies in pest control. By

elucidating the specific roles of key residues and employing computational simulations to explore the enzyme's behavior, our study lays the groundwork for designing targeted inhibitors or modulators that could potentially disrupt aphid myrosinase activity. Furthermore, the findings from our research can contribute to the broader field of enzymology and could inspire future studies aimed at manipulating enzymatic processes for agricultural or biomedical purposes.

2 Methods

2.1 The Protein

(*Brevicoryne brassicae* (1WCG PDB ID) [8] which is composed of two subunits (chain A and B). Each chain has 464 residues. Chain A has two glycerol molecules (GOL) and Chain B has three GOL molecules. Two residues from the N-terminal of both chains of the protein are not visible in the crystal structure (Met-1 and Asp-2). No attempts were made to model these missing residues because they are far from the active site and are disordered in the experimental structure.

There are 41 residues with two alternative conformations in the crystal structure that have the same occupancy. In all cases, we quite arbitrarily selected to include the A conformation in the calculations. For the QM/MM calculations, two GOL molecules of chain A and the entire chain B were removed. All crystal waters of chain A were kept in the calculations.

The protonation states of all residues were determined using the hydrogen-bond pattern, solvent accessibility, and the possible formation of ionic pairs. They were also checked using PROPKA [12–14]. All Asp residues were assumed to be negativity charged, while all Arg, and Lys residues were assumed to be positivity charged. All Cys residues were assumed to be protonated (neutral). While all the Glu residues were initially assumed to be negatively charged, we conducted further investigation into the protonation states of Glu-167 and Glu-423 through MD simulations (cf. section 3.1). From this study, we decided that they are neutral and negatively charged, respectively. A thorough manual investigation of all His residues gave the following protonation assignment: His-191 and His-308 were assumed to be protonated on the ND1 atom, His-39, His-59, and His-257 were presumed to be protonated on both the ND1 and NE2 atoms (and therefore positively charged), whereas His-196, His-202, and His-431 were modeled with a proton on the NE2 atom. The protonation state of His-122 needed further investigation. Thus, to choose the accurate protonation state of His-122, Glu-167, and Glu-423, we performed 27 sets of molecular dynamics (MD) simulations with different protonation state combinations of these residues (cf.

Table 2). Details of the simulations are described in the next subsections. Similar methods have been used to determine preferred protonation states of the active site glutamates of glyoxalase I [15,16], His residues in three proteins [17], and homocitrate and its nearby residues in nitrogenase [18].

2.2 MD Simulations

All MD simulations were performed using the GPU-accelerated *pmemd* code [19–21] of AMBER 18 [22]. The protein was modeled by the Amber ff14SB [23] force field and ligands were described with the general Amber force field (GAFF) [24]. Hetero-molecules were optimized at the B3LYP/6-31G(d) [25–29] level of theory and electrostatic potentials were calculated at the Hartree−Fock/6-31G(d) level of theory with points sampled according to the Merz−Kollman scheme [30], which has been shown to give the most reliable charges in previous studies [J. Comp. Chem. 19, 377]. These were calculated by the Gaussian 16 program package [31]. Atomic charges were then fitted to these potentials using the RESP procedure [32], as implemented in the antechamber program [32], which also assigned GAFF atom types to the molecules. The force field parameters of the hetero groups are collected in the Supporting Information.

The setup of the MD simulations is similar to that in our recent works [15,16]. First, the prepared protein was immersed in a periodic truncated octahedral box of TIP3P water molecules [33], extending at least 10 Å from the solute using the *tleap* program in the Amber suite [22]. This made a final system containing ~125,000 atoms. Next, the system was subjected to 1000 cycles of minimization, restraining the heavy atoms toward the starting crystal structure with a force constant of 100 kcal/mol/Å^2. Then, two 20 ps equilibrations (one with constant-volume and one with constant-pressure) were performed with the same restraints, but the force constant was 50 kcal/mol/Å^2. After that, the system was equilibrated for 1 ns without any restraints. Finally, the production simulations were performed for 100 ns, and coordinates were sampled every 10 ps.

Root-mean-squared deviations (RMSDs) were calculated with the AMBER *cpptraj* module [34], analyzing trajectories with saved coordinates in the production simulations, and using the crystal structure as the reference. Reported values are averages over these 10000 sets of coordinates. A schematic representation of the MD processes is shown in Fig. S3 in the Supporting Information.

2.3 Sequence Analysis

The crystallographic article on aphid myrosinase did not identify the catalytic nucleophile and acid/base residues [8]. Therefore, we determined them by comparing the sequences of aphid and plant myrosinase, which have known catalytic residues. We transferred the sequence data from Bones et al. [9] to a spreadsheet and compared them (see Table S1 in the Supporting Information for the final comparison results). In Table 1, we show the equivalent residues of the two enzymes in

the glucose and aglycon binding sites of the active site. The comparison results suggest that Glu-167 is the acid/base residue in aphid myrosinase, as it corresponds to Gln-187 in plant myrosinase. This glutamine in turn is analogous to the acid/base residues in *O*-glycosidases [35–38]. This is consistent with our MD results that showed that this residue is protonated (see section 3.1), as expected for an acid/base residue. Moreover, Table 1 suggests that Glu-374 is the nucleophilic residue in aphid myrosinase, as it corresponds to Glu-409 in plant myrosinase. The sequence comparison also suggests that Glu-423 corresponds to Glu-464 in *Sinapis alba* myrosinase. However, our MD results showed that this residue is more stable when it is charged, unlike the equivalent residue from *Sinapis alba*, which must be protonated [10].

The active site of aphid myrosinase is located at the bottom of a deep cleft. The glucose binding site is well-conserved for all β-glucosidases and has been established using 2-F-glucose derivatives of substrates. The aglycon binding site is highly variable due to the diversity of substrates that β-glucosidases can catalyze. Table 1 shows that some residues in the aglycon binding site have no counterparts in plant myrosinase and some others are exchanged with different residues. For example, Lys-173 changes to an Arg residue, while another residue that is close to the sulfur atom of the aglycon moiety, Ala-170, changes to a serine (Ser-190 in plant myrosinase). Another difference in the aglycon binding site of aphid and plant myrosinases is that Trp-424 in aphid myrosinase is replaced by Phe-465 in plant myrosinase. This results in an extra hydrogen bond between the nitrogen atom of the indole ring and the hydroxyl group at position 3 of the glucose ring (see Fig. 1 for the numbering of carbon atoms in the glucose ring and the active site residues of aphid myrosinase). This may increase the contribution of the glucose group to the substrate affinity of aphid myrosinase.

Glycosidases require a water molecule to catalyze the hydrolysis of the glycosidic bond. To locate this water molecule, we analyzed the 1WCG PDB file graphically. It can be seen that HOH-2774A is a plausible candidate for a -OH donor in the catalytic mechanism (it is located close to the nucleophilic and the general acid/base residues; cf. Fig. 1).

Table 1. Equivalent residues in the insect and plant myrosinase in the glucose and aglycon binding sites. Dashes indicate that there is no equivalent residue. For a full comparison of all residues, see Table S1 in the Supporting Information.

Glucose binding site		Aglycon binding site	
Aphid	Sinapis alba	Aphid	Sinapis alba
Gln-19	Gln-39	Ile-169	Tyr-189
His-122	His-141	Ala-170	Ser-190
Trp-123	Trp-142	Lys-173	Arg-194
Asn-166	Asn-186	Tyr-180	---
Glu-167	Gln-187	---	Ile-257
Tyr-309	Tyr-330	Val-228	Arg-259
Glu-374	Glu-409	---	Phe-282
Trp-416	Trp-457	---	Phe-331
Glu-423	Glu-464	Arg-312	---
Trp-424	Phe-465	Phe-324	---
Phe-432	Phe-473	---	Phe-371
		Tyr-346	---

		Ile-347	---

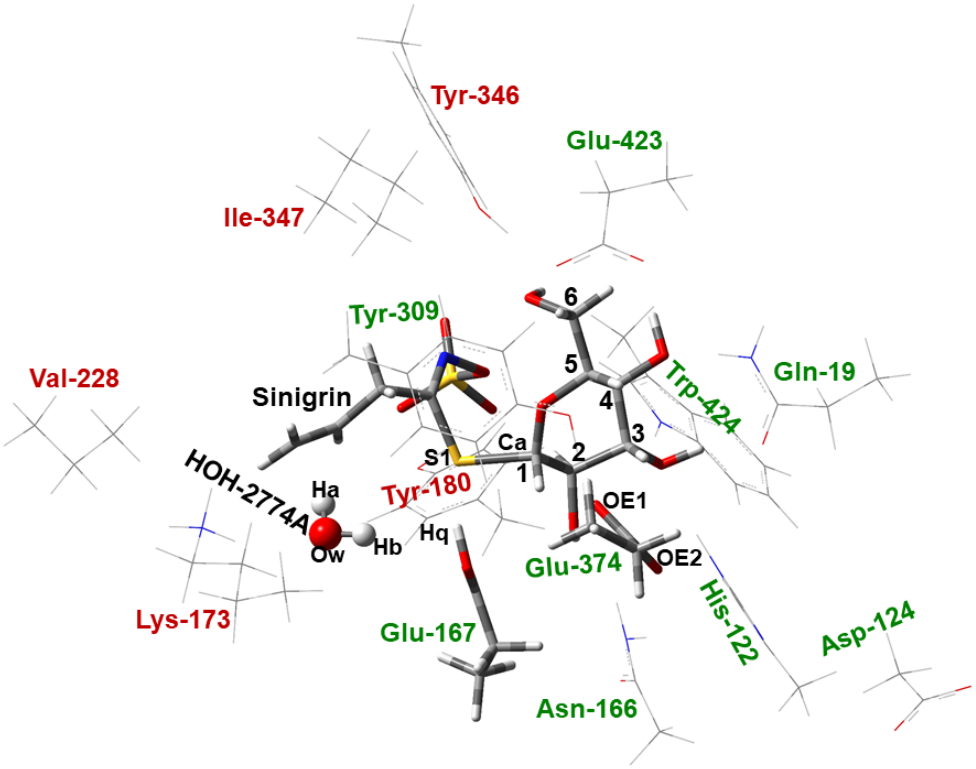


Fig. 1. The substrate and the active site residues of aphid myrosinase in a relative conformation. The carbon atoms of the glucose ring are numbered from 1 to 6. The residues that are in the glucose binding site are labeled in green, while the residues that are in the aglycon binding site are labeled in red. This figure also shows the QM system used in the QM/MM calculations.

2.4 Docking Sinigrin Into the Active Site

As mentioned above, the enzyme consists of two identical monomers. For QM/MM calculations, we removed all atoms and groups in chain B to reduce the computational effort, ~~economical~~ as was done in many similar studies [39,40]. We then superimposed the active site residues of the aphid myrosinase and the myrosinase of *Sinapis alba*, for which we already had a docked sinigrin [10]. The two active sites superimpose nicely with an RMSD of 0.98 Å (see Fig. 2.). Because the active sites are well imposed, the substrate should occupy a similar position in the aphid myrosinase. After that, we performed an RMS fit and found the position of sinigrin in aphid myrosinase. These analyses were done using *changepdb*, an application developed locally. After placing the substrate, we removed some crystal waters that came into short contact with the docked sinigrin or other residues (HOH-2193, 2449, 2074, 2395, 2722, 2308, 2172, 2452, 2721, 2629, 2161, 2386, 2448, 2451, and 2774).

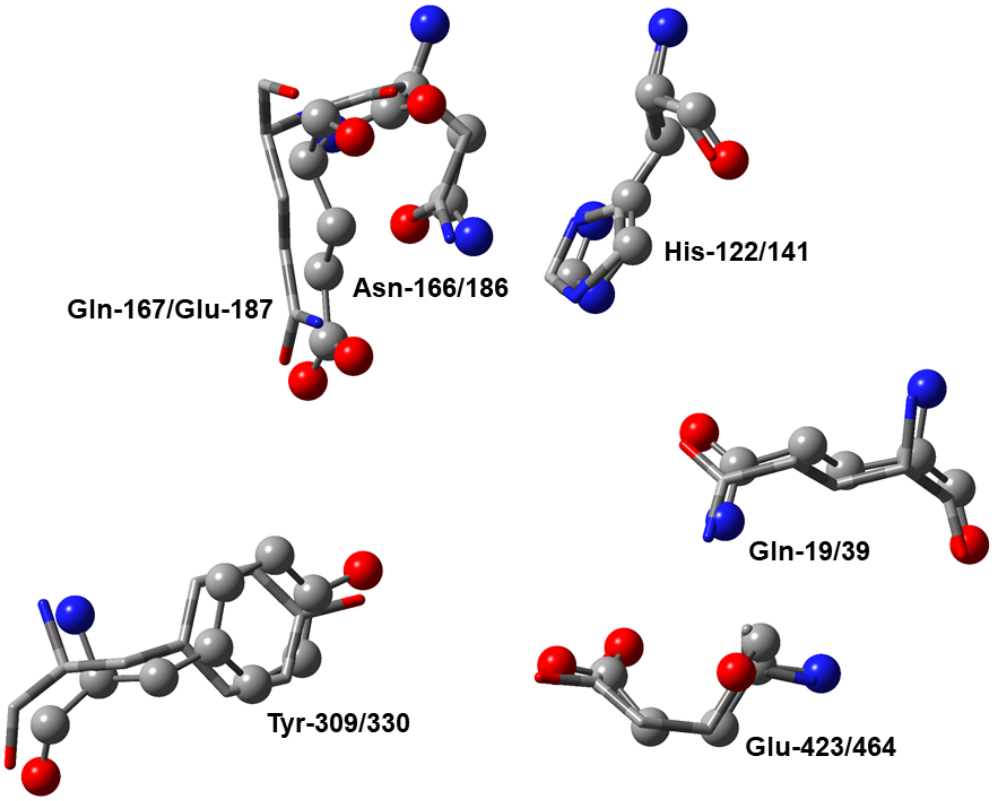


Fig. 2. Superimposition of some active site residues of aphid/plant myrosinase. Residues of the aphid myrosinase are shown with balls and those of *Sinapis alba* by tubes.

2.5 QM/MM Calculations

After docking the substrate into the 1WCG structure, the protein–ligand complex was solvated in a periodic truncated octahedral box of TIP3P [33] water molecules, extending at least 20 Å from the solute. After solvation, first, only the hydrogen atoms and then both the hydrogen atoms and added water molecules were subjected to 1000 cycles of minimizations with the heavy atoms of the protein restrained. This was followed by a 10 ps constant-volume equilibration with the same restraints. Finally, the system was equilibrated by a 1 ns constant-volume simulation and a 1 ns simulated annealing at constant pressure with the restraints (the force constant for the restraints in all steps was 1000 kcal/mol/Å^2). Bond lengths to hydrogen atoms were constrained by the SHAKE algorithm [41] in the equilibrations (but not in the minimizations), allowing for a time step of 2 fs during the simulations. The temperature was kept constant at 300 K using Langevin dynamics (the recommended method for temperature control in the AMBER software [22]) with a collision frequency of 2 ps^{-1} [42]. The pressure was kept constant at 1 atm using Berendsen's weak coupling isotropic algorithm with a relaxation time of 1 ps [43]. Long-range electrostatics were handled by particle-mesh Ewald summation [44] with a fourth-order B-spline interpolation and a tolerance of 10^{-5} (the recommended setting in the AMBER software [22]). The cut-off radius for Lennard–Jones interactions was set to 8 Å. After the final equilibration, the octahedral system was truncated to a spherical one with a radius of 38 Å from the geometric center of the protein.

We also performed calculations for two mutants of myrosinase (Arg312Met and Arg312Val). The mutations were built manually by changing the residue names, deleting atoms, or changing atom names. For the Arg312Met mutation, the CZ, NH1, and NH2 atoms were deleted, while CD and NE were renamed to SD and CE. For the Arg312Val mutation, the CD, NE, CZ, NH1, and NH2 atoms were deleted and CG to CG1. After changing the residue name, tleap in the AMBER software [22] automatically generate coordinates of missing atoms. The mutated proteins were also relaxed by MD simulations before QM/MM calculations, as described in the previous paragraph.

After the relaxation of the wild-type and mutated systems, QM/MM calculations were performed with the ComQum software (version 2023) [45,46]. In this approach [47,48], the protein and solvent are split into two subsystems: System 1 (the QM region) is treated by QM methods and System 2 (the MM region) is represented by MM methods. The latter system is fixed at the positions of the equilibrated structures. In the QM part of the QM/MM calculations, system 1 was represented by a wave function, whereas all the other atoms were represented by an array of partial point charges, one for each atom, taken from MM libraries. Thereby, the polarization of the QM system by the surroundings is included in a self-consistent manner. These calculations were performed at the TPSS-D3/def2-SV(P) level of theory [49–54], using the Turbomole software (version 7.2) [55,56] and were accelerated by the resolution-of-identity approximation [51,52]. The MM part of the QM/MM calculations was performed with the Amber software (version 14) [57], using the Amber ff14SB [23] and GAFF [24] force fields for the protein and the substrate, respectively. Water molecules were described by the TIP3P model [33]. When there were bonds between systems 1 and 2 (junctions), the hydrogen link-atom approach was employed. In this approach, the QM system was capped with hydrogen atoms (hydrogen link atoms, HL), the positions of which are linearly related to the corresponding carbon atoms (carbon link atoms, CL) in the full system [45,58]. All atoms were included in the point-charge model, except the CL atoms [59]. The point charges do not necessarily sum up to an integer, because the Amber force field does not employ charge groups [57]. The total QM/MM energy in ComQum is calculated as [45,46]

$$E_{\mathrm{QM/MM}} = E^{\mathrm{HL}}_{\mathrm{QM1+ptch2}} + E^{\mathrm{CL}}_{\mathrm{MM123},q_1=0} - E^{\mathrm{HL}}_{\mathrm{MM1},q_1=0} \quad (1)$$

where $E^{\mathrm{HL}}_{\mathrm{QM1+ptch2}}$ is the QM energy of system 1, truncated by HL atoms and embedded in the set of point charges representing system 2 (but excluding the self-energy of the point charges). $E^{\mathrm{HL}}_{\mathrm{MM1},q_1=0}$ is the MM energy of the QM system, still truncated by HL atoms, but without any electrostatic interactions. Finally, $E^{\mathrm{CL}}_{\mathrm{MM12},q_1=0}$ is the classical energy of all atoms in the system with CL atoms and with the charges of the QM system set to zero (to avoid double counting of the electrostatic interactions). By using this approach, which is similar to the one used in the Oniom method [60], errors caused by the truncation of the quantum system should partly cancel out. Thus, ComQum

employs a subtractive scheme with electrostatic embedding and van der Waals link-atom corrections [61].

2.6 The QM System

After assigning the best protonation states of the active site residues and docking the substrate into the enzyme, we equilibrated the final structure as described in section 2.5. The QM system consisted of the Gln-19, His-122, Asp-124, Asn-166, Glu-167, Lys-173, Tyr-180, Val-228, Tyr-309, Tyr-346, Ile-347, Glu-374, Glu-423, Trp-424, sinigrin as the substrate, and HOH-2774A. The amino acids were included up to their alpha carbons. The active site of aphid myrosinase and its coordinating ligands are shown in Fig. 1.

3 Results and Discussion

We first discuss the assignments of the proper protonation state of the aphid myrosinase's important residues. Then, we will investigate the catalytic mechanism for the wild-type enzyme, and after that, we will extend the study to examine the effects of two mutations on the catalytic reaction.

3.1 Assigning the Most Proper Protonation States of Protein Residues by MD Simulations

We have recently demonstrated that, in QM/MM calculations, assigning the proper protonation state of active site residues is more important than the size of the QM system [10]. Also, prior works have demonstrated that MD simulations are an effective technique for assigning protonation states of protein residues [10,17,18]. Therefore, we conducted MD simulations with various protonation states for His-122, Glu-167, and Glu-423 (active site residues with more than one possible protonation state; cf. Fig. 3 for the position of these residues with respect to the active site).

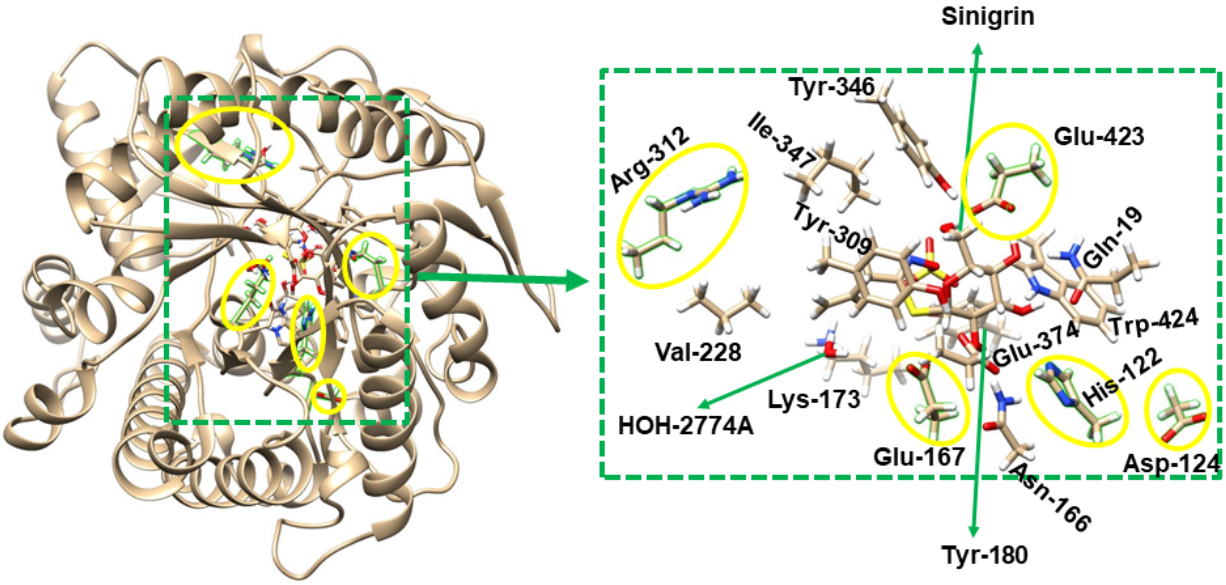


Fig. 3. The active site residues of aphid myrosinase in chain A. Chain B is not shown for clarity. The figure was

made with UCSF Chimera, a software developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311 [62].

To determine which protonation states are more accurate for these residues, we performed 27 (3^3) simulations because each of these residues has three different protonation states. The protonation states of these residues during the simulations are shown in columns 2–4 of

Table 2. We calculated RMSDs for the heavy atoms of the studied residue alone and for the heavy atoms of the studied residue and all residues with at least one atom within 3.5 Å of the residue. In the former, we see whether the studied residue stays in its crystallographic position, and the latter indicates whether any neighboring residues are moving to release any unfavorable interactions. We calculated these for each of the two subunits of the protein. If a protonation state is correct, MD simulations will reproduce the H-bonds observed in the crystal structure with low RMSDs; otherwise, the RMSDs will be higher. The last 9 rows in

Table 2 summarize the average RMSD values for each protonation state of each residue during the relevant simulations. For instance, the value next to OE2 in the last row (0.79 Å) indicates the average RMSD of Glu-423 over the simulations in which it is protonated on OE2 and over the two subunits of the protein (simulations 3, 6, 9, 12, 15, 18, 21, 24, and 27).

From the results in

Table 2 it is quite evident that Glu-167 is protonated on OE2 (the average RMSD of Glu-167 alone and that of the residue and its neighbors when it is protonated on OE2 is lower than the corresponding values when it is charged or protonated on OE1 (0.53 vs. 0.87 and 0.92 Å; compare the values in the 4th row from the end with those in the 5th and 6th rows from the end). Additionally, the results indicate that Glu-423 is charged (the average RMSD of Glu-423 alone and that of the residue and its neighbors when it is charged is lower than the corresponding values when it is protonated on OE1 or OE2 (0.56 vs. 1.00 and 0.79 Å); compare the values of the 3rd row from the end with those of the 2nd and 1st rows from the end). However, the RMSD values show that His-122 is more stable in the HIE state, but we cannot quite conclusively decide which protonation state is more proper for this residue (all protonation states of this residue have similar RMSD values; see the values in the 7th to 9th rows from the end in

Table 2). Hence, we will assign the protonation state of this residue using the more accurate QM/MM methods.

Table 2. RMSD values (in Å) of His-122, Glu-167, and Glu-423 with different protonation states in the 27 MD simulations. The table shows the protonation states of these residues in each simulation (columns 2-4) and the RMSDs of these residues and their neighboring residues (within 3.5 Å) in each subunit of the protein (A and B). The RMSDs are calculated for the heavy atoms only. The last nine rows show the average RMSD values for each residue and each protonation state over all relevant simulations. For example, the value next to OE2 in the last row (0.79 Å) means the average RMSD of Glu-423 when it is protonated on OE2 in both subunits of the protein (simulations 3, 6, 9, 12, 15, 18, 21, 24, and 27).

				Residue						Neighbors					
			Chain	A			B			A			B		
Stimulation	His-122	Glu-167	Glu-423	His-122	Glu-167	Glu-423	His-122	Glu-167	Glu-423	His-122	Glu-167	Glu-423	His-122	Glu-167	Glu-423
1	HID	NONE	NONE	0.42	1.04	0.35	0.61	0.82	1.02	0.97	0.83	1.30	1.02	0.75	1.50
2	HID	NONE	OE1	0.57	1.09	0.92	0.71	0.93	1.02	1.01	0.80	1.33	0.97	0.71	1.18
3	HID	NONE	OE2	0.22	0.95	0.88	0.84	1.04	0.64	0.89	0.77	1.34	1.08	0.94	1.24
4	HID	OE1	NONE	0.21	0.62	0.36	0.22	0.76	0.62	0.90	0.62	1.17	0.86	0.62	1.10
5	HID	OE1	OE1	0.27	0.85	1.02	0.30	0.79	0.99	0.80	0.61	1.30	0.84	0.62	1.21
6	HID	OE1	OE2	0.32	0.92	1.00	0.65	0.93	0.99	0.82	0.71	1.25	0.89	0.66	1.61
7	HID	OE2	NONE	**0.26**	0.24	0.86	**0.26**	0.71	0.81	**0.83**	0.57	1.66	**0.90**	0.63	1.52
8	HID	OE2	OE1	0.25	0.33	0.91	0.31	0.71	1.06	0.82	0.55	1.32	0.87	0.62	1.27
9	HID	OE2	OE2	0.23	0.51	1.12	1.01	0.54	0.56	0.78	0.55	1.37	0.98	0.60	1.09
10	HIE	NONE	NONE	0.42	1.24	0.64	0.48	0.96	0.72	1.05	0.89	0.90	1.06	0.78	1.14
11	HIE	NONE	OE1	0.59	0.67	0.94	0.44	1.04	0.88	1.05	0.86	1.06	1.04	0.78	1.40
12	HIE	NONE	OE2	0.40	1.05	1.12	0.30	0.85	1.04	1.05	1.01	1.30	0.88	0.64	1.24
13	HIE	OE1	NONE	0.41	1.10	0.35	0.31	0.78	0.35	0.89	0.81	1.09	0.82	0.71	0.94
14	HIE	OE1	OE1	0.38	1.04	0.96	0.41	1.12	1.03	0.88	0.77	1.27	1.00	1.01	1.21
15	HIE	OE1	OE2	0.25	1.05	0.27	0.47	1.09	0.53	0.78	0.87	1.11	1.04	1.02	1.41
16	HIE	OE2	NONE	**0.40**	0.40	0.66	**0.41**	0.64	0.22	**0.93**	0.68	1.24	**0.96**	1.01	1.01
17	HIE	OE2	OE1	0.35	0.31	1.01	0.26	0.34	1.08	0.82	0.58	1.12	0.82	0.65	1.15
18	HIE	OE2	OE2	0.49	0.30	0.99	0.41	0.58	0.40	0.96	0.76	1.31	1.01	0.97	1.40
19	HIP	NONE	NONE	1.04	0.83	0.61	0.82	0.92	0.34	1.06	0.76	1.23	1.08	0.82	1.08
20	HIP	NONE	OE1	0.21	0.66	1.10	0.28	0.76	1.10	0.84	0.61	1.18	0.91	0.64	1.15
21	HIP	NONE	OE2	0.53	0.78	0.99	0.20	0.94	0.29	0.96	0.71	1.20	0.93	0.68	1.14
22	HIP	OE1	NONE	0.30	0.44	0.50	0.64	0.76	0.38	0.79	0.63	0.90	0.90	0.72	1.07
23	HIP	OE1	OE1	0.21	0.89	1.12	0.24	0.88	0.91	0.74	0.64	1.20	0.90	0.69	1.49
24	HIP	OE1	OE2	0.21	0.67	1.06	0.92	1.03	1.04	0.91	0.70	1.52	0.96	0.72	1.29
25	HIP	OE2	NONE	**0.32**	0.36	0.41	**1.04**	0.68	0.80	**0.80**	0.53	0.90	**1.01**	0.77	1.20
26	HIP	OE2	OE1	0.74	0.44	0.95	0.22	0.61	0.93	0.94	0.65	1.14	0.86	0.64	1.38
27	HIP	OE2	OE2	0.31	0.77	0.30	0.21	0.63	1.06	0.78	0.69	1.11	0.81	0.68	1.37
		His-122	HID	0.43						0.90					
			HIE	0.40						0.95					
			HIP	0.47						0.90					
		Glu-167	NONE		0.92						0.78				
			OE1		0.87						0.73				
			OE2		0.53						0.68				
		Glu-423	NONE			0.56						1.16			
			OE1			1.00						1.24			
			OE2			0.79						1.29			

His-122 and Asp-124 are adjacent residues in the active site of aphid myrosinase, as shown in Fig. 3. Then, we tested different combinations of protonation states of these residues (A to E; cf. Table 3) and calculated their relative energies by QM/MM calculations using the QM system in Fig. 1. The protonation states A and B have the same number of atoms and electrons, so their QM/MM energies can be compared. The same applies to the protonation states C, D, and E. The details of the

protonation states and the QM/MM energies are given in Table 3.

The results summarized in Table 3 show that the protonation state A (Asp-124 is charged and His-122 is protonated on NE2) is much more stable than the protonation state B (Asp-124 is charged and His-122 is protonated on ND1) by 32 kcal/mol. Similarly, the protonation state C (Asp-124 is neutral and His-122 is protonated on both NE2 and ND1) is much more stable than the protonation states D and E (Asp-124 is neutral and His-122 is protonated on either NE2 or ND1) by 50 and 57 kcal/mol, respectively. Therefore, we can rule out the protonation states B, D, and E for Asp-124 and His-122, and only consider the protonation states A and C. However, we cannot compare the energies of A and C directly, because they have different numbers of atoms and electrons.

We compared the reaction energy profiles of aphid myrosinase using the protonation states A and C, which are shown in Fig. S5 (the details of the reaction profile are discussed in section 3.2). Fig. S5 shows that the overall energy barrier with A and C are 15.2 and 22.3 kcal/mol, respectively (the profiles are based on the more accurate ΔG_{tot} energy in Eq2). The experimental catalytic turnover number (k_{cat}) of native and recombinant aphid myrosinase for sinigrin as the substrate are 36.0 [6], and 21.0 s^{-1} [8], respectively. These correspond to ~15 kcal/mol at 300 K, assuming a pre-exponential factor of 6.2×10^{12} s^{-1} in the Arrhenius equation. Therefore, the calculated energy barrier with A is much closer to the experimental values [6], 8]. Thus, we chose the protonation state A, where Asp-124 is deprotonated and negatively charged, and His-122 is protonated on NE2, for the QM system in the following calculations. This choice is also consistent with the MD results that suggested His-122 is protonated on NE2 (see above). These results confirm the validity of the MD simulations in assigning protonation states of enzyme residues.

Table 3. The five different combinations of the protonation states of His-122 and Asp-124 and their relative QM/MM energies in kcal/mol. The QM/MM energies of A and B and those of C, D, and E combinations are comparable because they have the same numbers of atoms and electrons.

Combination	Asp-124	His-122	$E_{QM/MM}$
A	ASP	HIE	0.0
B	ASP	HID	32.5
C	ASP	HIP	0.0
D	ASH	HIE	50.0
E	ASH	HID	57.1

3.2 The Catalytic Reaction Mechanism

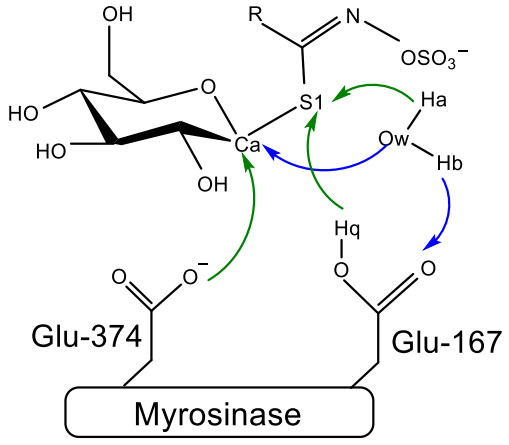
In this section, we study the glycosylation and deglycosylation reactions shown in Scheme 2 for wild-type aphid myrosinase. After declaring the reaction path and proposing a mechanism, we will study and calculate the reaction energy profiles of some mutants of this enzyme (cf. section 3.3). Reported relative energies in the following are QM/MM energies except for solid profiles in Fig. 9 which are ΔG_{tot} obtained according to Eq 2.

$$\Delta G_{tot} = E_{big\text{-}QM} + \Delta G_{QTCP} - E_{QM/MM} \qquad (2)$$

where, each energy term is relative to the starting state of the reaction, and the raw QM/MM, big-QM [63,64], and QM/MM thermodynamic cycle perturbation (QTCP) [65–67] energies are reported in Table S2 (details of the big-QM and QTCP calculations are given in the Supporting Information). The dashed profiles in Fig. 9 are obtained using QM-cluster calculations for the second reaction step shown in Scheme 2 (cf. the first paragraph of section 3.2.2).

3.2.1 The Glycosylation Step

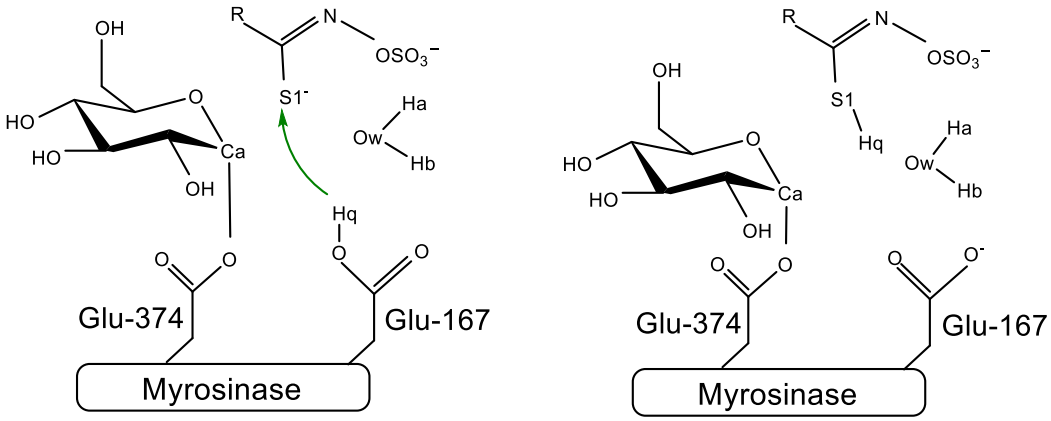
Fig. 1 shows the optimized structure of the reactant state of the reaction in the wild-type enzyme (**Re-WT**), and Scheme 3 shows it in a simplified form. In **Re-WT**, there are five possible reactions that can happen: Glu-374 or the water molecule can attack Ca, Hq (the hydrogen atom on OE2 of Glu-167) can move to S1, and Ha and Hb (the hydrogen atoms on the water molecule) can move to S1 and OE1 of Glu-167, respectively. Scheme 3 shows these reactions and the labels of the atoms involved.



Scheme 3. Schematic view of the reactant state of the wild-type enzyme, together with tested reactions. Reactions found to be possible, or with prohibitively high barriers are shown with arrows in green and blue, respectively. The optimized structure of **Re-WT** is shown in Fig. 1.

Fig. 4 shows the energy profiles of these reactions. The transfer of Hb to Glu-167 and the direct attack of the water molecule to Ca have high barriers (21.6 and 68.9 kcal/mol, respectively). However, the attack of Glu-374 to Ca has a low barrier (1.6 kcal/mol) and leads to **IM1'-WT+agl** (see Scheme 4a for its schematic structure), which is more stable than **Re-WT** by 8.5 kcal/mol. From **IM1'-WT+agl**, the only possible reaction is the transfer of Hq to S1 (the Hq-S1 distance is 1.89Å). This reaction has a low barrier (2.3 kcal/mol) and produces **IM1-WT+agl** (the product of the glycosylation step), where S1 is protonated by Hq, the Ca–S1 bond is broken (the Ca–S1 distance changes from 1.99 to 3.61 Å), and Glu-374 forms a bond with the glucose ring (see Scheme 4b). Alternatively, the transfers of Hq or Ha from **Re-WT** to S1 are also favorable and directly produce **IM1-WT+agl**. The transfer of Ha to S1 has a barrier of 13.3 kcal/mol, while the

transfer of Hq to S1 has a lower barrier of 3.9 kcal/mol. This means that the substrate is more basic than the water molecule and Glu-167, and that the water molecule is more basic than Glu-167. Therefore, there are three possible paths from **Re-WT** to **IM1-WT+agl**: path 1 is the attack of Glu-374 to Ca followed by the transfer of Hq to S1, path 2 is the transfer of Hq to S1, and path 3 is the transfer of Ha to S1. Path 1 has the lowest barrier energy (compare the green profile with the others in Fig. 4) and we only used this path when studying the reaction of some mutants of this enzyme (see section 3.3).



(a) **IM1'-WT+agl** (b) **IM1-WT+agl**

Scheme 4. Schematic representations of the (a) **IM1'-WT+agl** and (b) **IM1-WT+agl** intermediates. The atom labels and the bond distances are shown.

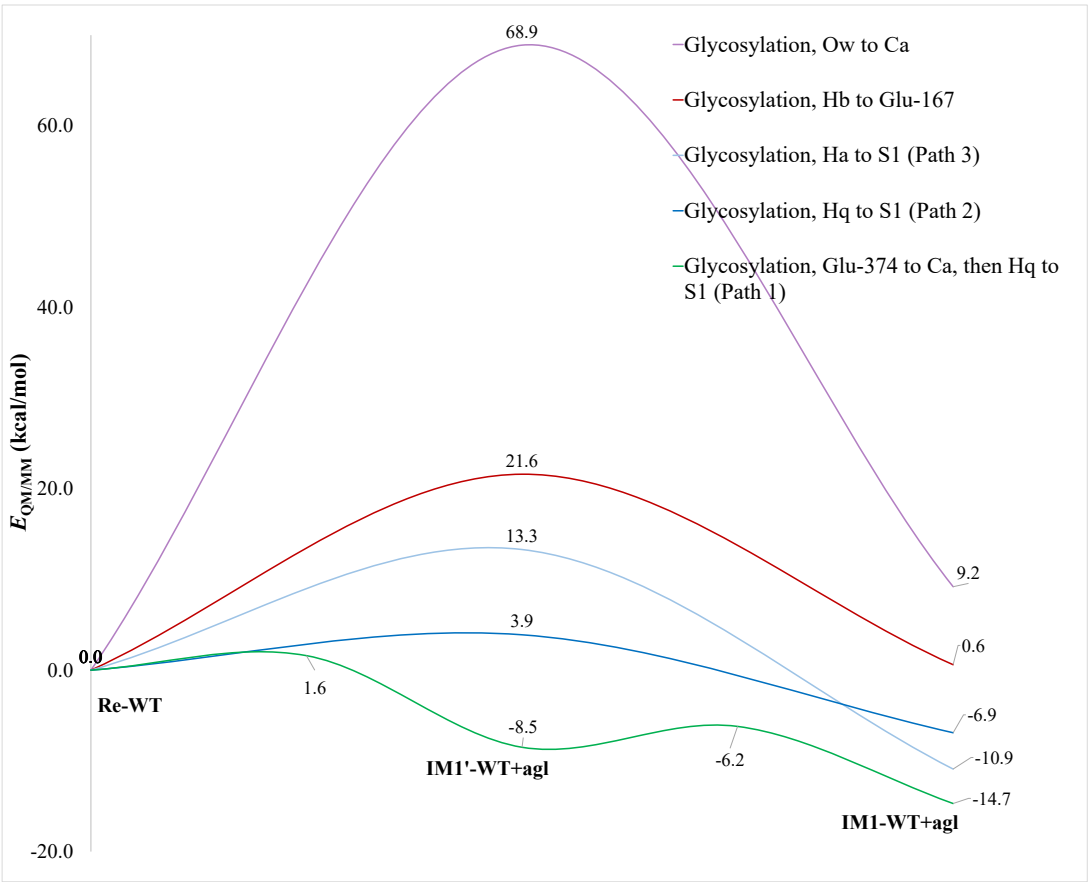


Fig. 4. QM/MM energy profile for all reasonable reactions of glycosylation reaction with **Re-WT**. The figure shows the relative energies of the reactant state (**Re-WT**), the first (**IM1'-WT+agl**), and the second (**IM1-**

WT+agl) intermediates for the different paths from **Re-WT** to **IM1-WT+agl**.

After the glycosylation step, the side chain of sinigrin leaves the active site. We optimized the structures of the first intermediate of the reaction in the wild-type enzyme with and without the side chain in the active site (**IM1-WT+agl** and **IM1-WT−agl**, respectively). Fig. 5 shows these structures. In the deglycosylation step of the reaction, an -OH group should join the Ca atom of the glucose ring, releasing Glu-374. We suggested that this -OH group comes from the water molecule, which is positioned by Glu-167 (see the next section).

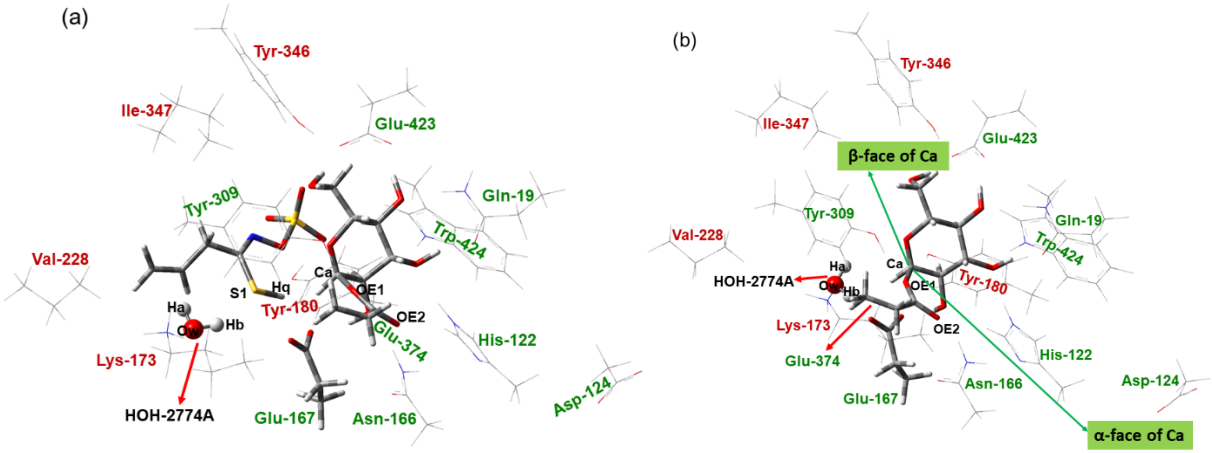


Fig. 5. Optimized structures of the intermediates formed after the glycosylation reaction in the wild-type enzyme, (a) when the aglycon moiety is still in the active site (**IM1-WT+agl**), (b) when the aglycon moiety has left the active site (**IM1-WT−agl**). The α-face of the anomeric carbon is occupied by Glu-374 and the water molecule can only attack the β-face of this carbon.

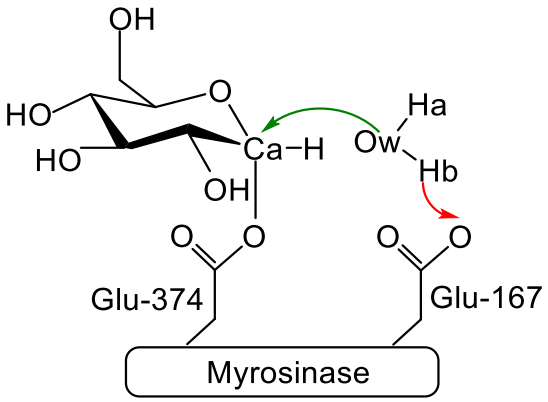
3.2.2 The deglycosylation step

In this section, we study the last reaction step, shown by the third double arrow in Scheme 2, for the wild-type enzyme. Nevertheless, we first need to connect the glycosylation and deglycosylation reaction profiles. We need to have an estimate of the reaction energy shown by the second double arrow in Scheme 2 (the dissociation of aglycon; **IM1+agl** → **IM1−agl** + aglycon). To calculate the energy of this reaction, we used QM-cluster calculations as we performed in recent work [10]. We extracted the structures of the enzyme states (**IM1+agl** and **IM1−agl**) from the QM/MM calculations and then optimized them without the point-charge model of the MM system, fixing the carbon atoms connected to the H-junction atoms in the crystallographic positions. However, the structure of the aglycon moiety was optimized without any fixed atoms. Then, we performed frequency calculations for all states and estimated zero-point and thermal corrections to the Gibbs free energies using a normal-mode harmonic-oscillator ideal-gas approximation. To these, we added solvation energies, which were calculated with a continuum COSMO solvation model [68,69], and single-point electronic energies in vacuum. The dielectric constant for the enzyme states and the ligand were 4 and 80, respectively. A similar method was used by Brás et al. to calculate the dissociation energy of a ligand in β-galactosidase [70]. All calculations were

performed at the TPSS-D3/def2-SV(P) level of theory. The calculations indicate that the second reaction step is endothermic by 13.2 kcal/mol. The profiles of the second reaction step are represented by dashed lines in the middle part of Fig. 6.

The starting structure for the deglycosylation step is **IM1-WT–agl**. As is shown in Scheme 5, there are two possible reactions from the **IM1-WT–agl** intermediate (Ow to Ca, and Hb to Glu-167). However, only the Ow to Ca reaction is possible, with a barrier and reaction energy of 18.8 and –2.5 kcal/mol, respectively (cf. the green profile on the right-hand side of Fig. 6). The reaction directly produces the final product of the myrosinase reaction. The optimized structure of the product state (**Pr-WT**; cf. Fig. 7) shows the extra proton of the water molecule is abstracted by Glu-167. In fact, Glu-167 which acted as the general acid in the glycosylation step, in this step activates the water molecule to attack the anomeric center (thus, this residue acts as a general acid in the glycosylation and a general base in the deglycosylation steps). As we mentioned in the introduction, this step is only possible with an ascorbate anion assistance in plant myrosinase [9,71], but the aphid myrosinase does not need an external anion to do [5,6]. Given the good agreement between our results and the experimental outcomes reported in references [5] and [6], which emphasize that aphid myrosinase is active without requiring assistance from an ascorbate anion, this suggests the suitability of our theoretical model in comprehensively elucidating the reaction mechanism of aphid myrosinase.

Also, as can be seen in the **Pr-WT** structure in Fig. 7, the enzyme retains the stereochemistry of the anomeric center, in line with experiments [72]. To invert the conformation of the anomeric center, the water molecule must attack the α-face of the anomeric carbon. But this face is occupied by Glu-374 in the **IM1-WT–agl** intermediate structure making an α-glycosidic bond (the OE1 atom of this residue is covalently bonded to Ca; cf. Scheme 5). Besides, the water molecule is positioned somewhere above Ca by Glu-167, which can attack the β-face only, and this retains the conformation of the anomeric center (cf. Fig. 5b).



Scheme 5. Schematic view of the **IM1-WT–agl** intermediate. Reactions found to be possible and impossible (it returns to the starting point after releasing any bond constraints) are shown with arrows in green and red, respectively.

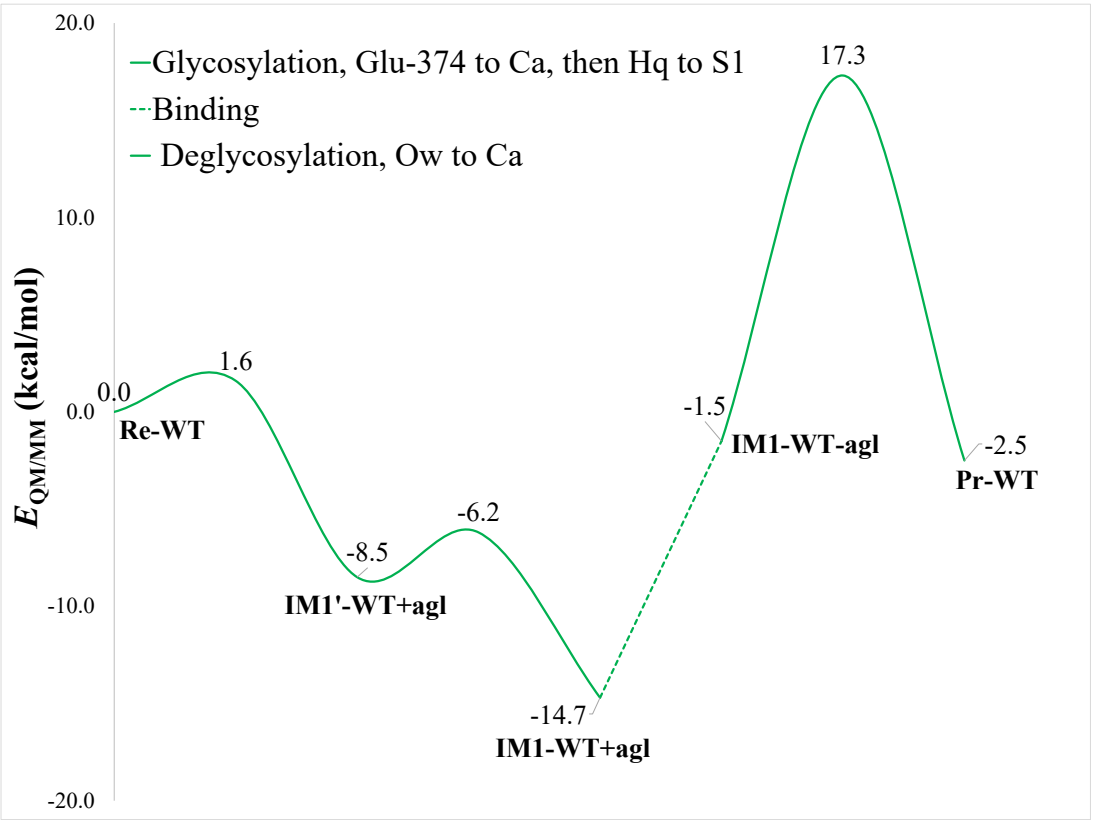


Fig. 6. QM/MM energy profile of the glycosylation (left-hand side) and deglycosylation (right-hand side) reactions of wild-type enzyme. The dashed lines represent the reaction energy for the second reaction step shown by the second double arrow in Scheme 2 (leaving aglycon moiety).

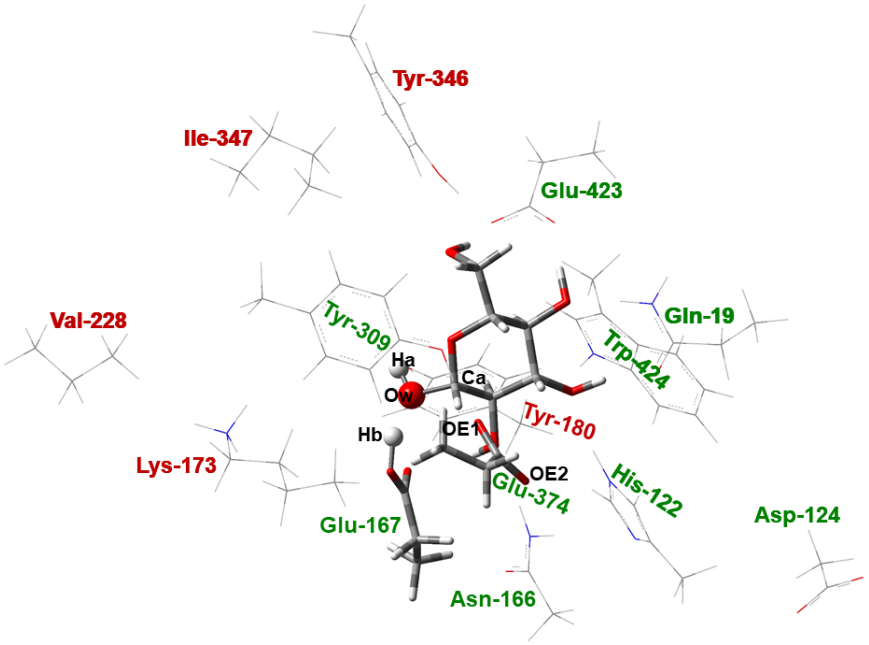


Fig. 7. Optimized structure of the **Pr-WT**.

3.3 Effect of Arg-312

Prior studies have shown that Arg-312 in aphid myrosinase possibly plays a key role in hydrolyzing glucosinolates and is critical for sulfur recognition [7] (cf. Fig. 3 for the position of Arg-312 with respect to the active site). To study the effect of this residue, we have set up two

mutants of the enzyme (R312V and R312M). We also tested the effect of neutralizing Arg-312 to find out the effect of the positive charge in the reaction (removing a proton from NH2 atom).

Our results show that the effect of neutralizing Arg-312 in the QM/MM system depends on the reaction step. For the glycosylation step, neutralizing Arg-312 increases the reaction barrier, as the blue profile is always higher than the green one in Fig. 8a. For the deglycosylation step, neutralizing Arg-312 decreases the reaction barrier, as the blue profile is always lower than the green one in Fig. 8b. This can be justified by the fact that Arg-312 is located in the aglycon binding site, far from the reaction center. A positively charged Arg residue can stabilize the negatively charged aglycon moiety that has left the glucose moiety of the substrate after the glycosylation step, but a neutralized Arg cannot. On the other hand, a neutralized Arg-312 leaves more negative charge on the oxygen atom of the water molecule than a positively charged one, making the water molecule a more powerful nucleophile and thus lowering the deglycosylation energy barrier. This is also in line with the results of the R312M and R312V mutations. These two mutations, which replace the positively charged Arg-312 with the hydrophobic and neutral methionine and valine amino acids, raise the reaction barrier of the glycosylation step but lower the barrier of the deglycosylation step. The valine side chain, which is smaller than those of methionine and neutralized arginine, has almost the same effect on the reactions, implying that the steric effect of Arg-312 is not a determining factor in the reaction. Overall, neutralizing Arg-312 (either by removing a proton from its side chain or by mutating it to valine or methionine) increases the barriers of the glycosylation reaction but lowers the barriers of the deglycosylation reaction.

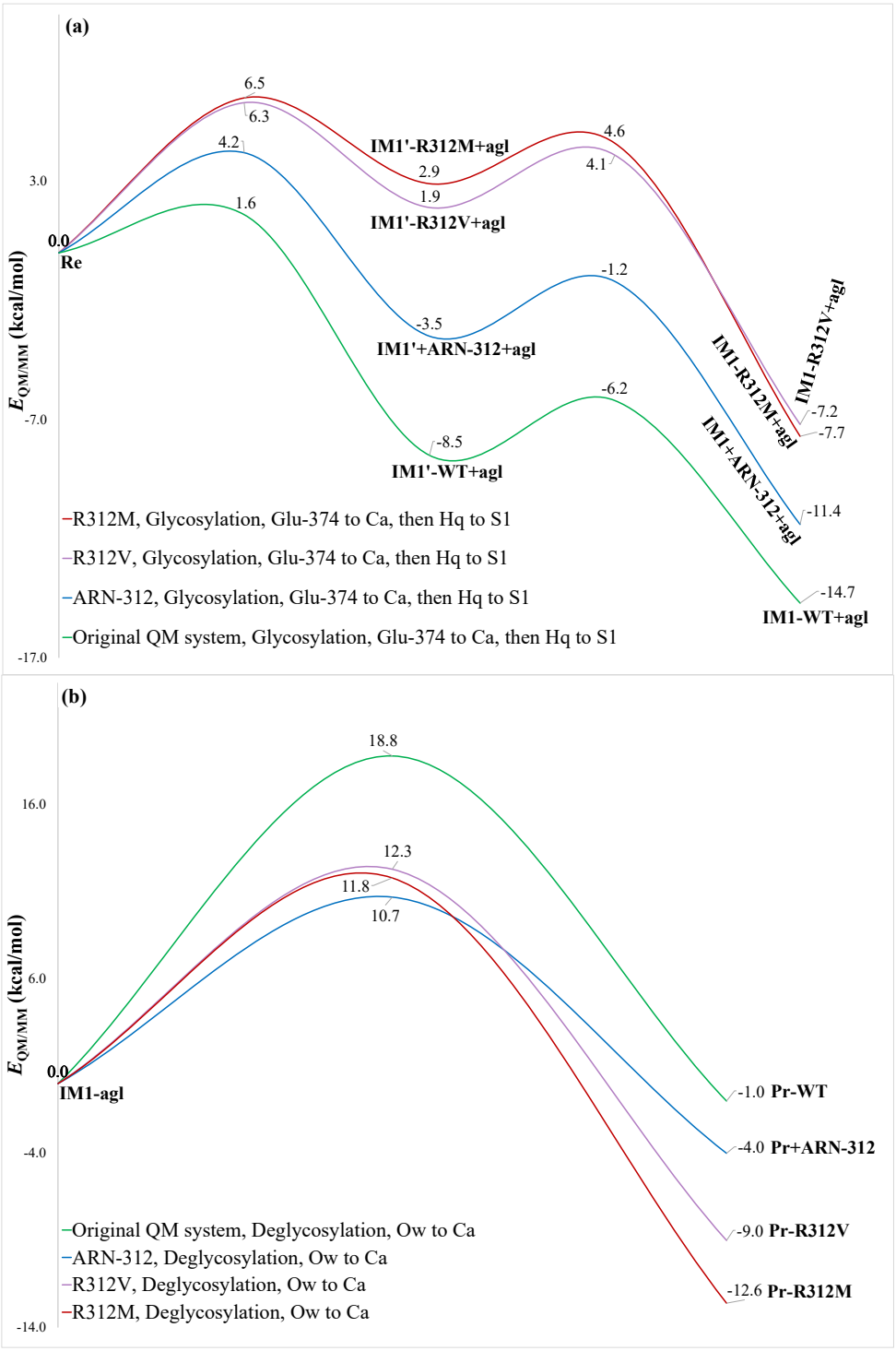


Fig. 8. Relative energy profiles of the (a) glycosylation and (b) deglycosylation steps of the reaction in the wild-type enzyme, the neutralized Arg-312 in the QM/MM system, and the R312M and R312V mutants.

4 Conclusions

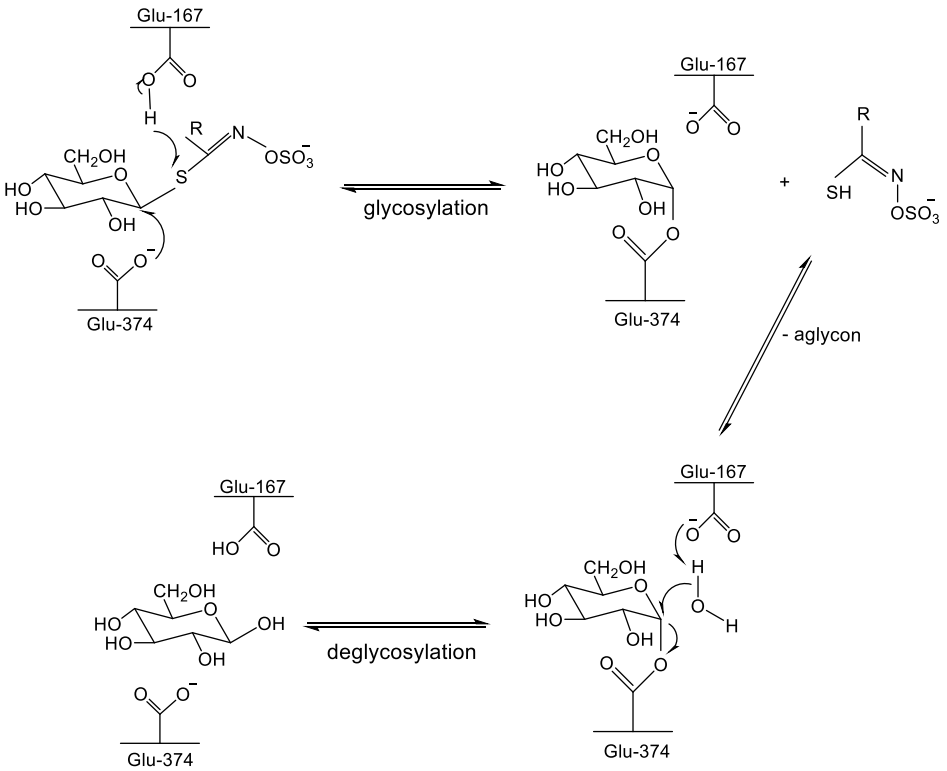
We have employed QM/MM calculations and MD simulations to study the catalytic reaction of Aphid myrosinase. An appropriate QM region to study the myrosinase reaction should contain the whole substrate, models of Gln-19, His-122, Asp-124, Asn-166, Glu-167, Lys-173, Tyr-180, Val-228, Tyr-309, Tyr-346, Ile-347, Glu-374, Glu-423, Trp-424, and a water molecule.

In our recent study, we found that assigning proper protonation states is more important than

the size of the model QM system in QM/MM calculations. As a result, we conducted MD simulations with various protonation states for active site residues with more than one possible protonation state (His-122, Asp-124, Glu-167, and Glu-423). However, the MD results did not definitively determine the proper protonation state for His-122. To resolve this uncertainty, we turned to more accurate QM/MM calculations. His-122 is located in close proximity to Asp-124 in the crystal structure of aphid myrosinase. We examined different combinations of protonation states for these residues (designated as A, B, C, D, and E; cf. Table 3) and computed their relative energies using QM/MM calculations with the QM system depicted in Fig. 1. Our results indicated that the protonation states represented by combinations B, D, and E should be discarded, leaving only combinations A and C under consideration. To make the final selection between the A and C combinations, we generated and compared the reaction energy profiles using these two sets of protonation states (cf. Fig. S5 in the Supporting Information). The calculated overall energy barrier with the A combination closely aligns with the experimental data [6]. Consequently, we adopted the QM system in which Asp-124 is charged, His-122 is protonated on NE2, Glu-167 is protonated on OE2, and Glu-423 is charged.

We used the QM/MM calculation with the QM system described above to propose the reaction mechanism for aphid myrosinase (see Scheme 6). To improve the accuracy of the calculated energies for the reaction path and to compare them with experimental data of other groups (reported in references ([5], [6] and [8]), we also performed big-QM and QTCP calculations for all stationary points along the reaction path and obtained total ΔG_{tot} energies according to Eq. 2. The energy profile is shown in Fig. 9. This is the first mechanism for this enzyme that is confirmed by computational methods. In the first step of our proposed mechanism (glycosylation), Glu-374 takes on the role of a nucleophile, making a nucleophilic attack on the anomeric carbon of the substrate, then the other carboxylic residue (Glu-167), acting as a general acid, protonates the glycosidic sulfur. This leads to the forming of a glycosyl–enzyme intermediate and the aglycon moiety leaves the reaction medium. In the subsequent step (deglycosylation), Glu-167, acting as a general base, activates a water molecule to attack the anomeric center of the intermediate, and the covalent bond between the enzyme and the glycosyl is broken, and the β-glucose molecule is released. Our proposed mechanism, supported by computational validation, helps us understand the enzyme in greater detail. The calculated overall energy barrier is 15.2 kcal/mol, using the more accurate ΔG_{tot} energy in Eq. 2 (cf. Fig. 9). This overall barrier is in agreement with the experimental turnover number (k_{cat}) of the native and recombinant aphid myrosinase which are of 36.0 [6] and 21.0 [8] s^{-1} for sinigrin as the substrate, respectively (these correspond to ~15 kcal/mol at 300 K, when the pre-exponential factor is equal to 6.2×10^{12} s^{-1} in the Arrhenius equation). Further, our model reproduced the anomeric retaining characteristic of myrosinase and indicated that the

deglycosylation reaction is the rate-determining step of the reaction in contrary to the reaction mechanism of *Sinapis alba* myrosinase that the rate-determining step of the reaction is the glycosylation reaction [10].



Scheme 6. Our proposed mechanism for the catalytic reaction of aphid myrosinase, based on the QM/MM calculations.

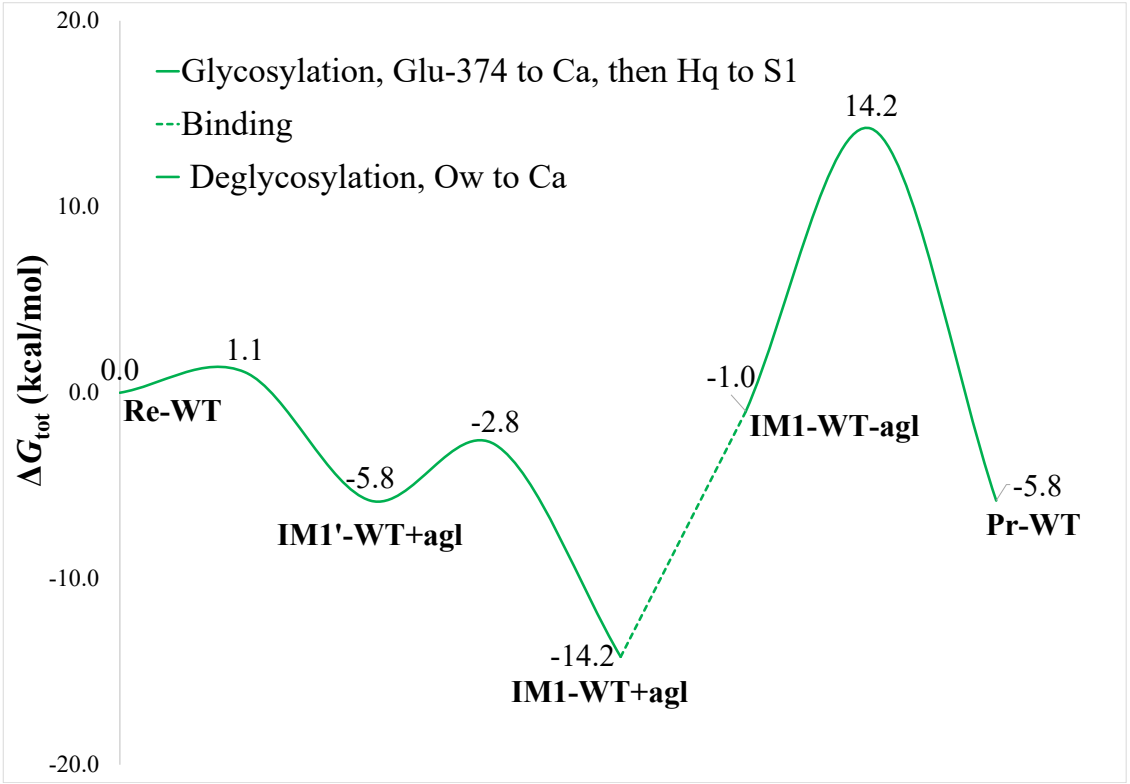


Fig. 9. The ΔG_{tot} energy profile for the glycosylation (left-hand side) and deglycosylation (right-hand side) reactions of wild-type aphid myrosinase. The dashed lines represent the reaction energy for the second reaction step shown by the second double arrow in Scheme 2 (leaving aglycon moiety).

We have also investigated the effect of mutating Arg-312 on the reaction barriers of aphid myrosinase. The results indicated that neutralizing Arg-312 by either removing a proton from its side chain or mutating it to valine or methionine increases the barriers of the glycosylation reaction but lowers the barriers of the deglycosylation reaction. This indicates that the role of this residue is to stabilize the aglycon moiety of the substrate in the glycosylation step. Overall, the mutants enhance the activity of the enzyme.

To date, no experimental data exists for the mutants of aphid myrosinase, limiting direct comparison with our simulation results. To validate our computational predictions and gain deeper insights into the functional significance of Arg-312 in aphid myrosinase catalysis, we suggest that experimental investigations focused on these mutants would be highly valuable. Specifically, empirical studies involving site-directed mutagenesis to create the proposed mutants could complement our computational findings. By conducting detailed biochemical assays on these mutants, would be possible to validate our predictions and unveil the precise role of Arg-312 in the catalytic mechanism of aphid myrosinase. Through this synergy between computational simulations and empirical experiments, we aim to reinforce our theoretical findings while contributing to a comprehensive understanding of the atomic-level details governing aphid myrosinase's reaction mechanism.

Supporting Information Available

Differences between retaining and inverting glycosidases, Equivalent residues of the aphid and plant myrosinases, The catalytic reaction mechanism of plant myrosinase, force field parameters for CGT and GLO and PDB structures showing their atom names, a schematic representation of our MD method, big-QM and QTCP calculations, the ΔG_{tot} energy profiles of the glycosylation and deglycosylation reactions of wild-type aphid myrosinase in A and C combinations, energy components of ΔG_{tot}, cartesian coordinates of the QM system in the different stationary structures.

Author contributions

S. J. Conceptualization; Computations; Data curation, Formal analysis; Validation; Visualization; Review & Editing

U. R. Conceptualization; Funding acquisition; Investigation; Project Administration; Supervision; Resources; Software; Methodology; Review & Editing

M. I. Conceptualization; Funding acquisition; Investigation; Project Administration; Supervision; Writing; Review & Editing.

Conflicts of interest

The authors declare no competing interests.

Acknowledgments

This investigation has been supported by grants from the Swedish Research Council (project 2018-5003) and the Vice President for Research and Technology, University of Kurdistan (project 01/9/3778). The computations were performed on computer resources provided by the Swedish National Infrastructure for Computing (SNIC) at Lunarc at Lund University and HPC2N at Umeå University, partially funded by the Swedish Research Council (grant 2018-05973).

5 References

[1] N. Agerbirk, C.E. Olsen, Glucosinolate structures in evolution, Phytochemistry. 77 (2012) 16–45.

[2] A.M. Bones, J.T. Rossiter, The myrosinase-glucosinolate system, its organisation and biochemistry, Physiol. Plant. 97 (1996) 194–208.

[3] B.A. Halkier, L. Du, The biosynthesis of glucosinolates, Trends Plant Sci. 2 (1997) 425–431.

[4] F. Francis, G. Lognay, J. Wathelet, E. Haubruge, Characterisation of aphid myrosinase and degradation studies of glucosinolates, Arch. Insect Biochem. Physiol. Publ. Collab. with Entomol. Soc. Am. 50 (2002) 173–182.

[5] A.M.E. Jones, M. Bridges, A.M. Bones, R. Cole, J.T. Rossiter, Purification and characterisation of a non-plant myrosinase from the cabbage aphid Brevicoryne brassicae (L.), Insect Biochem. Mol. Biol. 31 (2001) 1–5.

[6] B. Pontoppidan, B. Ekbom, S. Eriksson, J. Meijer, Purification and characterization of myrosinase from the cabbage aphid (*Brevicoryne brassicae*), a brassica herbivore, Eur. J. Biochem. 268 (2001) 1041–1048. https://doi.org/10.1046/j.1432-1327.2001.01971.x.

[7] A.M.E. Jones, P. Winge, A.M. Bones, R. Cole, J.T. Rossiter, Characterization and evolution of a myrosinase from the cabbage aphid Brevicoryne brassicae, Insect Biochem. Mol. Biol. 32 (2002) 275–284. https://doi.org/10.1016/S0965-1748(01)00088-1.

[8] H. Husebye, S. Arzt, W.P. Burmeister, F. V. Härtel, A. Brandt, J.T. Rossiter, A.M. Bones, Crystal structure at 1.1Å resolution of an insect myrosinase from Brevicoryne brassicae shows its close relationship to β-glucosidases, Insect Biochem. Mol. Biol. 35 (2005) 1311–1320. https://doi.org/10.1016/j.ibmb.2005.07.004.

[9] W.P. Burmeister, S. Cottaz, P. Rollin, A. Vasella, B. Henrissat, High resolution X-ray crystallography shows that ascorbate is a cofactor for myrosinase and substitutes for the function of the catalytic base., J. Biol. Chem. 275 (2000) 39385–93. https://doi.org/10.1074/jbc.M006796200.

[10] S. Jafari, U. Ryde, M. Irani, QM/MM Study of the Catalytic Reaction of Myrosinase; Importance of Assigning Proper Protonation States of Active-Site Residues, J. Chem. Theory Comput. 17 (2021) 1822–1841. https://doi.org/10.1021/acs.jctc.0c01121.

[11] J.T. Rossiter, A.M. Jones, A.M. Bones, Chapter six A novel myrosinase-glucosinolate defense system in, cruciferous specialist aphids, Recent Adv. Phytochem. 37 (2003) 127–142. https://doi.org/10.1016/S0079-9920(03)80021-7.

[12] D.C. Bas, D.M. Rogers, J.H. Jensen, Very fast prediction and rationalization of pKa values for protein-ligand complexes, Proteins Struct. Funct. Genet. 73 (2008) 765–783. https://doi.org/10.1002/prot.22102.

[13] M.H.M. Olsson, C.R. Søndergaard, M. Rostkowski, J.H. Jensen, PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK_a Predictions, J. Chem. Theory Comput. 7 (2011) 525–537. https://doi.org/10.1021/ct100578z.

[14] J.H. Jensen, H. Li, A.D. Robertson, P.A. Molina, Prediction and Rationalization of Protein pKa Values Using QM and QM/MM Methods, J. Phys. Chem. A. 109 (2005) 6634–6643. https://doi.org/10.1021/jp051922x.

[15] S. Jafari, N. Kazemi, U. Ryde, M. Irani, Higher Flexibility of Glu-172 Explains the Unusual

Stereospecificity of Glyoxalase I, Inorg. Chem. 57 (2018) 4944–4958. https://doi.org/10.1021/acs.inorgchem.7b03215.
[16] S. Jafari, U. Ryde, M. Irani, QM/MM study of the stereospecific proton exchange of glutathiohydroxyacetone by glyoxalase I, Results Chem. 1 (2019) 100011. https://doi.org/10.1016/j.rechem.2019.100011.
[17] J. Uranga, P. Mikulskis, S. Genheden, U. Ryde, Can the protonation state of histidine residues be determined from molecular dynamics simulations?, Comput. Theor. Chem. 1000 (2012) 75–84. https://doi.org/10.1016/j.comptc.2012.09.025.
[18] L. Cao, O. Caldararu, U. Ryde, Protonation States of Homocitrate and Nearby Residues in Nitrogenase Studied by Computational Methods and Quantum Refinement, J. Phys. Chem. B. 121 (2017) 8242–8262. https://doi.org/10.1021/acs.jpcb.7b02714.
[19] S. Le Grand, A.W. Götz, R.C. Walker, SPFP: Speed without compromise - A mixed precision model for GPU accelerated molecular dynamics simulations, Comput. Phys. Commun. 184 (2013) 374–380. https://doi.org/10.1016/j.cpc.2012.09.022.
[20] R. Salomon-Ferrer, A.W. Götz, D. Poole, S. Le Grand, R.C. Walker, Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh ewald, J. Chem. Theory Comput. 9 (2013) 3878–3888. https://doi.org/10.1021/ct400314y.
[21] A.W. Götz, M.J. Williamson, D. Xu, D. Poole, S. Le Grand, R.C. Walker, Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. generalized born, J. Chem. Theory Comput. 8 (2012) 1542–1555. https://doi.org/10.1021/ct200909j.
[22] D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R. Harris, N. Homeyer, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon-Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York, P.A. Kollman, Amber 18, University of California, San Francisco, (2018).
[23] J.A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, C. Simmerling, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB, J. Chem. Theory Comput. 11 (2015) 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.
[24] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general Amber force field, J. Comput. Chem. 25 (2004) 1157–1174. https://doi.org/10.1002/jcc.20035.
[25] C. Lee, W. Yang, R. Parr, G, Development of the Colic-Salvetti correlation-energy into a functional of the electron density, Am. Phys. Soc. 37 (1988) 785–789.
[26] A.D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior, Phys. Rev. A. 38 (1988) 3098–3100. https://doi.org/10.1103/PhysRevA.38.3098.
[27] A.D. Becke, Density-functional thermochemistry. III. The role of exact exchange, J. Chem. Phys. 98 (1993) 5648–5652. https://doi.org/10.1063/1.464913.
[28] G.A. Petersson, M.A. Al-Laham, A complete basis set model chemistry. II. Open-shell systems and the total energies of the first-row atoms, J. Chem. Phys. 94 (1991) 6081–6090. https://doi.org/10.1063/1.460447.
[29] G.A. Petersson, A. Bennett, T.G. Tensfeldt, M.A. Al-Laham, W.A. Shirley, J. Mantzaris, A complete basis set model chemistry. I. The total energies of closed-shell atoms and hydrides of the first-row elements, J. Chem. Phys. 89 (1988) 2193–2218. https://doi.org/10.1063/1.455064.
[30] B.H. Besler, K.M. Merz, P.A. Kollman, Atomic charges derived from semiempirical methods, J. Comput. Chem. 11 (1990) 431–439. https://doi.org/10.1002/jcc.540110404.
[31] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, G.A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V Marenich, J.

Bloino, B.G. Janesko, R. Gomperts, B. Mennucci, H.P. Hratchian, J. V Ortiz, A.F. Izmaylov, J.L. Sonnenberg, Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V.G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M.J. Bearpark, J.J. Heyd, E.N. Brothers, K.N. Kudin, V.N. Staroverov, T.A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A.P. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, J.M. Millam, M. Klene, C. Adamo, R. Cammi, J.W. Ochterski, R.L. Martin, K. Morokuma, O. Farkas, J.B. Foresman, D.J. Fox, Gaussian 16, (2016).

[32] C.I. Bayly, P. Cieplak, W. Cornell, P.A. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model, J. Phys. Chem. 97 (1993) 10269–10280. https://doi.org/10.1021/j100142a004.

[33] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935. https://doi.org/10.1063/1.445869.

[34] D.R. Roe, T.E. Cheatham, PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data, J. Chem. Theory Comput. 9 (2013) 3084–3095. https://doi.org/10.1021/ct400341p.

[35] B. Henrissat, I. Callebaut, S. Fabrega, P. Lehn, J.P. Mornon, G. Davies, Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases., Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 7090–4. https://doi.org/10.1073/PNAS.92.15.7090.

[36] J. Jenkins, L. Lo Leggio, G. Harris, R. Pickersgill, β-Glucosidase, β-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes wit 8-fold β/α architecture and with two conserved glutamates near the carboxy-terminal ends of β-strands four and seven, FEBS Lett. 362 (1995) 281–285. https://doi.org/10.1016/0014-5793(95)00252-5.

[37] D.N. Bolam, N. Hughes, R. Virden, J.H. Lakey, G.P. Hazlewood, B. Henrissat, K.L. Braithwaite, H.J. Gilbert, Mannanase A from Pseudomonas fluorescens ssp. cellulosa Is a Retaining Glycosyl Hydrolase in Which E212 and E320 Are the Putative Catalytic Residues, Biochemistry. 35 (1996) 16195–16204. https://doi.org/10.1021/bi961866d.

[38] Q. Wang, S.G. Withers, Substrate-assisted catalysis in glycosidases, J. Am. Chem. Soc. 117 (1995) 10137–10138. https://doi.org/10.1021/ja00145a035.

[39] S. Jafari, U. Ryde, A.E.A. Fouda, F.S. Alavi, G. Dong, M. Irani, Quantum Mechanics/Molecular Mechanics Study of the Reaction Mechanism of Glyoxalase I, Inorg. Chem. 59 (2020) 2594–2603. https://doi.org/10.1021/acs.inorgchem.9b03621.

[40] S. Jafari, U. Ryde, M. Irani, Two-Substrate Glyoxalase I Mechanism: A Quantum Mechanics/Molecular Mechanics Study, Inorg. Chem. 60 (2021) 303–314. https://doi.org/10.1021/acs.inorgchem.0c02957.

[41] J.P. Ryckaert, G. Ciccotti, H.J.. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, J. Comput. Phys. 23 (1977) 327–341. https://doi.org/10.1016/0021-9991(77)90098-5.

[42] X. Wu, B.R. Brooks, Self-guided Langevin dynamics simulation method, Chem. Phys. Lett. 381 (2003) 512–518. https://doi.org/10.1016/j.cplett.2003.10.013.

[43] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684–3690. https://doi.org/10.1063/1.448118.

[44] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089–10092. https://doi.org/10.1063/1.464397.

[45] U. Ryde, The coordination of the catalytic zinc in alcohol dehydrogenase studied by combined quantum-chemical and molecular mechanics calculations., J. Comput. Aided. Mol. Des. 10 (1996) 153–164. https://doi.org/10.1007/BF00402823.

[46] U. Ryde, M.H.M. Olsson, Structure, strain, and reorganization energy of blue copper models

in the protein, Int. J. Quantum Chem. 81 (2001) 335–347. https://doi.org/10.1002/1097-461X(2001)81:5<335::AID-QUA1003>3.0.CO;2-Q.
[47] U. Ryde, QM/MM Calculations on Proteins, Methods Enzymol. 577 (2016) 119–158. https://doi.org/10.1016/BS.MIE.2016.05.014.
[48] H.M. Senn, W. Thiel, QM/MM methods for biomolecular systems, Angew. Chemie - Int. Ed. 48 (2009) 1198–1229. https://doi.org/10.1002/anie.200802019.
[49] A. Schäfer, H. Horn, R. Ahlrichs, Fully Optimized Contracted Gaussian-Basis Sets for Atoms Li to Kr, J. Chem. Phys. 97 (1992) 2571–2577. https://doi.org/10.1063/1.463096.
[50] J. Tao, J.P. Perdew, V.N. Staroverov, G.E. Scuseria, Climbing the Density Functional Ladder: Nonempirical Meta–Generalized Gradient Approximation Designed for Molecules and Solids, Phys. Rev. Lett. 91 (2003) 146401. https://doi.org/10.1103/PhysRevLett.91.146401.
[51] K. Eichkorn, F. Weigend, O. Treutler, R. Ahlrichs, Auxiliary basis sets for main row atoms and transition metals and their use to approximate Coulomb potentials, Theor. Chem. Accounts Theory, Comput. Model. (Theoretica Chim. Acta). 97 (1997) 119–124. https://doi.org/10.1007/s002140050244.
[52] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, R. Ahlrichs, Auxiliary basis sets to approximate Coulomb potentials, Chem. Phys. Lett. 240 (1995) 283–289. https://doi.org/10.1016/0009-2614(95)00621-A.
[53] S. Grimme, S. Ehrlich, L. Goerigk, Effect of the damping function in dispersion corrected density functional theory, J. Comput. Chem. 32 (2011) 1456–1465. https://doi.org/10.1002/jcc.21759.
[54] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu, J. Chem. Phys. 132 (2010) 154104. https://doi.org/10.1063/1.3382344.
[55] M. Von Arnim, R. Ahlrichs, Performance of parallel TURBOMOLE for density functional calculations, J. Comput. Chem. 19 (1998) 1746–1757. https://doi.org/10.1002/(SICI)1096-987X(19981130)19:15<1746::AID-JCC7>3.0.CO;2-N.
[56] F. Furche, R. Ahlrichs, C. Hättig, W. Klopper, M. Sierka, F. Weigend, Turbomole, Wiley Interdiscip. Rev. Comput. Mol. Sci. 4 (2014) 91–100. https://doi.org/10.1002/wcms.1162.
[57] D.A. Case, V. Babin, J.T. Berryman, R.M. Betz, Q. Cai, D.S. Cerutti, T.E. Cheatham, T.A. Darden, R.E. Duke, H. Gohlke, A.W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T.S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K.M. Merz, F. Paesani, D.R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C.L. Simmerling, W. Smith, J. Swails, Walker, J. Wang, R.M. Wolf, X. Wu, P.A. Kollman, Amber 14 - University of California:, (2014). citeulike-article-id:14028955.
[58] N. Reuter, A. Dejaegere, B. Maigret, M. Karplus, Frontier bonds in QM/MM methods: A comparison of different approaches, J. Phys. Chem. A. 104 (2000) 1720–1735. https://doi.org/10.1021/jp9924124.
[59] L. Hu, P. Söderhjelm, U. Ryde, On the convergence of QM/MM energies, J. Chem. Theory Comput. 7 (2011) 761–777. https://doi.org/10.1021/ct100530r.
[60] M. Svensson, S. Humbel, R.D.J. Froese, T. Matsubara, S. Sieber, K. Morokuma, ONIOM: A Multilayered Integrated MO + MM Method for Geometry Optimizations and Single Point Energy Predictions. A Test for Diels−Alder Reactions and Pt(P(t -Bu) 3) 2 + H 2 Oxidative Addition , J. Phys. Chem. 100 (2002) 19357–19363. https://doi.org/10.1021/jp962071j.
[61] L. Cao, U. Ryde, On the Difference Between Additive and Subtractive QM/MM Calculations, Front. Chem. 6 (2018) 89. https://doi.org/10.3389/fchem.2018.00089.
[62] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera—a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612.
[63] S. Sumner, P. Söderhjelm, U. Ryde, Studies of reaction energies in proteins, J. Chem. Theory

Comput. 9 (2013) 4205–4214. https://doi.org/dx.doi.org/10.1021/ct400339c.
[64] L. Hu, P. Söderhjelm, U. Ryde, Accurate Reaction Energies in Proteins Obtained by Combining QM/MM and Large QM Calculations, J. Chem. Theory Comput. 9 (2013) 640–649. https://doi.org/10.1021/ct3005003.
[65] V. Luzhkov, A. Warshel, Microscopic models for quantum mechanical calculations of chemical processes in solutions: LD/AMPAC and SCAAS/AMPAC calculations of solvation energies, J. Comput. Chem. 13 (1992) 199–213. https://doi.org/10.1002/jcc.540130212.
[66] T.H. Rod, U. Ryde, Quantum mechanical free energy barrier for an enzymatic reaction, Phys. Rev. Lett. 94 (2005) 138302. https://doi.org/10.1103/PhysRevLett.94.138302.
[67] T.H. Rod, U. Ryde, Accurate QM/MM free energy calculations of enzyme reactions: Methylation by catechol O-methyltransferase, J. Chem. Theory Comput. 1 (2005) 1240–1251. https://doi.org/10.1021/ct0501102.
[68] A. Schäfer, A. Klamt, D. Sattel, J.C.W. Lohrenz, F. Eckert, COSMO implementation in TURBOMOLE: Extension of an efficient quantum chemical code towards liquid systems, Phys. Chem. Chem. Phys. 2 (2000) 2187–2193. https://doi.org/10.1039/b000184h.
[69] A. Klamt, G. Schüürmann, COSMO: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient, J. Chem. Soc. Perkin Trans. 2. (1993) 799–805. https://doi.org/10.1039/P29930000799.
[70] N.F. Brás, P.A. Fernandes, M.J. Ramos, QM/MM Studies on the β-Galactosidase Catalytic Mechanism: Hydrolysis and Transglycosylation Reactions, J. Chem. Theory Comput. 6 (2010) 421–433. https://doi.org/10.1021/ct900530f.
[71] W.P. Burmeister, S. Cottaz, H. Driguez, R. Iori, S. Palmieri, B. Henrissat, The crystal structures of Sinapis alba myrosinase and a covalent glycosyl–enzyme intermediate provide insights into the substrate recognition and active-site machinery of an S-glycosidase, Structure. 5 (1997) 663–676. https://doi.org/10.1016/S0969-2126(97)00221-9.
[72] S. Cottaz, B. Henrissat, H. Driguez, Mechanism-based inhibition and stereochemistry of glucosinolate hydrolysis by myrosinase, Biochemistry. 35 (1996) 15256–15259. https://doi.org/10.1021/bi9622480.