Catalysis Science & Technology

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/catalysis

Journal Name

ARTICLE

Semiempirical QM/MM calculations reveal a step-wise proton transfer and an unusual thiolate pocket in the mechanism of the unique arylpropionate racemase AMDase G74C

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

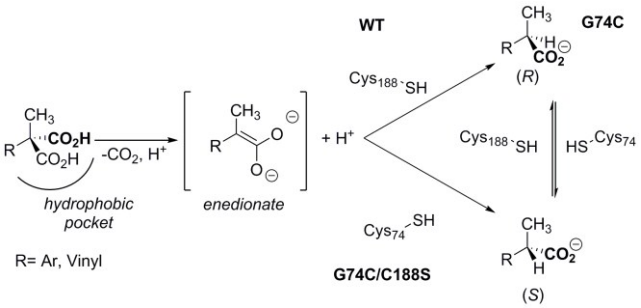
www.rsc.org/

F. Busch,[a] J. Enoki,[a] N. Hülsemann,[a] K. Miyamoto,[b] M. Bocola,[c] and R. Kourist[a]

The mechanism of the unique arylpropionate racemase AMDase G74C was investigated by a QM/MM approach. Molecular dynamics simulations showed that the mechanism is initiated by a deprotonation of the catalytic cysteine. The simulations revealed two thiolate pockets. While the first plays a role in the natural decarboxylative activity of AMDase, the second stabilizes the artificially introduced thiolate group of C74. The presence of the two structural motifs is a prerequisite for the promiscuous racemization reaction of AMDase G74C. QM/MM simulations show that the deprotonation and reprotonation proceed in a stepwise fashion, in which a planar enedionate intermediate is stabilized by a delocalized π-electron system on a vinylic or aromatic substituent of the substrate. The artificial racemase is thus a typical case of substrate-assisted catalysis.

Introduction

Chemical racemisation finds frequent application in organic chemistry.[1] Racemisation reactions are thermodynamically favoured by an increase in entropy. Yet, the inversion of an enantiomer requires overcoming a high energy barrier for the breaking of stable C-H bonds. This leads to very harsh reaction conditions. For instance, racemisation of carboxylic acids by sequential deprotonation and reprotonation of the α-proton requires elevated temperatures and strongly acidic or basic pH-values.[1] In contrast, enzymatic racemization can be conducted at neutral pH values and room temperature, which offers decisive advantages in terms of sustainability.[2] Efficient catalytic methods for the breaking of stable C-H bonds thus minimize energy, waste and the use of harmful reagents. An important application of racemization is the synthesis of pure enantiomers, which often relies on resolutions of racemic mixtures. The maximal yield limitation of 50% can be overcome by the catalytic recovery of the non-desired enantiomer by chemical or enzymatic racemisation. In addition, selective enzymatic racemisation reactions offer the possibility to racemise the substrate *in situ*, giving rise to dynamic resolutions with a theoretical yield of 100%.[3] Nature usually strives to create optical purity and to prevent racemization, which is consequently a rather infrequent reaction. Most racemases are restricted to specific natural substrates such as α- hydroxy aryl acetic acids[2] and amino acid derivatives.[4] The limited substrate spectrum and low activity of racemases for non-natural compounds represent obstacles for synthetic applications.[3] Protein engineering can alleviate these issues. Recently, a six-fold increase of the activity of N-acetyl amino acid racemase (NAAAR) considerably increased the conversion in the synthesis of optically pure amino acids.[3] Gu et al. were able to increase the activity of mandelate racemase six-fold by rational protein design.[5, 6]



Scheme 1. Reaction mechanism of AMDase

Inspired by its structural relation to amino acid racemases belonging to the the Asp/Glu racemase superfamily, Miyamoto and co-workers converted the bacterial arylmalonate decarboxylase (AMDase) into a racemase.[7] AMDase catalyzed the decarboxylation in a two-step mechanism. In the first step, the pro(R)-carboxylate binds in a hydrophobic pocket. This unfavourable interaction destabilizes the C-C bonds and leads to decarboxylation.[8] The delocalized π-electron system of the substituent stabilizes the resulting planar enedionate. Protona-

[a] Department of Biology and Biotechnology, Ruhr-University Bochum, Bochum, 44801 Bochum, Germany. Robert.Kourist@rub.de
[b] Department of Bioscience and Informatics, Keio University, Hiyoshi, Yokohama, Japan.
[c] Institute of Biotechnology, RWTH Aachen, 52062 Aachen, Germany.
† Footnotes relating to the title and/or authors should appear here.
Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

tion by C188 (in the wildtype) or C74 (in variant G74C/C188S) decides on the absolute configuration of the resulting enantiopure α-arylcarboxylic acids. In a recent study, Lind et al. were able to confirm this hypothetic mechanism by quantum mechanical calculations.[8] They also investigated the possibility of a concerted mechanism but were unable to find an energetically favourable transition state for this, which strongly points towards the existence of the planar intermediate. Introduction of a cysteine in position 74 opposite to C188 resulted in a catalytic machinery similar to that of co-factor free amino acid racemases. While these are restricted to α-amino acids and usually very specific for their natural substrate, AMDase G74C has the unique capability to racemize α-arylcarboxylic acids.[7] This class of compounds is highly interesting as it contains arylpropionates belonging to the class of non-steroidal anti-inflammatory drugs (NSAID), including the widely used active pharmaceutical ingredients naproxen, flurbiprofen, ketoprofen and ibuprofen.[9] Optically pure profens are usually produced by selective crystallization or kinetic resolution.[10] Both approaches require the racemization of the non-desired enantiomer under very harsh reaction conditions. This makes an enzymatic racemization of arylpropionates under mild reaction conditions very desirable.

While the racemizing activity of G74C is unique, the enzyme is rather slow. The turnover number of GluR from Bacillus anthracis is 400 times faster for glutamate[11] than that of AMDase G74C for naproxen (kcat of 0.12 s-1). Very recently, structure-guided saturation mutagenesis achieved a 40-fold activity increase of AMDase G74C,[12] thus reaching the performance of natural racemases. Despite sharing the same catalytic machinery, a closer look reveals striking differences between both GluR and AMDase G74C. GluR is an interesting target for antibiotics, which fuelled a thorough investigation of its mechanism. While GluR is restricted to α-amino acids, [4, 13, 14] AMDase G74C is confined to carboxylic acids with a vinyl or aryl substituent on the α-C-atom. Glavas and Tanner suggested that in glutamate racemase, two neutral thiol residues catalyse the reaction.[15] QM/MM simulations on GluR from B. subtilis by Puig et al. indicated a mechanism via four successive proton transfers in three different steps.[4] GluR uses co-catalytic residues for the activation of the Cys residues. The co-catalytic Asp10 deprotonates C74, which then initiated two proton transfers from the α-carbon to C74 and from C185 to the α-carbon. In the third step, H187 protonates the deprotonated C185. Puig et al. suggested that the key step, the second transfer, occurs in a concerted, yet asynchronic fashion.[4] AMDase G74C does not contain any co-catalytic residues that might serve this purpose. An attempt to introduce an aspartate and a histidineas cocatalytic residues to improve AMDase G74C failed. The question of the source of activation of the Cys residues thus remains open.[16] While the deprotonation and protonation of glutamate in GluR proceed concertedly, AMDase G74C requires a substituent with a π-electron system. This points towards the existence of an enedionate intermediate or partially charged transition state that is stabilized by resonance. Density functional theory calculations on the wildtype AMDase suggested the existence of an enedionate during the decarboxylative mechanism.[8] This makes it likely that a similar intermediate plays a role in the promiscuous racemization reaction. A further interesting observation is that AMDase G74C shows a high selectivity in the racemization of β-alkenyl carboxylic acids as substrates.[16] In base-catalyzed racemization, they undergo a conformational rearrangement to the thermodynamical preferred α-alkenyl isomers. This rearrangement was not observed after G74C-catalyzed conversion. A concerted mechanism similar to that of amino acid racemases would offer an explanation for the observed selectivity.

These open questions prompted us to investigate the mechanism of AMDase G74C, which can shed new light on the mechanism of cofactor-free racemases. In this work, we use molecular dynamics simulations and QM/MM calculations to elucidate the racemization mechanism of AMDase G74C. Focus lies on the clarification whether the reaction is initiated by a thiolate or by a thiol and on the possible involvement of an enedionate intermediate in the mechanism.

Results and discussion

Molecular dynamics simulations were carried out to shed light on the deprotonation of the catalytic Cys residues and to identify potential co-catalytic residues. The model of AMDase G74C was generated from a crystal structure of the (S)-selective variant G74C/C188S (PDB ID 3IXL). This structure contains a molecule of phenylacetic acid in the active site and shows a catalytic binding mode for the substrates. In the racemization of the (R)-enantiomers, C188 donates a proton and C74 abstracts a proton from the α-C-atom (Scheme 1). Both amino acids assume opposite roles in the conversion of the (S)-enantiomers. It is a thrilling question how the relatively weak thiol base of the catalytic cystein may break the C-H bond and abstract a proton. To identify whether the reaction is initiated by a thiol or a thiolate residue, we compared a system with two neutral thiol residues (Figure a) with a system in which one of the residues is in thiolate-form (Figure b). In simulations with two neutral thiol residues, the H-atom of the abstracting Cys intercalated frequently between the thiol atom and the α-H-atom. This prevents an initiation of the reaction. Figure a shows a snapshot, where the H-atom of the donating Cys is oriented away from the substrate, which was the most frequent configuration during the MD simulation (Figure c). The orientation of this H-atom towards the α-C-atom is an important factor for catalysis. As a measure for the catalytic productivity of the model, frame with the distance d1 smaller than 3.5 Å and a simultaneous smaller distance than 5 Å between the opposite thiol sulfur to the substrate (Figure a and b) were defined as catalytic or productive frames, while states with higher distances were considered unproductive.

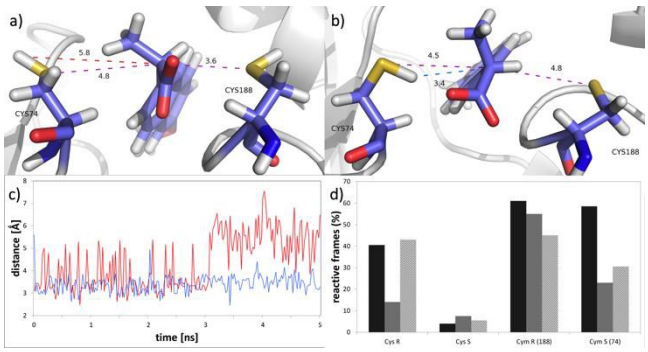


Figure 1. a) Model of the active site of AMDase with G74C with two neutral thiol residues ("Cys"); b) model with a thiol and a thiolate ("Cym"); c) distance d1 of cys (red) and of cym (blue) over the time course of MD simulation of (*R*)-naproxen, showing the higher stability of the catalytic cysteine in a system with a thiolate residue; d) frequency of catalytic frames in molecular dynamcis simulations of (*R*)- and (*S*)- naproxen (black), ketoprofen (dark grey) and flurbiprofen (bright grey) for both models.

In simulations with a deprotonated abstracting thiolate residue it was clearly seen that the H-atom of the donating Cys was oriented mainly towards the substrate (Figure b). Consequently, the frequency of catalytic frames in molecular dynamics simulations of of (S)-naproxen with the thiolate 74 and (R)-naproxen with the thiolate 188, respectively was significantly increased (Figure c, d). Most importantly, the positive effect on the thiolate on the number of catalytic frames was observed for both enantiomers for naproxen. Simulations of flurbiprofen and ketoprofen confirmed these observations (data not shown).

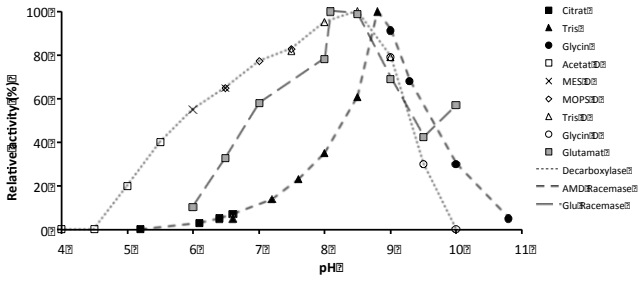


Figure 2. Effect of pH on the activity of AMDase decarboxylase (empty symbols) and racemase mutant G74C (black symbols). The pH-dependences of AMDase and GluR are adapted from Miyamoto *et al*. [17] and May et al.,[11] respectively; (□) AcOH/AcONa buffer, pH 4.0-5.5; (×) Mes/NaOH buffer, pH 6.0-6.5; (◊) Mops/NaOH buffer, pH 6.5-7.5; (⌂) Tris/HCl, pH 7.5-8.9; (○) glycine/NaOH buffer, pH 9.0 - 10.0. (■) AcOH/AcONa buffer, pH 5.2 – 6.6; (▲) Tris/HCl, pH 6.6-8.7; (●) glycine/NaOH buffer, pH 9.0 - 10.8.

While cysteine has a pKa of 8.1 in solution, the server PROPKA 3.0 predicted values of 13.2 and 12.3 for C74 and C188, respectively, for the closed configuration of the AMDase G74C (PDB ID 3IXL) and 10.7 and 10.8 for a open configuration (PDB ID 2VLB chain A). Figure shows the pH-spectrum of AMDase G74C in the racemization of naproxen spectra in comparison with literature data of glutamate racemiztion by GluR[11] and of the wildtype AMDase in the decarboxylation of phenylmalonate.[17] While GluR and wildtype AMDase have a broad pH-spectrum from ~6.5-9.5 with a maximum at 8, G74C shows a needle-like curve with a maximum at pH 9. The broad pH-spectrum of GluR has been seen as evidence for the existence of two neutral thiol residues. The assumed involvement of two basic co-catalytic residues for a highly asynchronous deprotonation mechanism during the catalytic process agrees well with this observation.[4] In the mechanism of the decarboxylation of phenylmalonate by wildtype AMDase, C188 acts as proton donor (Scheme 1), which explains the broad pH-spectrum.

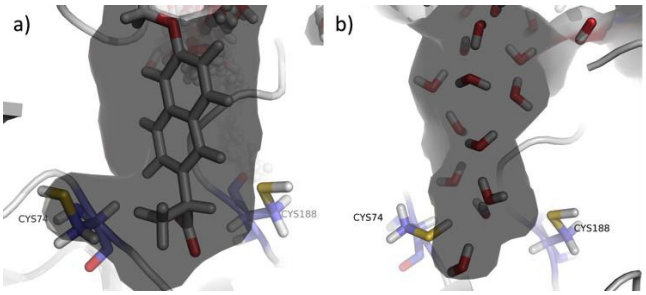


Figure 3. Snapshots after 5 ns of molecular dynamics simulations of (*R*)-naproxen a) and the empty AMDase G74C b) show the accessibility of the catalytic residues C74 and C188 for deprotonation by water molecules, whereby the solvent accessible surface of the enzyme is represented in grey.

While both racemases mediate the racemization with two Cys-residues, the striking differences in pH dependency are another indication for the different catalytic pathway and the putative role of the preformed thiolate in the mechanism of G74C. This assumption, however, opens up the question whether the thiolate is deprotonated before or after substrate binding. MD simulations of G74C with both enantiomers of naproxen (Figure), flurbiprofen and ketoprofen (not shown) did not show any water molecules in the close vicinity of the catalytic Cys that might initiate the deprotonation reaction. In the mechanism of the AMDase-catalyzed decarboxylation, Cys188 donates a proton to the enedionate intermediate and needs to be reprotonated afterwards. The simulations revealed that the ligand shields Cys188 effectively from the solvent. Figure shows that the reprotonation of Cys188 by water must therefore occur after the substrate leaves the active site. Interestingly, molecular dynamics simulations of the apo enzyme without ligand in closed configuration showed that the solvent can easily access the catalytic Cys residues. As the arylaliphatic acid, i.e. the product of AMDase, leaves the active site in the presence of the donating thiolate it is intuitive to assume that the arylaliphatic acid can also enter the same with the preformed abstracting Cys as thiolate. The strong pH-dependency of the racemization, the lack of any co-catalytic amino acid residues in the active site of G74C and accessibility of the shielding of the substrate bound enzyme to the solvent indicate that the racemization can only take place efficiently in the deprotonated form of the catalytic Cys before substrate binding.

How is the thiolate form of the catalytic cysteine stabilized? This question is relevant for the decarboxylative activity of the (*R*)-

selective WT AMDase (with C188 as H-donor), the (*S*)-selective variant G74C/C188G[18, 19] and the racemising G74C (with C74/C188 as H-donor and H-acceptor). Figure shows the direct vicinity of both cysteins in the thiolate form. For the stabilization of C188, the neighbouring residues G189, G190 and L191 can form backbone NH hydrogen bonds with the thiolate, thus strongly favouring the deprotonated form. This configuration strongly resembles the stabilization of the negatively charged oxanion of the tetrahedral intermediate in lipases,[20] in which backbone amide groups of neighbouring residues form a so-called oxyanion hole. During the natural decarboxylation mechanism of the wildtype AMDase C188 is supposed to be deprotonated. This makes it very likely that the thiolate hole was structurally optimized by evolution to stabilize this catalytic thiolate. Interestingly, a similar structural motif formed by 2 backbone amides and the sidechain of S76 and L77 stabilizes the artificially introduced C74 (Figure a). The simulations with deprotonated Cysteines show a high stability (H-bond distance <3 Å, cp. Figurec and d) of the thiolate anion pockets. The stabilization of the charged cysteines in thiolate pockets offers a mechanism via deprotonation of the Cys in the apoenzyme and subsequent substrate binding the most plausible hypothesis. This is exactly the reverse of the second half-reaction of the AMDase-catalyzed decarboxylation.

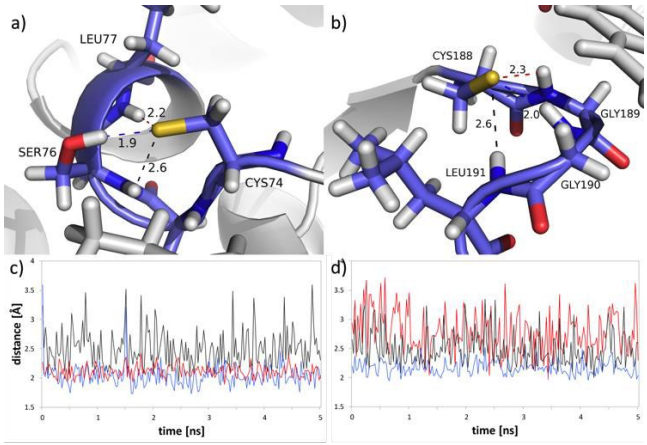


Figure 4. a) Stabilization of the thiolate residue C74 in a thiolate pocket; b) analogue stabilization of C188; c, d) distances of the hydrogen bonds of both pockets over a molecular dynamics simulation of 5 ns.

Combined quantum mechanical and molecular mechanical (QM/MM) calculations were performed to calculate the racemization reaction pathway using a combination of dispersion and H-bond corrected semiempirical Hamiltonian (PM6-DH+[21]) and Amber force field (ff03) utilising the AMBER 12 suite of programs. We chose a minimal unsaturated substrate (3-butenoic acid) and a minima chiral aromatic substrate (2-phenylpropionic acid) to investigate the geometry and free energy for the respective AMDase G74C-catalyzed isomerization pathways. According to the above shown MD-simulations with protonated and unprotonated C74 and C188 we started the isomeration reaction pathway calculations with the unprotonated thiolate anion, stabilized by the respective thiolate pocket. To asses the free energy barrier of this initial activation step and the isomerization reaction, we calculated the deprotonation of the two cysteines and the isomerization reaction pathways by steered QM/MM MD. In figure 5 the free energy barriers of the initial deprotonation of cysteines to yield the thiolate base for isomerization are shown.

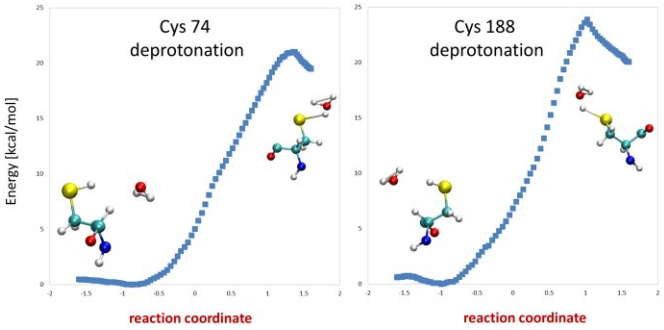


Figure 5. a) QM/MM MD free energy along the deprotonation reaction pathway of C74. As insert the starting structures and the final thiolate structure are shown. b) Deprotonation free energy of the Cys 188.

For both cysteine residues, a free energy barrier of >20 kcal mol^{-1} for the deprotonation of the thiolate anion by an adjacent water molecule were calculated. The isomerization pathway for the *R*-2-phenylpropionic acid was calculated using the same settings (deprotonated C188). A nearly symmetric reaction pathway was obtained, showing a planar endionate like structure in the transition region, with a free energy barrier of only 3.5 kcal mol^{-1}. The absolute free energy barriers of semiempirical calculations are not exact, and subsequent high level calculations using MP2 and CCSD(T) are needed to reach chemical accuracy.[22] Looking at the relative free energy barriers, we observed a lower barrier for the isomerization of the aromatic 2-phenylpropionic acid compared to the the barrier obtained for the nonaromatic small model substrate 3-butenoic acid (3.5 kcal mol^{-1}vs. 13 kcal mol^{-1}), for which a more disordered enediolate and product state after isomerization was observed (data not shown) and thus no symmetric pathway could be found. Interestingly, the planar endionate anion remained stable for about 10 ps (corresponding to a plateau of 20,000 MD time steps within 1 kcal mol^{-1}), before reprotonation from Cys 188 occurred, as is depicted in Figure 5b (see also supporting information). The first proton transfer is completed after 25 ps (reaction coordinate 0), whereas the second proton transfer occurred at 43 ps (reaction coordinate +1.5).[19]

Geometric analysis of the sampled geometries in the transition region revealed a planar endiolate structure after the completion of the first proton transfer at 25 ps until reprotonation occurred at 43 ps, indicating an elongated lifetime of the metastable planar enediolate for aromatic substrates. The stabilization and stepwise proton transfer for aromatic compounds is similar to the second part of the recently reported decarboxylation reaction mechanism[8] of AMDase. [21,22]

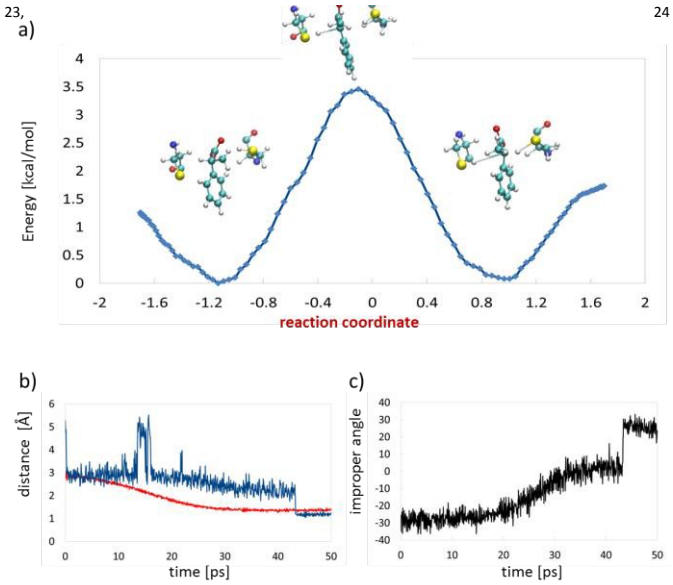


Figure 6. a) QM/MM MD free energy along the isomerization reaction pathway of 2-phenylpropionic acid. As insert the starting structure, a transition structure and the final structure are shown. b) distance for the hydrogen atoms from C74 in red (deprotonation step) and C188 in blue (reprotonation step) c) improper angle of the isomerized carbon atom, an angle near zero indicates the planar endiolate anion.

From the QM/MM simulations, a first racemization mechanism for AMDase G74C was formulated (Figure) The reaction is initiated by the deprotonation of the catalytic C74 or C188 by water, as AMDase has no cocatalytic residues like Asp 7 present in Glutamate racemase MurI[4] to deprotonate the cysteine. Instead the catalytic thiolate is stabilized by AMDase by backbone H-bonds in one of the two thiolate pockets. While the lower free energy barrier of the isomerization suggests a fast reaction, the initial deprotonation shows a higher free energy barrier. It should be noted that at a pH range of 7-9, only a small part of the AMDase molecules have the catalytic Cys as free base. In turn, the predicted pKa value about 10-11 of both catalytic Cys residues in the closed form can explain the narrow pH profile in comparison to GluR, which utilizes cocatalytic residues for Cys deprotonation. Abstraction of the α-proton then leads to a partial charge at the α-carbon atom, which attracts the H-atom of the other cysteine. A delocalized π-electron system on the vinylic or aromatic substituent of the substrate is required for the stabilization of this planar endiolate structure. This reaction is a typical example for substrate-assisted catalysis[23, 24] contributing to substrate specificity and catalytic activity of enzymes. Substrate assisted catalysis can facilitate enzyme evolution, as has been shown in the case of sugar isomerases[24] showing a new functionality by introducing only a single amino acid exchange. In deuterium exchange experiments of 3-butenoic acid, isomerization to 2-butenoic acid did not take place.[16] This was surprising, because *in situ* NMR studies showed 3-butenoic acid isomerizes under strongly basic conditions (5 M NaOD). Upon complete deprotonation in the active site of AMDase G74, an isomerization of the enedionate would be an obvious assumption. The simulations show that a partial deprotonation indeed occurs. Nevertheless, the isomerization requires a proton donor close to the δ-C-atom. As in the binding pocket of AMDase no amino acid residues with acidic protons are located in the close vicinity of this atom, lack of a suitable proton donor makes the reprotonation of the α-C-atom by the opposing catalytic cysteine the preferred reaction.

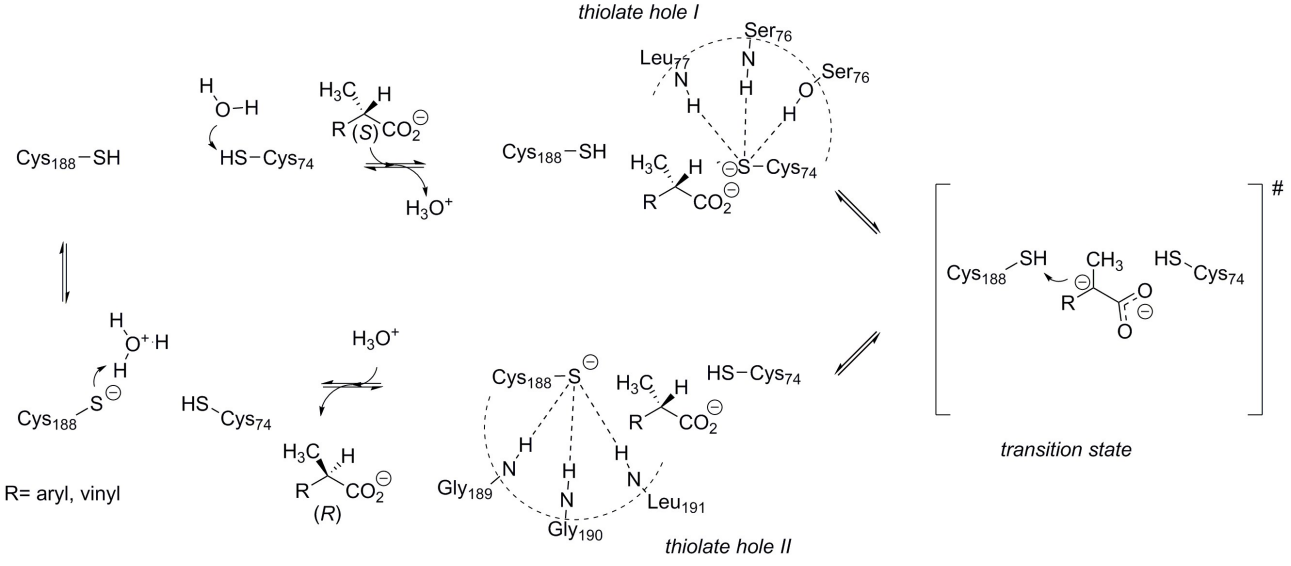


Figure 7 . Suggested mechanism for the AMDase G74C-catalyzed stereoinversion of (*S*)-α-arylcarboxylic acids. The racemization is initiated with the precatalytic deprotonation of the catalytic C74 by a water molecule. A thiolate pocket stabilizes the deprotonated catalytic C74 anion that deprotonates the α-C-atom, resulting in a planar enedionate intermediate. Reprotonation by C188 leads to a second thiolate-anion that is stabilized by a second pocket.

Journal Name

ARTICLE

Conclusions

Molecular dynamics simulations and semiempirical calculations revealed a mechanism for the enzymatic racemization of arylpropionates by AMDase G74C. The reaction is initiated by deprotonation of C188 or C74 to yield a thiolate residue as catalytic base. The lack of co-catalytic residues for the deprotonation of the catalytic cysteine explain the pronounced pH-dependency of the racemization. Indeed, the barrier for deprotonation is considerably higher than that of the isomerization.The MD simulations showed that the two thiolate residues before and after the proton transfer are stabilized by hydrogen bonds in two thiolate holes. While one of these is involved in the decarboxylation reaction of wildtype AMDase, the other was coincidentally created with the artificial introduction of C74. In glutamate racemase, the transfer of the two protons proceeds in a concerted way. In contrat, semiempirical calculations show a stepwise mechanism for AMDase G74C, which explains the requirement for substrate substituents with a delocalized π-electron systems. The striking suppression of isomerization in the enzymatic conversion of 3-butenoic acid can be explained by the lack of a suitable proton donor close to the δ-C-atom, which favours reprotonation in α-position. The presented isomerization mechanism is in line with the experimentally observed sharp pH-profile and substrate spectra of the designed catalyst AMDase G74C.

Experimental

Determination of pH profile

Purified AMDase G74C was produced as described.[25] To a solution of 1.1 mL (*S*)-(+)-naproxen (2 mM) in the appropriate buffer, 0.1 mg enzyme was added. After addition of the substrate, the pH of every sample was confirmed by using a micro pH-electrode. The mixture was shaken at 28°C. Samples are taken after 30 and 45 min and quenched by the addition of 250 µl HCl (2 M). The samples were extracted with methyl-*tert*-butylether, derivatized with TMS-diazomethane[12] and analysed with chiral GC-chromatography (Shimadzu 2010 plus) with the cyclodextrin column FS-Hydrodex-β-6TBDM (Macherey-Nagel, Germany). Using an isotherm method with a column temperature of 160°C, the two enantiomers of naproxen eluted at 31.5 min. (*S*) and 32.2 min. (*R*), respectively. Three different buffer systems were used to cover the range from pH 5-11. This is Citrate-buffer (50 mM, pH 5, 6, 6.3, 6.7, 7), Tris-HCl (50 mM pH 7, 7.3, 7.7, 8, 8.5, 9) and Gylcin-NaOH (50 mM, pH 9, 10, 11).

Molecular dynamics simulations

For the preparation and implementation of MD Simulations YASARA 12.10.3 and AMBER03 force field were used.[26, 27] Parameterization of the substrate was done by YASARA AutoSMILES.[28] The initial structure was prepared starting with the X-ray crystal structure of AMDase G74C/C188S mutant containing a 2-phenylacetic acid in the active site (PDB ID 3IXL).[29] Protons were added according to calculated pk_a-values and pH set for 7.4 using 'Cell neutralization and pk_a prediction' tool of YASARA.[30] YASARA uses a pk_a-value of 8.7 for Cys, which differs from the calculated values from PROPKA. As the active site cavity does not contain protic residues effected by a pH shift from 9 to 7.4, a pH of 7.4 was used for all simulations in order to avoid an artificial deprotonation of all cysteins. Into a periodic box surrounding the protein structure by 5 Å water molecules were added as well as Na^+ and Cl^- ions in a concentration of 0.9 % to neutralize the cell. The density was set for 0.997 g L^{-1}. Afterwards a bound sulfate ion and glycerol molecule were removed, followed by energetic minimization of the solvent while the protein and its ligand were fixed. Additionally in position 148 the 2-mercaptoethanol was removed from the cysteine. Another energy minimization was performed allowing all residues in a perimeter of 5 Å to relax, whereas other atoms were fixed and relaxed by a second minimization. By similar procedure the serine 188 was mutated back to cysteine. Starting with this prepared structure, both enantiomers of 2-phenylacetic acid were modeled to yield starting geometries for flurbiprofen, naproxen and ketoprofen. Blocking water molecules were removed. Energy minimizations were performed, allowing the added phenyl ring to relax with fixed solvent and protein. Atom by atom the relaxing part of the substrate was extended. Afterwards the residues next to the substrate were also relaxed by energy minimization, extending the radius of the considered atoms for the minimization step by step. These structures were the starting point for MD simulations with all cysteines protonated. Furthermore the thiolate was modeled according to the substrate chirality. For (*R*)-substrates cysteine 188 was deprotonated, whereas in case of (*S*)-substrates cysteine 74 was modified. Energy minimization was performed as previously described for residue mutation. The MD Simulations[31] were performed at a temperature of 298 K with a cutoff of 8 Å considering long range coulomb electrostatics[32]. With a time step of 2.5 fs every 25 ps a snapshot was taken over a total simulation time of 5 ns. The stability of naproxen in the active site was confirmed with a MD run over 20 ns.The recorded trajectories reaction pathways were visualised and geometrically analysed using VMD 1.8.7.[33]

Semiempirical QM/MM calculatons

QM/MM MD simulations were performed by using the AMBER12 suite of programs[34, 35] and employing force fields Amber03 for proteins[27] and GAFF for ligands.[28] The simulations were based on the crystal structure of *Bordetella bronchiseptica* arylmalonate decarboxylase G74C/C188S mutant with bound 2-phenylacetic acid. The structure were initially protonated using PROPKA 3.0,[36, 37] the acid moiety of the model substrates 3-butenoic acid and 2-phenylpropionic acid were treated as deprotonated. The protein/substrate complexes were solvated in a cubic box filled with Tip3[38] molecules. The substrates were parameterized according to the standard GAFF procedure, assigning partial charges by using a HF/631G* RESP[39] procedure. A stepwise minimization/equilibration procedure was applied to all structures with and without bound substrate, minimizing first the solvent and then the entire system for 2500 steps using periodic boundaries and a cutoff of 10 Å. During heating from 0 to 300 K over 20 ps, the protein backbone and the ligand were restrained by a force 10 of kcal mol^{-1}, before an additional heating of the unrestrained system over 100 ps was performed. All bonds to hydrogen atoms were restrained using SHAKE[40] during minimization and MD simulations. Productive QM/MM MD simulations under NPT-ensemble conditions[41] were performed at 300 K and 1 bar pressure using Langevin temperature control (ntt = 3) and isotropic position scaling to maintain the pressure (NTP = 1) using a relaxation time of 2 ps (TAUP =2.0). The equilibrated system was then further subjected to QM/MM SMD simulation for 25 ps at 300 K with a time step of 0.5 fs to calculate the barriers involved in the reaction process. The hydride or proton was transferred with 50000 steps and 50 integration windows (1000 steps per window) along the reaction coordinate from +2.0 to −2.0.QM/MM simulations of substrate racemization reactions in pure solvent and in the active site were performed using QM/MM PM6-DH+ Hamiltonian[21, 42] including dispersion and H-bond correction for the substrate atoms, the two reactive Cys residues and the sidechains of all adjacent active-site residues (C74, T75, S76 Y126, C188) or water for the deprotonation pathways. To calculate the QM/MM electrostatic interactions, a cutoff of 8 Å was used. The energy along the reaction coordinate was scanned by Adaptively Biased MD[43] using a linear combination of distances (LCOD) of the formed and broken bonds for the respective proton transferred for all reaction pathways. The steered molecular dynamics simulation (SMD) approach implemented in Amber12 was used to force proton transfer from the substrate to the adjacent deprotonated Cys residue.The simulations were carried out with the QM/MM PM6-DH+/ff03 Hamiltonian using LCOD for driving the reaction coordinate (RC). RC for the proton transfer step is defined as the distance difference (D1 − D2) of the transferred hydrogen (H) toward C of the phenylpropionate substrate (D1) and H to the deprotonated SG of the respective Cysteine residue (D2). The H atom of the substrate was moved step by step toward the carbonyl carbon (C) of the substrate by applying harmonic constraint (1000 kcal mol^{-1} $Å^{-2}$) to drive the reaction The Jarzynski averaged potential of mean force was calculated to estimate the free energy of the respective reaction. Since proton transfer reactions were analyzed, QMSHAKE was turned off. The recorded trajectories reaction pathways were visualised and geometrically analysed using VMD.[33]

Acknowledgements

R.K. thanks the Federal Ministry for Innovation, Science and Research of North Rhine-Westphalia (grant number PtJ-TRI/1411ng006) and R.K, K.M. and F.B. thank the German Academic Exchange Service (grant number 57154401) for financial support. M.B. is indebted to the Deutsche Forschungsgemeinschaft (grant number SPP 1170:BO 3066/1-1,1-2) for financial support.

Notes and references

1. E. J. Ebbers, G. J. A. Ariaans, J. P. M. Houbiers, A. Bruggink and B. Zwanenburg, *Tetrahedron*, 1997, **53**, 9417-9476.
2. U. Felfer, M. Goriup, M. F. Koegl, U. Wagner, B. Larissegger-Schnell, K. Faber and W. Kroutil, *Adv. Synth. Catal.*, 2005, **347**, 951-961.
3. S. Baxter, S. Royer, G. Grogan, F. Brown, K. E. Holt-Tiffin, I. N. Taylor, I. G. Fotheringham and D. J. Campopiano, *J. Am. Chem. Soc.*, 2012, **134**, 19310-19313.
4. E. Puig, E. Mixcoha, M. Garcia-Viloca, A. Gonzalez-Lafont and J. M. Lluch, *J. Am. Chem. Soc.*, 2009, **131**, 3509-3521.
5. J. Gu and H. Yu, *J. Biomol. Struct. Dyn.*, 2012, **30**, 585-593.
6. J. Gu, M. Liu, F. Guo, W. Xie, W. Lu, L. Ye, Z. Chen, S. Yuan and H. Yu, *Enzyme Microb. Tech.*, 2014, **55**, 121-127.
7. Y. Terao, K. Miyamoto and H. Ohta, *Chemical Commun.*, 2006, 3600-3602.
8. M. E. S. Lind and F. Himo, *ACS Catal.*, 2014, **4**, 4153-4160.
9. R. Kourist, P. Domínguez de María and K. Miyamoto, *Green Chem.*, 2011, DOI: 10.1039/c1gc15162b, 2607-2618.
10. L. Steenkamp and D. Brady, *Process Biochem.*, 2008, **43**, 1419-1426.
11. M. May, S. Mehboob, D. C. Mulhearn, Z. Wang, H. Yu, G. R. J. Thatcher, B. D. Santarsiero, M. E. Johnson and A. D. Mesecar, *J. Mol. Biol.*, 2007, **371**, 1219-1237.
12. S. K. Gassmeyer, H. Yoshikawa, J. Enoki, N. Hülsemann, R. Stoll, K. Miyamoto and R. Kourist, *ChemBioChem*, 2015, **16**, 1943-1949.
13. E. Puig, M. Garcia-Viloca, A. Gonzalez-Lafont and J. M. Lluch, *J. Phys. Chem. A*, 2006, **110**, 717-725.
14. E. Mixcoha, M. Garcia-Viloca, J. M. Lluch and A. Gonzalez-Lafont, *J. Phys. Chem. B*, 2012, **116**, 12406-12414.
15. K. A. Gallo, M. E. Tanner and J. R. Knowles, *Biochemistry*, 1993, **32**, 3991-3997.
16. R. Kourist, Miyauchi, Y., Uemura D., Miyamoto, K., R, *Chem. Eur. J.*, 2011, **17**, 557-563.
17. K. Miyamoto and H. Ohta, *Eur. J. Biochem.*, 1992, **210**, 475-481.
18. S. Yoshida, J. Enoki, R. Kourist and K. Miyamoto, *Biosci. Biotech. Biochem.*, 2015, **79**, 1965-1971.

19. Y. Miyauchi, R. Kourist, D. Uemura and K. Miyamoto, *Chem. Commun.*, 2011, **47**, 7503-7505.
20. A. M. Brzozowski, U. Derewenda, Z. S. Derewenda, G. G. Dodson, D. M. Lawson, J. P. Turkenburg, F. Bjorkling, B. Huge-Jensen, S. A. Patkar and L. Thim, *Nature*, 1991, **351**, 491-494.
21. M. Korth, *J. Chem. Th. Comp.* , 2010, **6**, 3808-3816.
22. R. A. Mata, H.-J. Werner, S. Thiel and W. Thiel, *J. Chem. Phys.*, 2008, **128**, 025104.
23. W. Dall'acqua and P. Carter, *Protein Sci.*, 2000, **9**, 1-9.
24. B. Reisinger, M. Bocola, F. List, J. Claren, C. Rajendran and R. Sterner, *Protein Eng. Des. Sel.*, 2012, **25**, 751-760.
25. M. Bartsch, S. K. Gassmeyer, K. Igarashi, M. Miyamoto, M. Rögner, M. Nowaczyc and R. Kourist, *Microb. Cell. Fact.*, 2015, **14**, 53.
26. E. Krieger and G. Vriend, *Bioinformatics*, 2014, **30**, 2981-2982.
27. Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang and P. Kollman, *J. Comput. Chem.*, 2003, **24**, 1999-2012.
28. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **26**, 114.
29. R. Obata and M. Nakasako, *Biochemistry*, 2010, **49**, 1963-1969.
30. E. Krieger, J. E. Nielsen, C. A. E. M. Spronk and G. Vriend, *J. Mol. Graph. Model.*, 2006, **25**, 481-486.
31. E. Krieger, T. Darden, S. B. Nabuurs, A. Finkelstein and G. Vriend, *Proteins*, 2004, **57**, 678-683.
32. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577-8593.
33. W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33-38, plates, 27-28.
34. D.A. Case, J.T. Berryman, R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K.M. Merz, G. Monard, P. Needham, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, R. Salomon-Ferrer, C.L. Simmerling, W. Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, D.M. York and P.A. Kollman (2015), AMBER 2015, University of California, San Francisco.
35. R. Salomon-Ferrer, D. A. Case and R. C. Walker, *WIREs Comput. Mol. Sci.*, 2013, **3**, 198-210.
36. C. R. Sondergaard, M. H. M. Olsson, M. Rostkowski and J. H. Jensen, *J. Chem. Theory Comput.*, 2011, **7**, 2284-2295.
37. M. H. M. Olsson, C. R. Sondergaard, M. Rostkowski and J. H. Jensen, *J. Chem. Theory and Comput.*, 2011, **7**, 525-537.
38. B. Zhang, V. B. C. Tan, K. M. Lim and T. E. Tay, *J. Chem. Inf. Model.*, 2007, **47**, 1877-1885.
39. P. Cieplak, W. D. Cornell, C. Bayly and P. A. Kollman, *J. Comput. Chem.*, 1995, **16**, 1357-1377.
40. S. Miyamoto and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 952-962.
41. G. J. Martyna, A. Hughes and M. E. Tuckerman, *J. Chem. Phys.*, 1999, **110**, 3275-3290.
42. J. J. P. Stewart, *J. Mol. Model.*, 2007, **13**, 1173-1213.
43. V. Babin, C. Roland and C. Sagui, *J. Chem. Phys.*, 2008, **128**, 134101/134101-134101/134107.

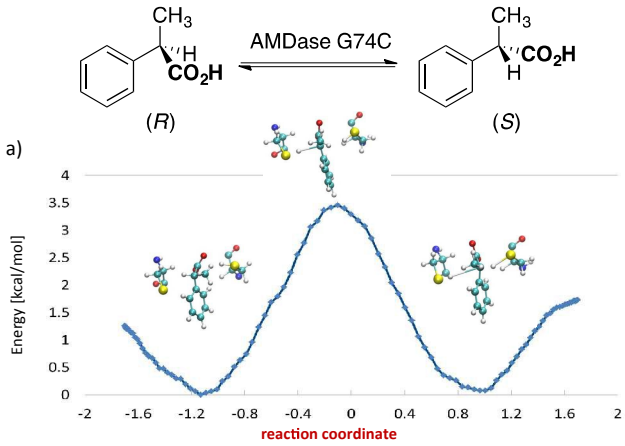
1

Journal Name

ARTICLE

Semiempirical QM/MM calculations reveal a step-wise proton transfer and an unusual thiolate pocket in the mechanism of the unique arylpropionate racemase AMDase G74C

F. Busch, J. Enoki, N. Hülsemann, K. Miyamoto, M. Bocola, and R. Kourist



Semiempiric calculations on the mechanism of the arylpropionate racemase AMDase G74C reveal a step-wise mechanism involving a planar-enedionate intermediate, which explains the specificity of the enzyme for unsaturated substrates.