6. The catalytic mechanism of arylsulfatase A and its theoretical investigation

When the crystal structure of arylsulfatase A was solved, a remarkable structural analogy to another hydrolytic enzyme, the alkaline phosphatase was encountered. The analogy could even be extended to the active sites of both enzymes. This finding lead to the conclusion that ASA could apply the same mechanism as AP, which is well understood. For this reason a mechanism for the sulfate ester hydrolysis by ASA was suggested by the authors of the crystal structure on the basis of the active site structure and the analogy to alkaline phosphatase. In this chapter I will give a short comparison of the suggested mechanism of ASA and the mechanism of AP followed by the outline of a strategy how one part of this suggested mechanism could be verified by theoretical methods.

6.1. The proposed mechanism of arylsulfatase A

The key role in the catalytic mechanism of ASA is played by a cysteine residue (Cys69) that undergoes a post-translational modification and exists then as a formylglycine residue (FGly69).[162,163] The chemical modification of Cys69 to a formylglycine is shown in Figure 6.1.

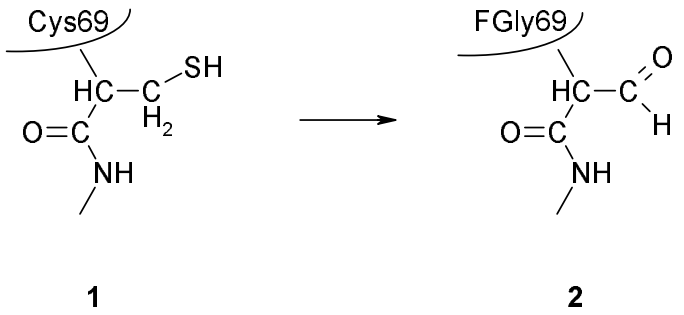


Figure 6.1.: The post-translational modification of Cysteine residue 69 (**1**) conserved among all eukaryotic sulfatases, leads to a formylglycine (**2**, FGly69). In this reaction the cysteine is oxidized at the C_β-atom yielding L-2-amino-3-oxopropionic acid with an aldehyde group as the functional group of the side chain. It is unknown by which enzyme this modification is performed.

The cysteine residue is conserved among all eukaryotic sulfatases. The rare disorder Multiple Sulfatase Deficiency (MSD) results from a defect of the sulfatases in this post-translational modification as already mentioned in section 5.2. Additionally it was shown that typical aldehyde reagents as sulfite, hydrazine, hydroxylamine and cyanide inhibit ASA.[146,164,165] This fact also indicates that it is in fact an aldehyde residue that exerts a prominent task in the catalytic mechanism.

The second important feature of the ASA active site is the presence of a magnesium ion. This ion is coordinated by amino acids that are conserved in all sulfatases: The three carboxylate side chains of Asp29, Asp30 and Asp281 coordinate the Mg^{2+} ion with four of their oxygens. The distorted octahedral coordination sphere is completed by Asn282O_δ and the side chain of FGly69 as shown in Figure 6.2.

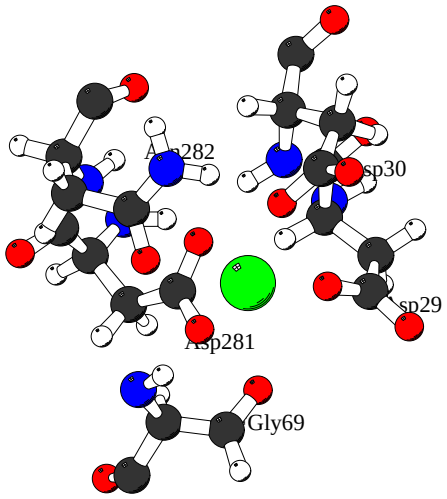


Figure 6.2.: The metal binding site in ASA. The Mg^{2+}-ion is coordinated by six oxygen atoms. Asp281 coordinates with two carboxylate oxygen atoms and Asp29 and Asp30 have each one oxygen atom of their side chain involved in the complexation. In addition the oxygen atoms of the Asn282 amide side chain and the aldehyde oxygen atom of FGly69 coordinate the ion.

The electron density found at the aldehyde was interpreted as a two-fold disordered aldehyde group with possible contribution of aldehyde hydrate. One hydroxyl oxygen of the geminal diol coordinates Mg^{2+}, the other coordinates His125N_δ. The structure of the active site with the unhydrated aldehyde is shown in Figure 6.3.

The Mg^{2+}-ion and the Lys302N_ζ atom are found in positions where in AP two Zn^{2+}-ions are located. Both Zn^{2+}-ions in AP are responsible for the binding of the phosphate oxygen atoms of the phosphate ester substrate. Therefore Mg^{2+} and Lys302N_ζ are ascribed the function to bind the sulfate ester oxygen atoms of the ASA substrates. Lys123N_ζ and His229N_ε in ASA may

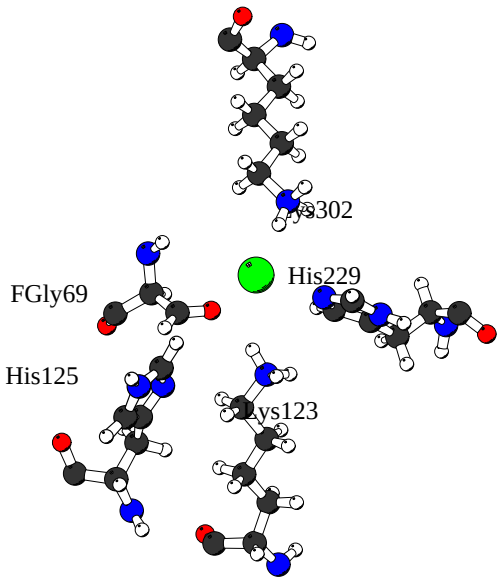


Figure 6.3.: The active site of ASA. The aldehyde group of FGly69 can be hydrated. Then one oxygen atom of the geminal diol coordinates the $Mg^{2}+$ ion, whereas the other hydroxyl group coordinates His125N_{δ} via a hydrogen bond. Lys302, Lys123, His229 and Mg^{2+} are involved in substrate binding.

substitute for Arg166N_ε and Arg166N_η in AP that bind an oxygen ester atom in bidentate mode. The mechanism of AP is shown in Figure 6.4

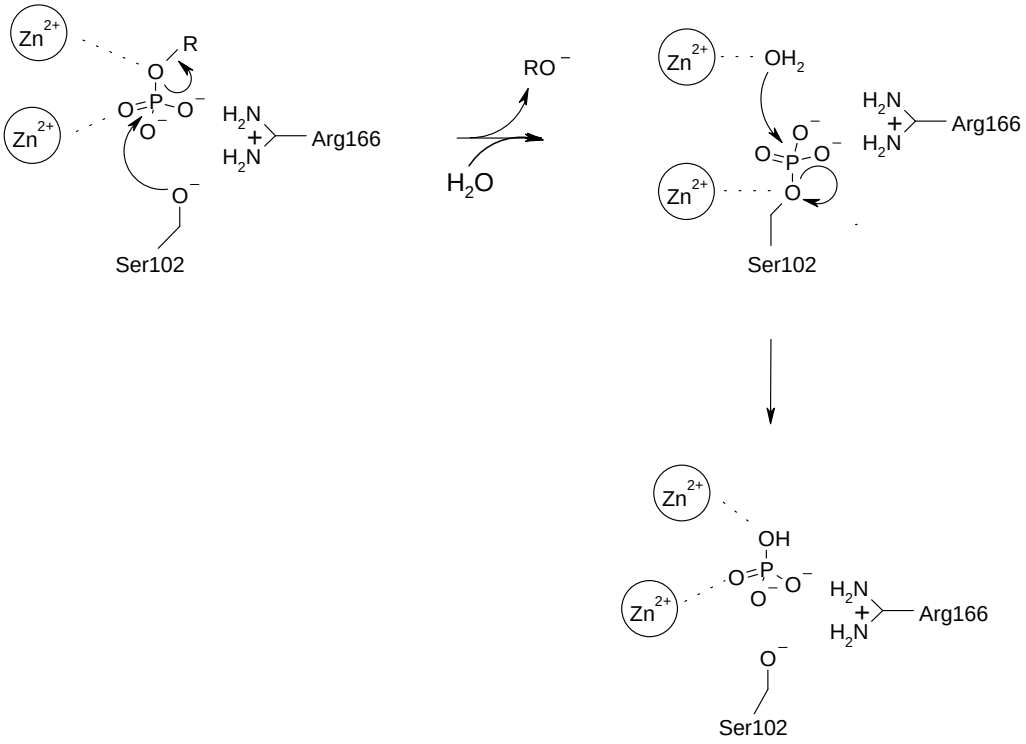


Figure 6.4.: The mechanism of alkaline phosphatase. In the active site of AP Ser102 acts as nucleophile. It attacks the phosphor atom of the phosphate group. The phosphorylated serine is then hydrolyzed by a water molecule. The Zinc ions and Arg166 are responsible for the binding of the phosphate ester.

The active site structure of ASA and the comparison with the mechanism of AP lead to the suggestion of the mechanism, that is shown in Figure 6.5:

The first step is the addition of a water molecule to the aldehyde function of FGly69 to form the diol. The formation of the diol is in analogy to the reaction in AP, where the oxygen of a serine side chain acts as a nucleophilic atom that attacks a phosphorous atom. In ASA this role can be adopted by a hydroxyl oxygen atom of the diol which can be considered as a mild acid.[166] In principle, it would also be possible that the aldehyde is not converted into a hydrate and that its carbonyl carbon atom is the target of a nucleophilic attack by one of the non-ester oxygens of the sulfate ester. But the diol hypothesis is supported by an investigation with a modified ASA:[167] By site directed mutagenesis the residue FGly69 is replaced by a serine. That serine attacks the sulfur atom of the ester to form a sulfate ester derivate of ASA. Therefore, it is most likely that in ASA it is a nucleophilic attack of the diol that initiates the hydrolysis of the sulfate ester.

The next step is the binding of the sulfate ester. The sulfur atom of the ester is then attacked to bind to one oxygen atom of the diol. The alcohol moiety of the ester is the leaving group of this nucleophilic substitution. The following step consists in the elimination of the SO_4^{2-} group, that is driven by the reformation of the aldehyde group and is initiated by the second hydroxyl

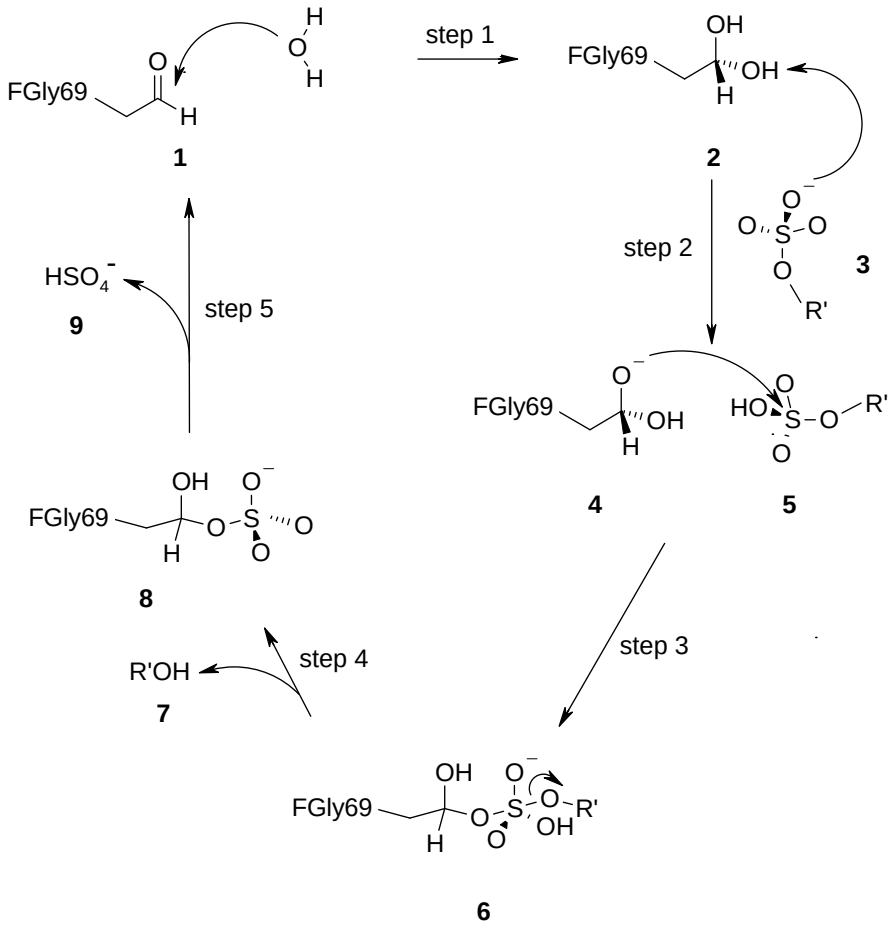


Figure 6.5.: The proposed mechanism of ASA. In the first step the aldehyde group of FGly69 (**1**) is hydrated by a water molecule and forms the aldehyde hydrate (**2**) The sulfate ester (**3**) binds to the enzyme and activates the diol (**2**) by abstracting a proton. The negatively charged diol (**4**) attacks now the protonated ester (**5**) and a pentavalent intermediate (**6**) is formed. The alcohol (**7**) splits from the intermediate and a new sulfate ester (**8**) is formed. This ester is instable due to the hydroxyl group next to the sulfate ester group. So a hydrogen sulfate (**9**) leaves the ester and the aldehyde group of FGly69 is reformed.

group. A remarkable support for the involvement of the second hydroxyl group is the fact that in the case of the Cys69Ser mutation discussed above, the sulfate ester derivate of ASA is not processed further, because a hydroxyl group that could eliminate the sulfate ion is missing there.

6.2. The simulation of the ASA mechanism

The suggested mechanism for the ester hydrolysis by ASA includes some interesting steps that are not very common to other enzyme reactions and should be investigated further to clarify the processes in detail:

1. The formation of an aldehyde hydrate, shown as step 1 in Figure 6.5, in an enzyme environment is to my knowledge not reported from any other enzyme and its feasibility should be proven be additional experimental and theoretical work.

2. The role of the diol as nucleophilic agent is also unique to ASA. As the chemical properties of diols are not very well characterized, the acidity of the diol in the enzyme and the nucleophilicity of the corresponding anion should be investigated. These properties are significant for steps 2 and 3 in Fig. 6.5.

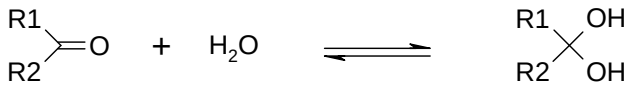
3. The elimination of the sulfate ion by a geminal hydroxide group in step 5 in Figure 6.5 can be explained by the stability of the resulting aldehyde group. But it is still unclear, which mechanism is employed to recreate the aldehyde.

The EVB method in combination with the titration of the protein is an ideal tool to assess these open questions. In contrast to earlier applications of the EVB method the catalytic reaction in ASA is however somewhat more demanding. There is not such a profound basis of experimental data on the reactions in solution which is prerequisite for the straightforward usage of the EVB method. Therefore, the investigation of the ASA mechanism requires gas phase *ab initio* calculations, that are necessary to provide data for the parameterization of the EVB framework. In the following section, I will describe the possible mechanism of the diol formation in ASA and introduce the possible theoretical methods to simulate the suggested reactions.

6.2.1. The formation of the diol

Step 1 in the ASA mechanism is the hydration of an aldehyde. Some general facts of this reaction type have to be considered before analyzing the possible mechanism in the enzyme: A water molecule can add to the carbonyl groups of aldehydes and ketones as shown in Figure 6.6. While the equilibrium constants for ketones and aryl aldehydes are much less than unity for this reaction, aliphatic aldehydes are hydrated preferably in aqueous solution. The equilibrium constant for formaldehyde is 2×10^3, for acetaldehyde it is 1.3 and for acetone it is 2×10^{-3}.

The diol formation is subject to general acid and general base catalysis.[168,169,170] As a model for the hydration of FGly69 the corresponding reaction of acetaldehyde seems to be suitable, as the methyl group next to the aldehyde group in acetaldehyde should show comparable effects as the methylene group in FGly69. The reaction rate for the hydration of acetaldehyde is $7.9 \times 10^{-3} s^{-1}$ in aqueous solution, the acid catalyzed reaction has a rate constant of $930 M^{-1} s^{-1}$ and the base catalyzed reaction has a rate constant of $8 \times 10^4 M^{-1} s^{-1}$.
The general acid catalyzed mechanism is shown in figure 6.7:
The general base catalyzed mechanism is shown in fig 6.8.



R1, R2 = H, aryl or alkyl

Figure 6.6.: The hydration of the carbonyl group of aldehydes or ketones.

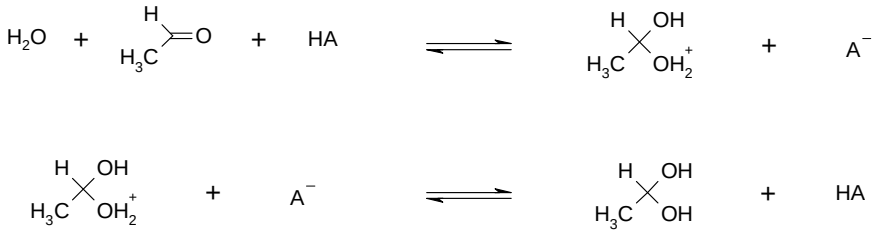


Figure 6.7.: The acid catalyzed mechanism of diol formation. HA denotes the general acid, that protonates the carbonyl oxygen. The carbonyl group undergoes the nucleophilic attack then.

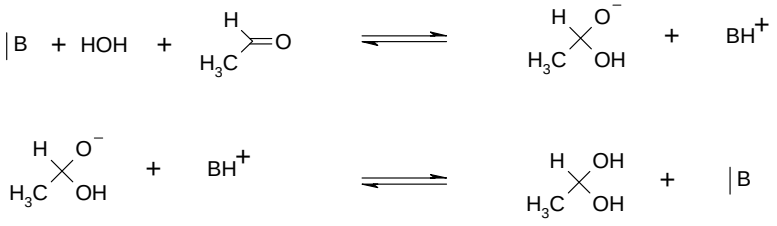


Figure 6.8.: The base catalyzed mechanism of diol formation. B is the general base that accepts a proton from the water molecule, resulting in a hydroxide ion that attacks the carbonyl group of the aldehyde.

The kinetic data of the catalyzed and uncatalyzed hydration of acetaldehyde in aqueous solution have to be compared with rate constants of the whole reaction that is catalyzed by ASA: The overall rate of the catalyzed sulfate ester cleavage in ASA is $k_{cat} = 16.8s^{-1}$ for 4-nitrocatechol sulfate, $k_{cat} = 0.8s^{-1}$ for 4-methylumbelliferyl sulfate and $k_{cat} = 1.5s^{-1}$ for cerebroside sulfate.[171] That means that the turnover rate for the whole reaction in the enzyme is at least larger by a factor of 1000 than the diol formation in aqueous solution without catalysis which is $7.9 \times 10^{-3}s^{-1}$ as stated above. The diol formation in ASA is only one step of the complete mechanism. Its rate must not be slower than the total mechanism. That indicates that the formation of the diol has to be accelerated by the enzyme compared to the uncatalyzed reaction in solution, such that the overall rate of the whole catalytic cycle is able to yield the measured rates. The enzyme should thus provide a means to accomplish an acid or base catalyzed diol formation of FGly69. The visual inspection of the active site shows that a water molecule, which is needed for the hydration, would fit into a position between FGly69 and Lys123. The side chain of Lys123 could play a role as acid or base depending on its protonation state. The titration studies of ASA, carried out to investigate the dimer-dimer association that were described in chapter 5, revealed that Lys123 is unprotonated even at lower pH values around 5. This suggests that Lys123 catalyzes the diol formation as a base. This mechanism is shown in Figure 6.9.

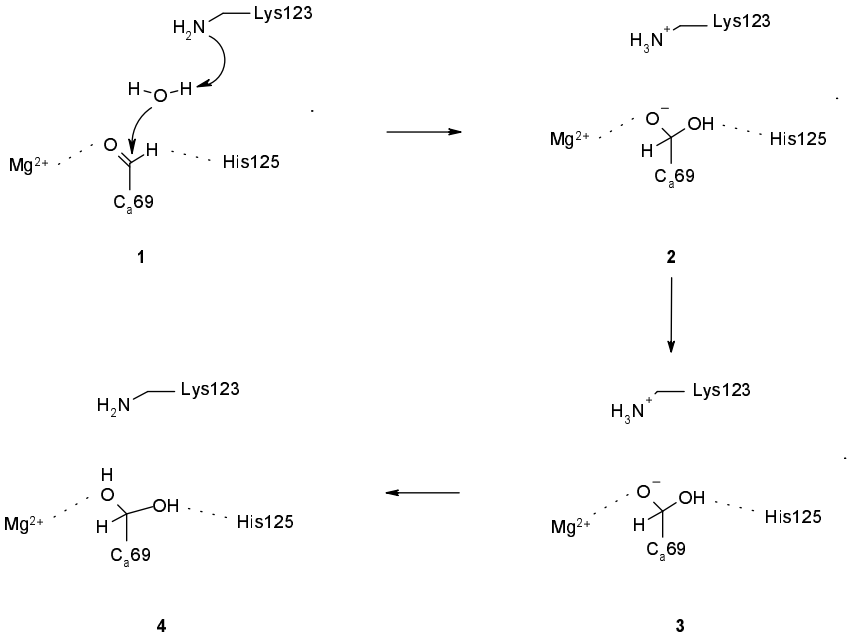


Figure 6.9.: The formation of the diol. The aldehyde oxygen coordinates the Mg^{2+}-ion, the hydrogen forms a hydrogen bond to His125. The water molecule is activated by Lys123. The formation of the aldehyde hydrate is base catalyzed.

An experimental mutation study on ASA showed that the Lys123Ala mutant has a rate reduced by 90% compared to the wild type[172] rate. In this study, it was found that also K_M of the Lys123Ala mutant is increased by a factor of around 15. Therefore Lys123 was supposed to play an important role in substrate binding. The authors tested mutations of nine residues in the active site. Most of them showed a decreased rate constant comparable to the one of

the Lys123Ala mutant. Therefore, from this study it cannot be deduced straightforwardly, that Lys123 is directly involved in the catalytic mechanism as suggested above. The rate reduction of the Lys123Ala mutant could also result from conformational changes due to the replacement of a putative hydrogen bond donor in the active site.

As it is not clear how the critical role of Lys123 can be enlighted by further experiments, a theoretical study on the suggested function is an important contribution.

6.2.2. Simulation of the diol formation

The mechanism for the diol formation, that was suggested in the previous section, should be analyzed by a theoretical method to investigate its feasibility. A tool would be the EVB method that is described in section 3.3. As mentioned above, the experimental data needed to adjust the empirical parameters of the EVB method, are not available in the literature in the same way as for the study on acetylcholinesterase that was described in chapter 4. In the following, I will describe how a typical simulation of the diol formation could be done within the EVB framework. There, I will indicate the missing experimental data, which impose restrictions on this method.

The hydration of the diol consists of the conversion of structure **1** to structure **2** and then to structure **4** in Figure 6.9. It consists of a proton transfer from a water molecule to a lysine residue and a nucleophilic attack of the water oxygen atom on the carbonyl carbon atom of the aldehyde, which is then followed by another proton transfer from lysine to the diolate anion. For this reaction four resonance structures can be defined as shown in Figure 6.10.

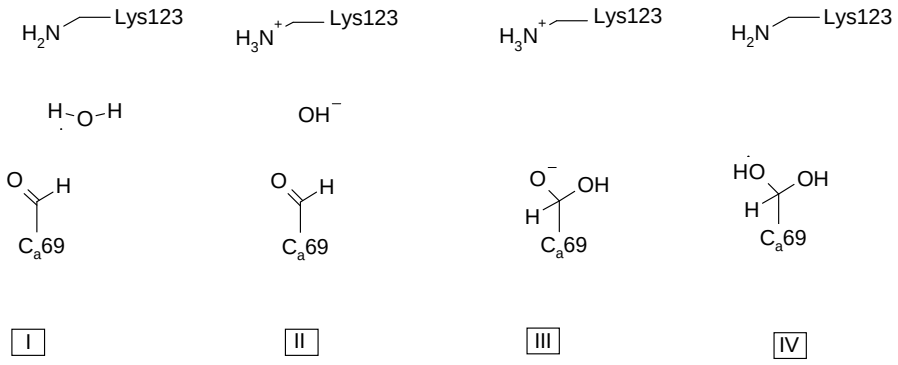


Figure 6.10.: The resonance states of the diol formation. State I is the reactant state containing the water molecule, the aldehyde and lysine in the unprotonated state. The second state (II) shows the system after proton transfer from the water molecule to lysine, which leads to a hydroxide ion. In state III the nucleophilic attack of the hydroxide ion to the aldehyde has occurred and the negatively charged diolate is obtained. The last resonance state (IV) shows the product state of the reaction, where the proton has been transferred from lysine to the diolate and the diol is completely formed. Combinations of these resonance states can be used to perform EVB simulations of the generation of the diol.

The resonance structures in Figure 6.10 can be used to simulate the reaction that leads to the diol. It is possible to simulate a stepwise mechanism by following the reaction path from structure I via structure II and III to structure IV. In addition the reaction could be simulated in

a concerted manner *e. g.* by applying a transformation from state I to state III and then to state IV. This procedure is comparable to the application on acetylcholinesterase in section 4.3.2

For each step the EVB parameters α and and the parameters for the off-diagonal function H_{ij} (see section 3.3 for an explanation) have to be adjusted such that the free reaction energy and the activation barrier of a chosen reference reaction, for which experimental or theoretical data are available, are reproduced.

The free energy of the proton transfer reaction in aqueous solution can be deduced from the pk_a values: For the diol one could use the pk_a-value of water which is 15.7 The pk_a-value of Lysine is 10.4. With the usage of Eq. 4.1 the free reaction energy of the proton transfer reaction in solution would amount to Δ G = 7.3 kcal/mol. The activation barrier is deduced corresponding to the description in section 4.3.2 to $\Delta G^{\ddagger} = 11.1$ kcal/mol.

The nucleophilic attack of the hydroxide ion on the carbonyl carbon of the aldehyde should also be parameterized by experimental values of this reaction preferably in aqueous solution. But such values could not be found in the literature. Therefore, the parameterization has to be done by a theoretical approach. A suggestion will be made in the next section.

The last step is another proton transfer from lysine to the diolate anion. The energetics of this step can again be deduced via the pK_a values. The pK_a value of acetaldehyde hydrate, which can be used as a model, is 13.6.[166] Together with the pK_a value of 10.4 for lysine this leads via Eq. 4.1 to a free reaction energy of ΔG = -4.4 kcal/mol. The barrier of this downhill reaction should can be approximated via the rate of diffusion controlled reaction ($k = 10^{10} M^{-1} s^{-1}$) and is thus 2.5 kcal/mol.

The schematic picture of the free energies of the various steps in the diol formation reaction are shown in Figure 6.11. This picture lacks the energetics of the nucleophilic attack step, for which no experimental data are available and for which a theoretical approach is discussed in the following section.

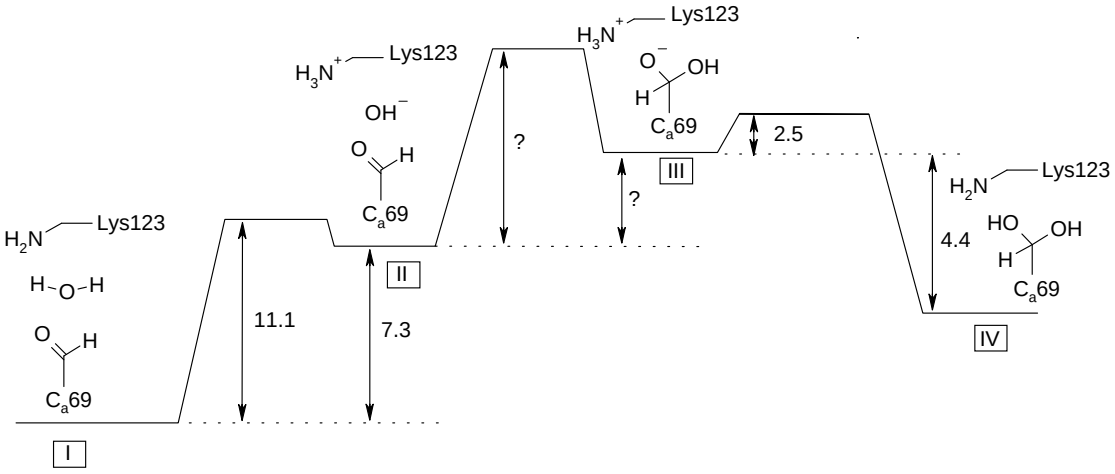


Figure 6.11.: Energy Scheme of the reference reaction for the diol formation. All values are in kcal/mol. The deduction of all values is given in the text. The energetics of the nucleophilic attack reaction is not deducible from experimental values, which is expressed by question marks in this picture.

6.2.3. Ab initio reference reactions

The catalytic efficiency of enzymes can only be determined by comparison with a proper reference reaction in solution. That is, because the definition of catalysis is based on this comparison. In addition, the EVB model relies on the energetics of well defined reference reactions to calibrate the parameters of this method.

As could be seen in the last section, there are reactions or single steps of reactions for which no experimental data of the free reaction energy or the activation barrier are available. In such cases a theoretical ab initio model could provide a way to establish a reference reaction. A recent paper by Štrajbl, Florián and Warshel introduced a method, that showed how *ab initio* calculations in gas phase can be used as a first step to evaluate the energetics of reactions in aqueous solution.[173] In their study they determined the gas phase potential surface of a model reaction for the acylation step in serine proteases. The solvation free energies of all geometries along the reaction path were then calculated by a langevin dipole (LD)[59] solvation model together with a polarized continuum solvation model (PCM)[174,175] implemented in the software package CHEMSOL.[176] This calculation of solvation energies together with the gas phase potential surface yields the free energy surface of the reaction in solution. The reaction surface in solution can differ considerably from the gas phase surface due to the different interaction of charges in the aqueous medium. Due to the large solvation energies a reliable and elaborate method has to be applied to calculate the solvation energy of the different geometries. Errors in the calculation of solvation energies can lead to misleading assumptions on the solution energy surface. The solvation model in CHEMSOL has been applied to several systems before.[177,178,179,180]

The procedure applied by Štrajbl et al.[173] could be used to establish the energetics of the nucleophilic attack of the OH^- atom on the carbonyl carbon of acetaldehyde that is missing in the scheme of the reference reaction for the diol formation in section 6.2.2. With the complete reference reaction in solution for all steps of the diol formation the EVB parameters could be determined and the suggested mechanism could be investigated.