Julius-Maximilians-Universität Würzburg

Fakultät für Chemie und Pharmazie

Computational study on the catalytic mechanism of *mt*KasA

Dissertation zur Erlangung des naturwissenschaftlichen Doktorgrads
der Julius-Maximilians-Universität Würzburg

vorgelegt von

Wook Lee

aus Seoul, Korea

Würzburg 2013

Eingereicht am: _____

bei der Fakultät für Chemie und Phamrazie.

1. Gutachter: _____

2. Gutachter: _____

der Dissertation

1. Prüfer: _____

2. Prüfer: _____

3. Prüfer: _____

des öffentlichen Promotionskolloquiums.

Tag des öffentlichen Promotionskolloquiums: _____

Doktorurkunde ausgehändigt am: _____

Die vorliegende Arbeit wurde unter Anleitung von Prof. Dr. Bernd Engels von September 2008 bis Oktober 2013 am Institut für Physikalische und Theoretische Chemie der Julius-Maximilians-Universität Würzburg angefertigt.

Teilergebnisse dieser Arbeit waren Gegenstand von Publikationen sowie von Postern und Kurzvorträgen.

Lee, W.; Luckner, S. R.; Kisker, C.; Tonge, P. J.; Engels, B. Elucidation of the Protonation States of the Catalytic Residues in mtKasA: Implications for Inhibitor Design. *Biochemistry* **2011**, *50*, 5743–5756.

Lee, W.; Engels, B. Clarification on the Decarboxylation Mechanism in KasA Based on the Protonation State of Key Residues in the Acyl-Enzyme State. *J. Phys. Chem. B* **2013**, *117*, 8095–8104.

Lee, W.; Engels, B. The Protonation State of Catalytic Residues in the Resting State of KasA Revisited: Detailed Mechanism for the Activation of KasA by Its Own Substrate. Submitted to *Biochemistry* **2013**

Hiermit erklare ich an Eides statt, dass ich die Dissertation

“Computational study on the catalytic mechanism of *mt*KasA”

selbststandig angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt habe.

Ich erklare außerdem, dass diese Dissertation weder in gleicher oder anderer Form bereits in einem anderen Prufungsverfahren vorgelegen hat.

Ich habe fruher außer den mit dem Zulassungsgesuch urkundlich vorgelegten Graden keine weiteren akademischen Grade erworben oder zu erwerben versucht.

Würzburg, 24. 09. 2013

Wook Lee

Contents

Chapter 1 Introduction

Tuberculosis (TB) is an infectious disease which is known to be caused by *Mycobacterium tuberculosis*. It is spread through air when people who are infected with TB expel the bacteria by coughing or sneezing. TB typically attacks lungs (pulmonary TB), but it can also affect other sites of the body as well (extrapulmonary TB). Since World Health Organization (WHO) declared TB as a global health emergency in 1993, it has remained a significant threat to global health. In that year, it was estimated that 7-8 million incident cases and 1.3-1.6 million deaths occurred each year. In 2010, there were 8.5-9.2 million incident cases and 1.2-1.5 million deaths (including deaths from TB among HIV-positive patients) estimated, and this made TB become the second leading cause of death from an infectious disease globally. The most of those incident cases in 2010 took place in Asia (59%) and Africa (26%), and a small portion of the cases occurred in the Eastern Mediterranean Region (7%), the European Region (5%), and the Region of the Americas (3%) (Figure 1.1).[1]

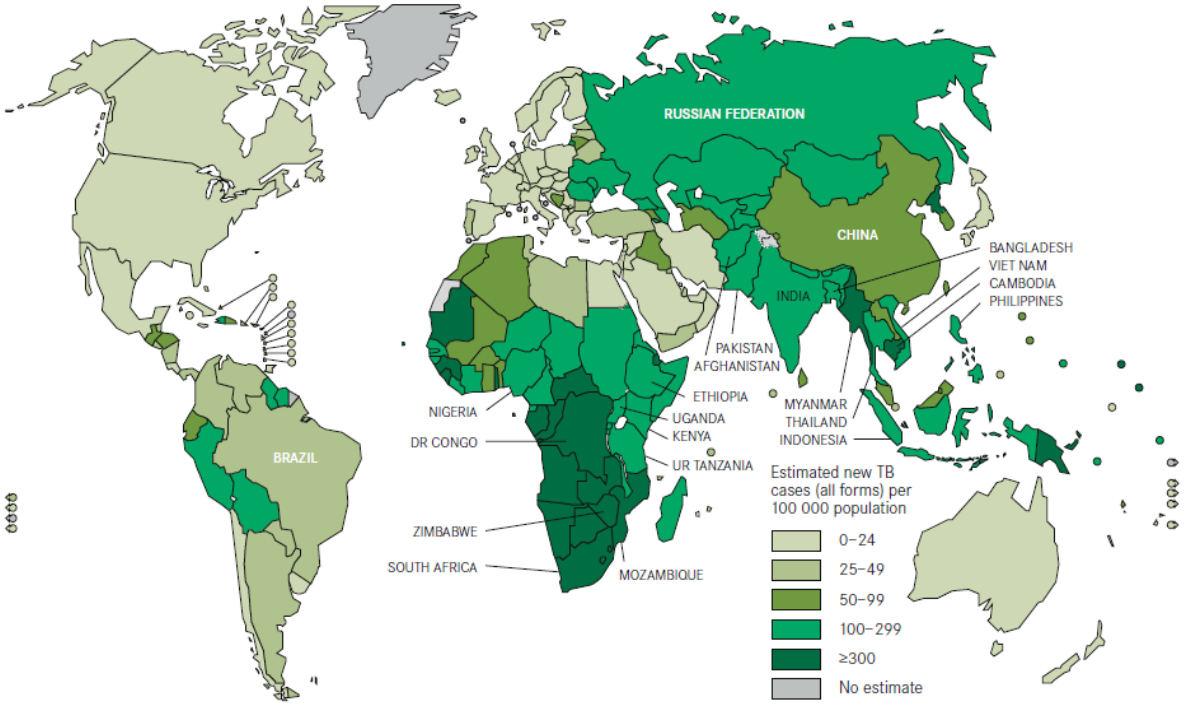


Figure 1.1. Estimated TB incidents rate in 2010. Reprinted with permission from ref.[1] Copyright 2011 World Health Organization

To make matters worse, the emergence of the strains which acquire drug resistance made TB a more intractable disease. The origin of these drug resistant strains is mainly attributed to the treatment difficulties such as noncompliance and administration of inadequate treatment regimes rather than the advent of new resistant mechanisms. The molecular mechanism of the drug resistance is that, when *Mycobacterium tuberculosis* is under the pressure of inadequate drug treatment, it accumulates mutations in the way of overproducing or altering the drug target protein, thereby attenuating the effect of antitubercular drugs.[2]

Drug-resistant TB can be categorized into two groups according to the number and the type of drugs against which they have resistance. One is multidrug-resistant tuberculosis (MDR-TB) and the other is extensively drug-resistant tuberculosis (XDR-TB). MDR-TB can be defined as the TB acquired resistance to the two most commonly used front-line antituberculosis agents, isoniazid and rifampicin. The treatment of MDR-TB is far from trivial since it requires intensive resources. Especially in the regions where the prevalence of MDR-TB is high, the recommended therapeutic strategies consists of the combination of second-line drugs which are more expensive, more toxic and less effective than the drugs

used in standard therapy.[3–6] The emergence of XDR-TB was first reported in 2005.[7] The strains belonging to XDR-TB were discovered during the treatment of MDR-TB and found to be resistant to virtually all of the second-line drugs. The term XDR-TB was first used in the joint report by US Centers for Disease Control and Prevention (CDC) and WHO.[8] According to the definition used in this report, XDR-TB is the strains of *Mycobacterium tuberculosis* which are resistant to isoniazid and rifampicin like MDR-TB as well as at least three of the six classes of second line antituberculosis agents, which are aminoglycosides, polypeptides, fluoroquinolones, thioamides, cycloserine and para-aminosalicylic acid.

Another great concern about TB is its co-epidemics with HIV. According to the WHO report, if HIV positive patients are also infected with TB, then the chance to develop TB among them are about 21-34 times higher than those who are HIV negative. The fact that TB incidence is decreasing or stable worldwide except for Africa where TB incidence is increasing with high prevalence of HIV in that region mirrors this aspect very well.[9] The HIV epidemic had been a main cause of the upsurge in TB cases and mortality in many countries since 1980s, and it persisted throughout 1990s up to around 2004, especially in southern and east Africa. In 2010, 33% of TB patients are diagnosed as HIV positive.[1] Since this percentage is only among the TB patients who know their HIV status, the actual number of HIV positive cases among TB patients would be larger than it is known.

Due to the advent of XDR-TB and its association with HIV mentioned above, TB has emerged again as a great threat to the public health. Therefore, it is now necessary to identify new drug targets and their inhibitors to treat drug resistant TB. In this regard, understanding the catalytic mechanism of possible target enzymes at the atomistic level is of significant importance because it can provide us with valuable insights into the design of new inhibitors as pharmaceutical leads. However, determining the correct catalytic mechanism of enzymes is far from trivial since some important information which is essential for establishing the catalytic mechanism of enzymes, for example, the structure of a transition state or reaction intermediate, is not accessible through experiments. At this point, computational modeling can play a significant role to complement the weakness of experimental techniques. For the last several decades, the techniques for modeling enzymes have advanced rapidly together with the developments in both computer hardware and the theories in the field of computational chemistry. As a result, the contributions from molecular simulations to the understanding of the catalytic mechanism of enzymes have increased simultaneously. The

increasing number of computational studies concerning enzymatic reaction mechanisms in recent years reflects this aspect very well.

In the context mentioned in the above paragraph, this thesis presents a computational study to elucidate the catalytic mechanism of beta-ketoacyl ACP synthase I (KasA) from *Mycobacterium tuberculosis*, which is a promising drug target against tuberculosis. Since the catalytic mechanism of KasA is closely related to the protonation state of catalytic residues in each reaction step, the possible protonation states in the resting state and acyl-enzyme state are mainly investigated in this thesis. Molecular dynamics (MD) simulations with a force field were used for the comparison of geometrical arrangement of the active site between the X-ray crystal structures and model systems. For simulations including chemical reactions where electronic rearrangement involved in bond forming or breaking has to be modeled, the combined quantum mechanical / molecular mechanical (QM/MM) scheme was used. This scheme is very useful especially for the modeling of enzymes because it enables us to describe chemical reactions occurring in the active site with QM as well as the effects from the enzyme environment with MM at the same time, thereby providing computational efficiency with reasonable accuracy. The QM/MM scheme was applied to geometry optimizations and MD simulations depending on the purpose of computations. In addition, free energy calculation methods such as umbrella sampling and the free energy perturbation method were employed to compute the free energy profiles for chemical reactions and the perturbed pKa values of catalytic residues in the enzyme environment.

Chapter 2 Background

2.1 Beta-ketoacyl ACP synthase I (KasA)

2.1.1 Fatty acid synthesis (FAS) system in mycobacteria

Fatty acid synthesis is an essential process for the survival of microbes since fatty acids are a major component of the cell wall. Mycolic acids are very long-chain fatty acids that constitute the cell wall in mycobacteria. The main components of mycolic acids are a meromycolate chain, which is up to 56 carbon units (C_{56}), and a long saturated α-branch, which is between 24 (C_{24}) and 26 (C_{26}) carbon units. These two components can be distinguished from each other by the chemical modifications found in the meromycolate.[10]

Mycolic acids have many important characteristics such as resistance to chemical injuries,[11] resistance to dehydration,[12] biofilm formation,[13] acid-fast staining,[14] the ability to

persist within the host,[12,14–16] and the low permeability to hydrophobic antibiotics,[17] which contributes to the intrinsic drug resistance of the pathogen. From those characteristics and its role as a major component of the cell wall, it can be easily inferred that the synthesis of the mycolic acids is an indispensable part for the living of mycobacteria. The experimental result showing that the depletion of the enzymes involved in the biosynthetic pathway of mycolic acids leads to cell death clearly indicates the importance of this process in the survival of mycobacteria.[18]

The biosynthetic pathway of mycolic acid consists of two systems, namely FAS I and FAS II (Figure 2.1). FAS I is similar to the system found in mammals and consists of a single gene referred to as *fas* gene.[19] This gene encodes a polypeptide containing all requiring enzymatic activities and carrier functions, which are acyltransferase, enoyl reductase, dehydratase, malonyl/palmitoyl transferase, acyl carrier protein, β-ketoacyl reductase and β-ketoacyl synthase. During the elongation process of FAS I, all intermediates produced in each step keep bound to the enzyme and are transferred from one active site to another within the enzyme. FAS I conducts *de novo* synthesis of fatty acids and produces two types of fatty acids.[19] One is C_{16} fatty acid, which will be eventually handed over to FAS II system by CoA and used as a substrate for further elongation to generate meromycolic acid. The other is C_{26} fatty acid, and it is used as an α-branch of the final mycolic acid.[10]

On the other hand, FAS II is analogous to the system found in other bacteria, plants, and parasites. It has separate genes encoding separate enzymes each of which plays a distinct role in the pathway of elongating the growing acyl chain.[20] Due to this feature, FAS II is often considered as less efficient than FAS I because, unlike FAS II, the intermediates generated from each step do not need to diffuse to another enzyme in FAS I. However, while FAS I usually produces only palmitate, FAS II can generate various types of fatty acids which are different in length or degree of saturation. Additionally, the intermediates generated in FAS II also can be used as precursors for further synthesis of important cellular components such as lipoic acid and quorum-sensing molecules. Those intermediates are not transferred by itself but attached via a thioester linkage to the prosthetic group of an acyl carrier protein (ACP) and are transferred from one enzyme to another by this carrier protein.[21]

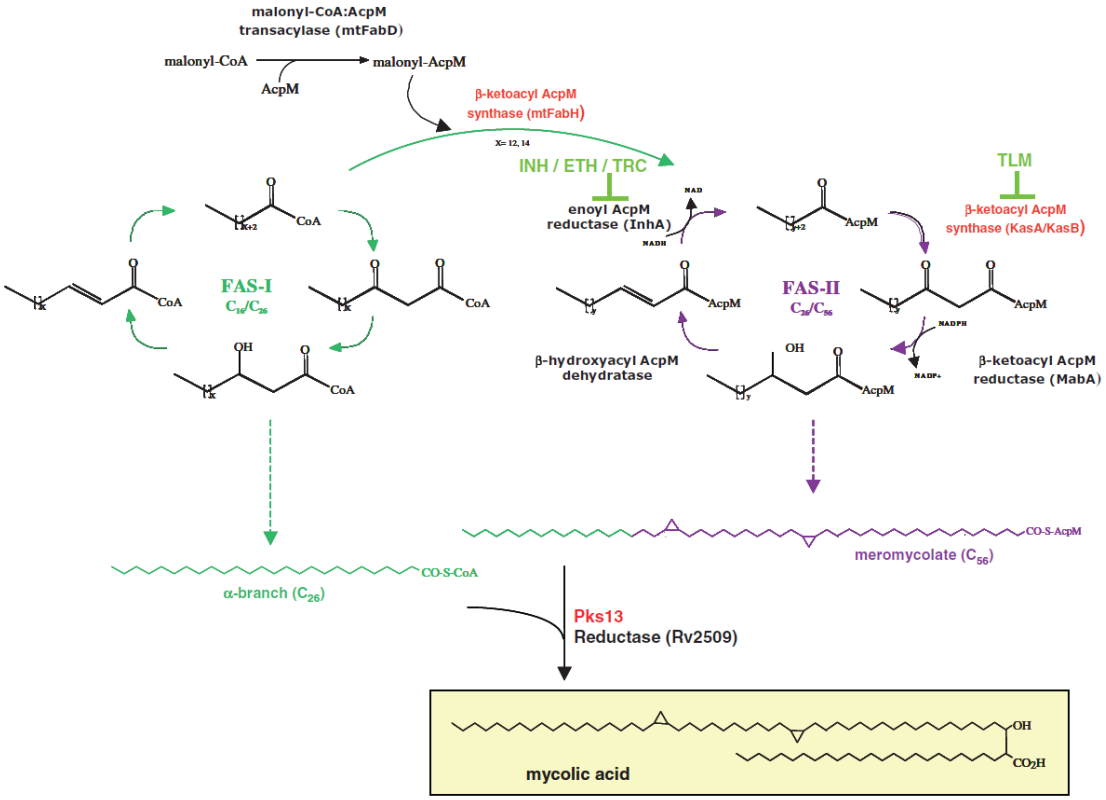


Figure 2.1. Biosynthetic pathway of mycolic acid. The enzymes belonging to FAS II are indicated in black except the condensing enzymes, which are represented in red. The fatty acids are colored according to the contribution from FAS I or FAS II. The green indicates the contribution from FAS I, and the purple indicates the contribution from FAS II. Several inhibitors, isoniazid (INH), ethionamide (ETH), triclosan (TRC), and thiolactomycin (TLM) are also represented with the targets they act on. Reprinted with permission from ref.[10] Copyright 2007, John Wiley and Sons

Unlike FAS I, FAS II is not able to perform *de novo* synthesis of fatty acids but accepts acyl-CoA generated by FAS I and elongate it further. Since ACP plays a role as a carrier protein in FAS II, the acyl-CoA from FAS I should first undergo the conversion from acyl-CoA to acyl-ACP, and mtFabH has been proposed to perform this task.[22] Then, acyl-ACP undergoes a cycle of keto reduction by keto-reductase MabA,[23] dehydration by unidentified dehydratase, and enoyl reduction by enoyl-reductase InhA.[24] The product from this cycle now undergoes several iterative cycles where the reaction by mtFabH is replaced by the condensing reaction by KasA and KasB.[25] Both KasA and KasB catalyze condensing reaction and show substantial identities in sequence and structure.[26] However, KasA is

assumed to catalyze the initial elongation reaction while KasB catalyzes the elongation reaction to full length mycolates.[27] The iteration cycle is repeated until meromycolate is synthesized. Subsequently, merocycolate is condensed with the saturated straight fatty acid chain from FAS I,[28] the precursor of the α-branch, by the termination condensing enzyme, polyketide synthase Pks13.[29–31] The product from this reaction further undergoes the final reduction step to produce mycolic acid.[32]

2.1.2 Catalytic mechanism of KasA

The condensation reaction catalyzed by KasA consists of three steps as shown in Figure 2.2. The first step is the acyl-transfer step where an acyl group is delivered by ACP to the active site and covalently transferred to the catalytic cysteine, forming an acyl-enzyme intermediate. Then, the second substate, malonyl-ACP, binds to the active site and the decarboxylation takes place, generating a carbanion. In the last step, the carbanion formed from the decarboxylation step attacks the carbonyl group of the acyl-enzyme intermediate, and the elongated acyl chain with two additional carbons is released.

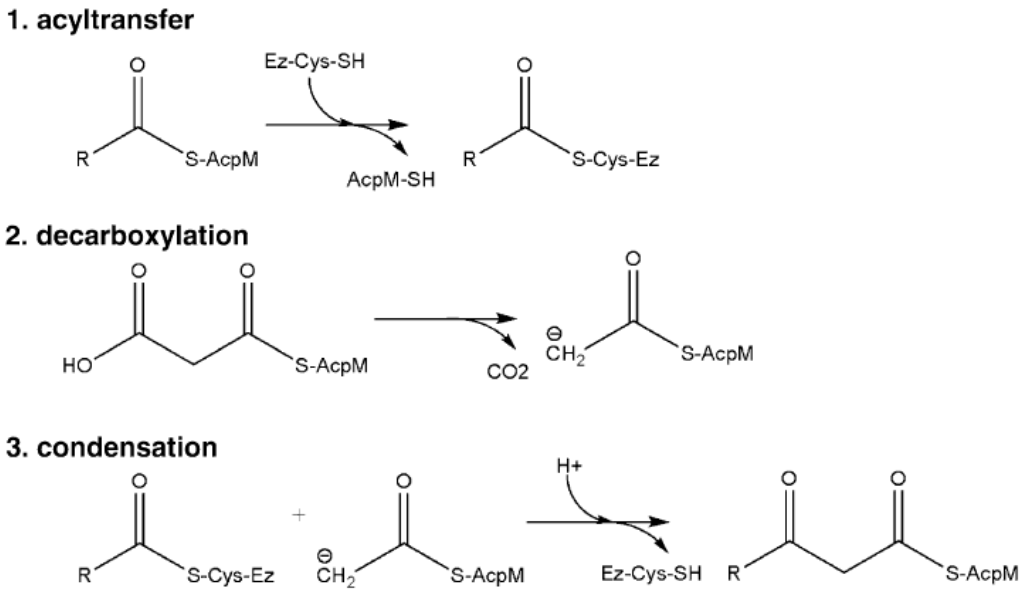


Figure 2.2. Schematic drawing of three catalytic steps in condensation reaction catalyzed by KasA. Reprinted with permission from ref.[10] Copyright 2007, John Wiley and Sons

In KasA, Cys171, His311, and His345 have been considered as catalytic triad (Figure 2.3). These residues were identified by the X-ray crystal structure of *Ec*FabB complexed with

inhibitors,[33] and it was shown that the mutations of these residues were critical to the overall forward reaction.[34] However, the functions of those residues in each step of the reaction are not completely understood. Additionally, two ordered water molecules are also found in the vicinity of the triad in most of the X-ray crystal structure of KasA and assumed to play an important role in the catalysis due to the positions where they are found. One water molecule (Wat1) is positioned between Cys171 and His311 and bridges these two residues. It was also suggested by Olsen et al.[35] that His311 acts as a base to abstract the proton to activate Cys171 at the beginning of the catalysis via this water molecule. The other water molecule (Wat2) is located in the center of the hydrogen bond network beside the active site, but the function of this hydrogen bond network is still not known.

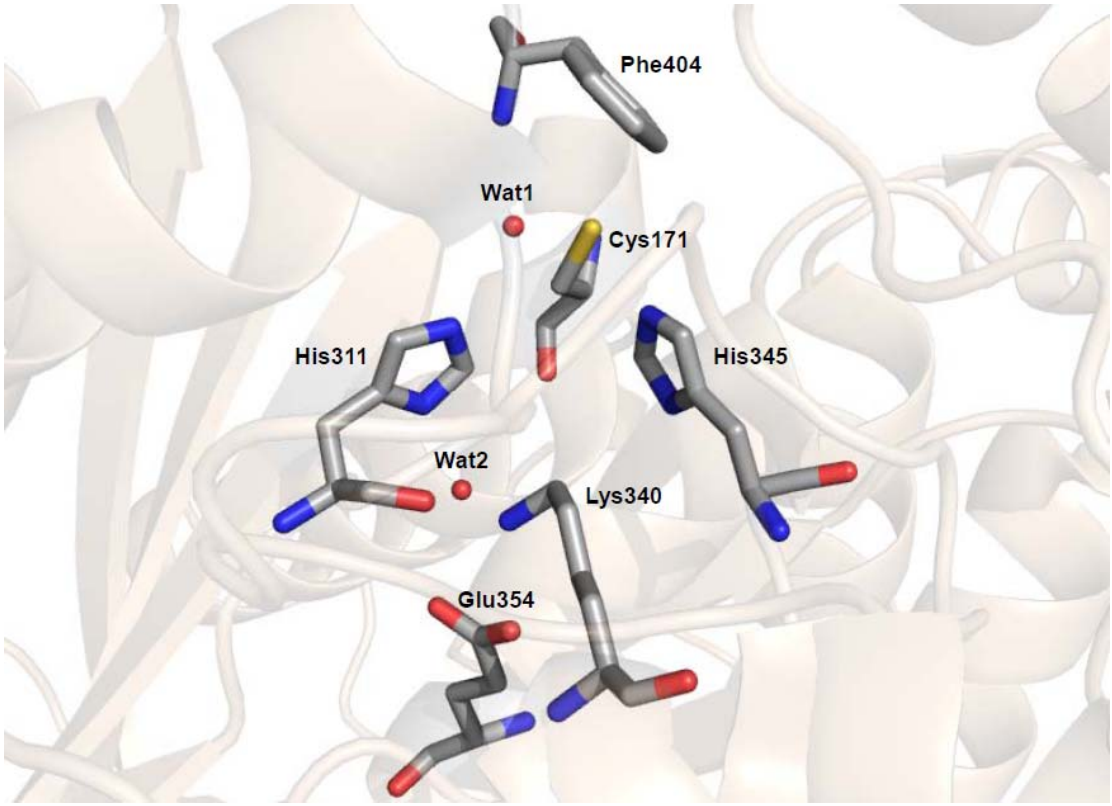


Figure 2.3. Active site of wild-type KasA. (Protein Data Bank (PDB) entry: 2WGD) The important conserved residues and ordered water molecules are shown as a ball and stick model, and the protein environment is represented as a cartoon model.

There are also several other conserved residues which are expected to have significant contributions to the catalysis. Lys340 and Glu354 are such a case since they not

only form an ionic bond just beside the catalytic site but also participate in a hydrogen bond network including one of the catalytic histidines with the ordered water molecule (Wat2) in its center. Furthermore, some biochemical studies showed that the mutation of the conserved lysine residue to alanine gave rise to severe impairment in catalytic activity of the enzyme.[34,36] The X-ray crystal structure of this mutant (PDB entry: 1OEO) showed that the absence of this lysine residue results in the rotation of one of the catalytic histidines, and therefore this lysine plays a structural role to maintain the structure of the catalytic site.[37] Nevertheless, it is still not clearly explained whether and how they explicitly contribute to the catalysis.

Another important conserved residue is Phe404. The role of this residue was first proposed by Huang et al. based on the X-ray crystal structure of KasII,[38] and they suggested that this residue would act as a gatekeeper. The aromatic side chain of this conserved residue points into the active site and restricts the access of the substrate to the catalytic cysteine by partially blocking it. Therefore, the conformational change of the aromatic side chain of Phe404 seems to be essential for the substrate to gain full access to the catalytic cysteine. The conformational flexibility given by three neighboring conserved glycine residues further underpins this assumption. Furthermore, this residue was also assumed to play a role in the regulation of the order of the catalytic reaction. It was suggested that, after the acyl transfer step is complete, the side chain of Phe404 is rotated to become an open conformation so that the entrance of the binding pocket can be widened, and this eventually leads to the binding of the second substrate much easier. This assumption was proposed by Luckner et al. on the basis of the comparison of X-ray crystal structures between the wild-type and Cys171Gln mutant of KasA and their complexes with TLM inhibitor (PDB entries: 2WGD, 2WGE, 2WGF, and 2WGG).[39] In that study, The Cys171Gln mutant was considered as identical to the acyl-enzyme state since it was reported that the mutation of the catalytic cysteine into glutamine can mimic the structural change caused by the binding of the acyl group.[40] Based on this assumption, the X-ray crystal structure was examined, and the formation of the hydrogen bond between the backbone amide and the carbonyl oxygen from the acyl group was suggested as a driving force to make Phe404 become an open conformation.

In order for the condensation reaction to be initiated, the very first step to be carried out is the nucleophilic activation of the catalytic cysteine, Cys171. There have been several suggestions about how this step takes place. One suggestion is that one of the catalytic

histidines can abstract a proton from the cysteine.[38,41] However, His345 is not suitable for this task because N_δ of His345 accepts a hydrogen bond from the backbone amide group of Ile347, so N_ε of His345, which is located in the side of Cys171, is not able to abstract a proton. Consequently, the other catalytic histidine, His311, is suggested as a base to accept a proton from Cys171. However, several mutagenensis experiments have demonstrated that neither catalytic histidine is essential for the acyl-transfer activity.[36,42] Therefore, a new idea was necessary to explain the nucleophilic activation of the catalytic cysteine, and the effect of the macro helix dipole was suggested as a source to activate the catalytic cysteine since this cysteine is located at the N terminus of the α-helix.[41] However, this explanation is still controversial because one theoretical study demonstrated that the zwitterionic state of catalytic Cys-His dyad in cathepsin B is stabilized only partially by the helix dipole moment but mostly by the complex hydrogen bond network with neighboring residues.[43]

Another step, which is also on a debate, is the decarboxylation reaction. The reason for the difficulty to reach the consensus in this case is mainly due to the protonation state of His311. As it is already mentioned above, the protonation state of His345 is obvious because it receives a hydrogen bond from the backbone amide of Ile347. Unlike His345, His311 is not so unambiguous concerning its protonation state. Furthermore, it is involved in the hydrogen bond network whose function is still unknown. Despite of these difficulties, there have been several mechanisms suggested so far, and it can be narrowed down to three as shown in Figure 2.4.

The first mechanism (Figure 2.4 A) was suggested by Olsen *et al.*[35] From the X-ray crystal structure of *Ec*FabB (PDB entry: 1EK4), they observed electron density that can be successfully modeled as two oxygen atoms in very short distance. They interpreted one of these oxygen atoms that is found in the side of the N_δ atom of His298 (His311 in KasA) as a hydroxide ion and insisted that His298 should be protonated in N_δ in order to donate a hydrogen bond to this hydroxide ion. They also assumed that the malonyl residue of the substrate is likely to exist as a malonic acid and suggested that the decarboxylation reaction would start by the abstraction of a proton from the malonic acid by His298.

The second mechanism (Figure 2.4 B) was first suggested by Witkowski *et al.*[44] They came up with this idea based on the measurement of the reaction products from the decarboxylation reaction by the Cys161Gln (Cys171Gln in KasA) mutant of mammalian fatty acid synthase (FAS) I. In that measurement, they found that the bicarbonate is released

as a final product instead of carbon dioxide. Based on this observation, they argued that N_ε of His293 (His311 in KasA) would abstract a proton from an ordered water molecule and produce an hydroxide ion which subsequently attacks C3 in malonyl-ACP to release bicarbonate. This mechanism was further adopted by the Rock group[21,36] and applied to the case of bacterial type FAS. They argued that the lone pair electrons should normally reside on N_ε of His303 (His311 in KasA) because His303 has an interaction with the Lys/Glu ionic pair in a similar way observed in other enzymes such as serine proteases. Furthermore, they found that there is an appropriately positioned ordered water molecule in the vicinity of the N_ε atom of His303 in *Sp*FabF (His311 in KasA) from the X-ray crystal structure (PDB entry: 1OX0) and insisted that this finding of the ordered water molecule in this position further support this idea. However, this mechanism is not compatible with the X-ray crystal structures of the enzyme complexed with cerulenin and thiolactomycin (PDB entries: 1FJ4 and 1FJ8) because it was observed that the carbonyl oxygen of the inhibitor accepts a hydrogen bond from N_ε of His298 (His311 in KasA) in both structures. They suggested the repositioning of His298 upon binding of the inhibitor and expect that this repositioning would cause the movement of the proton from N_δ to N_ε of His298.

The last mechanism (Figure 2.4 C) proposed by Price *et al.*[33] is based on the X-ray crystal structures of *Ec*FabB-inhibitor complexes with cerulenin and thiolactomycin. In both X-ray crystal structures, they observed that two catalytic histidines donate a hydrogen bond to the carbonyl oxygen of the inhibitor and expected that this mimics the condensation transition state. From this interpretation, they suggested the decarboxylation mechanism where two catalytic histidines stabilize the negative charge developed in the carbonyl oxygen of the inhibitor during the condensation transition state by forming a second oxyanion hole. They also suggested that the decarboxylation reaction is initiated by the movement of the electrons away from the carboxyl group of the inhibitor promoted by the interaction with the conserved phenylalanine residue.

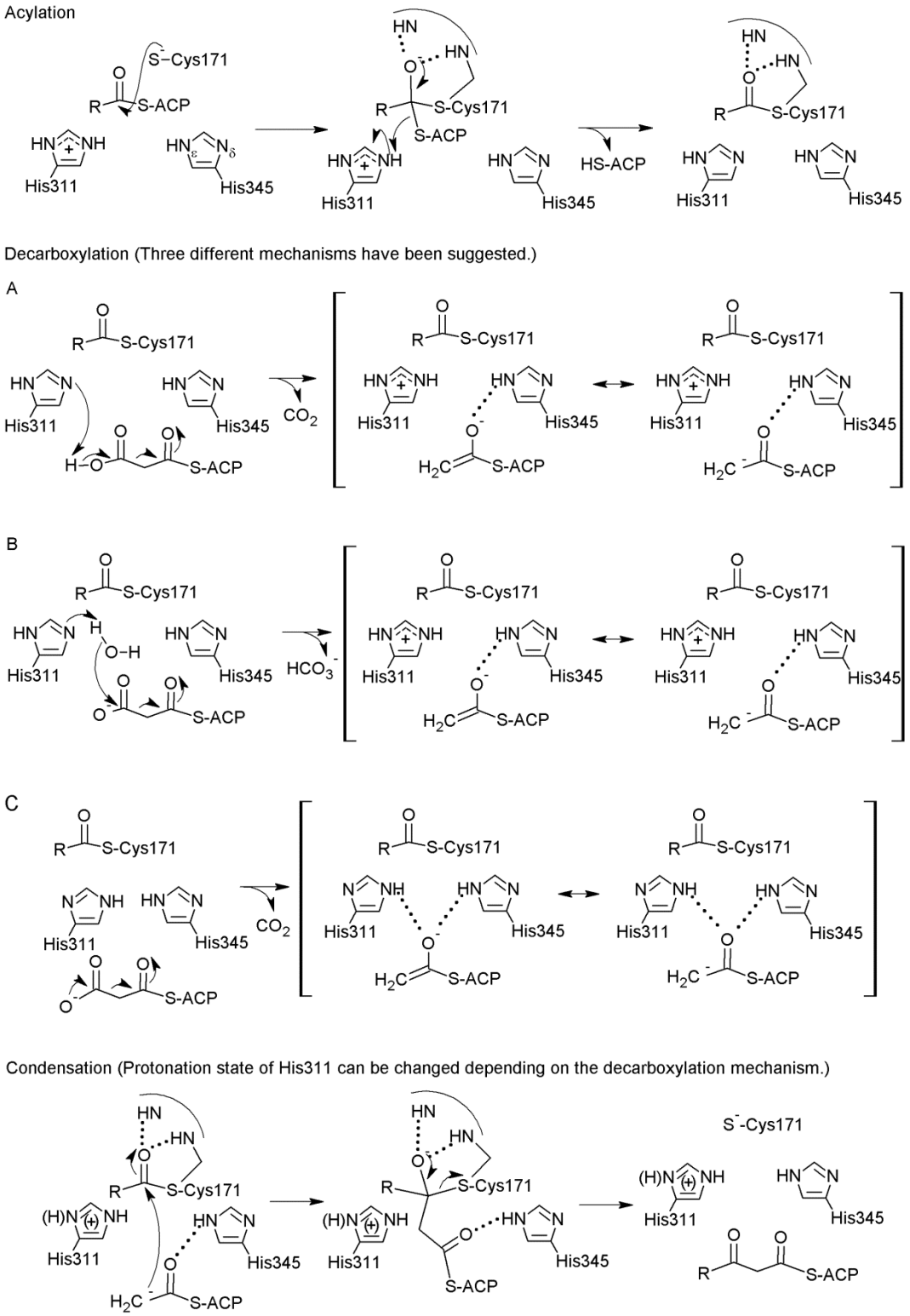


Figure 2.4. Suggested mechanisms for catalytic reaction of KasA. The types of nitrogen according to its position in histidine are only marked in the first drawing and omitted in the rest for clarity.

2.2 Quantum mechanical (QM) methods

This chapter mainly refers to ref.[45], ref.[46], and ref.[47]

2.2.1 Density functional theory (DFT)

2.2.1.1 Motivation and early attempt

Although we can access any information of a given system through a wave function, it is a very complicated quantity that cannot be investigated experimentally. Furthermore, since wave function based methods depends on 4N variables (3N spatial coordinates and N spin coordinates), the cost of calculations becomes formidable when it comes to the system with a large size. With regard to this fact, the question was raised, whether a wave function contains any redundant and irrelevant information for obtaining energy and other properties, because the Hamiltonian operator only contains operators which act on one or at most two particles at a time. As a result, there had been attempts to find a new quantity that can replace the wave function, and this endeavor led to the birth of DFT.

Unlike wave function based methods, DFT uses electron density as a central quantity, which gives the total number of electrons when it is integrated over all space.

$$N = \int \rho(\mathbf{r}) d\mathbf{r} \tag{2.1}$$

In DFT, operators are supposed to act on the electron density, and density itself is also a function of three-dimensional coordinates. This is why it is called density functional theory since functional refers to the function which also has a function as an argument. The first attempt to employ electron density instead of wave function for obtaining information was done by Thomas and Fermi in 1927.[48] They used fermion statistical mechanics to derive the kinetic energy from electron density and treated the nuclear-electron and electron-electron interaction in a completely classical manner.

$$E_{TF}[\rho(\vec{r})] = \frac{3}{10}(3\pi^2)^{\frac{2}{3}} \int \rho^{\frac{5}{3}}(\vec{r}) d\vec{r} - Z \int \frac{\rho(\vec{r})}{r} d\vec{r} + \frac{1}{2} \iint \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{r_{12}} d\vec{r}_1 \, d\vec{r}_2 \tag{2.2}$$

In this way, they succeeded to formulate the energy equation completely in terms of electron density with no reference to a wave function. However, this equation gives very poor results where all molecules are unstable with respect to the dissociation into their constituted atoms, so it is not possible to find any use of this equation in modern chemistry.

2.2.1.2 Kohn-Sham (KS) approach

In 1964, Hohenberg and Kohn proved two theorems, and these established a foothold for DFT to become a legitimate quantum chemical methodology.[49] The first theorem is an existence theorem which proves that ground state electron density uniquely determines the Hamiltonian operator and thus all properties of the system. Due to this first theorem, the key idea of DFT where electron density is sufficient to obtain all properties of interest could now be based on a rigorous physical foundation. However, the first theorem doesn't give any clue how we can obtain the ground state density. The second theorem gives a hint regarding this problem, that is, the only functional which has true ground state density delivers the lowest energy.

Although the theorems mentioned above provided a solid foundation for DFT, yet nothing was provided about how the functional that delivers the ground state energy should be constructed. In 1965, Kohn and Sham made a breakthrough by suggesting a Hamiltonian operator for the fictitious non-interacting system of electrons which consists of one electron operators only.[50]

$$\hat{f}_i^{KS}\varphi_i = \varepsilon_i\varphi_i \tag{2.3}$$

$$\hat{f}_i^{KS} = -\frac{1}{2}\nabla_i^2 - \sum_k^{nuclei}\frac{Z_k}{|r_i - r_k|} + \int\frac{\rho(r')}{|r_i - r'|}dr' + V_{XC} \tag{2.4}$$

The operator is called Kohn-Sham (KS) operator, and they also employed orbitals which are associated with this operator. To distinguish these orbitals from their counterparts in the Hartree-Fock theory, they are usually termed KS orbitals. The important property of the KS orbitals, although they were devised for a fictitious non-interacting system, is that the density resulting from the summation of their squares equals to the exact density if the exact functionals are employed. Thanks to the KS orbitals, the kinetic energy term could be now expressed as

$$T = -\frac{1}{2}\sum_i^N\langle\varphi_i|\nabla_i^2|\varphi_i\rangle \tag{2.5}$$

Since the reason of very poor performance of Thomas-Fermi model was mainly attributed to its kinetic energy term which is an explicit functional of electron density, $\int\rho^{\frac{5}{3}}(\vec{r})d\vec{r}$, the fatal defect observed in Thomas-Fermi model is not appeared in the KS approach. Furthermore, though the kinetic energy derived in the KS approach is based on a non-interacting system, it

still gives reasonable result because the kinetic energy computed from a non-interacting system comprises the large fraction of the kinetic energy of the actual system.

2.2.1.3 Exchange-correlation energy functional

Using the KS approach, the whole energy expression now can be represented as

$$E[\rho(r)] = \sum_{i}^{N}\left(\left\langle\varphi_i\left|-\frac{1}{2}\nabla_i^2\right|\varphi_i\right\rangle - \left\langle\varphi_i\left|\sum_{k}^{nuclei}\frac{Z_k}{|r_i - r_k|}\right|\varphi_i\right\rangle\right) + \sum_{i}^{N}\left\langle\varphi_i\left|\frac{1}{2}\int\frac{\rho(r')}{|r_i - r'|}dr'\right|\varphi_i\right\rangle + E_{xc}[\rho(r)] \quad (2.6)$$

It should be noted that the electron density $\rho(r)$ also appears in the right hand side of the equation, so this equation should be solved iteratively like the Hartree-Fock equation. The terms in right hand side represent, respectively, kinetic energy for a non-interacting system, the nuclear-electron interaction, the classical electron-electron repulsion, and exchange-correlation energy (E_{xc}) which is the functional containing everything that is unknown, namely, the correction for the kinetic energy (the kinetic energy difference between the non-interacting system and the real one), exchange energy, correlation energy, and the correction for the self-interaction error. The self interaction error in the KS approach arises because the classical electrostatic repulsion term in the KS operator contains the spurious interaction of the density with itself. This error is exactly cancelled out by the exchange term in the case of the Hartree-Fock scheme. However, there is no such term in the KS approach, so the correction for this error should be explicitly considered in the exchange-correlation energy functional.

The explicit form of the exchange-correlation energy functional is still unknown. However, if we could obtain the exact functional, the KS approach would provide the exact energy since it is in principle exact. Unlike the Hartree-Fock theory, the approximation was only adopted when the form of the exchange-correlation energy functional has to be decided in the KS scheme. Therefore, finding a better approximation of this functional is a central goal of modern density functional theory.

2.2.2 Semi-empirical methods

2.2.2.1 Introduction

Applying an *ab initio* method such as the Hartree-Fock theory to a large system is normally a challenging task. This is mainly due to the computations of two-electron integrals which scales fourth power of the number of basis functions. As a result, the target of *ab initio* computations had been restricted to the system of few atoms. However, the growing demands to apply molecular orbital theory to a large system which is inaccessible by *ab initio* methods led to the development of semi-empirical approaches.

Semi-empirical methods adopt many approximations in order to reduce the cost of the computations, and the most fundamental approximation on which all semi-empirical methods are grounded is the zero differential overlap (ZDO) approximation. The ZDO approximation neglects all products of basis functions depending on the same electron coordinates if they are located on different atoms. This approximation results in three consequences. First, the overlap matrix becomes a unit matrix. Second, one electron integrals involving basis functions from two different atomic centers are set to zero. Third, All two electron integrals involving three- and four center are neglected.

In addition to the ZDO approximation, semi-empirical methods consider only the valence electrons explicitly using minimal basis set. Therefore, the core electrons are accounted implicitly by means of reducing the nuclear charge or introducing additional functions to describe the combined repulsion by the nuclei and core electrons. The approximations mentioned so far are already too many, so employing all those approximations inevitably deteriorates the quality of the results. To compensate for these approximations, semi-empirical methods adopt parameters obtained from calculations or experimental data in replacement of some or all of the remaining integrals

Among many semi-empirical methods, we will mainly deal with the Recife Model 1 (RM1),[51] which is used in this work, and its related methods, namely the modified neglect of diatomic overlap (MNDO)[52] and the Austin model 1 (AM1).[53] All those models mentioned here are different from each other in how the core-core repulsion is computed and how the parameters are assigned. However, they are based on the same basic approximations called neglect of diatomic differential overlap (NDDO). NDDO contains approximations mentioned above and no further approximation, and they can be defined by following equations.

Overlap integrals

$$S_{\mu\nu} = \langle\mu|\nu\rangle = \delta_{\mu\nu}\delta_{AB} \tag{2.7}$$

One electron operator

$$h = -\frac{1}{2}\nabla^2 - \sum_{a}^{N_{nuclei}} \frac{Z'_a}{|R_a - r|} = -\frac{1}{2}\nabla^2 - \sum_{a}^{N_{nuclei}} V_a \tag{2.8}$$

Here Z' represents the reduced nuclear charge due to the core electrons.

One electron integrals

$$\langle\mu_A|h|\nu_A\rangle = \delta_{\mu\nu}\left\langle\mu_A\left|-\frac{1}{2}\nabla^2 - V_A\right|\nu_A\right\rangle - \sum_{a\neq A}^{N_{nuclei}} \langle\mu_A|V_a|\nu_A\rangle \tag{2.9}$$

$$\langle\mu_A|h|\nu_B\rangle = \left\langle\mu_A\left|-\frac{1}{2}\nabla^2 - V_A - V_B\right|\nu_B\right\rangle \tag{2.10}$$

$$\langle\mu_A|V_C|\nu_B\rangle = 0 \tag{2.11}$$

Two electron integrals

$$\delta_{AC}\delta_{BD}\langle\mu_A\nu_B|\lambda_C\sigma_D\rangle = \langle\mu_A\nu_B|\lambda_A\sigma_B\rangle \tag{2.12}$$

2.2.2.2 Modified neglect of diatomic overlap (MNDO)

Based on the NDDO approximations, a diagonal Fock matrix element in MNDO can be expressed as following.

$$F_{\mu\mu} = U_\mu - \sum_{B\neq A} Z'_B\langle\mu s_B|\mu s_B\rangle + \sum_{\nu\in A} P_{\nu\nu}[\langle\mu\nu|\mu\nu\rangle - \frac{1}{2}\langle\mu\mu|\nu\nu\rangle] + \sum_{B}\sum_{\lambda\in B}\sum_{\sigma\in B} P_{\lambda\sigma}\langle\mu\lambda|\mu\sigma\rangle \tag{2.13}$$

The first term is the atomic orbital ionization potential which is defined as

$$U_\mu = \left\langle\mu\left|-\frac{1}{2}\nabla^2 - V_A\right|\mu\right\rangle \tag{2.14}$$

The second term describes the attraction to the other nuclei where the reduced nuclear charge is proportional to the repulsion with the valence s electron on that nucleus. The third term is the one-center two-electron integral which represents the Coulomb and exchange interactions with the other electrons in the same center, and the final term is the two-center two-electron integral that represents the Coulomb repulsion with the other electrons in different center. In the same way, an off-diagonal Fock matrix element can be expressed as

$$F_{\mu\nu} = -\sum_{B\neq A} Z'_B\langle \mu s_B|\nu s_B\rangle + P_{\mu\nu}\left[\frac{3}{2}\langle\mu\mu|\nu\nu\rangle - \frac{1}{2}\langle\mu\nu|\mu\nu\rangle\right] + \sum_B\sum_{\lambda\in B}\sum_{\sigma\in B} P_{\lambda\sigma}\langle\mu\lambda|\nu\sigma\rangle \quad (2.15)$$

where basis function μ and ν are on the same center. On the other hand, when μ and ν are on different centers, the Fock matrix element is defined as

$$F_{\mu\nu} = \frac{1}{2}(\beta_\mu + \beta_\nu)S_{\mu\nu} - \frac{1}{2}\sum_{\lambda\in A}\sum_{\sigma\in B} P_{\lambda\sigma}\langle\mu\nu|\lambda\sigma\rangle \quad (2.16)$$

The first term is the two-center one-electron integral which is the average of the resonance integral, the one-electron kinetic energy and nuclear attraction term, multiplied by the overlap element. This overlap element is computed explicitly in MNDO, which violates the ZDO approximation, and this is the reason why this method is called “Modified” neglect of diatomic overlap. The second term is the two-center two-electron integral which describes favorable exchange interactions.

MNDO adopts five different parameters obtained from atomic spectra for one-center two-electron integrals which are survived from the NDDO approximation. Using a sp-basis, they can be defined as

$$\begin{aligned}\langle ss|ss\rangle &= G_{ss}\\ \langle sp|sp\rangle &= G_{sp}\\ \langle ss|pp\rangle &= H_{sp}\\ \langle pp|pp\rangle &= G_{pp}\\ \langle pp'|pp'\rangle &= G_{p^2}\end{aligned} \quad (2.17)$$

p′ is used when two different type of p functions should be represented. The G-type parameters are Coulomb terms, and H parameter is an exchange integral.

In case of two-center two-electron integrals, total 22 cases are survived from the NDDO approximation within a sp-basis. Unlike the one-center two-electron integral, it is not easy to obtain universal values from spectroscopic data in this case. Therefore, this type of integrals is evaluated by replacing the continuous charge clouds with classical multipoles. Namely, when electron 1 is in an <sμ|sμ> type integral, then it is modeled as a point charge. When it is in an <sμ|pμ> type integral, it is modeled as a dipole, and in a <pμ|pμ> type integral, it is modeled as a quadrupole. The magnitudes of the moments, which are one-center in nature, are related to the parameterized integrals from the one-center two-electron case.

In order to complete energy evaluation, the nuclear repulsion energy should also be computed. In MNDO, nuclear repulsion energy is computed in the core-core repulsion term which is expressed as

$$V_N = \sum_{k<l} Z'_k Z'_l \langle s_k s_l | s_k s_l \rangle (1 + \tau e^{-\alpha z_k r_{kl}} + e^{-\alpha z_l r_{kl}}) \qquad (2.18)$$

where Z is the reduced nuclear charge, α is a parameter specific for each atom type, r is the inter-atomic distance, and τ is the adjustable value which is normally 1, but has a specific value rXH when the interaction involves O-H and N-H bonds. As shown in this term, the core-core repulsion is proportional to the repulsion between valence s electrons on each atom. In MNDO, all parameters except those for one-center two electron integrals (G_{ss}, G_{sp}, H_{sp}, G_{pp}, and G_{p2}) are fitted to molecular data. So far, the atoms for which the MNDO parameters are available are H, He, Li, Be, B, C, N, O, F, Mg, Al, Si, P, S, Cl, Zn, Ge, Br, Sn, I, Hg, and Pb.

2.2.2.3 Austin Model 1 (AM1)

The major defect of MNDO is that it evaluates the repulsion between two atoms of inter-atomic distance of 2-3Å too high. As a results, it overestimates the activation energy too large and, most crucially, is very poor at predicting the energy and geometry of hydrogen bonds. To remedy this problem of MNDO, Gaussian functions are added to the core-core repulsion term in AM1.

$$V_N(A,B) = V_{AB}^{MNDO} + \frac{Z'_A Z'_B}{r_{AB}} \sum_{i=1}^{N} [a_{A,i} e^{-b_{A,i}(r_{AB} - c_{A,i})^2} + a_{B,i} e^{-b_{B,i}(r_{AB} - c_{B,i})^2}] \qquad (2.19)$$

The number of Gaussian functions added to core-core repulsion (N) is varied depending on the atom. For example, oxygen and the halogens have two Gaussian functions, hydrogen, nitrogen, phosphorus, and sulfur have three Gaussian functions, and carbon has four Gaussian functions. $a_{A,i}$ is a Gaussian multiplier for the i^{th} Gaussian of atom A, and $b_{A,i}$ is a Gaussian exponent multiplier for the i^{th} Gaussian of atom A, and $c_{A,i}$ is a radial center of the i^{th} Gaussian of atom A. These constants and other parameters are fitted to molecular data except those for the one-center two electron integrals, which are taken from atomic spectra like MNDO. In AM1, C, H, O, N, B, F, Mg, Al, Si, P, S, Cl, Zn, Ge, Br, Sn, I, and Hg have been parameterized. Due to many improvements over MNDO and its reasonable robustness over a

broad range of chemical functionality, AM1 is still reported in the chemical literature and included in a number of computational quantum chemistry programs.

2.2.2.4 Recife Model 1 (RM1)

RM1 is simply a reparameterization of AM1. It has the same analytical construct and the same number of parameters for each atom as AM1, so it can be easily implemented in any quantum chemistry program where AM1 is already available by just introducing new parameters without modifying any line of code. To achieve more accurate model than AM1, RM1 optimized 191 different parameters for 10 elements (C, H, O, N, P, S, F, Cl, Br, and I) based on the training set of 1736 species which mainly consists of atoms and molecules having significant importance in organic and biochemistry. The experimental properties they referred to for parameterization are enthalpies of formation, dipole moments, ionization potentials and geometrical variables, namely, inter-atomic distances, bond, and dihedral angles. In RM1, the parameters for one-center two electron integrals, which were taken from atomic spectra in AM1, have been also optimized as well as other parameters. Although RM1 still displays some qualitative problems appeared in AM1 such as the ϕ-ψ barriers in peptides, it clearly shows the improvement in quantitative accuracies over AM1.

2.3 Free energy calculations

This chapter mainly refers to ref.[45], ref.[54], ref.[55], and ref.[56]

2.3.1 Ensemble average

In a classical system, any state can be completely described by the positions and momenta of all particles constituting the system. Therefore, a system composed of N particle in three dimensions can be determined by 6N coordinates, and the space defined by these 6N coordinates is called 'phase space' of that system. The position and momentum coordinates can be expressed as

$$\mathbf{q} = (x_1,y_1,z_1,x_2,y_2,z_2,x_3,y_3,z_3,\ldots.)$$
$$\mathbf{p} = (p_{x,1},p_{y,1},p_{z,1},p_{x,2},p_{y,2},p_{z,2},p_{x,3},p_{y,3},p_{z,3},\ldots.) \quad (2.20)$$

and a point in phase space can be expressed as

$$\mathbf{X} = (\mathbf{q},\mathbf{p}) \quad (2.21)$$

Using this notation, we can further formulate the ensemble average of a property A of certain system at equilibrium (expectation value), where the temperature, volume, and number of particles are constant.

$$\langle A \rangle = \iint A(q,p)P(q,p)dqdp \quad (2.22)$$

where P is the probability of being at a certain phase point. The P can be defined as

$$P(q,p) = \frac{e^{-E(q,p)/k_B T}}{Q} \quad (2.23)$$

where E is the total energy composed of kinetic and potential energies, k_B is Boltzmann's constant, and T is the temperature. Q is the system partition function, which is a canonical partition function here since the temperature, volume, and number of particles are constant, that is defined as

$$Q = \iint e^{-E(p,q)/k_B T}dqdp \quad (2.24)$$

so it can be considered as the normalization constant for P.

It is almost impossible to evaluate $\langle A \rangle$ analytically in most cases, so it should be solved numerically. In order to do that, we should try to obtain the values of integrand at as many phase points as possible until $\langle A \rangle$ converges to a particular value. In this context, $\langle A \rangle$ can be reformulated as

$$\langle A \rangle = \frac{1}{M}\sum_{i}^{M} A(p,q) \tag{2.25}$$

and M is the number of points where A is sampled.

2.3.2 Free energy perturbation (FEP) method

The free energy perturbation method was first introduced by Zwanzig in 1954[57] based on statistical mechanics. This method allows one to compute free energy difference between two different states through MD or MC simulations. This means we can obtain macroscopic property (Helmholtz free energy) through microscopic properties such as a potential energy of a single molecule. Therefore, in order to derive the core equation, which is also called the Zwanzig equation, in the FEP method, we need to adopt the partition function because it exactly plays such a role to establish a connection between macroscopic properties and microscopic properties. The relationship between Helmholtz free energy (A) and the partition function (Q) can be expressed as

$$A = -k_B T \ln Q \tag{2.26}$$

This can be further rewritten as

$$\begin{aligned} A &= k_B T \ln \frac{1}{Q} \\ &= k_B T \ln \frac{\iint e^{E(p,q)/k_B T} e^{-E(p,q)/k_B T} dq dp}{\iint e^{-E(p,q)/k_B T} dq dp} \\ &= k_B T \ln \iint e^{E(p,q)/k_B T} P(q,p) dq dp \end{aligned} \tag{2.27}$$

by the equation (2.23) and (2.24). Using this equation and the relationship shown in the equation (2.22), the ensemble average of the Helmholtz free energy can be defined as

$$\langle A \rangle = k_B T \ln \langle e^{E(p,q)/k_B T} \rangle \tag{2.28}$$

Therefore, the Helmholtz free energy difference between two states A and B can be written as

$$\begin{aligned} \langle A \rangle_B - \langle A \rangle_A &= k_B T \ln \langle e^{E(p,q)/k_B T} \rangle_B - k_B T \ln \langle e^{E(p,q)/k_B T} \rangle_A \\ &= k_B T \ln \frac{\langle e^{E(p,q)/k_B T} \rangle_B}{\langle e^{E(p,q)/k_B T} \rangle_A} \\ &= k_B T \ln \frac{\frac{1}{M_B}\sum_i^{M_B} e^{E(p,q)/k_B T}}{\frac{1}{M_A}\sum_i^{M_A} e^{E(p,q)/k_B T}} \end{aligned} \tag{2.29}$$

Now the Helmholtz free energy difference between two states A and B can be computed through MD or MC simulations using this equation. However, this equation is still immature for practical use because the convergence of two ensemble average of Helmholtz free energy from two different systems which evolve differently is computationally very expensive. This problem can be circumvented if the ensemble average can be taken from an identical ensemble. To achieve that, equation (2.29) should be reformulated as

$$\langle A\rangle_B - \langle A\rangle_A = k_B T \ln \frac{1}{M_A} \sum_{i}^{M_A} e^{(E_B - E_A)/k_B T} \tag{2.30}$$

In this way, the ensemble average will be sampled based on the system A, and only the energy difference arising from the part where atoms belonging to the system B interact with the atoms belonging to the system A will be taken into account. The interactions taking place among atoms belonging to the system A will cancel out in energy difference because the ensembles are identical.

Although we can have a big advantage in terms of computational expenses in this manner, it also contains a serious problem as well. That is when E_B is sampled from system A, an atom that originally belongs to the system B can materialize in close proximity to the atom belonging to the system A, or they can be even on top of each other, thereby resulting in extremely high in energy. To avoid this problem, coupling parameter λ is introduced.

$$E(\lambda) = \lambda E_B + (1 - \lambda) E_A \tag{2.31}$$

where λ can be varied from 0 to 1. By adopting this parameter, the system where sampling takes place can be slowly switched from the system A to the system B. By inserting equation (2.31) into (2.30), we can finally obtain the equation for FEP calculations.

$$\langle A\rangle_B - \langle A\rangle_A = \sum_{\lambda=0}^{1} k_B T \ln \frac{1}{M_\lambda} \sum_{i}^{M_\lambda} e^{(E_{\lambda+\Delta\lambda} - E_\lambda)/k_B T} \tag{2.32}$$

The FEP calculation is especially useful for evaluating the pKa value of amino acids in the protein environment. To evaluate the pKa value, it is necessary to evaluate the solvation free energies of the neutral and ionized species in both water and protein. However, these energies cannot be directly obtained by the FEP calculations, and therefore a thermodynamic cycle such as the one shown in Figure 2.5 should be devised first for that.

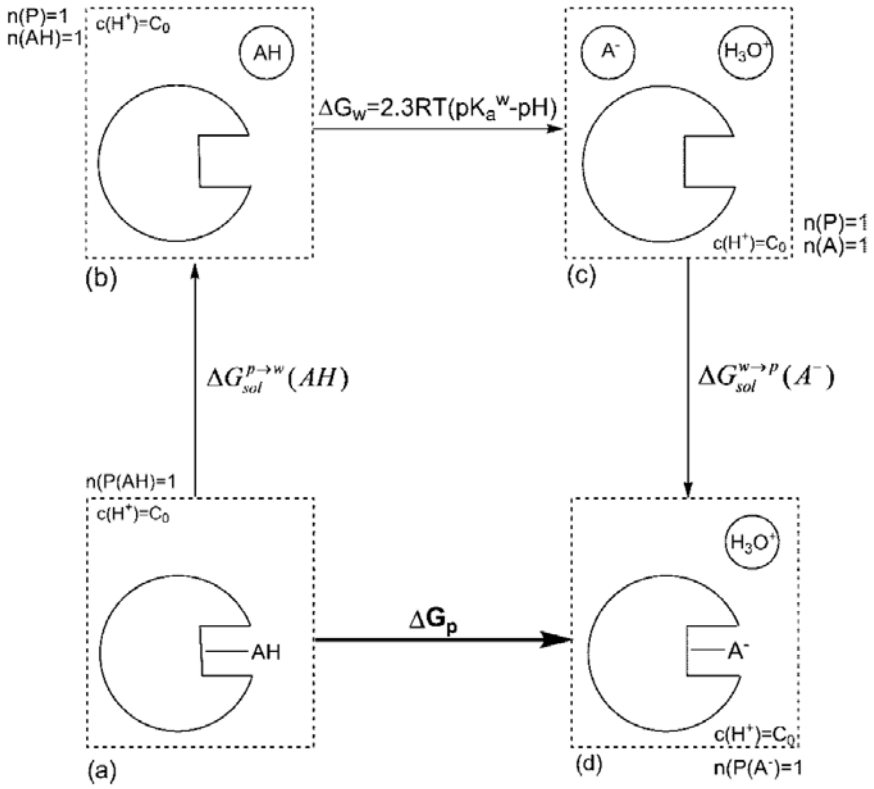


Figure 2.5. Thermodynamic cycle to evaluate a pKa value of amino acids in enzyme environment. ΔG_p and ΔG_w indicate the free energy of ionization of the functional group in protein and water, respectively. $\Delta G_{solv}^{p\rightarrow w}$ and $\Delta G_{solv}^{w\rightarrow p}$ represent the change of the solvation free energy as the indicated species moves from water to protein environment or *vice versa*. Reprinted with permission from ref.[56] Copyright 2009 American Chemical Society

According to the thermodynamic cycle shown in Figure 2.5, the free energy change corresponding to the ionization of the functional group in protein can be represented as

$$\Delta G_p = \Delta G_{solv}^{p\rightarrow w}(AH) + \Delta G_w + \Delta G_{solv}^{w\rightarrow p}(A^-) \quad (2.33)$$

Here, the ionization free energy (ΔG_p and ΔG_w) of the functional group can be defined using its pKa value depending on the environment and the pH.

$$\Delta G_p = 2.3RT(pKa^p - pH) \quad (2.34)$$

$$\Delta G_w = 2.3RT(pKa^w - pH) \quad (2.35)$$

Also,

$$\Delta G_{solv}^{p\rightarrow w}(AH) = -\Delta G_{solv}^{w\rightarrow p}(AH) \quad (2.36)$$

Using these relations, we can rewrite the equation (2.33) as

$$pKa^{p} = pKa^{w} - \frac{q}{2.3RT}\Delta\Delta G_{solv}^{w\to p}(AH \to H^{-}) \quad (2.37)$$

where q represents the charge of the ionized form of the functional group, that is -1 for acids and +1 for bases. Although it is not possible to compute the term, $\Delta\Delta G_{solv}^{w\to p}(AH \to H^{-})$, directly from the FEP calculation, we can still obtain pKa^{p} value because $\Delta\Delta G_{solv}^{w\to p}(AH \to H-$ is equivalent to ΔG_w - ΔG_p according to the thermodynamic cycle, and we can compute this value by the FEP calculation. It is also noteworthy that we can only compute the double free energy difference (ΔG_w - ΔG_p) by the FEP calculation, not a single free energy difference (ΔG_w or ΔG_p alone), since the energy differences we can obtain from the FEP calculation involve an unphysical process.

2.3.3 Umbrella sampling

Umbrella sampling is a method to improve the sampling of the system in the region of high energy on its free energy landscape by introducing a biasing potential, which is also called an umbrella potential. This method was first suggested by Torrie and Valleau in 1977.[58] The free energy landscape can be described by the potential of mean force (PMF) which is the free energy expressed as the function of a coordinate q. The PMF was first introduced by Kirkwood in 1936 [59] and can be defined as

$$W(q) = -k_B T \ln \pi(q) \quad (2.38)$$

where q is the coordinate, and π is the probability of coordinate being at a certain point in phase space. Using a partition function Q, the probability π can be defined as

$$\pi(q) = \frac{1}{Q}\iint \delta[q'(\mathbf{q}) - q]e^{-E(\mathbf{p},\mathbf{q})/k_B T}d\mathbf{q}d\mathbf{p} \quad (2.39)$$

where δ is the Dirac delta function which has the value of 1 only when $q'(\mathbf{q})$ equals to q and zero in elsewhere. q′ is the function of a positional coordinates **q** in phase space and can take the value equivalent to any arbitrary point on the PMF coordinate.

Umbrella sampling is very useful for computing $\pi(q)$ especially when there is a large energy barrier along the PMF coordinate. Since high energy region is always associated with the low probability according to the equation (2.38), it is not feasible for normal MD simulations to sample the configurational space with high energy accurately within the

available computer time. However, in the presence of the biasing potential, this problem can be solved by enhancing the sampling in the region with high energy.

In principle, there is no unique function which can be used as a biasing potential. However, the most reasonable choice and also the most widely used is the harmonic function which is defined as

$$U(q) = \frac{1}{2}k(q - q_0)^2 \tag{2.40}$$

where q_0 is the point we want to sample on the PMF coordinate. Since the sampling will be done accurately only for a confined region on the PMF coordinate, if we use a harmonic function as a biasing potential, only the estimated PMF for the corresponding part will have the enough accuracy. Therefore, in order to obtain the accurate PMF over the whole range of the coordinate, it is necessary to carry out a number of simulations at sufficient points which are enough to cover the entire range using biasing potentials with different q_0. Each simulation performed here can be termed as a 'window', so the biasing potential is also called as a window potential.

When the biasing potential is added, the probability obtained from the sampling will be biased as well, so the PMF computed directly from the biased probability using the equation (2.38) will not be correct. Therefore, to obtain the correct PMF from the *i*th window, we should reformulate the equation (2.38) with the biased probability as

$$W(q) = -k_B T \ln \pi^*(q) - U_i(q) - k_B T \ln \langle e^{-U_i(q)/k_B T} \rangle \tag{2.41}$$

where $\pi^*(q)$ is the biased probability, $U_i(q)$ is the biasing potential in the *i*th window, and $-k_B T \ln \langle e^{-U_i(q)/k_B T} \rangle$ is the free energy constant F_i associated with adding the biasing potential in the *i*th window and can be derived as follows.

$$\begin{aligned} F &= -k_B T \ln Q \\ &= -k_B T \ln \frac{\int e^{-U(q)/k_B T} dq}{\int e^{-U(q)/k_B T} e^{U(q)/k_B T} dq} \\ &= -k_B T \ln \int \frac{1}{e^{U(q)/k_B T}} \pi(q) dq \\ &= -k_B T \ln \langle e^{-U(q)/k_B T} \rangle \end{aligned} \tag{2.42}$$

The equation (2.22) and (2.27) are used for this derivation.

The procedure to generate the complete PMF by patching up the individual PMF fragment from each window can be illustrated as shown in Figure 2.6. Each PMF fragment

shown in the lower part only represents statistically useful region, and the regions at either end of each fragment which go to infinite due to the artifacts of biasing potential are omitted for clarity. In this procedure, the sufficient overlaps between each PMF fragment are essential. Several approaches have been suggested to accomplish this overlap, and the most widely used one is the weighted histogram analysis method (WHAM), which was suggested by Kumar et al.[60]

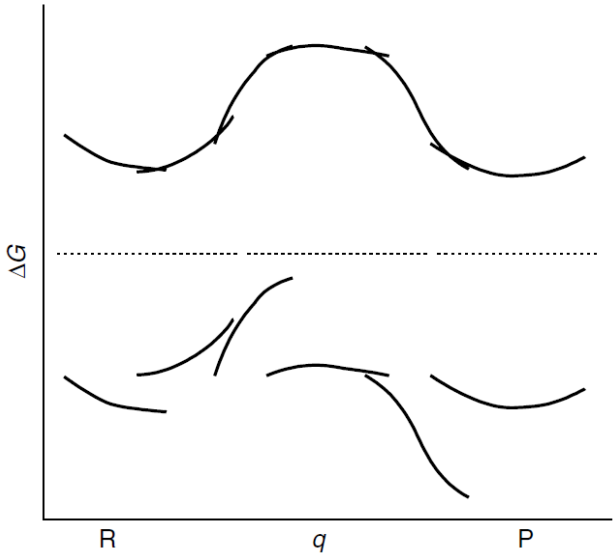


Figure 2.6. A schematic drawing of the procedure to generate a complete PMF from individual PMF fragments. The complete PMF shown in the upper part and the fragments are shown in the lower part. Reprinted with permission from ref.[45] Copyright 2004 John Wiley and Sons

In WHAM, it adds up the individual unbiased probability obtained from each window using a coordinate dependent weight as a coefficient to compute the $\pi(q)$ for the coordinate q.

$$\pi(q) = \sum_{i=1}^{N_w} \pi_i(q) \times \left[\frac{n_i e^{-\frac{[U_i(q)-F_i]}{k_B T}}}{\sum_{j=1}^{N_w} n_j e^{-\frac{[U_j(q)-F_j]}{k_B T}}} \right] \tag{2.42}$$

where n_i is the number of independent data points collected to compute the biased probability, and N_w is the number of windows. Furthermore, the equation (2.42) can be slightly modified to give

$$e^{-F_i/k_BT} = \langle e^{-U_i(q)/k_BT} \rangle \tag{2.43}$$

Using this equation together with (2.38) and (2.41), we can derive the relationship between the unbiased probability and biased probability in the *i*th window as follows.

$$\pi_i(q) = e^{\frac{[U_i(q)-F_i]}{k_BT}} \pi_i^*(q) \tag{2.44}$$

By substituting this equation into the equation (2.42), we can obtain

$$\pi(q) = \sum_{i=1}^{N_w} n_i \pi_i^*(q) \times \left[\sum_{j=1}^{N_w} n_j e^{-\frac{[U_j(q)-F_j]}{k_BT}} \right]^{-1} \tag{2.45}$$

The biased probability $\pi_i^*(q)$ in the *i*th window can be expressed as

$$\pi_i^*(q) = \frac{1}{n_i} \sum_{l=1}^{n_i} \delta[q'(q) - q_{i,l}] \tag{2.46}$$

The equation (2.43) can be rewritten as

$$e^{-F_i/k_BT} = \int e^{-U_i(q)/k_BT} \pi(q) dq \tag{2.47}$$

and if we insert the equation (2.45) into (2.47), we obtain

$$e^{-F_i/k_BT} = \int e^{-U_i(q)/k_BT} \sum_{i=1}^{N_w} n_i \pi_i^*(q) \times \left[\sum_{j=1}^{N_w} n_j e^{-\frac{[U_j(q)-F_j]}{k_BT}} \right]^{-1} dq \tag{2.48}$$

again inserting the equation (2.46) into (2.48) gives

$$\begin{aligned} e^{-F_i/k_BT} &= \int \sum_{i=1}^{N_w} \frac{e^{-U_i(q)/k_BT}}{\sum_{j=1}^{N_w} n_j e^{-\frac{[U_j(q)-F_j]}{k_BT}}} \sum_{l=1}^{n_i} \delta[q'(q) - q_{i,l}] \, dq \\ &= \sum_{i=1}^{N_w} \sum_{l=1}^{n_i} \frac{e^{-U_i(q_{i,l})/k_BT}}{\sum_{j=1}^{N_w} n_j e^{-\frac{[U_j(q_{i,l})-F_j]}{k_BT}}} dq \end{aligned} \tag{2.49}$$

Since the free energy constant F_i appears in both side of the equation, it must be solved in an iterative manner. The iteration cycle starts from an initial guess for set of free energy constants $\{F_i\}$ and is repeated until $\{F_i\}$ are converged. With the set of free energy constants $\{F_i\}$ in hand, we can compute $\pi(q)$ and therefore the PMF.

2.4 Salt bridges

Salt bridges are one of the interactions studied thoroughly in protein science and can be frequently observed in protein environment. It has properties of two non-covalent interactions, namely, electrostatic and hydrogen bond interaction. Simple salt bridges found in protein environment are formed by two oppositely charged residues, but they can also form salt bridge networks composed of three or more interacting charged residues. Here, the negatively charged residues involve Asp, Glu, Tyr, Cys, and the C-terminal carboxylate group, and positively charged residues include His, Lys, Arg, and the N-terminal amino group.[61] The salt bridges play various roles related to protein function and structure, which incorporate substrate binding in enzyme,[62–64] mediating molecular recognition,[65–67] regulation of allosteric behavior,[68,69] guiding the protein folding pathway,[70,71] and assembly of protein subunits.[72–74] All those roles mentioned here are closely linked to the tendency of salt bridges to stabilize folded structure of proteins. One explicit example about this fundamental nature of salt bridges can be the finding that salt bridges are more frequent in thermophilic proteins than their mesophilic counterparts.[75,76] Additionally, the observation that removing a charge from a residue involved in a buried salt bridge resulted in the loss of the stability by 3-5kcal/mol also exemplifies the stabilizing effect of salt bridges.[77,78] However, whether salt bridges contribute to protein stability or not seems closely related to how much they are desolvated from solvents. Several studies showed that solvent-exposed salt bridges contribute very little to protein stability at physiological ionic strengths,[79–82] and this implies that the stabilizing effect of salt bridges is mainly attributed to buried ones. Yet other studies demonstrated that buried salt bridges can relatively destabilize proteins compared to hydrophobic residues,[83–86] and these findings raised a question about the stabilizing role of buried salt bridges. Another important point to note is that salt bridges consist of ionizable residues. Therefore, depending on whether the constituting residues are in ionic or neutral form, the interaction also can become either salt bridge or neutral hydrogen bonded form. According to some theoretical studies, the character of the interaction is affected by the polarity of the environment, and it was shown that the neutral form was preferred in low polar solvents.[87,88] However, when those constituting residues lie in a heterogeneous dielectric system such as proteins, the character of the interaction strongly depends on the surrounding microenvironment. Thus, it is not straightforward to predict whether the

interaction is a salt bridge or a neutral hydrogen bonded form if the constituting residues are buried inside a protein.

Chapter 3 Results and discussion

3.1 The protonation state of catalytic residues in the resting state

In this chapter, the protonation state of catalytic residues in the resting state is mainly addressed. Since the information on the positions of protons cannot be obtained from X-ray crystal structure, it should be investigated computationally. The protonation state of catalytic residues in the resting state is important because it is closely related to the mechanism of the catalytic reaction. Furthermore, this information can also provide a valuable clue in designing potent inhibitors against KasA. In addition to the protonation state in the resting state, the other important related issues are also addressed here. The unusual conformation of one of the catalytic histidines (His345) from X-ray crystal structure is corrected using MD simulations, and the molecular reason for the unexpected acyl-transfer activity in the

His311Ala mutant is addressed based on the results from MD simulations and FEP computations.

3.1.1 Computational details

The initial structure for molecular dynamic (MD) simulations was obtained from the crystal structure of wild type *M. tuberculosis* KasA (PDB entry: 2WGD). All MD simulations were performed using the NAMD 2.6 simulation package[89] in combination with the CHARMM 22 force field[90] with the CMAP correction.[91] The force field parameters for the thiolate ion in anionic cysteines were obtained from Foloppe. et al.[92] TIP3P water molecules[93] were used for explicit solvent conditions. Each initial structure was solvated in a water shell with a radius of 45 Å. The water molecules were forced to stay inside the sphere during the simulation using spherical boundary conditions. After solvation, all systems were neutralized by adding an appropriate number of ions according to the total charge of the system using the Autoionize package version 1.1. Then, energy minimization was performed for 500 steps using the conjugate gradient method[94] to avoid any huge steric hindrance. A time step of 2 fs was used, and all chemical bonds containing hydrogen atoms were constrained using the RATTLE algorithm.[95] Non-bonded interactions were truncated at a cutoff distance of 12 Å, and the switching function was used to smoothly reduce the potential to zero at the cutoff distance. The temperature was gradually increased after minimization from 50 K to 310 K while constraining the protein in its initial position with a force constant of 24 kcal $mol^{-1}Å^{-2}$. Afterwards, the force that was applied to the protein gradually decreased in a stepwise manner until it finally reached zero. Langevin dynamics was used to maintain the temperature constant. For a productive MD run, the first 1 ns was discarded for equilibration, and the following 5 ns were used for data gathering.

The FEP method[96] was employed to measure the relative free energy difference between different protonated systems. Since all FEP calculations were performed in an NVT ensemble, the obtained free energies are given as Helmholtz free energies. The devised thermodynamic cycle involves two alchemical transformations. The first calculation is for the reference system where an individual residue with a neutral blocking group at both termini solvated in the water shell with a radius of 22 Å is gradually transformed from the protonated state to the deprotonated state. The second calculation simulates the same transformation but in the protein environment. A total of 118 windows of uneven width were introduced by

adopting a set of intermediate mapping potentials between the start and the end point. For each window, 10 ps of equilibration and 30 ps of data collection were carried out, resulting in a total of 4.72 ns of running time. Within a range of 0.1 from both end points of λ, a total of 84 windows were introduced to avoid an end-point catastrophe.[97] To estimate the accuracy of the calculation, each FEP calculation was repeated four times in one direction,[98] and 20 Å as a cutoff value was used for all FEP calculations in order to minimize any possible artifact from a short cutoff value.

During the process of proton transfer from one residue to the other, the cleavage and formation of covalent bonds is inevitable. For calculating the energy profile of this process occurring in the enzyme, the QM/MM optimization is a suitable choice because it can not only deal with the breakage and formation of a bond by quantum mechanics but, at the same time, also takes the enzyme environment into account by classical mechanics. The initial structures for QM/MM optimizations were taken from the MD simulation trajectory and subsequently minimized for 2500 steps using the conjugate gradient method.[94] In the QM/MM calculation, the entire enzyme and water molecules only within a distance of 25 Å to the active site were included. All QM/MM optimizations were conducted using the Chemshell program.[99] Cys171, His311, and a water molecule were involved in the QM region. The Turbomole program package[100] was used at the level of density functional theory for QM calculations. MM calculations were done by the DL_POLY code [101] in combination with the CHARMM force field. At the boundary of the QM and MM part, the charge-shift scheme was applied to prevent overpolarization.[102,103] A hybrid delocalized internal coordinate (HDLC) optimizer[104] was used for geometry optimization, and only water molecules and residues within 10 Å of the active site were involved in the optimization process while the remaining part was held fixed. Geometry optimizations in the QM part were performed with RI-DFT/B-LYP in combination with TZVP basis sets. For single point calculations, the DFT/B3-LYP level was used with the same basis sets.

Because of the strong dependency of QM/MM optimization on starting structures, sampling of the system was required. For this reason, we performed umbrella sampling in the context of the QM/MM scheme to compute the potential of mean force (PMF) surfaces for the proton transfer. They were performed using the Recife Model 1 (RM1) method[105] for QM, and the AMBER force field (FF99SB)[106] for the MM part. A total of 448 windows were employed for umbrella sampling, and the weighted histogram analysis method (WHAM) was

applied to generate a PMF surface.[107] 20 ps of equilibration and 30 ps of productive run were carried out for each window, and the force constants used were in the range of 800 to 900 kcal $mol^{-1}Å^{-2}$. To test the reliability of this semi empirical approach we performed additional model computations involving the QM part of our QM/MM computations and neglecting environmental effects. These computations were performed with the Sander module of the AMBER 11 program package[108,109] and the Turbomole program package.[100]

To test the importance of residues for the energy difference between the zwitterionic and the neutral state the charge deletion analysis was used.[110,111,112] For this calculation, the partial charges of the corresponding residues in the MM part were set to zero, and additional single point QM/MM calculations were performed.

3.1.2 Determining the orientation of the imidazole ring of His345

With respect to the orientation of the imidazole ring of His345 Luckner et al.[39] interpreted their data in wild type KasA (PDB entry: 2WGD) as shown in Figure 3.1.1. This conformation is flipped by 180° in comparison to other X-ray structures of the KasA enzyme from *Mycobacterium tuberculosis* (PDB entries: 2WGE, 2WGF, 2WGG) or other KAS enzymes (PDB entries: 1DD8, 1KAS, 1J3N, 1E5M, 1OX0). This conformation of His345 seems to be unfavorable because it does not allow the formation of a hydrogen bond between the δ-nitrogen of His345 and the backbone amide, which was postulated on the basis of other crystal structures. However, it would allow a hydrogen bond to the Cys171 residue which is not possible in the other conformation. Because a reliable designation of the conformation on the basis of X-ray data alone is not possible an incorrectly assigned conformation might have resulted. Hence, we performed MD simulations for each conformation in which either the δ- or the ε-nitrogen of His345 was protonated. The resulting tautomers are indicated in Figure 3.1.2. Since the protonation states of Cys171 and His311 may also affect the orientation of His345 their protonation states have to be taken into account as well. According to usual pKa-values one would assume that the sulfur center of Cys171 and the δ-nitrogen of His311 are protonated while the ε nitrogen of His311 is expected to be deprotonated. Since the His311/Cys171 dyad is neutral in this case this situation will be called the neutral state. However, in the resting states of various enzymes with Cys/His pairs the proton is moved from the sulfur of the Cys to the ε nitrogen of the His residue ($His311H^{+}/Cys171^{-}$). This situation will be called the zwitterionic state. It is proposed in many mechanistic studies since

the neutral state is not sufficiently reactive for the enzymatic reactions.[113] To take mutual interactions into account we performed MD simulations for all resulting protonation states.

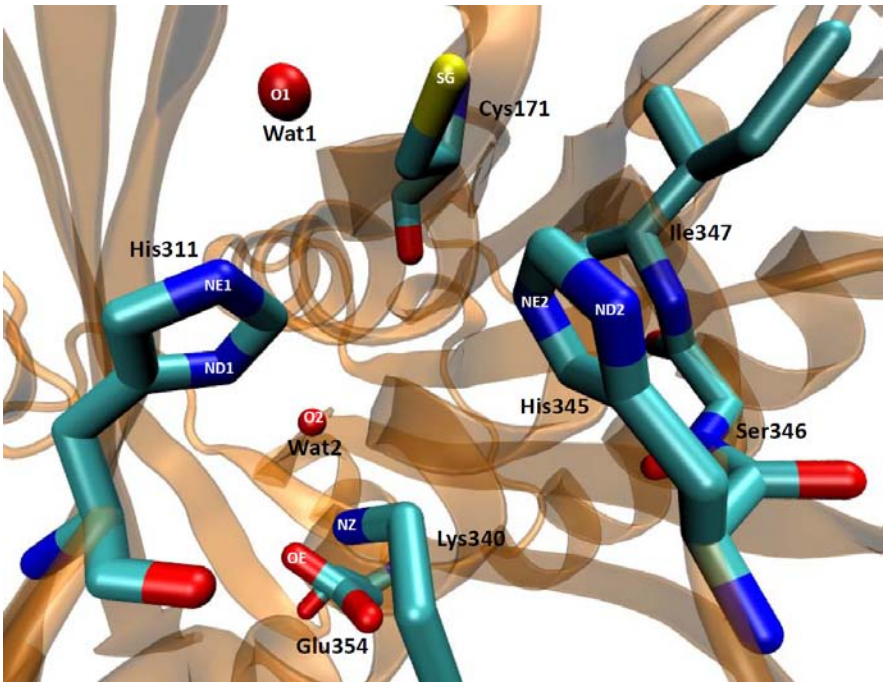


Figure 3.1.1. The active site of wild type KasA from the X-ray structure by Luckner et al.[39] (PDB entry: 2WGD) The active site residues and water molecules are shown in stick representation, and the protein surroundings are represented in a cartoon model. Residue numbers and atoms are depicted in black and white, respectively.

Pair	Distance (Å)
SG – NE2	3.55
SG – O1	3.19
O1 – NE1	2.79
ND1 – O2	2.78
O2 – NZ	3.00
O2 – OE	3.02

Table 3.1.1. Selected distances of important pairs depicted in Figure 3.1.1. All distances were extracted from the crystal structure of wild type KasA.

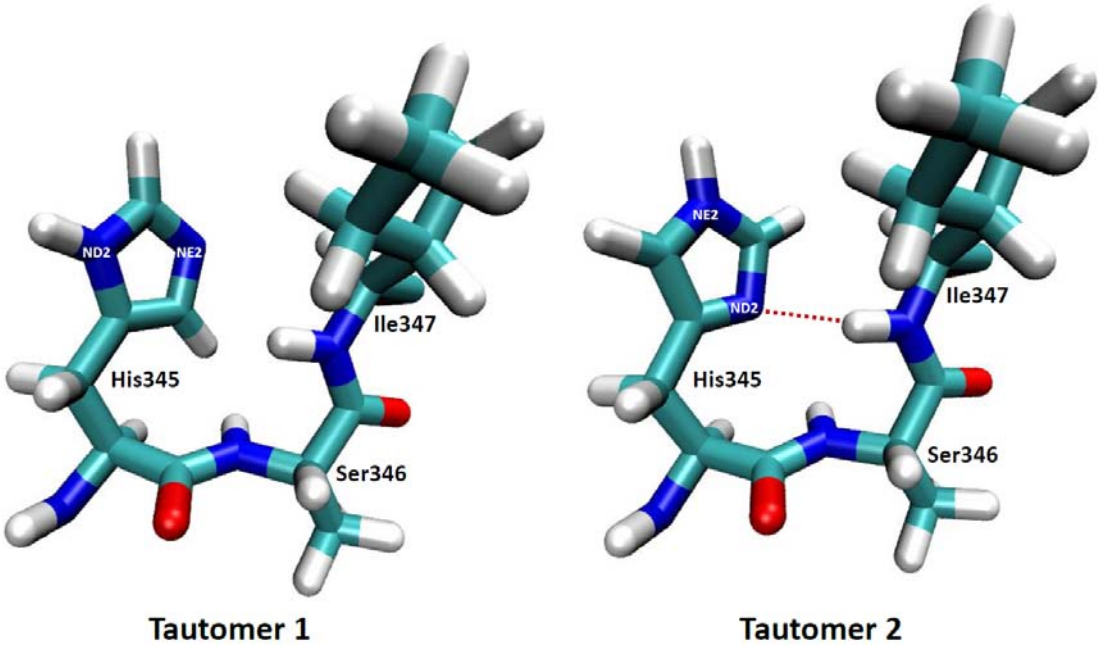


Figure 3.1.2. Two possible tautomeric forms of His345. Tautomer 1 is derived from the unusual conformation of His345 in the crystal structure of wild type KasA. Tautomer 2 describes the situation found in other crystal structures of KasA. In addition to His345, Ser346 and Ile347 are also shown to indicate possible interactions between His345 and the backbone of Ile347. The interaction is represented as a red dotted line.

If the δ-nitrogen is protonated (Figure 3.1.2, tautomer 1) the imidazole ring remained in the unusual conformation given by Luckner et al.[39] during the entire simulation. For tautomer 2 in which the ε-nitrogen of His345 is protonated, the imidazole ring flipped during the simulation and turned into the conformation indicated in Figure 3.1.2. This conformation corresponds to the orientation described in previous crystal structures. This rotation is due to repulsive interactions between the hydrogen at the ε-nitrogen of His345 and the hydrogen of the backbone amide group. It is important to note, that the different behaviors of the two tautomers were always observed regardless of whether the other two catalytic residues, His311 and Cys171, are in the neutral or in the zwitterionic state.

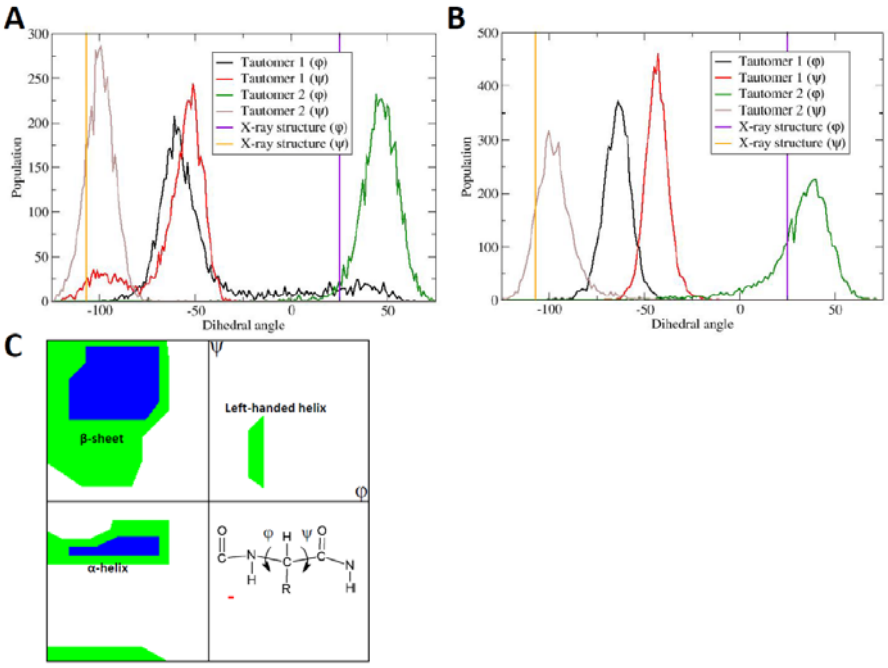


Figure 3.1.3. The backbone φ and ψ angle distributions of Ile347 during MD simulation of the two tautomeric forms of His345 given in Figure 3.1.2. The measurements were performed for the neutral **A** and the zwitterionic state **B**. The φ and ψ angles from the crystal structure are also shown as vertical lines for comparison. The Ramachandran plot of Ile347 from the crystal structure is represented as a red dot in **C**, and the positions of the φ and ψ angles in the protein backbone are shown in the inset.

To determine which of the two tautomers is more likely we monitored the movements of all neighboring residues during the MD simulation and compared them with the experimentally measured orientations. All of them adjust to the rotated conformation but the position of the backbone of Ile347 is the decisive property towards the relative orientation of the imidazole ring of His345. This backbone rotation can be measured as changes in the φ and ψ dihedral angles of Ile347. The distributions of these dihedral angles during the MD simulations are plotted in Figure 3.1.3A for the neutral state of the His311/Cys171 dyad and in Figure 3.1.3B for the corresponding zwitterionic state. The crystal structure by Luckner et al.[39] shows that the backbone of Ile347 adopts an unusually strained conformation (φ = 25˚, ψ = -107˚) as shown in the Ramachandran plot (Figure 3.1.3C). This unusual conformation is only found for tautomer 2 for which the maxima of the computed distributions are about 50° (φ) and -100° (ψ) if His311 and Cys171 adopt the neutral state and about 40° (φ) and -100° (ψ) if His311 and Cys171 are in the zwitterionic state. Figure 3.1.2 reveals that the hydrogen

bond between His345 and the backbone amide of Ile347 acts as one of the constraints to maintain this strained conformation. Due to the protonation of the δ-nitrogen this bond cannot be formed in tautomer 1. Thus, the strained backbone turns into the conformation which belongs to the regular α-helical region in the Ramachandran plot. The resulting maxima of the computed distribution for tautomer 1 are about -60° (φ) and -50° (ψ) for the neutral state and about -70° (φ) and -50° (ψ) for the zwitterionic state. They deviate from the measured values by about - 80° (φ) and 50° (ψ). Based on the agreement with the unusual conformation of Ile347 we conclude that tautomer 2 represents the situation from the crystal structure given by Luckner et al.[39] This means that the imidazole ring is flipped by 180° in comparison to the original position. However, both conformations would have equally well explained the electron density maps and could not have been discerned in the crystal structure.

3.1.3 The preferred protonation state of catalytic residues in the resting state

MD simulations to identify the preferred protonation state of His311 - So far we could substantiate the orientation and protonation of His345 but due to the weak influence on the protonation state of the catalytic dyad His311/Cys171 their state remains to be determined. For the investigation of the protonation status of His311 we took three different tautomers into account. In the zwitterionic state of the His311/Cys171 dyad described above both nitrogen centers of the imidazole ring of His311 are protonated (Figure 3.1.4A). In the corresponding neutral state only the δ-nitrogen center is protonated. Additionally, we considered the situation in which only the ε-nitrogen center is protonated (Figure 3.1.4B). We used the same strategy as employed for His345 and monitored the movements of all neighboring residues and compared the data with their experimental counterparts. The distance between the oxygen center of the second water molecule (Wat2) and the ε-nitrogen of His311 allowed a first determination of the protonation state of His311. According to the crystal structure the distance is about 2.8 Å which indicates a hydrogen bond. During the MD simulations of the ionic and neutral state in which the δ-nitrogen of His311 is protonated the distances between these two atoms fluctuated around this value (Figure 3.1.4C). Note that the distance distribution found for the zwitterionic state agrees better with the X-ray data than the neutral counterpart. However, the difference between both is too small for a definitive answer.

If only the ε-nitrogen center is protonated (Figure 3.1.4B) the maximum of the distance distribution increases to about 3.3 Å and becomes considerably broader. The increase in the distance and the broadening take place because the hydrogen bond between both centers is no longer possible (Figure 3.1.4B). The strong deviations between measured and computed data indicate that a protonated ε-nitrogen of His311 can be ruled out. The protonation of His311 corresponds either to the ionic (both nitrogen centers protonated) or neutral state (only the δ-nitrogen is protonated). This is also in line with the observation that the δ-nitrogen of the corresponding histidine residue in *Sp*FabF showed substantial electron density in the 1.3 Å crystal structure.[41]

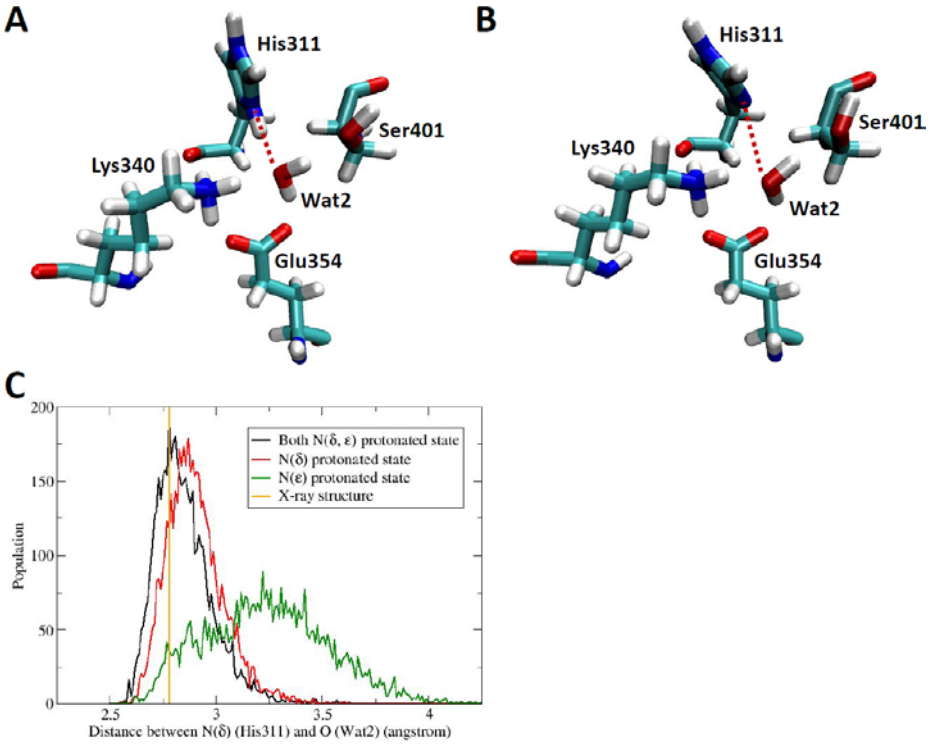


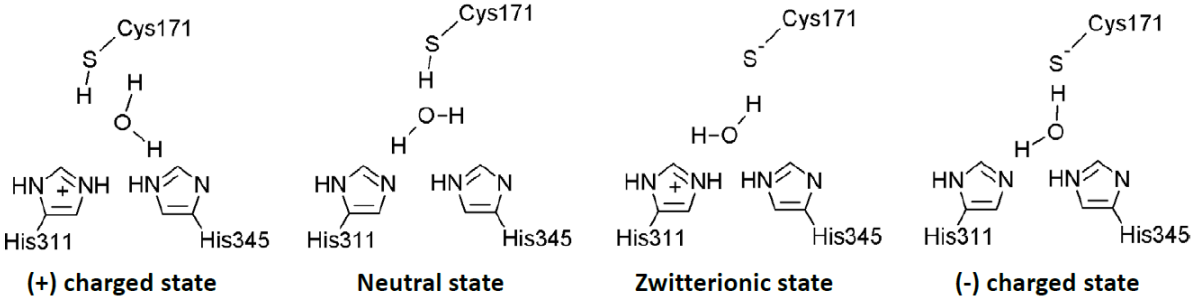
Figure 3.1.4. The interaction of His311 with a water molecule (Wat2) according to the change in protonation state of His311. Structure **A** sketches the geometrical arrangement when the δ and ε nitrogen atoms of His311 are protonated, and sketch **B** indicates the structure if only the ε-nitrogen is protonated. Each picture was created from a representative frame of MD simulation results. (selected by a visual inspection) The distances between the δ-nitrogen of His311 and the oxygen of the water molecule (dotted red line in **A** and **B**) recorded during the MD simulation for each protonation state are plotted as corresponding distributions in **C**. The distance from the crystal structure is also given as a vertical line for comparison.

The investigation also sheds some light onto the role of the Wat2 molecule conserved in most Kas enzymes and the functions of Glu354, Lys340, and Ser401. Wat2 represents an important anchor for His311. It seems to fix His311 in a position which is necessary for the function of the enzyme. To execute this function Wat2 must be fixed as well. Wat2 is located in the center of a hydrogen bonding network which comprises Glu354, Lys340, and Ser401. The former two form strong hydrogen bonds to Wat2 through the ammonium group of Lys340 and the carboxylate moiety of Glu354. The salt bridge between these two groups further strengthens this network and rigidifies it. The final interaction of Wat2 results from the hydrogen bond to Ser401. To test this observation we performed an MD simulation without Wat2. During this simulation the imidazole ring of His311 and various other residues changed their orientation so strongly that the enzyme would become inactive. This indicates that Wat2 represents an important factor for the stabilization of the active site. The hydrogen bond between the ammonium moiety of Lys340 and the oxygen center of the backbone carbonyl group of His311 represents a second anchor for the imidazole ring of His311. However, Wat2 seems to be more important since it is in direct contact to the imidazole ring. In contrast, the backbone carbonyl group of His311 is separated from the imidazole ring by two CH_2 groups. Hence this interaction can induce less rigidity.

MD simulations to identify the preferred protonation state of the Cys171/His311 dyad - So far we could clarify the protonation status of His345 and substantiate that the δ-nitrogen of the His311 is also protonated. As discussed above, the protonation state of the remaining ε-nitrogen of His311 is closely related to the protonation state of Cys171. The structural arrangement in the crystal structure (Figure 3.1.1 and Table 3.1.1) indicates that both residues interact with each other through a water molecule (Wat1). Since the δ-nitrogen of His311 was shown to be protonated the question about the protonation status of His311 and Cys171 is narrowed down to the situations depicted in scheme 3.1.1

The (+) charged state (scheme 3.1.1 left hand side) is expected for environments at lower pH-values. The (-) charged state (scheme 3.1.1 right hand side) will appear at high pH-values. Neutral and zwitterionic states will be found in between but the actual protonation state of the resting state is unknown. It can be estimated from the crystallographic data since they were obtained at pH 7.5. Considering the standard pK_a values of the involved residues (6.04 and 8.14 for the imidazole ring of His311 and the thiol moiety of cysteine, respectively)

at this pH-value a neutral state is expected. For the active site of the KasA enzyme however, the question remains open since the protein environment quite often affects the pK_a values considerably.



Scheme 3.1.1. Schematic representation of the four different protonation states.

To identify possible structural differences between the various protonation states we used the same strategy as before and compared structural changes observed during the MD simulations with the corresponding experimental data. In this case the orientation of the neighboring His345 (Figure 3.1.5) allows the determination of the protonation state of the His311/Cys171 dyad. If the sulfur center of Cys171 is protonated (thiol functionality) which is the case for the neutral and the (+) charged state the N_ε-H group of His345 is oriented towards the oxygen center of Wat1 (Figure 3.1.5A) and forms a hydrogen bond with the water molecule. If Cys171 is deprotonated (thiolate functionality) which is found for the zwitterionic and the (-) charged state, the N_ε-H group of His345 is directed towards the negatively charged sulfur center. The resulting hydrogen bond is considerably stronger due to the negative charge on the sulfur (Figure 3.1.5B). A visual inspection of the crystal structure (Figure 3.1.1) indicates that His345 is pointing towards Cys171. Hence we expect that the crystal structure indicates that the KasA enzyme is in the zwitterionic or in the (-) charged state.

A quantification of the orientation of His345 is best obtained through the distance between the ε-nitrogen of His345 and the sulfur of Cys171. Other structural parameters, e.g. dihedral angles involving parts of the imidazole ring and the neighboring CH_2 group turned out to provide less definite answers. The corresponding distance distributions for the four different states are given in Figure 3.1.5C. The peak of the distribution found for the zwitterionic state agrees perfectly with the corresponding measured distance of 3.36 Å. Also for the (-) charged state Cys171 possesses a thiolate group. It is therefore not surprising that

the distance distribution resembles the one found for the zwitterionic state. Note that not only the distance but the whole orientation of His345 is quite similar in both protonation states. However, the maximum of the distance distribution is at a larger distance (≈3.5 Å). The maximum of the distance distribution computed for the neutral state is at about 3.6-3.7 Å. It differs significantly by about 0.3-0.4 Å from the experimentally measured distance. Due to these differences and on the basis of the overall orientation of His345 (Figure 3.1.4 A, B) we rule out the neutral state. Note that the distance distribution function computed for the neutral state is broader than the distance distributions found for the zwitterionic and the (-) state. The increased broadness results from the hydrogen bond between the N_{ε}-H group of His345 and the oxygen center of Wat1. This bond is considerably weaker than the bond formed in the other two states. On the basis of the distance distributions the (+) charged state can be ruled out. Moreover, for this state the water molecule which stayed between Cys171 and His311 at the beginning of the MD simulation moved away. This result would also be in conflict with the X-ray data where a water molecule builds a bridge between His345 and Cys171.

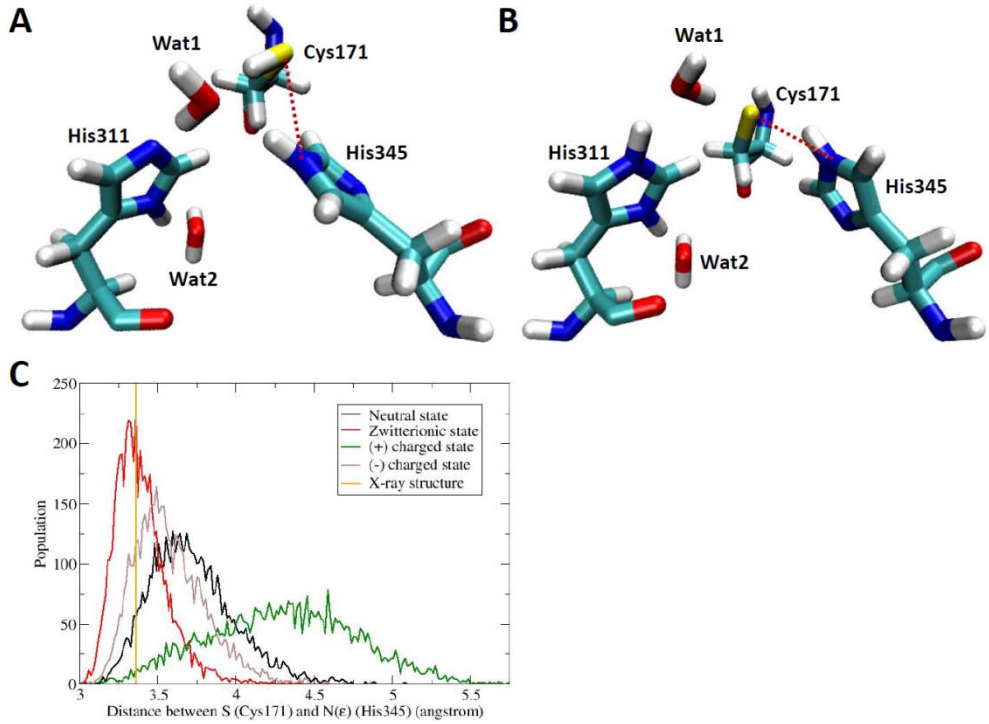


Figure 3.1.5. The change of the relative orientation of Cys171 and His345 according to the change of their protonation state. **A** provides a representative structure of the active site in the neutral state and **B** sketches the situation for the zwitterionic state. (selected by a visual inspection) The distances between the sulfur of Cys171 and the ε-nitrogen of His345, which is shown as a red dotted line in **A** and **B**, are plotted as corresponding distributions in **C**.

FEP calculations to evaluate the perturbed pKa value of His311 - The geometrical data indicate so far that the zwitterionic state is the most likely in the crystal structure. However, the data computed for the (-) charged state do not differ significantly from the experimental data. To substantiate that indeed the zwitterionic state is present, we performed free energy perturbation calculations (FEP) to estimate the pK_a value of His311 in the protein environment. This in turn enables us to determine which of both states is most likely populated at a given experimental condition. The results from these calculations are gathered in the first column of Table 3.1.2 while the thermodynamic cycles designed for these calculations are shown in the upper part of Figure 3.1.6. In the upper process we alchemically protonated the N_ε center of His311 to N_ε-H to obtain a positively charged His311. In the standard approach the influence of the enzyme environment on the pKa value is estimated by $\Delta pKa = (q*\Delta\Delta G) / (2.303*k*T)$ with $\Delta\Delta G=\Delta G_1^P-\Delta G_1^R$ and q represents the charge of the ionized form of the relevant residue. ΔG_1^P and ΔG_1^R are the Gibbs free energies which correspond to the free Helmholtz energies depicted in Figure 3.1.6. Gibbs and Helmholtz free energies are connected by the term $p*\Delta V$ with ΔV being the volume change in the examined processes. The ΔV of both described processes should be very similar. So, the use of $\Delta\Delta A$ instead of $\Delta\Delta G$ should not enlarge the error bars of the computations considerably.

	Upper cycle	Lower cycle
ΔA^P	12.5 ± 1.0	-66.0 ± 0.5
ΔA^R	7.5 ± 1.6	-68.3 ± 0.9
$\Delta\Delta A$	5.0 ± 1.9	2.3 ± 1.0

Table 3.1.2. Free energy difference obtained from the free energy perturbation method. Each column represents calculated values for corresponding thermodynamic cycles shown in Figure 3.1.6. Four independent runs were performed for each calculation, and the standard deviations from these runs are also presented. All energies are given in kcal/mol.

The FEP computations predict a $\Delta\Delta A$ value of 5.0 kcal/mol which corresponds to a pKa shift of 3.5. Due to the resulting pK_a value of 9.5 His311 is expected to be completely

protonated at a pH of 7.5. Even if ΔΔA has been overestimated by a factor of two for a pH of 7.5 His311 would still be mainly protonated. Hence the FEP computations are in line with the geometrical data indicating that the zwitterioinic state is present in the X-ray experiment performed by Luckner et al.[39]

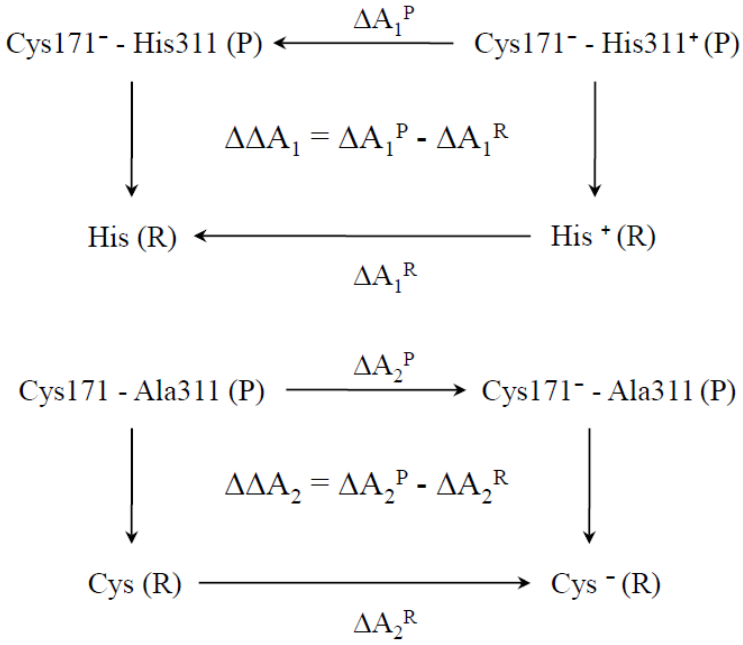


Figure 3.1.6. Thermodynamic cycles devised to obtain perturbed pKa value in the protein environment. The upper cycle is for computing the shifted pKa value of His311 to find the preferred state between the zwitterionic and (-) charged state, and the lower cycle is for the shifted pKa value of Cys171 in the His311Ala mutant. (R) represents the reference system, and (P) represents the protein environment.

Our investigations indicate that the His311/Cys171 dyad adopts a zwitterionic situation. However, if the zwitterionic and the neutral situation differ only slightly in energy the enzyme can switch between both quite easily by a proton transfer mediated by Wat1. Considering the strong changes in the geometrical orientation of His345 its movement could induce the switch. Hence, the energy difference between the neutral and the zwitterionic situation is important for mechanistic considerations. It is also of interest for the development of new drugs since the zwitterionic and the neutral state represent considerably different environments for potential binders. If only one state is energetically accessible the binders should be optimized for this state. If, however, the enzyme can easily switch between both states the binders should be somehow adequate for both situations. The energy difference

between the neutral and zwitterionic state is indirectly accessible from FEP if the neutral state is also compared with the (-) charged state. We performed such calculations but the lack of precision[98] made it impossible to resolve this issue.

Computations to estimate the energy difference between neutral and zwitterionic state - Since the corresponding FEP computations were not sufficiently reliable we decided to substantiate the energy difference between the neutral and the zwitterionic state by means of independent methods. We used QM/MM optimization, which in previous studies were often able to provide reliable information about this topic.[114] As a first step we characterized the proton transfer by computing potential energy surfaces (PESs) using the distance between the ε-nitrogen of the His311 and the hydrogen of Wat1 (N_{His311}-H_{Wat1}) and the distance between the oxygen of Wat1 and the hydrogen of Cys171 as reaction coordinates (O_{Wat1}-H_{Cys171}). Both internal reaction coordinates were varied independently while all other coordinates (in the following called secondary coordinates) are simply relaxed for both reaction coordinates. The results which are summarized in Figure 3.1.7 show that for the KasA enzyme the relative energetic positions of the zwitterionic and the neutral state strongly depend on the preparation of the system. These preparations involve extensive MD simulation to adopt the whole system to the respective state. If the computations are initiated from the system prepared for the zwitterionic state, the zwitterionic state is lower in energy than the neutral state by about 9 kcal/mol. In contrast, if the computations are started from the system prepared for the neutral state, the neutral state is lower in energy by 5 kcal/mol. This indicates that the relative energy of both states strongly depends on the surroundings, therefore these QM/MM optimizations are not able to determine the energy difference between both states reliably. However, they allow some important conclusions about the KasA enzyme. They clearly indicate that both states are quite close in energy and in addition the QM/MM optimization also suggests that the KasA enzyme is able to switch between both states by adapting the orientation of the various residues and water molecules to the respective situation.

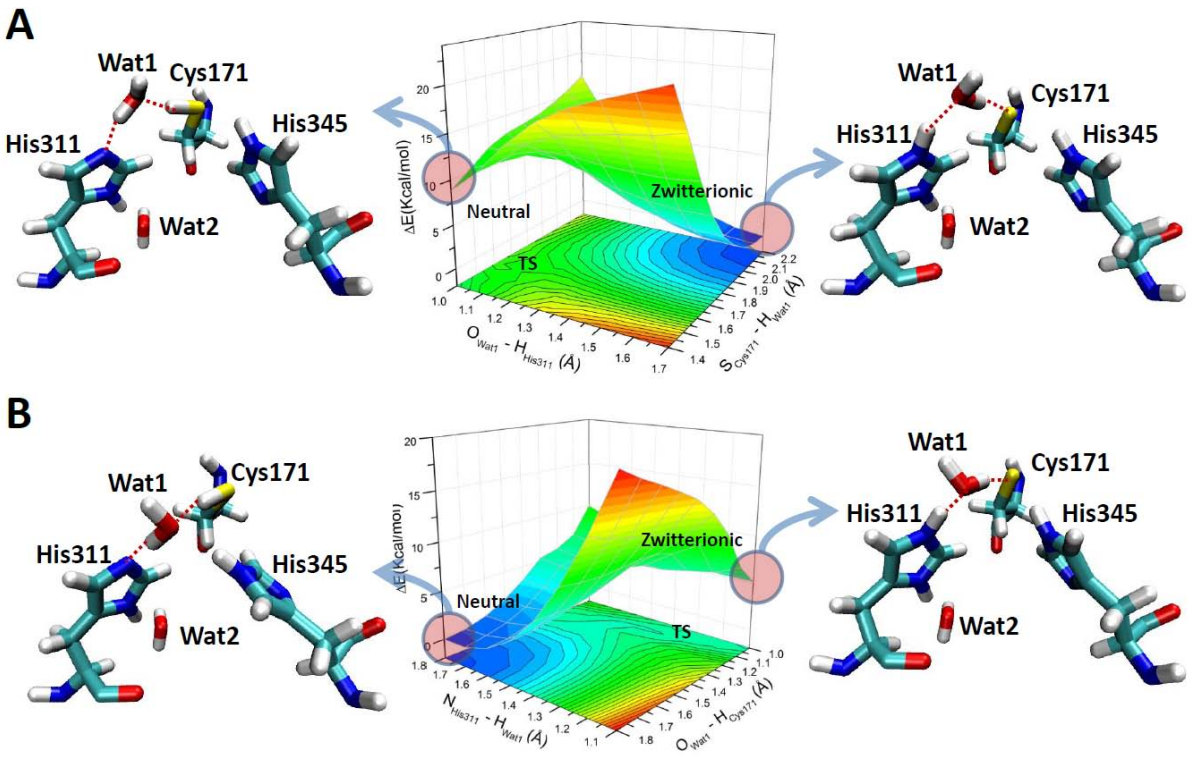


Figure 3.1.7. Potential energy surface (PES) for proton transfer from QM/MM calculations. The starting frames to generate PES in **A** and **B** were taken from the MD simulation of the zwitterionic and the neutral state, respectively (selected by a visual inspection). The structural arrangement of the active site in both end states from each PES is also given, and the hydrogen bonds between QM residues are shown as a red dotted line in the pictures.

An analysis of the geometry variations allows a deeper insight into this capability. The variations in the geometries involve the QM part of the computation containing the bridging water molecule Wat1, His311, and Cys171 (Figure 3.1.8) and the surrounding residues which were taken into account using MM (Figure 3.1.7). Let us first concentrate on the variations found within the QM part (Figure 3.1.8). If the computations are started from the system prepared for the zwitterionic state, one hydrogen atom of the bridging water molecule points towards the thiolate group of Cys171 while the oxygen atom is directed towards the N_{ε}-H group (Figure 3.1.8 blue part on the right hand side). If we now move along the surface given in Figure 3.1.7 part A towards the neutral state the orientation of the Wat1 molecule adopts to the situation (Figure 3.1.8 blue part on the left hand side). It is rotated slightly so that now one hydrogen atom points towards the N_{ε} center of His311 while the

oxygen is directed towards the thiol group of Cys171. It is important to note that the relative orientations of His311, Cys171 and the center of mass of Wat1 do not change considerably.

If the system is prepared for the neutral state we obtain a different geometrical orientation within the QM part (Figure 3.1.8: comparing blue and red part on the left hand side). While the orientations of Wat1 with respect to the N_ε center of His311 and the thiol group of Cys171 closely resemble in the blue and red sketched geometries the relative orientation of His311 and Wat1 with respect to the Cys171 residue is different. Both are shifted downwards. The fact that this orientation is not obtained in the computations starting from the zwitterionic state shows that both geometrical orientations are separated by a reaction barrier. If we now start from the red orientation (Figure 3.1.8 red part on the left hand side) and move along the surface given in Figure 3.1.7 part B towards the zwitterionic state (Figure 3.1.8 red part on the right hand side) the water reorients but again the relative orientations of His311, Cys171 and the center of mass of Wat1 do not adopt to the situation found if the whole system was completely optimized for the zwitterionic state (Figure 3.1.8 blue part on the right hand side). Again blue and red orientations are separated by a potential energy barrier. Please note that the orientation of the water molecule Wat1 also differs.

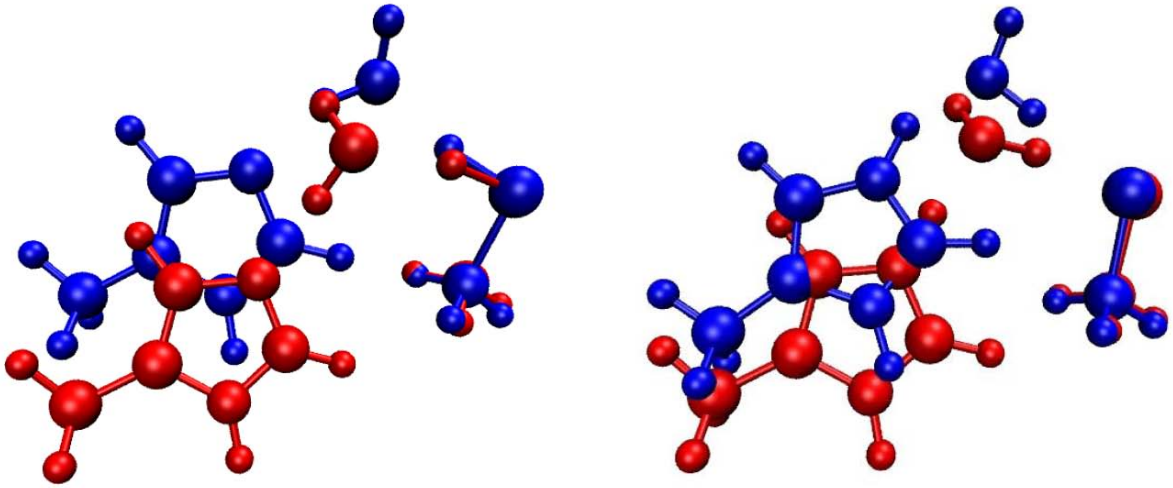


Figure 3.1.8. Geometrical orientations of the QM calculations for the neutral (left) and the zwitterionic states (right). Blue and red indicate the orientations which are obtained if the computations start from the zwitterionic and the neutral state, respectively.

From the surrounding residues His345 shows a quite distinct behavior. If the system is prepared for the zwitterionic state (Figure 3.1.7 A right hand side) the N_ε-H group of His345

points towards the thiolate group of Cys171. Moving along the computed surface to the neutral state the N_{ε}-H group of His345 is still oriented towards the sulfur center of Cys171. However, if the system is prepared for the neutral situation (Figure 3.1.7 B left hand side) His345 is oriented differently. In this geometry it points towards the oxygen center of Wat1. Again both geometrical orientations sketched in Figure 3.1.7 on the left hand side (A *vs.* B) are obviously separated by a barrier so that the lower one is not reachable by a simple relaxation of all secondary parameters. If we now start from the system prepared for the neutral state and move towards the zwitterionic state (Figure 3.1.7 B) His345 reorients. But perhaps due to its influence Wat1 does not adopt the orientation which is found if the system was prepared for the zwitterionic state (Figure 3.1.7 A *vs.* B on the right hand side).

The analysis of the various geometries indicates that KasA can switch between the zwitterionic and the neutral state quite easily since it only needs slight modifications to stabilize one state with respect to the other. The analysis indicates that the movements of His345 and Wat1 play a central role, however, the orientations of other residues also contributes and remain to be analyzed in details.

Theory	ΔE_1	ΔE_2
AM1	32.3	64.7
PM3	37.4	50.6
PDDG-PM3	39.1	61.9
SCC-DFTB	Not converged	-
RM1	25.0	41.5
B3LYP	22.9	61.2
B3LYP-D	22.8	60.8

Table 3.1.3. Comparison of the results from single point calculations in gas phase between semi-empirical and DFT methods. ΔE_1 represents the energy difference between the neutral and the zwitterionic state. ΔE_2 represents the energy difference between the neutral state and the intermediate as indicated in Figure 3.1.9. All energies are given in kcal/mol.

The analysis also indicates that the potential energy surface describing the proton transfer from Cys171 to His311 is very rugged so that a projection of the multi-dimensional PES on the subspace spanned by the distances N_{His311}-H_{Wat1} and O_{Wat1}-H_{Cys171} is not sufficient to estimate the energy difference between the neutral and the zwitterionic state. To overcome this limitation we used the umbrella sampling method in combination with a QM/MM scheme. It is able to overcome the mentioned barriers since it does not only relax all secondary parameters but performs a MD-sample for each point of the hypersurface. Furthermore it can be used to compute the free energy. Due to the necessary sampling currently available *ab-initio* methods are too expensive for this type of calculations, therefore semi-empirical methods have to be employed for the QM part. To estimate the accuracy of available approaches we compared the predictions of several semi-empirical methods for the energy difference between the neutral and the zwitterionic state with the corresponding result of the B3LYP/TZVP approach. The computations were performed for the QM part of our previous QM/MM optimizations neglecting further environmental effects (Table 3.1.3). The results indicate that the RM1 method seems to be well suited to estimate the energy difference between the zwitterionic and the neutral form. The comparison also underlines that RM1 strongly overestimates the stability of intermediate states which involve hydronium ions. We started the umbrella sampling at the system prepared for the neutral state to exclude a bias towards the zwitterionic state.

The corresponding PMF surface for proton transfer is shown in Figure 3.1.9. The reaction coordinates in this computation were defined as the difference between two internal coordinates. For the first reaction coordinate, the difference between the internal coordinates S_{Cys171} - H_{Cys171} and $H_{Cys171} - O_{Wat1}$ was used. The second reaction coordinate represents the difference between $O_{Wat1} - H_{Wat1}$ and $H_{Wat1} - N_{His311}$. The predicted stable intermediate represents an artifact of the RM1 method. As already mentioned, this method strongly overestimates the relative stability of this part of the surface (Table 3.1.3) but should be well suited to estimate the energy difference between neutral and zwitterionic states. The computations predict that the zwitterionic state is about 0.4 kcal/mol lower in energy than the neutral state. However, the zwitterionic state should be slightly more stable. RM1 seems to underestimate the relative stability of the zwitterionic state with respect to the neutral one. In vacuum computations summarized in Table 3.1.3, this underestimation amounts to about 2 kcal/mol in comparison to the B3LYP/TZVP approach. Transferring this error to the situation

within the enzyme the zwitterionic state would be 2-3 kcal/mol more stable than the neutral state. This may represent an upper bound for the energy difference since the error may be less in the enzyme due to error cancellations.

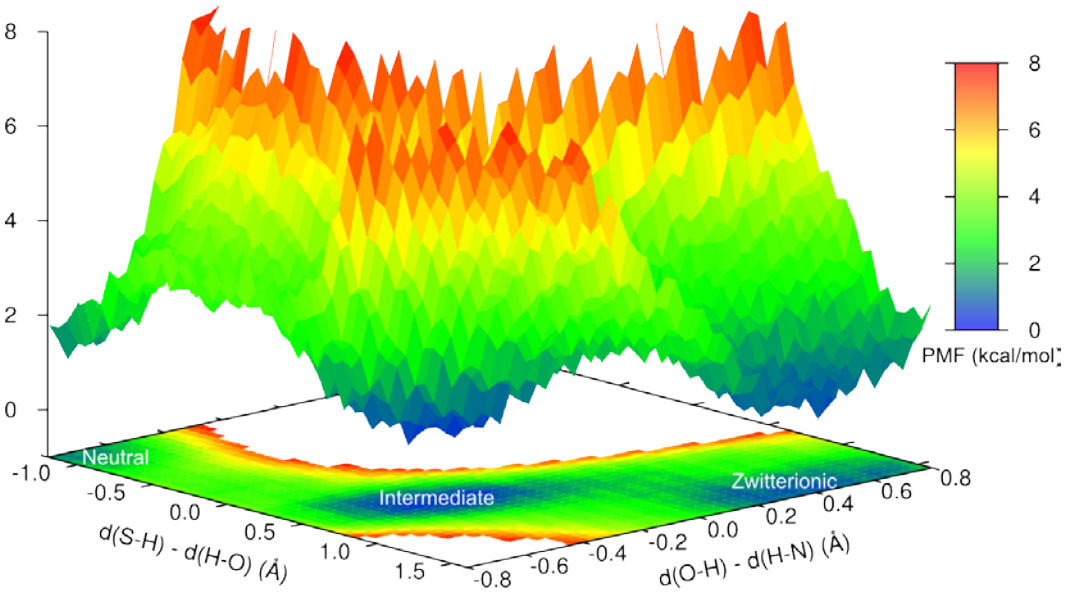


Figure 3.1.9. PMF surface for proton transfer generated from QM/MM computations in combination with the umbrella sampling method. Those parts of the surface which are more than 8 kcal/mol higher in energy than the zwitterionic state were omitted for a better description of the surface of interest.

Nevertheless, such small energy differences are still within the uncertainties of the employed methods. Hence, it is important to note that the predictions obtained from the PMF surface agree nicely with those based on our simulation of the geometrical parameters. These simulations indicated that the zwitterionic state is observed in the X-ray data and they point out that neutral and zwitterionic states differ in several geometrical parameters. Due to these differences a population of both states should be seen in the X-ray data if the neutral state is populated by at least 10 – 20 %. Using the Boltzmann distribution for the temperature at which the X-ray experiments were performed (100 K) such populations can be converted to an energy difference of 0.4-0.3 kcal/mol between the zwitterionic and the neutral states. Hence, since the X-ray experiments do not indicate that both states are populated their energy difference must be at least 0.3-0.4 kcal/mol. These values represent the lower bound for the energy difference and again agree nicely with our computations. Please note, that both states

should be populated at a body temperature of 320 K. For the lower bound the neutral state will be populated by nearly 40 %. At an energy difference of 2 kcal/mol (upper bound) the neutral state would still be populated by about 5 %.

Charge deletion analysis to estimate the contributions from each surrounding residue to stabilize the zwitterionic state - The stability of the zwitterionic state with respect to the neutral state is determined by additional residues. This is summarized in Table 3.1.4 with the results of a charge deletion analysis on the relative energies of both states. Note that this table can not include the influence of His311 and Wat1 since they belong to the QM part of the QM/MM calculations. The moieties of Glu322, Lys340, and His345 seem to have the biggest effect. The influence of Glu322 and His345 are obvious since they are in direct contact to either Cys171 or His311. His345 stabilizes the negatively charged Cys171 through a hydrogen bond. Glu322 seems to enhance the basicity of His311, thereby promoting the abstraction of the proton from Cys171 by His311. Lys340 destabilizes the zwitterionic state since it is positively charged itself. The size of this effect could result from a direct interaction of Lys340 with the carbonyl backbone group of His311 and indirectly via Wat2. One would expect a stronger influence of the oxyanion hole which consists of the NH-backbone groups of Phe404 and Cys171. However, if these charges are switched off, the missing influence could be compensated by other residues in close contact with the thiolate group. This would be in line with the polarizability of the sulfur center. It is also possible that the stabilization effect of the oxyanion hole is similar for the zwitterionic and the neutral state.

Residue switched off	ΔE (Zwitterionic surroundings)	ΔE (Neutral surroundings)
α – helix except Cys171	0.7	0.5
NH group of Cys171	2.6	3.5
Thr313	-1.0	-0.2
Glu322	7.6	6.9
Lys340	-5.7	-6.1
His345	6.2	4.9
Glu354	2.2	4.5
Ser401	0	0.4
Phe402	0.2	-1.9
Phe404	1.8	0.9
Wat2	1.5	3.5

Table 3.1.4. The contributions from surrounding residues in the MM part of the QM/MM computations to stabilize the zwitterionic state relative to the neutral state. These energy contributions were calculated using the charge deletion method. All energies are given in kcal/mol. In both columns, the positive sign means that a given residue contributes to the stabilization of the zwitterionic state.

3.1.4 The molecular reason of the acyl-transfer activity in the His311Ala variant

MD simulations to investigate the unexpected acyl-transfer activity in the His311Ala variant – All MD simulation results so far support the idea that the zwitterionic state is most probably the resting state. Additionally QM/MM computations showed that the zwitterionic state is easily achievable by proton transfer from Cys171 to His311. However, previous mutational experiments in FabF showed that the His303Ala (corresponds to His311Ala in KasA) variant remains active as an acyl-transferase.[36] This result implies that Cys171 in KasA, could also be deprotonated regardless of His311. To address this issue, we performed MD simulations for the His311Ala mutant and compared them with the simulations on the

wild type enzyme in the zwitterionic state. Originally, there are four surrounding residues which form hydrogen bonds with the thiolate ion of Cys171 in wild type KasA. Those are His345, the backbone amides of Cys171 and Phe404, and His311 via a water molecule. In the MD simulation results of the His311Ala mutant, the thiolate ion in Cys171 frequently forms temporal interactions with surrounding water molecules. These interactions will also stabilize the negatively charged form but there was no stable state that lasted for a noticeable time period as in the wild-type protein. The other hydrogen bonds provided by His345 and the backbone amide of Cys171 maintained almost the same distance distributions as in the wild type protein as shown in Figure 3.1.10. However, a remarkable shift to shorter distances was observed in the distance distribution of the hydrogen bond formed between the thiolate and the backbone amide of Phe404. This hydrogen bond stabilizes the thiolate form in the His311Ala variant considerably more compared to the wild-type protein and hence compensates for the missing His311. This strong hydrogen bond was possible due to a conformational change of the side chain of Phe404 as shown in Figure 3.1.10.

FEP calculations to evaluate the perturbed pKa value of Cys171 in the His311Ala mutant – To further analyze the above mentioned hypothesis we used the FEP method to compute the pKa shift of Cys171 in the His311Ala variant. Since the conformational change of the side chain of Phe404 was expected to be critical for shifting the pKa value of Cys171, one of the frames after this conformational change had taken place was used as a starting structure for the FEP calculations. The corresponding thermodynamic cycle is provided in the lower part of Figure 3.1.6. The results are shown in the second column of Table 3.1.2. The resulting pKa value of Cys171 in the His311Ala mutant is around 6.7. If Phe404 stays in a similar position as in the wild type protein the corresponding pKa value is around 8.3. In line with our structural results this indicates that the mutant can be active since the stabilization of the thiolate is compensated by Phe404.

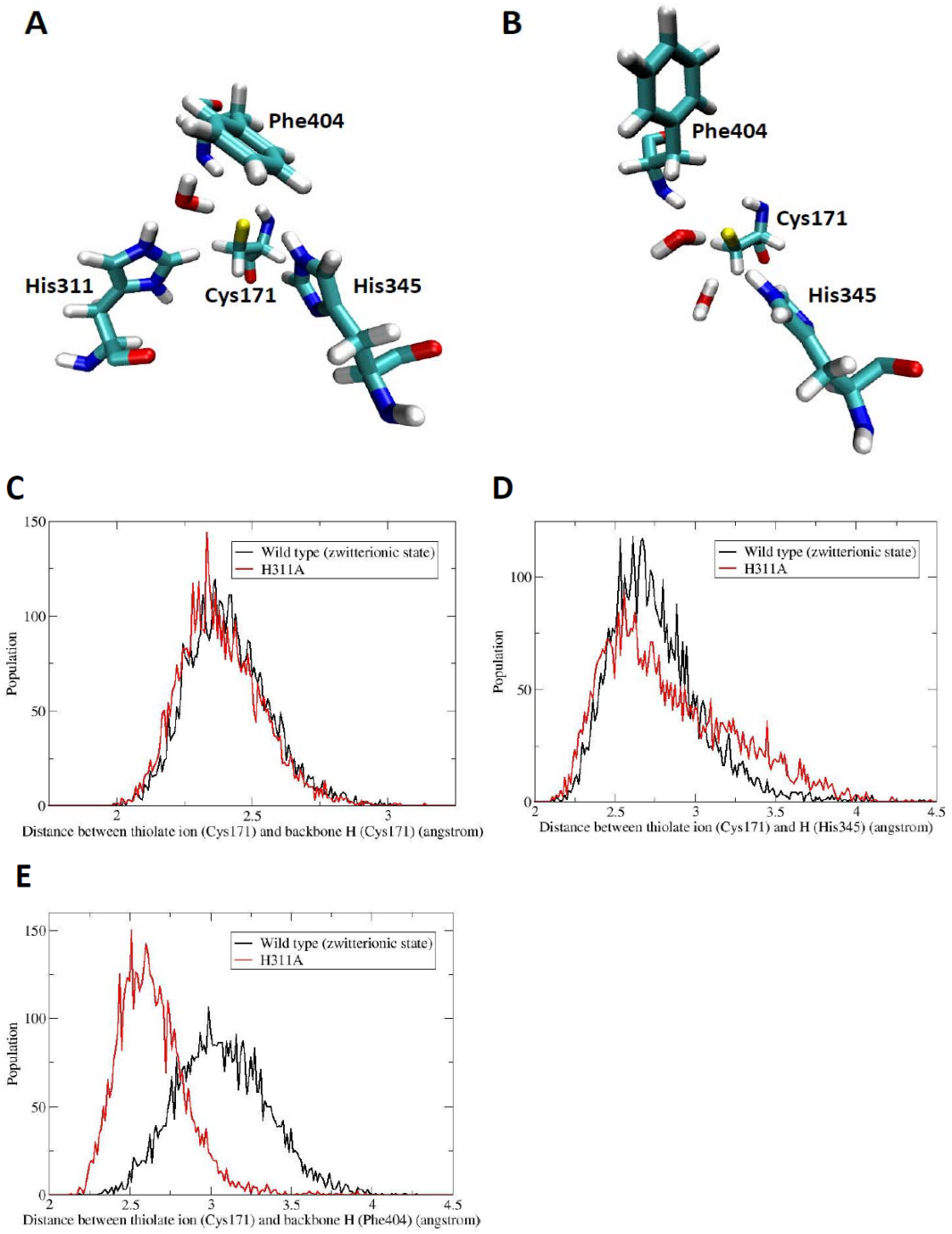


Figure 3.1.10. The difference in the hydrogen bonding network between wild type KasA (zwitterionic state) and the His311Ala mutant. The geometrical arrangement of Cys171 and surrounding residues stabilizing the thiolate form of Cys171 in wild type KasA and His311Ala are shown in **A** and **B**, respectively. The distance distributions for hydrogen bonds provided by the backbone amide of Cys171, His345, and the backbone amide of Phe404 are plotted in **C**, **D**, and **E**, respectively.

3.1.5 Discussion

KasA represents an attractive drug target against tuberculosis since its depletion gives rise to cell lysis and cell death. The development of new agents requires detailed information about the structure and the function of the active site. Structural data are the methods of choice for this analysis but the positions of hydrogen atoms are in most cases not determinable and the crystal structure can only characterize the energetically lowest conformer. The present study provides additional information by means of MD simulation, FEP and QM/MM computations in combination with the umbrella sampling method. Using this combined methodology, we were able to shed light on the protonation state of the active site of KasA. Furthermore, a picture emerges which indicate how different moieties (water molecules and residues) influence the functionality of the enzyme.

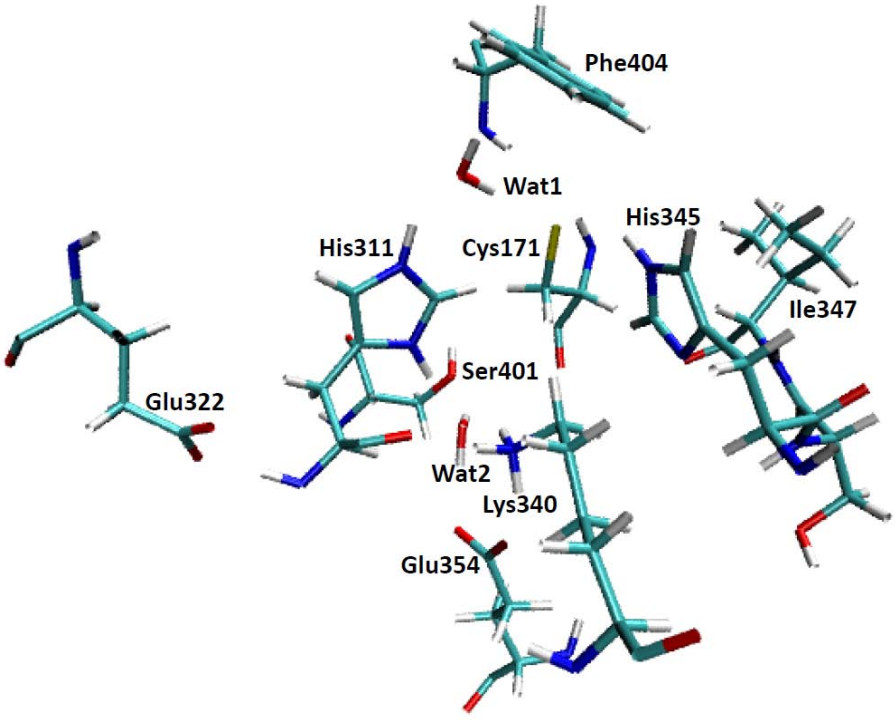
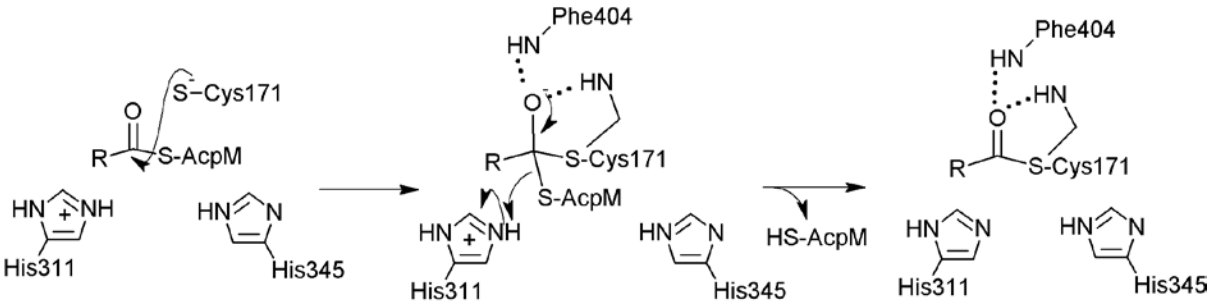


Figure 3.1.11. Geometrical orientation of the zwitterionic resting state of KasA predicted by the present work.

The simulations were initiated from the X-ray data provided by Luckner et al.[39] To obtain information about the protonation status we performed MD simulations for all possible tautomers and compared the computed structural behavior with the experimental X-ray data. The best agreement was obtained for the protonation status of the zwitterionic state as

indicated in Figure 3.1.11. The data obtained for the (-) charged state (Figure 3.1.5) also do not deviate strongly but FEP computations reliably showed that the zwitterionic state is lower in energy. In line with the structural data QM/MM computations in combination with the umbrella sampling method also predict the zwitterionic state to be lower in energy than the neutral state. However, the energy difference seems to be so small that both states should be populated at body temperature. An analysis of the structural changes also indicates that the enzyme can easily switch between both states.

The MD simulation also allowed insights into the role of the different moieties in the vicinity of the active site. They reliably showed that a hydrogen bonding network including Glu354, Lys340, Ser401, and Wat2 fixes the imidazole ring of His311 in the position needed for the enzymatic reaction. In this network Wat2 represents the anchor for His311 while the other residues keep the Wat2 molecule in its position. An additional hydrogen bond between Lys340 and the backbone carbonyl group of His311 provides extra stability. However, MD simulations showed without doubt that Wat2 is most important. If it is removed, the orientation of the active site changes in a way that in all probability the enzyme becomes inactive.



Scheme 3.1.2. Schematic illustration of the mechanism for the acyl transfer involving a tetrahedral KasA I-fatty acid-ACP complex intermediate.

Our computational analysis predicts that the zwitterionic state given in Figure 3.1.11 represents the resting state of the KasA enzyme. This supports the mechanism for the acyl transfer step of the Claisen condensation suggested by Olsen et al.[35] The acyl transfer starts with an attack of the negatively charged Cys171 sulfur center which is present in the zwitterionic but not in the neutral state. Furthermore, the zwitterionic state can catalyze the elimination of the S-AcpM group by providing the proton from the N_{ε} center of His311

(Scheme 3.1.2). This framework makes the overall reaction thermodynamically more favourable since a negatively charged thiolate (Cys171) attacks while a neutral thiol (HS-AcpM) is eliminated. Due to the proton transfer to the leaving group the imidazole ring of His311 can accept the proton from Malonyl-ACP. This transfer represents the initial step of the decarboxylation step of the Claisen condensation.

The charge deletion analysis shows that the contribution from the helix macrodipole to stabilize the zwitterionic state was very small in comparison to the effects arising from other residues. This is contrary to the idea that this helix macrodipole is the main reason why the catalytic cysteine residues are often deprotonated. However, a similar case has been reported for cathepsin B.[43] The remarkable contribution of the Glu322 residue to the stabilization of the zwitterionic state is another surprising result since this residue is positioned near the backbone of His311, i.e. far from Cys171 (Figure 3.1.11). In the wild type protein Glu322 does not directly stabilize the thiolate form of Cys171. It enhances the basicity of His311 so that an abstraction of the proton from Cys171 by His311 becomes possible. Through its water mediated hydrogen bond His311 in turn stabilizes the thiolate moiety of Cys171. In the His311Ala mutant this stabilization is no longer possible but the negatively charged thiolate remains since it is stabilized through its interaction with Phe404. Thus, the acyl-transfer remains feasible in His311Ala.

3.2 The mechanism of the decarboxylation reaction in KasA

The decarboxylation reaction is a controversial step in the catalytic reaction of KasA. However, the correct protonation state of His311 in the acyl-enzyme state is expected to provide a decisive clue in determining the correct decarboxylation reaction because it is a key to narrow down the possible decarboxylation mechanisms suggested so far. Therefore, the protonation state of His311 in the acyl-enzyme state is mainly investigated in this chapter using various computational methods, thereby determining the most probable mechanism for the decarboxylation reaction. In addition, each suggested mechanism is also critically reviewed in this chapter on the basis of known experimental results which were obtained for various mutants and the new computed data which were obtained from the simulations of the real acyl-enzyme state.

3.2.1 Computational details

The model systems for the acyl-enzyme state were built from the X-ray crystal structure of the KasA mutant (Cys171Gln) complexed with the inhibitor thiolactomycin (TLM) (PDB entry: 2WGG). The inhibitor was removed and Gln171 was modified into an acylated cysteine group. The acyl chain has been approximated by a methyl group. The partial charges for this acyl group were generated with AM1-BCC method using the antechamber module[115] of AMBERTOOLS 1.4. The corresponding force field parameters were taken from General AMBER Force Field[116] (GAFF) and assigned by antechamber. All MD simulations were performed using the sander module of AMBER 11[117] in combination with ff99SB parameters.[106] The surrounding water molecules were taken explicitly into account as TIP3P water.[93] The modified protein structures were solvated in a water shell with a radius of 45 Å using spherical boundary conditions to retain the water molecules inside the shell. After solvation, the system was neutralized by adding sodium ions. Solvation and neutralization have been done with the help of the leap module of AMBERTOOLS 1.4. The created systems were subsequently minimized for 500 steps to avoid any steric clashes. After the minimization, the equilibration for the solvent was carried out by gradually increasing temperature with a solute restrained until the temperature reached 310 K. Then, the restrain applied to the solute was gradually reduced to zero, and equilibration dynamics was followed for 1 ns. The productive MD was performed for 1 ns after the equilibration dynamics.

Langevin thermostat was used for temperature control. For all MD simulations, we used a time step of 1 fs. For non-bonded interaction, a cut-off radius of 12 Å was employed. During force field MD simulations performed for a neutral Lys340/Glu354 pair, the hydrogen bond broke apart leading to a collapse of the hydrogen bond network. This was not found in the corresponding computations involving an ionic Lys340/Glu354 pair. To arrive at a reliable description in the force field simulations of the neutral Lys340/Glu354 pair, we applied a harmonic restraint to the Lys340/Glu354 neutral pair. This restraint was applied in FEP computations as well.

The pKa value of the catalytic residues were determined by means of the alchemical FEP method[96] employing NAMD 2.7 simulation package[89] in combination with the CHARMM 22[90] force field including CMAP correction.[91] Force field parameters and partial charges for acylated Cys171 were borrowed from a thioester group.[118] As reference systems, we used a single histidine residue capped with a neutral blocking group solvated in a water shell with a radius of 22 Å. Minimizations and equilibrations were performed in the same way as mentioned above. A time step of 2 fs was used in combination with the RATTLE algorithm[95] to constrain the bonds involving hydrogen atom. A cut-off radius of 18 Å was used for non-bonded interaction. Total 100 windows with even width were introduced for intermediate mapping potentials between the start and the end point, and the dual topology approach[119,120] was employed to describe each end point. At each window, 15 ps of equilibration and 40 ps of productive run for data collection were performed. To address the accuracy of the computations,[98,121] each FEP computation was conducted in one direction and repeated four times. To circumvent the end-point catastrophe at both end points, the soft core potential was employed.[122–124]

For QM/MM MD simulations, we used the developmental version of AMBER 11 and Terachem (v1.46-dev).[125,126] We used the same conditions as employed for the force field MD simulations mentioned above. For the QM region, we used the BLYP functional[127–129] with dispersion corrections[130] in combination with 6-31G** basis set.[131–134] In assigning the QM region for amino acids, the link atom approach using hydrogen was employed. The QM/MM boundary was set at the bond between the α and β carbon of histidine, the β and γ carbon of glutamate, and the δ and ε carbon of lysine. The MM excess charge of the QM region is distributed equally over all of the MM atoms in the QM/MM simulation to conserve the original total system charge. For the MM region, the AMBER ff99SB parameter set was

employed. For the umbrella sampling for the proton transfer between Lys340 and Glu354, we used the same condition as for the QM/MM MD simulations. The initial structures for umbrella sampling were obtained from 1 ns long MM/MD runs. The reaction coordinate was defined as the difference between the bond lengths between the nitrogen atom of Lys340 and the transferred proton, and the bond length between the oxygen atom of Glu354 and the transferred proton. Each window comprised a 10 ps equilibration and a subsequent 10 ps productive run for data collection. After biased simulations, the potential of mean force (PMF) profiles were obtained with weighted histogram analysis (WHAM)[135–137] using the code by Grossfield.[107]

3.2.2 The influence of the Lys340/Glu354 pair on His311

The correct protonation state of His311 in the acyl-enzyme state is a prerequisite to determine the right mechanism for the decarboxylation step. After the acylation step, His311 is supposed to be neutral with a protonated N_δ because it is expected to hand over the proton of the N_ε to the leaving ACP in order to facilitate the acylation step (Figure 2.4).[138] However, since KasA functions by a ping pong mechanism, it is quite probable that His311 changes its protonation state before the second substrate is bound to the active site. To determine the preferred protonation state of His311 in the acyl-enzyme state, we performed alchemical FEP under the generally believed assumption that the Lys340/Glu354 pair is zwitterionic.[21,37,41,139] In this computation, the pKa values of N_δ and N_ε of a positively charged His311 in the acyl-enzyme state were evaluated separately using the two thermodynamic cycles shown in Figure 3.2.1, and the computed free energy differences are gathered in the left part of the Table 3.2.1.

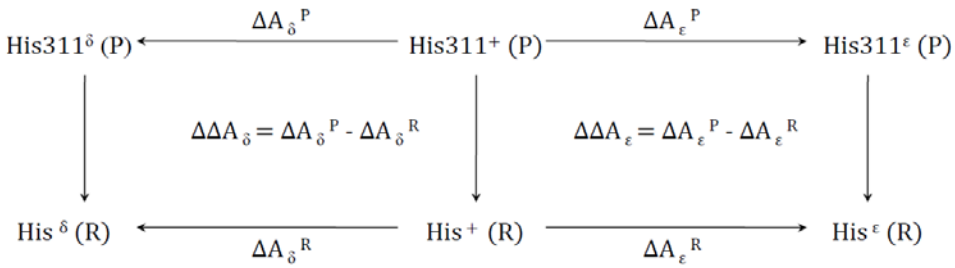


Figure 3.2.1. Thermodynamic cycle to obtain pKa values at different protonation sites of His311. (R) represents the reference system where a histidine residue is solvated in water shell, and (P) represents the protein environment. The cycle at the left is devised to compute the pKa value of N_ε of His311, and the cycle at the right is for the pKa value of N_δ.

	N_δ position	N_ε position
ΔA^P	28.9 ± 2.3	38.0 ± 0.9
ΔA^R	37.3 ± 0.8	38.1 ± 1.0
$\Delta\Delta A$	-8.4 ± 2.4	-0.1 ± 1.3
Approximate pKa	0	6.0

Table 3.2.1. FEP computation results to evaluate pKa values of N_δ and N_ε of His311 with an ionic Lys340/Glu354 pair. All energies are given in kcal/mol.

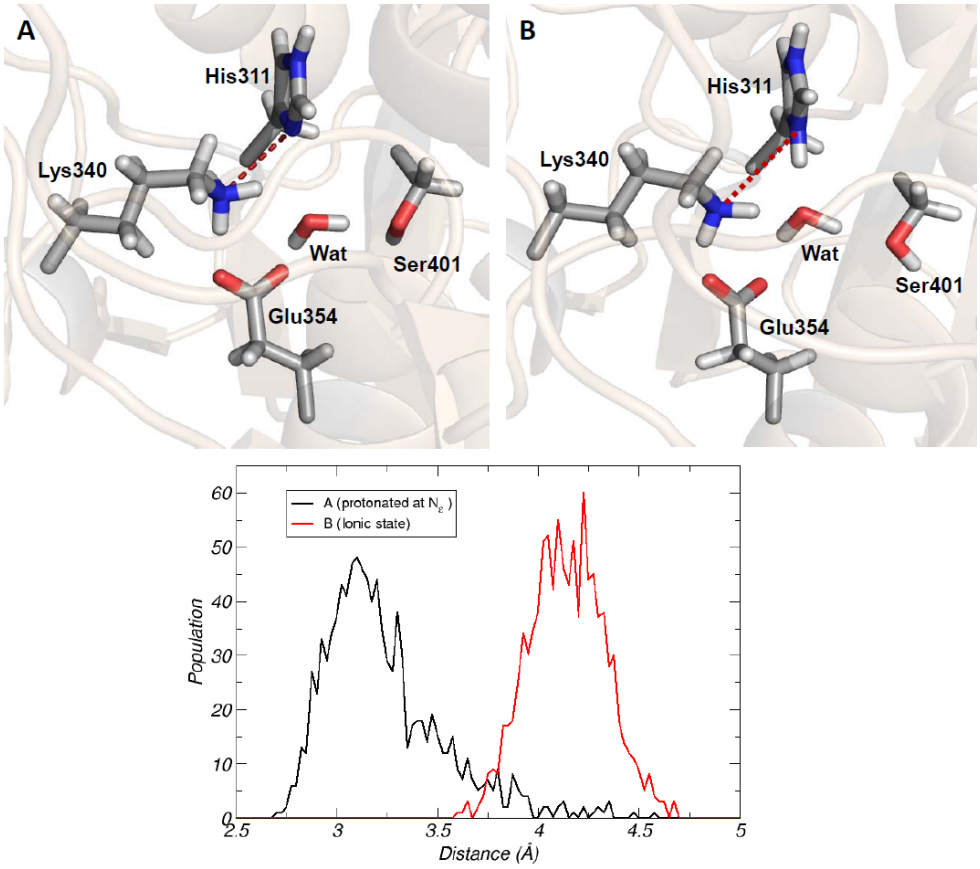


Figure 3.2.2. Interaction of N_δ of His311 with N_ζ of Lys340 in the acyl-enzyme state. The curves in the bottom represent the distributions of the distances between N_δ of His311 and N_ζ of Lys340, which are also marked as red dotted lines in the pictures above. Picture **A** and **B** correspond to the curve **A** and **B** shown in the graph, respectively.

The data indicate that, for an ionic Lys340/Glu354 pair, the deprotonation of the doubly protonated His311 at N_δ (pKa≈0) is considerably more favorable than at N_ε (pKa ≈ 6.0). This implies that for a neutral environment, (pH≈7) His311 is predicted to be neutral and only protonated at N_ε. Figure 3.2.2., which plots the distances between N_δ of His311 and N_ζ of Lys340 during MD simulations (1 ns) of the appropriate systems, reveals that the low pKa value of N_δ of His311 results since a hydrogen bond can be formed between the deprotonated N_δ center and the protonated N_ζ center of Lys340. Both centers repel each other when the N_δ of His311 is protonated.

Our 1ns of MD simulations clearly show that the pKa value of N_δ of His311 is strongly influenced by the electronic structure of the Lys340/Glu354 pair which was assumed to be zwitterionic. To investigate the reliability of this assumption, we performed two QM/MM MD simulations with different QM regions using the BLYP functional[127–129] in combination with 6-31G** basis sets.[131–134] We used this DFT-based approach because semi-empirical approximations were recently found to be too inaccurate for reliable predictions of relative energies of neutral and zwitterionic states connected by proton transfer.[138,140] Furthermore, the BLYP functional has been extensively used in a large number of computations to study the proton transfer occurring within biological molecules.[141–149] The use of even more reliable approaches failed because they are way too expensive. Standard QM/MM computations of the proton transfer potential, which perform simple geometry optimization instead of short MD simulations but employ more accurate approaches for the QM region, do also not offer ways out. They lead to large errors in the relative energies because the adaptation of the enzyme surroundings to the varying charge distribution of the Lys340/Glu354 pair is not described properly. The QM region of the first simulation comprises His311, Glu354, and the bridging water (Figure 3.2.3A). For the second computations, the Lys340 was further added to the QM part (Figure 3.2.3B). Ser401 is also part of this hydrogen bond network, but its role seems to be less important. The hydrogen bond between Ser401 and the water molecule turned out to be not stable during MD simulation due to the excessive flexibility of its hydroxyl side chain. Furthermore, Ser401 is not conserved in the KasA enzyme family. For these reasons, Ser401 has not been included in the QM part. The results of the simulations are given in Figure 3.2.3 which monitors the distances of interest during the QM/MM MD simulation (20ps).

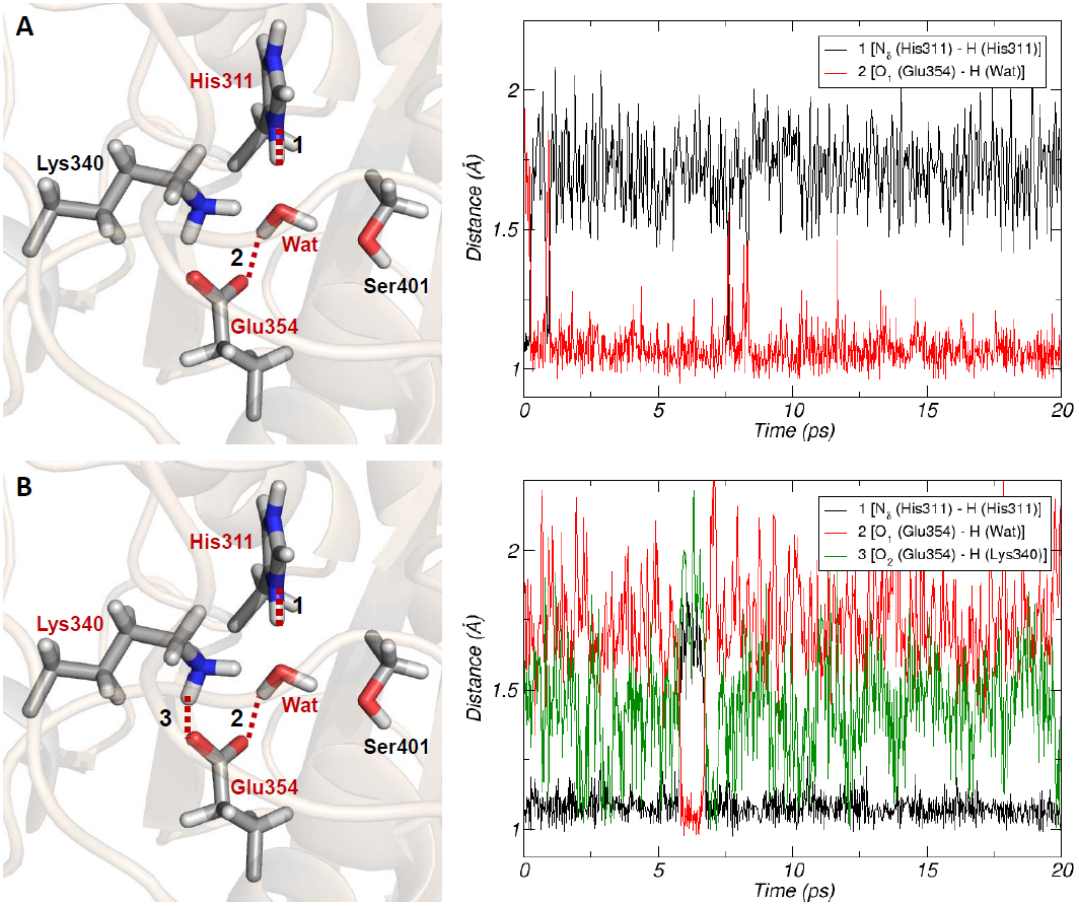


Figure 3.2.3. Trajectories of protons involved in hydrogen bond network in the acyl-enzyme state. Each picture corresponds to each graph in the same row. In the pictures shown left, the monitored distances are shown as red dotted lines and numbered according to the legends in the graphs. The names of the residues involved in QM region are represented in red in the pictures.

If Lys340 is not included in the QM region (Figure 3.2.3A), the proton is directly transferred from N_δ of the ionic His311 to Glu354 using the water molecule as a bridge. Until the end of the simulation, no back transfer takes place. This result implies that the pKa value of N_δ of His311 is considerably lower than that of Glu354 and supports the FEP computation results in which the pKa value of N_δ of His311 is perturbed considerably by the protonated Lys340 group. In contrast, if the Lys340 is included in the QM region, the proton remains at the N_δ of the ionic His311. Instead, the simulation shows a frequent hopping of the proton between Lys340 and Glu354 indicating that this prevents the proton hopping from His311 to Glu354. This proves that the protonation state of His311 relies on the electronic structure of

the Lys340/Glu354 pair (zwitterionic vs. neutral). Taking their quality into account, these computations also strongly question the generally accepted zwitterionic situation for Lys340 and Glu354.[21,37,41,139] For reliable conclusions about the decarboxylation mechanisms, it is therefore necessary to determine the pKa values of His311 for a neutral Lys340/Glu354 pair.

The evaluation of the pKa values of His311 with a neutral Lys340/Glu354 pair was achieved by repeating the FEP calculations discussed in Figure 3.2.1 but assuming a neutral Lys340/Glu354 pair. The computed free energy differences are gathered in the Table 3.2.2. Comparing these results with the data obtained for a zwitterionic Lys340/Glu354 pair (Table 3.2.1), the situation for the N_ε of His311 obviously does not change considerably (6.0 vs. 6.5). For N_δ, however, the computed $\Delta\Delta A$ value corresponds to a pKa value of about 2.9, i.e. the pKa value increases by three units in comparison to the ionic Lys340/Glu354 pair (pKa≈0). Nevertheless, N_δ still remains quite acidic so that these computations also predict that His311 should be neutral and protonated at N_ε regardless of the electronic structure of the Lys340/Glu354 pair.

	N_δ position	N_ε position
ΔA^P	32.8 ± 2.0	38.8 ± 2.5
ΔA^R	37.3 ± 0.8	38.1 ± 1.0
$\Delta\Delta A$	-4.5 ± 2.2	0.7 ± 2.7
Approximate pKa	2.9	6.5

Table 3.2.2. FEP computation results to evaluate pKa values of N_δ and N_ε of His311 with a neutral Lys340/Glu354 pair. All energies are given in kcal/mol.

3.2.3 The influence of His311 on the Lys340/Glu354 pair

Our QM/MM MD simulations indicate that the Lys340/Glu354 pair frequently switches between the neutral and zwitterionic situation, but delivers no information about the energy difference between both situations. To compute the missing data, we performed QM/MM umbrella sampling to generate the PMF profile for the proton transfer between Lys340 and Glu354 in the acyl-enzyme state. For these calculations, we also used the

BLYP/6-31G** approach for the QM region. To address the adequacy of the employed level of theory for simulating proton transfer between Lys340 and Glu354, we performed a benchmark calculation first using QM/MM optimizations prior to the QM/MM umbrella sampling (Figure 3.2.4). The reference level of theory employed was LCCSD(T)/cc-pVQZ, and His311, Lys340, Glu354 and a water molecule were included in the QM region. Each point on the reaction coordinate was first optimized with the BLYP/6-31G** level of theory, and then single point calculations were performed with the reference level of theory. The result is shown in Figure 3.2.4, and overall, the BLYP/6-31G** approach shows a quite good agreement with the reference level of theory along the reaction coordinate.

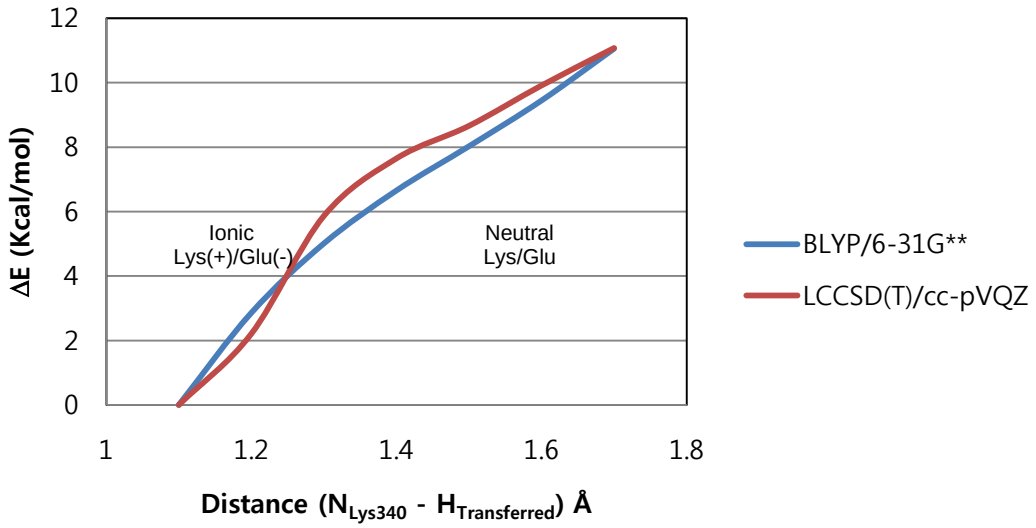


Figure 3.2.4. The benchmark calculation for the proton transfer between Lys340 and Glu354.

Subsequently, the PMF profiles were generated for all possible protonation states of His311. The QM region incorporates His311, Lys340, Glu354 and a water molecule while the influence of the remaining enzyme was treated with MM. For all protonation states of His311, the computed PMF for the proton transfer (Figure 3.2.5) has its minimum close to the ionic form of the Lys340/Glu354 pair. With the ionic form of the Lys340/Glu354 pair, our FEP calculations predict pKa values of 0 and 6.0 for N_δ and N_ε in His311, respectively. The neutral form of the Lys340/Glu354 pair is predicted to be less than 1kcal/mol higher in energy. Due to this small energy difference, Boltzmann statistics give an ionic vs. neutral population of about 60:40. Nevertheless, even in the case where the Lys340/Glu354 pair is neutral, the difference in the pKa values between both nitrogen centers of His311 is still large

indicating that His311 should be protonated in the acyl-enzyme state in any case at N_ε but not at N_δ .

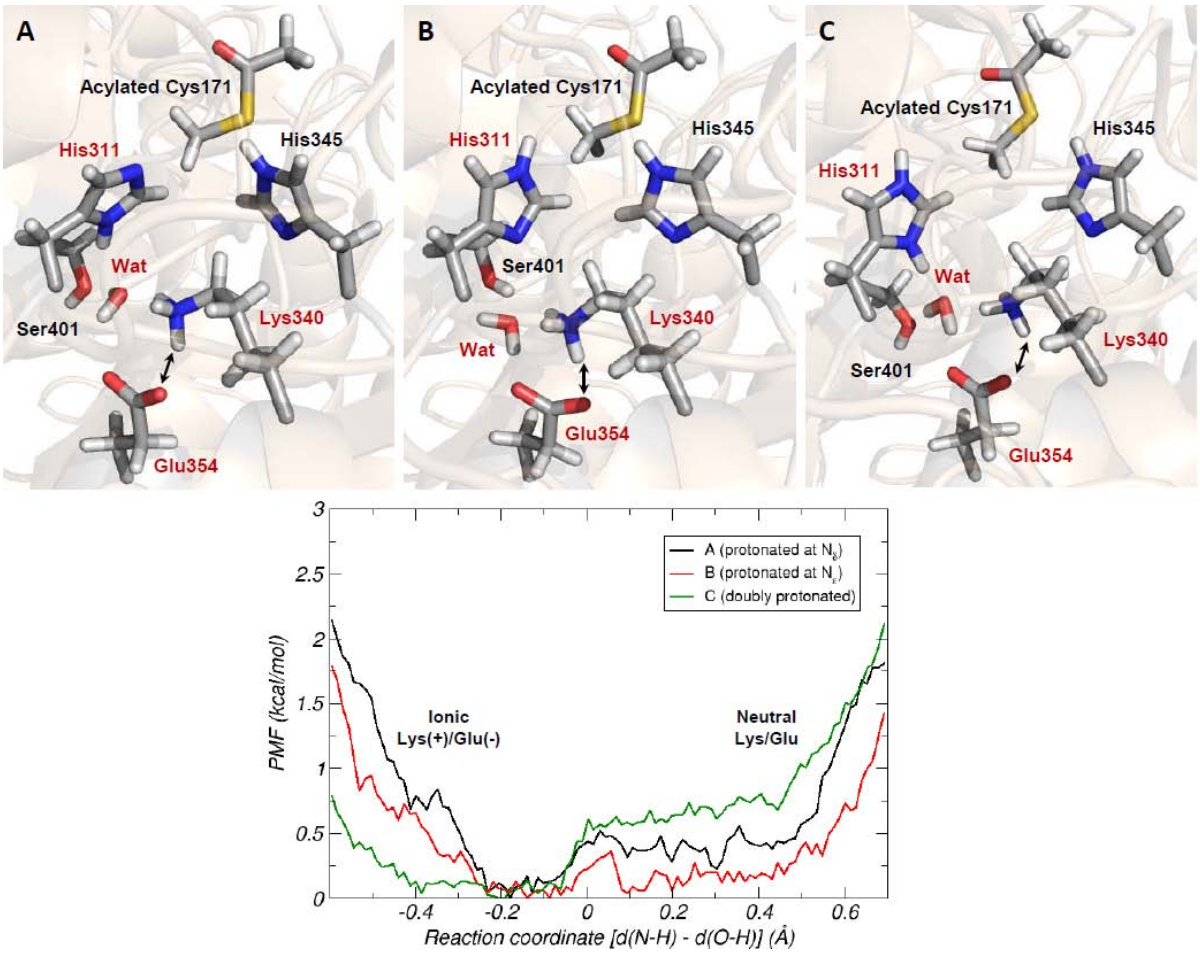


Figure 3.2.5. PMF profiles for the proton transfer between Lys340 and Glu354. In the pictures shown above, the proton transfer of interest is represented as double headed arrow. The name of the residues involved in the QM region is colored with red. The below PMF profiles marked as **A**, **B** and **C** correspond to the picture **A**, **B** and **C** above, respectively.

3.2.4 Discussion

Depending on the protonation state of His311, the mechanisms of the decarboxylation reaction suggested so far can be narrowed down to three, which are shown in Figure 2.4. The first mechanism (Mechanism A in Figure 2.4) suggested by Olsen *et al.*[35] is based on the X-ray crystal structure of *Ec*FabB. (PDB entry: 1EK4) They found electron density of two oxygen atoms separated by 2.5Å in the vicinity of N_δ of His298 (His311 in

KasA) in their electron density map. They interpreted one of the electron densities found in close proximity to the N_δ atom of His298 as a hydroxide ion since the short bond between two oxygen atoms indicates a very strong hydrogen bond, which normally involves one charged group.[150] From this reasoning, they concluded that His298 (His311 in KasA) must be protonated at N_δ. Assuming a protonated malonyl group, they suggested that His298 (His311 in KasA) initiates the decarboxylation as a general base by abstracting the proton from the malonic acid. Therefore, this mechanism assumes that His298 (His311 in KasA) should be neutral and protonated at N_δ in the acyl-enzyme state.

However, several points question this mechanism. Our hand-waving argument is connected with the stabilization of the emerging carbanion intermediate and the protonation state of the malonyl substrate. In mechanism A, the emerging carbanion can only be stabilized by His333 (His345 in KasA) because during the decarboxylation reaction, the His298 (His311 in KasA) interacts with the dissociating CO_2 group of the malonyl substrate. Only after the CO_2 group has left the active site an additional stabilization through the His298 (His311 in KasA) would be possible. Nevertheless, mechanisms in which the emerging carbanion intermediate is stabilized by both histidines from the very beginning seem to be more favorable. Another argument against mechanism A is based on the pKa value of malonic acid. The malonyl substrate is expected to be protonated at the beginning of the decarboxylation step, but based on the pKa values of malonic acid, (pKa_1=2.83, pKa_2=5.69) it should be deprotonated in neutral pH. More solid argument against mechanism A comes from two X-ray studies. Price *et al.*,[41] for example, who crystallized *Sp*FabF with high resolution of 1.3 Å, failed to model the above mentioned hydroxide ion in the vicinity of N_δ of His303. (His311 in KasA) The same holds for the X-ray data which we used to set up our model.[39] Our computations about the protonation state of the enzyme at the beginning of the decarboxylation step do also not support mechanism A since they predict the reverse protonation state for the His311 residue, i.e. according to our computations, it is mainly protonated at N_ε but deprotonated at N_δ.

The second mechanism (Mechanism B in Figure 2.4) is based on measurements by Witkowski *et al.*[44] who observed bicarbonate instead of carbon dioxide as reaction products of the decarboxylation reaction of the Cys161Gln (Cys171Gln in KasA) mutant of mammalian fatty acid synthase (FAS) I. To explain this observation, they suggested that the C3 of malonyl-ACP is attacked by an activated water molecule or hydroxide ion, producing

acetyl carbanion and bicarbonate directly. This mechanism has been further elaborated and supported by White *et al.*[21] based on X-ray measurements (PDB entry: 1OX0) which showed one appropriately positioned ordered water molecule in the vicinity of N_ε of His303 in *Sp*FabF (His311 in KasA). They further argued that the conserved Lys/Glu ionic pair controls the electronic state of the imidazole ring of His303 in a similar way as the well known case of serine proteases. Thus, His303 (His311 in KasA) is expected to be protonated at N_δ but deprotonated at N_ε according to their reasoning.

This mechanism seems to be questioned by the X-ray crystal structures of *Ec*FabB complexed with thiolactomycin (TLM) or cerulenin (PDB entries: 1FJ4 and 1FJ8). These structures indicate a hydrogen bond between N_ε of His298 (His311 in KasA) and the carbonyl oxygen of the inhibitors, and surely, the histidine has to be protonated at N_ε to form this hydrogen bond. In this protonation state, an activation of the water molecule cannot take place. White *et al.*[21] resolved this inconsistency to mechanism B by suggesting that His298 adjusts its position upon inhibitor binding causing the movement of the proton from N_δ to N_ε. However, Mechanism B is also questioned by the experimental results from Zhang et al.[36] By proving a decarboxylation activity for the *Sp*FabF Cys163Ala/His303Ala mutant, (Cys171Ala/His311Ala in KasA) they excluded an important involvement of His303 in the catalytic process. The initiation of the process through the activation of a water molecule seems unrealistic if the decarboxylation activity remains without His303. The contradictions between the various investigations probably result since this mechanism is only valid for the mammalian fatty acid synthase. One further argument against this mechanism results from the stabilization of the carbanion intermediate. Like in the first mechanism, the carbanion intermediate can only be stabilized by one catalytic histidine, i.e. also this mechanism is not able to take full advantage of the catalytic power of the enzyme. Our new computations also do not support mechanism B. The FEP computations indicate that N_ε of His311 (pKa≈6.0) is considerably less acidic than the N_δ (pKa≈0-3 depending on the electronic structure of the Lys340/Glu354), i.e. the N_δ will be deprotonated first. Due to the high pKa value of about 15 for the N_ε for a histidine already deprotonated at N_δ, no further deprotonation of His311 will take place so that the N_ε cannot activate a water molecule as assumed in mechanism B.

The last mechanism (Mechanism C in Figure 2.4) has been proposed by Price *et al.*[151] This mechanism is based on the X-ray crystal structures of *Ec*FabB-inhibitor complexes with TLM or cerulenin, (PDB entries: 1FJ4 and 1FJ8) and they assumed that the binding

modes of these inhibitors mimic the transition state of the condensation reaction. This assumption is quite reasonable since enzyme inhibitors frequently resemble the transition state of the catalytic reaction. Furthermore, it has been reported that TLM binds preferentially to the acyl-enzyme state which possesses the full decarboxylase activity for malonyl substrates rather than to the apo-state in KasA which possess only a very poor activity.[36,39] The similarity between TLM and malonyl substrate further support the assumptions. Based on this interpretation, Price *et al.*[151] suggested that both catalytic histidines (His311 and His345 in KasA) form an oxyanion hole which maximally stabilize the negative charge of the carbonyl oxygen of the emerging carbanion after decarboxylation through two hydrogen bonds. This interpretation is further supported by the X-ray crystal structure of the Cys171Gln mutant of KasA complexed with TLM which has been crystallized by Luckner *et al.*[39] (PDB entry: 2WGG). They used the Cys171Gln mutant of KasA because it was reported to mimic the structural changes induced by the acylation.[152] Assuming that TLM closely resembles the malonyl-ACP substrate, the structure of the TLM-Cys171Gln mutant complex should closely reflect the real structure of the malonyl substrate bound to the acylated enzyme. Indeed in the TLM-Cys171Gln mutant complex crystal structure, the carbonyl oxygen of TLM accepts hydrogen bonds from both His311 and His345 which is in line with the argumentations of Price *et al.*[151] Additionally, this mechanism also can be compatible with the experimental results from Zhang et al.,[36] which is not the case for the mechanism B. In this mechanism, both His311 and His345 are involved in the decarboxylation reaction by donating a hydrogen bond, so even if His311 is mutated into alanine, His345 still gives a hydrogen bond, and the decarboxylation activity can be retained. However, the His345 seems to be considerably more important for the stabilization of the intermediate since the Cys163Ala/His337Ala mutant (Cys171Ala/His345Ala in KasA) does not exhibit decarboxylation activities. While our computational study is not in line with mechanism A and B, it strongly supports mechanism C because it predicts the corresponding protonation state to be the most probable one at the beginning of decarboxylation step. Since our computations have been performed for the acyl-enzyme state which is not yet available experimentally, it also indicates that the structure of the TLM-Cys171Gln mutant complex resembles the situation in the real acyl-enzyme state.

3.3 The detailed mechanism of the activation in KasA

The protonation state of the resting state addressed in chapter 3.1 is revisited in this chapter because not only it was shown in a recent benchmark study that the employed semi-empirical method (RM1) is not reliable[140] but also the complicate electronic structure of the Lys340/Glu354 pair addressed in chapter 3.2 was not correctly taken into account in chapter 3.1. Contrary to the previous results, the results from this reinvestigation predict the catalytic residues to be rather neutral than zwitterionic in the resting state. Therefore, it is further investigated how the neutral catalytic residues can be activated, i.e., the switch from the neutral state to the zwitterionic state, since the catalytic reaction is difficult to be initiated in the neutral state. As a result, a more plausible mechanism for the activation of KasA is suggested in this chapter.

3.3.1 Computational details

The initial structure for the various QM/MM MD simulations was taken from the X-ray crystal structure of wild type KasA from *Mycobacterium tuberculosis*. (PDB entry: 2WGD) The protein was solvated in a water shell with a radius of 45 Å using the TIP3P water model,[93] and the system was neutralized by adding sodium ions employing the leap module in AMBERTOOLS 1.4. Using the sander module of AMBER 11[117] in combination with ff99SB parameter set,[106] the solvated and neutralized system was subsequently energy minimized for 500 steps to relax steric clashes. Subsequently, the system was gradually heated up from 50K to 310K employing positional restraints on the protein. The restraints were gradually decreased to zero once the temperature reaches 310K. For final preparation, the system was equilibrated for more than 50ps. These preparation steps were carried out on MM level, and spherical boundary conditions were applied.

QM/MM MD simulations were performed with the developmental version of AMBER 11 in combination with Terachem (v1.46-dev).[125,126] A time step of 1fs was used for those simulations, and the cutoff value of 18 Å was employed for non-bonded interactions. To maintain the temperature constant during simulations, Langevin thermostat was used. The Becke-Lee-Yang-Parr (BLYP) exchange-correlation functional[127–129] in combination with 6-31G**[131–134] basis set was employed for the QM part since semi-empirical methods were shown to lead to severe errors.[140] Dispersion interaction was taken into account via Grimme's

correction scheme.[130] For the MM part, we used the AMBER ff99SB force field parameter set. For the various questions, we used QM parts of different sizes. For the amino acids included in the QM part, the boundaries were set at the bond between the α and β carbon of cysteines and histidines, the β and γ carbon of glutamates, and the δ and ε carbon of lysines.

Due to the computational cost, only the BLYP/6-31G** level of theory was feasible. This approach was already successfully adopted in a large number of theoretical studies[141–149] and also showed comparable results as LCCSD(T)/cc-pVQZ in the benchmark study conducted in the previous chapter. Further insights into the accuracy for the present questions are also offered by Table 3.3.1 which compares computed energy differences for the proton transfer between Cys171 and His311 via a water molecule. The COSMO approach[103] was used to simulate environmental effects. According to these results, BLYP/6-31G** is in excellent agreement with the SCS-MP2/aug-cc-pVTZ approach. The high accuracy of SCS-MP2/aug-cc-pVTZ was proven in a recent benchmark paper.[140] These computations were carried out using the Turbomole program package.[100]

Level of theory	ΔE
BLYP/6-31G**	3.6
BLYP/6-311G**	2.0
BLYP/6-311++G**	3.4
SCS-MP2/cc-pVTZ	4.8
SCS-MP2/aug-cc-pVTZ	3.4

Table 3.3.1. Benchmark computation results for the energy difference (ΔE) between the zwitterionic (Cys171(-) / His311(+)) and the neutral (Cys171 / His311(δ)) state. All energies are given in kcal/mol.

To gain insights into the structural changes in the active site upon substrate binding, a force field based MD simulation was performed for a tetrahedral intermediate state. The initial structure of the tetrahedral intermediate was built manually using MOLDEN[153] and

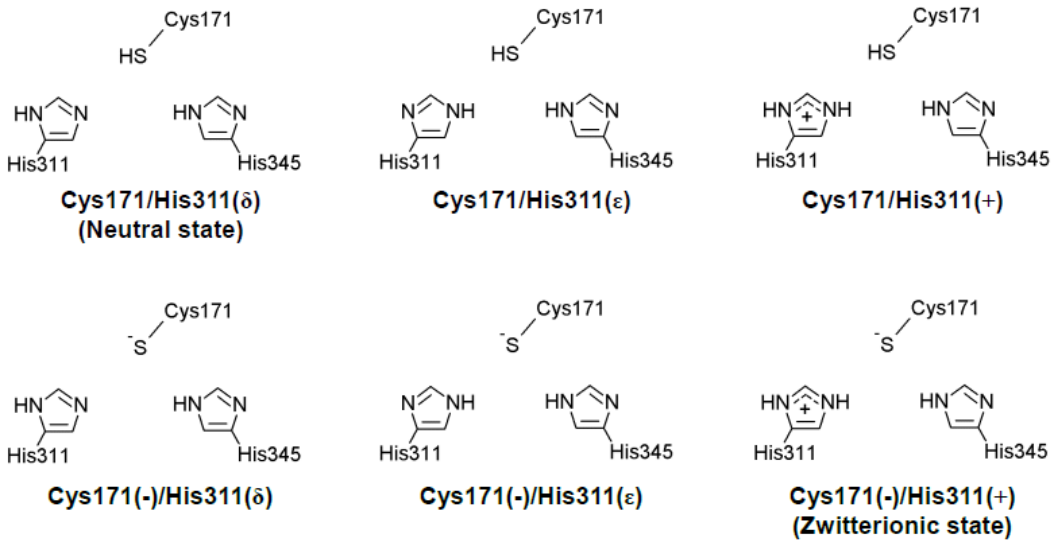
then optimized at the B3LYP/TZVP level of theory. The electrostatic potential of this optimized structure was further computed by HF/6-31G*. These calculations were done using Gaussian03.[154] Subsequently, partial charges were fitted from the computed electrostatic potential using the RESP method.[155] The force field parameters for the tetrahedral intermediate were taken from General AMBER force field[116] (GAFF) and assigned by antechamber.[115] The X-ray crystal structure of the Cys171Gln mutant complexed with the inhibitor thiolactomycin (PDB entry: 2WGG) was used as a starting structure for the enzyme. The inhibitor was removed from the active site, and Gln171 was modified into the tetrahedral intermediate. The heating and the equilibration were conducted as described above. For the 1ns of productive run, we used the same conditions used for the other MD simulations.

To compute the pKa values of His311 with the alchemical FEP method,[96] we employed the NAMD 2.7 simulation package[89] in combination with the CHARMM 22 force field[90] with the CMAP corrections.[91] The preparation of model systems and simulation conditions were identical to those of all other MD simulations, but a time step of 2fs was used in combination with the RATTLE algorithm[95] to constraint the bonds including hydrogen atoms. As reference systems for the thermodynamic cycle (Figure 3.3.3), we used a single histidine residue. Both terminals were capped with a neutral blocking group, and the capped histidine was solvated in a water shell with a radius of 22 Å. The alchemical transformation was stratified into 100 windows with even width. The dual topology paradigm was adopted. For each window, we performed a 15 ps of equilibration and a 40 ps of productive run. Soft core potentials were employed to avoid end-point catastrophe.[122–124] Each FEP computation was repeated four times to address the accuracy of the computations.[98,121] A harmonic restraint was applied to the model systems with the neutral Lys340/Glu354 pair to maintain their hydrogen bond, thereby preventing a collapse of the hydrogen bond network. This problem was not observed in the model systems with an ionic Lys340/Glu354 pair.

The PMF profile for the conformational change of Phe404 was obtained by force field based umbrella samplings using the same conditions mentioned above. Each window consists of a 50 ps of equilibration and a 50 ps of productive run. The PMF profile was computed with the weighted histogram analysis method (WHAM)[135–137] as implemented in the software developed by Grossfield.[107]

3.3.2 The protonation state of the resting state revisited

In this chapter, we reinvestigate the protonation state of catalytic residues in the resting state of KasA from scratch. KasA has three catalytic residues, namely, Cys171, His311, and His345 which formally can take 18 different protonation states. In these possibilities, states with negatively charged histidines were not considered. The protonation state of His345 is obvious from X-ray structures which show that its N_δ accepts a hydrogen bond from the backbone amide of Ile347. These considerations reduce the number of possible states to 6 which are depicted in scheme 3.3.1. From these states, the protonation states Cys171/His311(ε) and Cys171/His311(+) can also be ruled out because they are not able to attack the substrate. As a result, the protonation states we have to actually consider can be narrowed down to four.



Scheme 3.3.1. The six possible protonation states of catalytic residues in KasA. A conserved water molecule positioned between Cys171 and His311 is omitted for clarity.

In chapter 3.1, the zwitterionic state (Cys171(-)/His311(+)) is predicted to be the most probable protonation state in the resting state mainly based on structural arguments but also supported by a semi-empirical based PMF surface. However, this seems to be questionable in the light of a benchmark study[140] and the results from chapter 3.2. First of all, the benchmark study revealed that the semi-empirical method (RM1) employed in chapter 3.1 artificially favors the zwitterionic state with respect to its neutral counterpart.[140] Another source of error stems from the electronic character of a Lys340/Glu354 pair which strongly

influences the pKa values of His311. The Lys340/Glu354 pair was assumed as purely ionic in chapter 3.1. However, as shown by the high-level QM/MM MD simulations in chapter 3.2, the proton can easily hop between the constituting residues.

To come to more reliable estimates about the protonation state of the catalytic residues in the resting state of KasA, we performed a 100ps long QM/MM MD simulation. The QM part (Figure 3.3.1) which was described at the BLYP/6-31G** level of theory comprises Cys171, His311, and the bridging Wat1. Additionally, we included Lys340 and Glu354 in the QM part to take their influence on His311 better into account and to describe the proton hopping between them. The conserved Wat2 was also added to the QM part to account for a possible proton transfer from His311 to Glu354.

Since PMF profiles for proton transfers were too expensive to compute, the energy difference between the neutral and the zwitterionic state could not be determined explicitly. However, the population of both states can be estimated from the distribution of the positions of the transferred proton during QM/MM MD simulations. Figure 3.3.1C shows the corresponding histogram which monitors the position of the proton along the distance α (Figure 3.3.1A). Fortunately, due to the small barriers, the proton hops so often during the simulations that the data are statistically valid. The time length of the simulation (100ps) also seems to be sufficient for our purpose because the motion of the side chain of amino acids or proton transfers take place in picosecond or lower time scale. Figure 3.3.1C indicates that the neutral state is considerably more populated than the zwitterionic state, i.e. the catalytic residues are predicted to prefer the neutral rather than the zwitterionic state.

Due to the small energy difference between both states, our prediction in chapter 3.1 was mainly based on the structural criteria, namely the orientation of His345 (Figure 3.3.1A and 3.3.1B) which is known from the X-ray crystal structure (PDB entry: 2WGD).[138] In the force field based simulations performed in chapter 3.1, His345 only took the similar orientation (Figure 3.3.1A) as in X-ray crystal structure if a zwitterionic state was assumed. In this orientation, His345 points at the sulfur center of Cys171. Assuming a neutral state, it points at the oxygen center of Wat1 as indicated in Figure 3.3.1B. However, in the present chapter in which all involved residues are treated quantum mechanically, His345 takes the similar orientation as in the X-ray crystal structure also in the neutral state. This is indicated in Figure 3.3.1D which plots the distance β (Figure 3.3.1A) from the trajectory of the corresponding QM/MM MD simulation in comparison to the value taken from the X-ray

crystal structure (green vertical line). The distribution of the distance γ (Figure 3.3.1E) gives information about the preferred protonation state of the Lys340/Glu354 pair. According to our computations, both states are similarly populated indicating the small energy difference between both states.

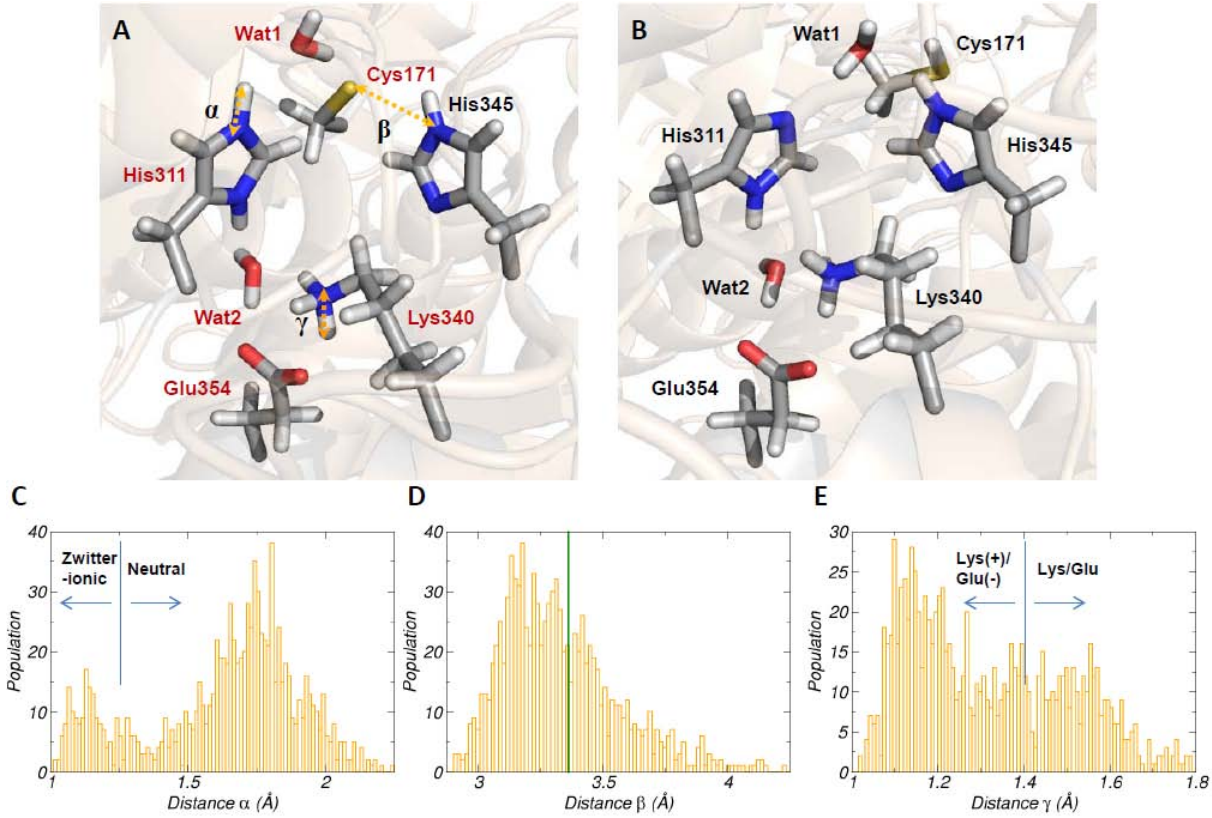


Figure 3.3.1. Distributions of the distances of interest in the resting state. Picture A and B show a representative frame from the force field based MD simulation of the zwitterionic state and the neutral state, respectively. (Those frames were selected by a visual inspection.) The name of residues involved in the QM region during QM/MM MD is shown in red in the picture A. The corresponding distributions for the distance marked as α, β, and γ in the picture A are C, D, and E, respectively.

The QM/MM MD simulation discussed so far cannot compare the neutral state with the states Cys171(-)/His311(δ) and Cys171(-)/His311(ε) since the number of protons differ. In previous force field based simulations in chapter 3.1, we excluded Cys171(-)/His311(ε) since the computed distribution of the distance δ (Figure 3.3.2A) did not agree with the X-ray data. In the X-ray crystal structure, the distance between the oxygen center of Wat2 and N_δ of

His311 is in the range of hydrogen bond formation (=2.78Å), but this hydrogen bond was not observed in the simulation of Cys171(-)/His311(ε), and Wat2 took the orientation as shown in Figure 3.3.2B. To exclude errors arising from the shortcomings discussed earlier, we also performed improved QM/MM MD simulation for this state. In these computations, the QM part included the residues His311, Lys340, and Glu354, and the bridging Wat2. The residue Cys171 and Wat1 were excluded to ensure that His311 remains in this protonation state.

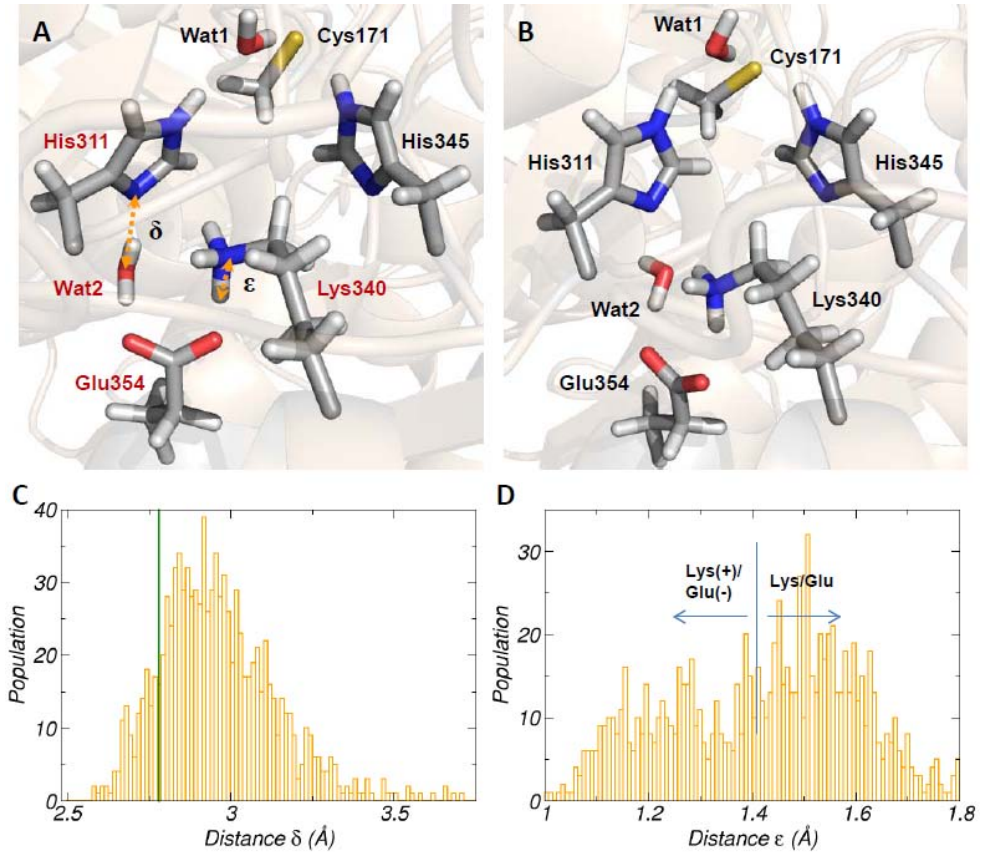


Figure 3.3.2. Distributions of the distances of interest in the system with the protonation state Cys171(-)/His311(ε). The picture A shows a representative frame from the QM/MM MD, and the picture B shows a representative frame from the force field based MD (Those frames were selected by a visual inspection.). The name of residues involved in the QM region during the QM/MM MD is shown in red in the picture A. The corresponding distributions for the distance marked as δ and ε in the picture A are C and D, respectively.

In this improved QM/MM MD description, Wat2 is observed to form a hydrogen bond with N_δ of His311 as shown in Figure 3.3.2A. Additionally, the distribution of the

distance δ also shows a peak in the vicinity of the experimental value (green vertical line in Figure 3.3.2C). Figure 3.3.2D, which analyzes the bonding character of the Lys340/Glu354 pair by monitoring the distance ε in Figure 3.3.2A, indicates again its mixed zwitterionic and neutral character of the Lys340/Glu354 pair. The quantum mechanical description of the Lys340/Glu354 pair seems to be the main reason of the differences between our previous force field based simulation and present results because the Lys340/Glu354 pair was assumed to be purely ionic in the force field description. Therefore, the new QM/MM MD simulation indicates that the protonation state Cys171(-)/His311(ε) cannot be excluded on the basis of its structure.

The states Cys171(-)/His311(δ) and Cys171(-)/His311(ε) cannot be directly compared to the neutral state but can be connected to the zwitterionic state Cys171(-)/His311(+) via force field based FEP simulations. The corresponding thermodynamic cycles are shown in Figure 3.3.3. The processes in the upper horizontal connections represent alchemical transformations from doubly protonated His311 to either the N_δ protonated (Left side of Figure 3.3.3) or the N_ε protonated His311 (Right side of Figure 3.3.3) in the protein environment. The lower part represents the corresponding alchemical transformations in the water shell. Since these FEP simulations were force field based, the calculations were performed separately for neutral or ionic Lys340/Glu354 pair as extreme cases. The computed free energy differences together with the derived pKa values are gathered in Table 3.3.2.

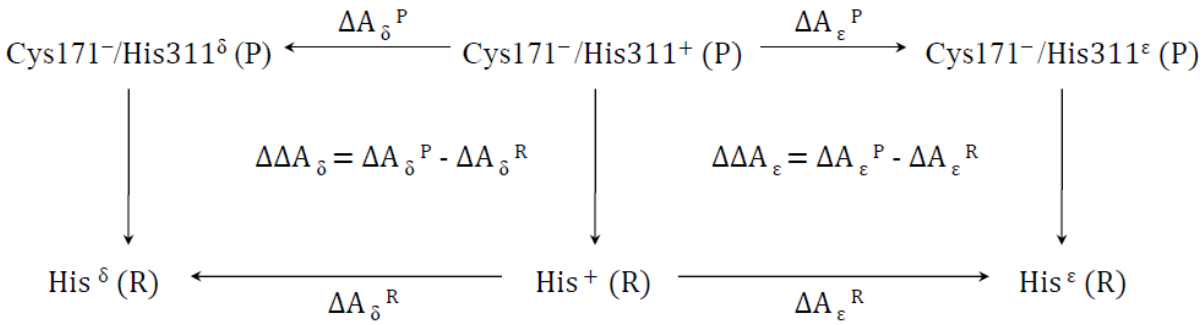


Figure 3.3.3. Thermodynamic cycles devised to compute pKa values at both protonation sites of His311. (P) represents the protein environment, and (R) represents the reference system where simulations are performed in water shell.

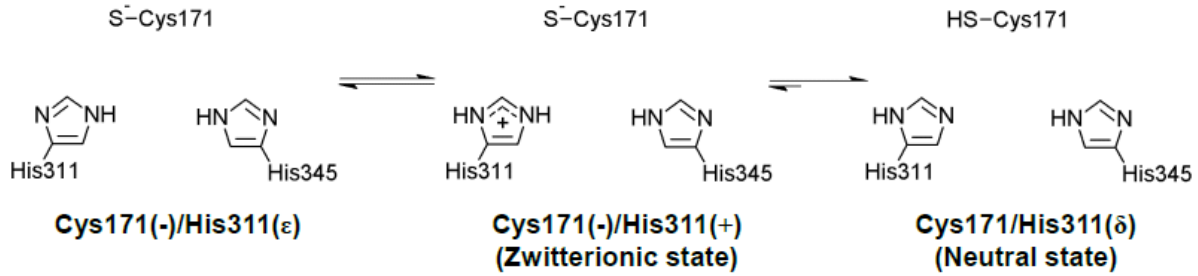
	Lys340(+) / Glu354(-)		Lys340 / Glu354	
	N_δ position	N_ε position	N_δ position	N_ε position
ΔA^P	38.0 ± 1.7	43.9 ± 1.0	41.1 ± 1.1	48.6 ± 0.6
ΔA^R	37.3 ± 0.8	38.1 ± 1.0	37.3 ± 0.8	38.1 ± 1.0
$\Delta\Delta A$	0.7 ± 1.9	5.8 ± 1.4	3.8 ± 1.4	10.5 ± 1.2
pKa-value	6.5	10.1	8.7	13.4

Table 3.3.2. Computed free energies and derived pKa values for both protonation sites of His311. All energies are given in kcal/mol.

For N_ε of His311, the computations predict pKa values of around 10 and 13.4 for an ionic and a neutral Lys340/Glue354 pair, respectively. For N_δ of His311, the corresponding values are 6.5 and 8.7. The considerably higher acidity of the N_δ center rules out the Cys171(-)/His311(δ) state from possible protonation states, i.e., the N_ε center is supposed to be always protonated. The computed pKa values of N_δ also indicate an equilibrium between Cys171(-)/His311(ε) and the zwitterionic state for an environment with pH 7. Depending on the protonation state of the Lys340/Glu354, the ratio between the zwitterionic state and Cys171(-)/His311(ε) in this equilibrium varies between 32:68 and 98:2.

The results of these computations described so far are summarized in Scheme 3.3.2. We found an equilibrium between Cys171(-)/His311(ε) and the zwitterionic state, and the latter was predicted to be less populated than the neutral state. Hence, in its resting state, the enzyme is predicted to mainly populate the neutral state. However, this contradicts the generally accepted catalytic cycle in which the zwitterionic state is assumed in the resting state (Figure 2.4). This raises two questions. The first question results from the reactivity of the neutral state. Its thiol group was computed to be too unreactive for a reaction with carbonyl groups.[113] Therefore, this reaction is only possible with a thiolate group, but this group is only present in the zwitterionic state and Cys171(-)/His311(ε). The second question is connected with the recovery of the enzyme at the end of the catalytic cycle. On the basis of the computations in chapter 3.2, the catalytic cycle ends in the protonation state Cys171(-)/His311(ε) (Figure 2.4). According to the present calculations, it starts in the neutral state.

The answer to this question is given by Scheme 3.3.2 which establishes an equilibrium between the end state and the starting state via the zwiterionic state.



Scheme 3.3.2. Schematic drawing to show the shift between species of interest.

3.3.3 The role of Phe404 in the activation of KasA

Among two questions raised above, the first question is very complicated to answer. The reaction could take place due to the equilibrium between the neutral and the zwitterionic state, but this is unexpected taking the high efficiency of enzymes into account. Hints to an alternative mechanism came from the observation that the His303Ala mutant of FabF (His311Ala in KasA) retains its acyl-transfer activity.[36] Due to the mutation of His311, this activity cannot result from the zwitterionic state but relies solely on a cysteine residue which cannot transfer its proton to an accepting partner. The simulation in chapter 3.1 explained this unexpected reactivity of Cys171 by the formation of a strong hydrogen bond between the sulfur center of Cys171 and the backbone amide of Phe404, and this bond is induced by the rotation of the aromatic ring of Phe404. Through the accompanying stabilization of the emerging thiolate, this hydrogen bond considerably increases the acidity of the thiol group so that Cys171 is predicted to be deprotonated in the resting state. In the wild type, this hydrogen bond is much weaker since the steric demands of the His311 restrict the orientation of Phe404. In the His311Ala mutant, the smaller size of the alanine allows the rotation of Phe404 to the stabilizing position.

On the basis of the results in chapter 3.1, here we hypothesize that the neutral catalytic residues could be activated by the same conformational change of Phe404 observed in the simulation of the His311Ala mutant. We will refer to this conformation as an “open” conformation since Cys171 becomes fully accessible in this conformation. To test whether this conformational change is energetically feasible, we computed the corresponding PMF

profile using force field based umbrella sampling. The computed profile is given in Figure 3.3.4D which also sketches the rotation of Phe404. The closed form which is equivalent to the X-ray crystal structure of wild type KasA (PDB entry: 2WGD) is depicted in Figure 3.3.4A, while the geometrical arrangement of the computed open conformation is indicated in Figure 3.3.4C. The energy barrier for this rotation is computed to be 7-8 kcal/mol, while the relative energy difference between the open and the closed conformation is predicted to be about 6 kcal/mol. Both values are certainly in the range of enzymatic activation barriers. Additionally, the fact that the local minimum exists at the point which exactly corresponds to our suggested open conformation also implies that this open conformation is not an artifact but an intrinsic property of KasA.

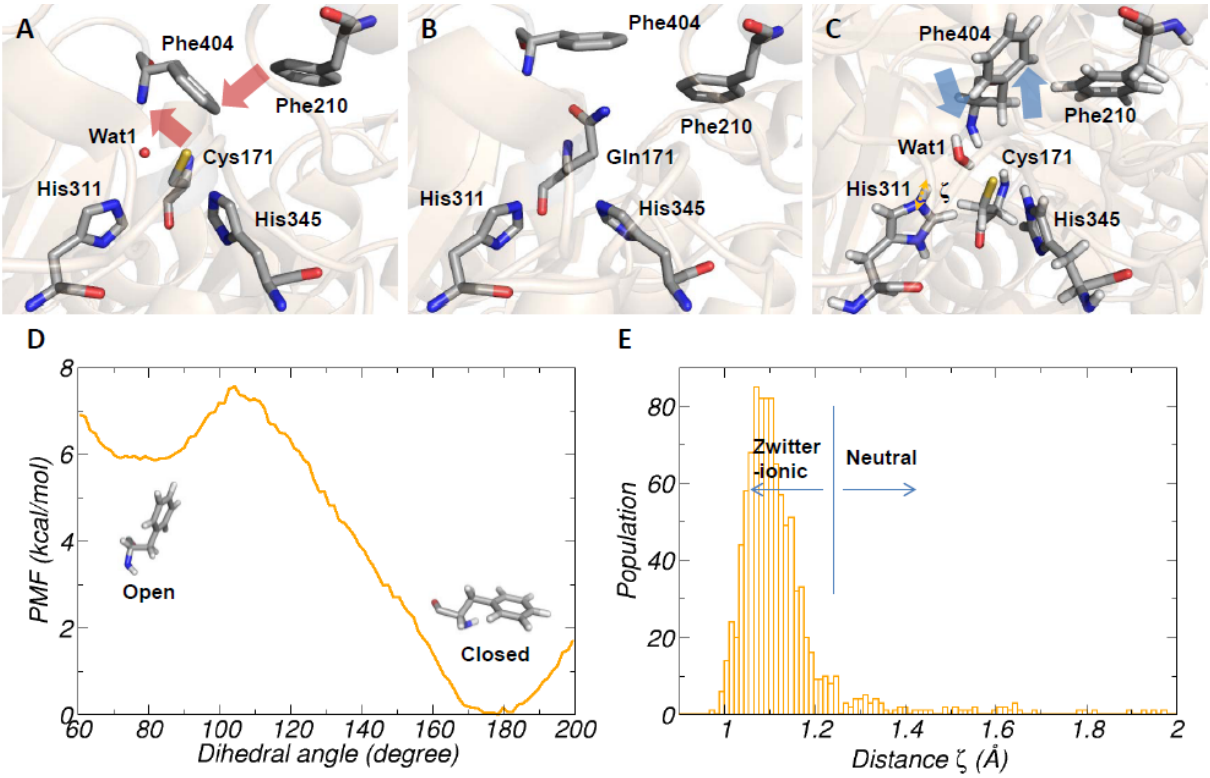


Figure 3.3.4. The PMF profile along the rotation of the aromatic ring of Phe404 illustrating the switch from closed to open conformation (D) and the distribution of the distances to evaluate the ratio between the populations of the neutral and zwitterionic state. (E) The picture A and B are created from X-ray crystal structures (PDB entries: 2WGD and 2WGG for A and B, respectively), and C is created from our simulation. The distribution of the distance marked as ζ in the picture C is shown in E.

To test whether KasA becomes indeed activated by the open conformation of Phe404, we carried out the same QM/MM MD simulation as employed at the beginning of the study for the proton transfer between Cys171 and His311. However, we restricted Phe404 to its open conformation with a harmonic restraint at this time. To determine if the zwitterionic or the neutral state is populated at a given point of the MD simulation, we monitored the distance between the transferred proton and N_ε of His311 (marked as ζ in Figure 3.3.4C). The computed distribution (Figure 3.3.4E) indicates that, in the open conformation, the catalytic residues indeed exclusively populate the zwitterionic state. The remarkable difference to the distribution obtained for the closed conformation (Figure 3.3.1C) implies that the conformational change of Phe404 undoubtedly causes a switch from the neutral to the zwitterionic state of the catalytic residues. The formation of a strong hydrogen bond between the backbone amide of Phe404 and the sulfur center of Cys171 is demonstrated by the average bond length between both centers. It decreases from 2.73 Å for the closed conformation to 2.37 Å for the open conformation. The molecular reason of the formation of this strong hydrogen bond in the open conformation of Phe404 can be explained by the interactions between Phe210 and Phe404. In the closed conformation, Phe210 and Phe404 are in van-der-Waals distance, and Phe210 pushes Phe404 from above so that the backbone amide of Phe404 is moved away from Cys171 (red arrows in Figure 3.3.4A). Such interactions are missing in the open conformation. Hence, the backbone amide of Phe404 can move closer to Cys171 (blue arrows in Figure 3.3.4C). Please note that the water molecule (Wat1) mediating the proton transfer is also involved in the stabilization.

Based on the X-ray crystal structure of the Cys171Gln mutant (PDB entry: 2WGG), Luckner et al. also suggested an open conformation of Phe404 in KasA (Figure 3.3.4B).[39] However, this conformation deviates from the one predicted by our computations (Figure 3.3.4C). Figure 3.3.5, which sketches the involved residues with their van-der-Waals radii, compares those different conformations. The open conformation suggested by Luckner et al. (Figure 3.3.5C) is assumed to represent the acyl-enzyme state, while ours (Figure 3.3.5B) is expected to resemble the state before the acyl-transfer reaction takes place. Luckner et al. used the term “open” because they observed that the corresponding conformational change widens the entrance of the malonyl binding pocket in the acyl-enzyme state compared to the resting state. On the other hand, our “open” conformation is devised for the acyl substrate to gain full access to the catalytic Cys171 at the beginning of the reaction. This is clearly

indicated when comparing Figure 3.3.5A (closed conformation) to Figure 3.3.5B (open conformation). From Figure 3.3.5C, it is also obvious that the open conformation by Luckner et al. cannot provide sufficient space above Cys171 for the acyl substrate (Figure 3.3.5E), while this space would be available in the conformation predicted by our computations (Figure 3.3.5B). The difference between two conformations originates from the rotational axes (Figure 3.3.5D). The open conformation suggested by our computations is obtained through a rotation of the aromatic ring around the axis χ_1, while the open conformation suggested by Luckner et al. is achieved by a rotation around axis χ_2.

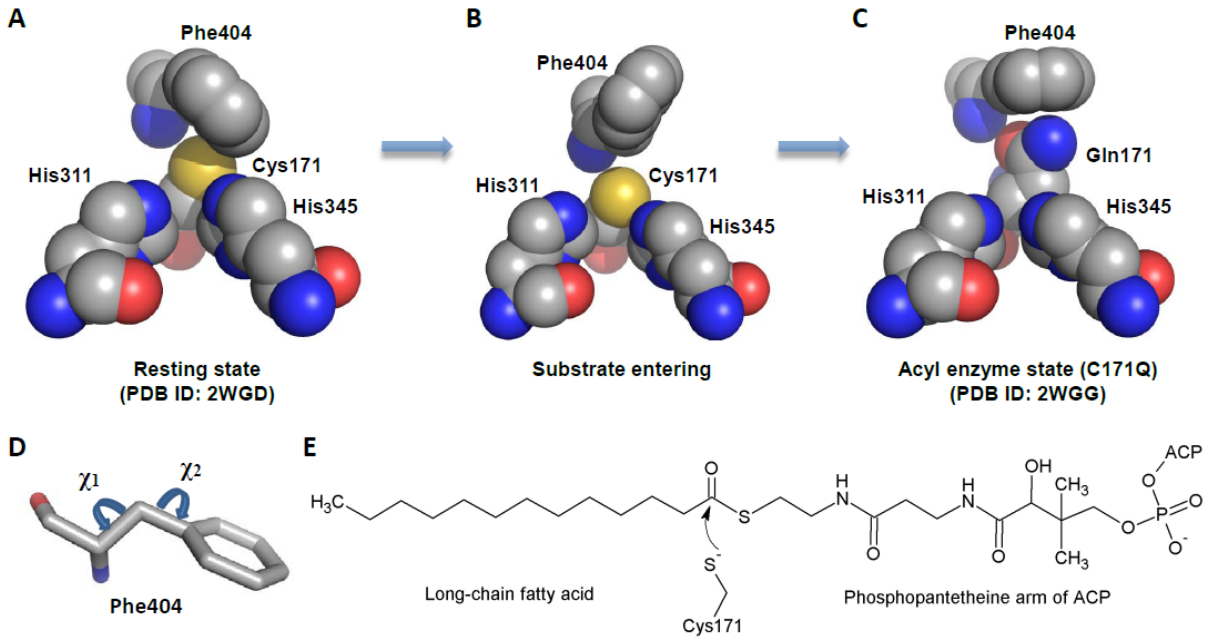


Figure 3.3.5. Putative series of conformational change occurring in Phe404 along the acyl-transfer reaction and the chemical structure of the substrate. The residues in picture A, B, and C are represented as their Van der Waals radii. The picture A and C are created from X-ray crystal structures, and B is created from our simulation. The possible axes of rotation for conformational changes of Phe404 are shown in D. The chemical structure of the acyl substrate and the nucleophilic attack of Cys171 on it are shown in E.

The X-ray crystal structure of the wild type further supports our hypothesis. In the case of wild type, the aromatic ring of Phe404 is below Phe210 (Figure 3.3.4A), while it is the other way round in the Cys171Gln mutant (Figure 3.3.4B). Assuming that the X-ray crystal structure of the Cys171Gln mutant correctly reflects the accompanying structural

change by the acylation,[152] the movement of the aromatic ring of Phe404 from below to above Phe210 has to be feasible in the resting state. By the rotation around the axis χ_1, Phe404 can easily shift its aromatic ring from below to above Phe210. However, as observed in restrained MD simulations (data not shown), this is not possible by a rotation around the axis χ_2 due to the steric hindrance between Phe210 and Phe404.

3.3.4 Discussion

In a widely accepted mechanism, the cysteine residues are activated by helix macro dipoles which especially stabilize the emerging thiolate at the N-terminus. In contrast, recent investigations employing first principle based QM/MM computations revealed a minor contribution of the helix macro dipoles in stabilizing the emerging thiolate.[156] The charge deletion analysis performed for KasA in chapter 3.1 also supports this conclusion so that the activation mechanism based on the helix macro dipoles can be excluded. The activation could also result from a concerted mechanism in which the proton transfer from Cys171 to His311 takes place simultaneously with the nucleophilic attack of Cys171 on the substrate. This type of activation has been reported for serine hydrolases and proteases.[157,158] Such a mechanism is also easily conceivable for KasA since our QM/MM MD results indicate a very low activation barrier for the proton transfer. However, unlike serine hydrolases or proteases, the proton transfer between the catalytic residues is mediated by a bridging water molecule in KasA. This means that the proton transfer cannot take place if the substrate shoves aside the water molecule by entering the active site in advance of the reaction.

Insights into such effects would be provided by simulations of an enzyme-substrate complex, but the indispensable structural information is missing. Since the substrates are expected to change their position only slightly during the bond formation between Cys171 of the enzyme and the carbonyl carbon of the substrate, simulation of the tetrahedral intermediate of the acylation step (Figure 2.4) also offers some information. We performed a 1ns of force field based MD simulation for the tetrahedral intermediate state approximating the long-chain fatty acid and the phosphopantetheine arm of the substrate by a methyl group. Figure 3.3.6 compares a representative frame from the simulation with the neutral resting state. It shows that the thioester group connected to phosphopantetheine arm occupies the same region of the active site which the bridging water molecule (Wat1) occupies in the

resting state. This indicates that the water molecule will be kicked out of the active site before the reaction takes place, so the activation in a concerted manner seems unfeasible in KasA.

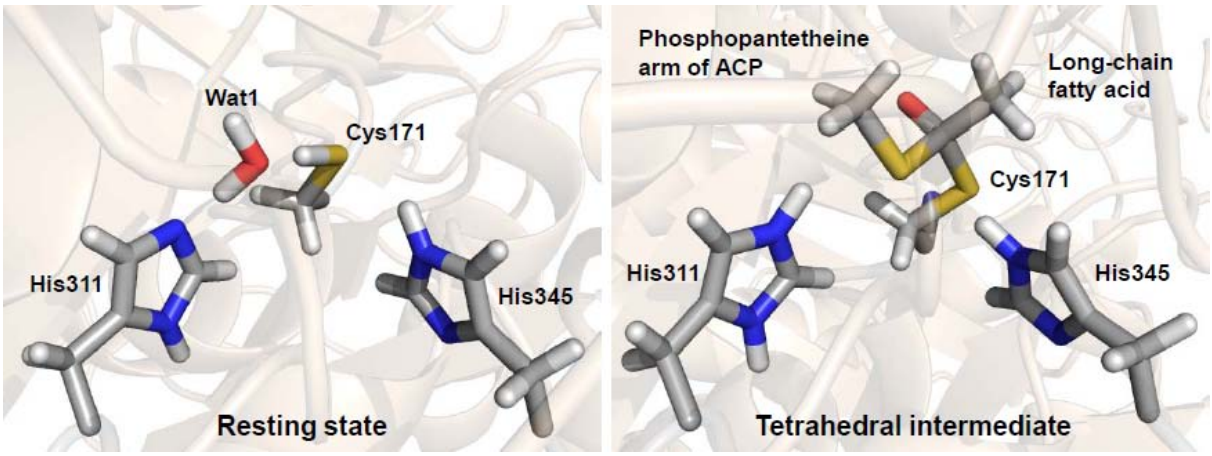


Figure 3.3.6. The comparison between the resting and tetrahedral intermediate state. The resting state is assumed to be neutral here.

Combining our results with the positions of the conserved residues in the malonyl binding pocket of KasA (Figure 3.3.7), we can provide some new insights into the mechanism about the formation of the KasA-substrate complex and the activation of KasA. Considering the connecting point where ACP is attached in the acyl substrate (Figure 3.3.5E), it would enter the active site through the malonyl binding pocket, and the long-chain fatty acid of the substrate is supposed to enter first. When the substrate fully binds to the enzyme, the fatty acid is expected to be accommodated in the acyl binding channel which is located behind Cys171 (blue circle in Figure 3.3.7).[39] The various steps in between are less known. Important information about this process comes from the distribution of conserved hydrophilic and hydrophobic residues in the malonyl binding pocket (Figure 3.3.7). The left side of the pocket contains only hydrophilic residues, while hydrophobic residues only appear in the right side of the pocket. Two conserved threonine residues comprise the borderline laying their hydroxyl group on the hydrophilic side and the methyl group on the hydrophobic side. Due to this distribution, the malonyl binding pocket provides a hydrophobic (red arrow in Figure 3.3.7) and hydrophilic (white arrow in Figure 3.3.7) route. With the long-chain fatty acid (hydrophobic) and the phosphopantetheine arm (hydrophilic), the substrate also divides into two parts with different electrostatics.

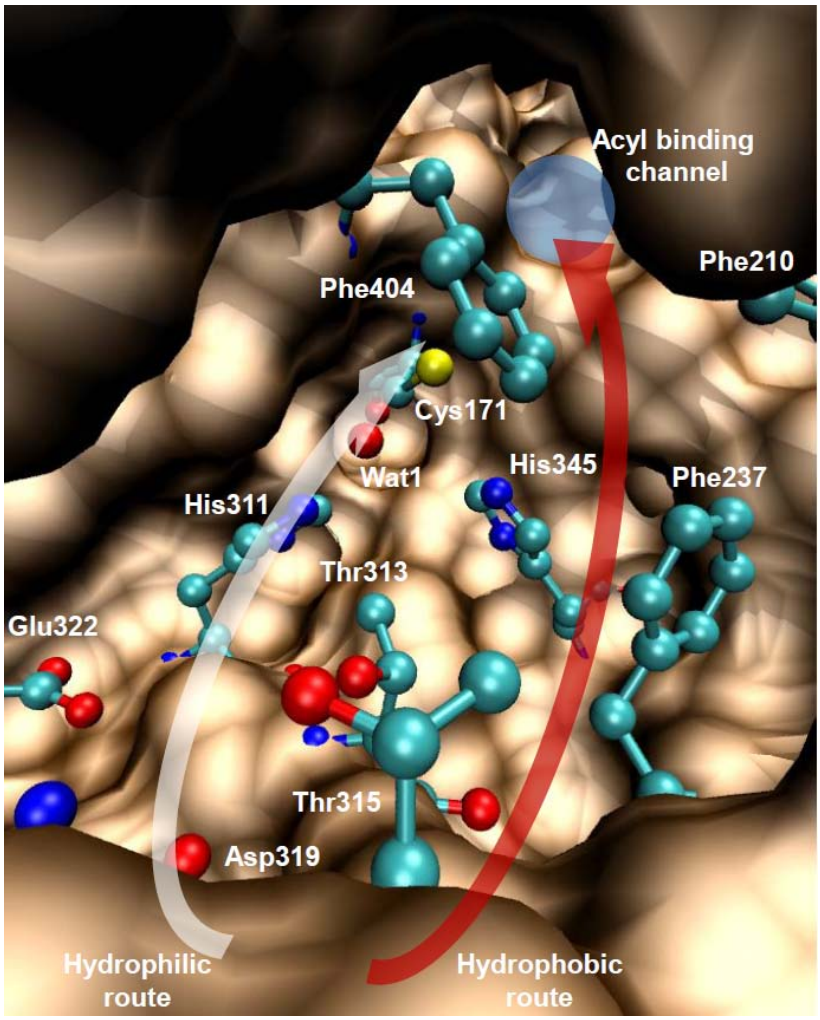


Figure 3.3.7. The malonyl binding pocket and conserved residues therein. All residues shown here are conserved. The location of the acyl binding channel is marked as a blue circle. The hydrophilic and hydrophobic route are shown in white and red arrow, respectively.

This suggests the following course of activation. The long-chain fatty acid which enters the malonyl binding pocket first takes the hydrophobic route. The catalytic residues, Cys171 and His311, are supposed to be neutral at this stage, and since the hydrophobic route is far from the bridging water molecule Wat1, it can keep its bridging position. When the fatty acid reaches the entrance of the acyl binding channel, it pushes up Phe404 to the open conformation to enter the channel. The open conformation of Phe404 in turn promotes the proton transfer from Cys171 to His311 so that the catalytic residues become activated. At this point, the emerging thiolate, which is found to be very reactive in its isolated form, is

stabilized through the hydrogen bond network consisting of the positively charged His311 via Wat1, the neutral His345, and the backbone amide of Cys171 and Phe404. Subsequently, the long-chain fatty acid starts to be accommodated in the acyl binding channel, and simultaneously the phosphopantetheine arm starts to enter the active site along the hydrophilic route.

In the following step, the reactivity of the thiolate can be increased through a destruction of the stabilizing hydrogen bond network. The water molecule, Wat1, which is positioned on the hydrophilic route is expected to be pushed away by the phosphopantetheine arm of the substrate. According to the MD simulation results of the tetrahedral intermediate (Figure 3.3.6), His311 is expected to form a hydrogen bond to the sulfur center of the thioester group of the phosphopantetheine arm. This would predetermine the direction of the proton transfer, i.e., from the protonated His311 to the leaving ACP (Figure 2.4), thereby facilitating the dissociation of the thiol group at the end of the acylation step. Finally, after the long-chain fatty acid is completely accommodated in the acyl binding channel, Phe404 possibly turns back to the closed conformation so that its stabilizing hydrogen bond to the thiolate is also weakened. The accompanying increase of the reactivity of the thiolate would in turn promote the attack on the substrate which is now in the right position for the reaction. This complicated mechanism, which relies on an active manipulation of the electronic structure of the active site through the substrate, also nicely explains the high specificity of KasA and why KasA is only able to catalyze elongated fatty acids but cannot contribute to the *de novo* synthesis.

Our work also can add some information to the design of improved inhibitors. Predicting the resting state of KasA to be neutral, we fully support the work of Sotriffer and cowokers[159] who offered a suggestion for the inhibitor design based on MD simulations for the neutral state. For the design of covalent inhibitors, the role of Phe404 for the shift to the zwitterionic state is important. Since such inhibitors can only react with the zwitterionic state, they need a component which can induce the open conformation of Phe404 described in Figure 3.3.4C. As in the substrate, this component should be non-polar, while the reacting electrophilic warhead, which is supposed to react with the thiolate, should possess polar groups. They will guide the warhead along the hydrophilic route so that it can push away the bridging water molecule, thereby destroying the stabilizing hydrogen bond network of the thiolate. For binding affinity, it would be also desirable for warhead to possess a hydrogen

bond acceptor in the vicinity of its electophilic center to accept hydrogen bonds from both catalytic histidines.

Chapter 4 Summary

KasA is a key enzyme which plays an essential part in the biosynthetic pathway of mycolic acids, the building block of cell wall in *Mycobacterium tuberculosis*. Its importance was demonstrated by the finding that the depletion of KasA leads to the cell lysis of *Mycobacterium tuberculosis*.[18] Since *Mycobacterium tuberculosis* is a pathogen of tuberculosis, the second leading cause of death from an infectious disease worldwide,[1] KasA has drawn attention as one of the attractive drug targets against tuberculosis. Due to the emergence of extensively drug-resistant strains which make most of the known antibiotics for treating tuberculosis ineffective,[7] it became an urgent issue to develop new drugs against tuberculosis.

The knowledge about the detailed catalytic mechanism of a target enzyme is essential in designing effective inhibitors. As a first step to develop new drugs against tuberculosis, in present study, the catalytic mechanism of KasA was investigated computationally. In chapter

3.1, the protonation state of the catalytic residues in the resting state was mainly addressed. Prior to this, it was necessary to confirm the correct conformation of His345 in the X-ray crystal structure first because its conformation is different from that of the corresponding residues found in other KasA X-ray crystal structures. The MD simulation results revealed that the original conformation found in the X-ray crystal structure is not correct. Based on the corrected structure, the protonation state of the catalytic residues was investigated. The FEP computation and MD simulations were employed for this investigation, and the results showed that the zwitterionic state is most probable. To underpin this conclusion with more solid data, The PESs for the proton transfer between the neutral and zwitterionic state were computed in the context of QM/MM. However, due to the strong dependency of the QM/MM optimization on the initial structure, it was not possible to obtain consistent results from these computations. To circumvent this problem, QM/MM based umbrella sampling was carried out with a semi-empirical method (RM1), and the resulting PMF surface indicated that the zwitterionic state is more stable than the neutral state. Additionally, in order to explain the molecular reason of the acyl-transfer activity of the His311Ala mutant, the MD simulations and FEP computations were further performed for this mutant. The results demonstrated that the ionized catalytic cysteine can be significantly stabilized by the hydrogen bond given by the backbone of Phe404. As a result, the catalytic cysteine remains ionized even when His311 is not available.

In chapter 3.2, the protonation state of significant residues in the acyl-enzyme state was investigated. Unlike other catalytic residues, the protonation state of His311 is ambiguous in the acyl-enzyme state, and different decarboxylation mechanisms can be derived depending on the protonation state of His311 in the acyl-enzyme state. Therefore, FEP computations were carried out to find most probable protonation state of His311 in terms of free energy, and the results showed that the pKa value at N_δ is considerably lowered by the enzyme environment while that of N_ε is not. The molecular reason for this severe perturbation at N_δ is due to the hydrogen bond from Lys340. We could confirm this effect by observing the proton transfer from N_δ of doubly protonated His311 to Glu354 in the following QM/MM MD simulation. However, another QM/MM MD simulation with the larger QM part including Lys340 demonstrated that the proton transfer actually occurs between Lys340 and Glu354, thereby preventing the proton transfer between His311 and Glu354. This result also indicated that the Lys340/Glu354 pair, which has been considered to

form a salt bridge in previous studies, is not purely ionic. Therefore, FEP computations were repeated to check the effect of the neutral Lys340/Glu354 pair on the pKa perturbation of His311 and found that N_δ of His311 is not as much perturbed as the case with the ionic Lys340/Glu354 pair. However, the newly computed pKa value at N_δ of His311 is still quite low (around 3) while N_ε is close to the intrinsic pKa value. This indicates that His311 is supposed to be protonated at N_ε but deprotonated at N_δ. Additionally, the PMF profiles for the proton transfer between Lys340 and Glu354 were computed using QM/MM based umbrellas sampling method, and the results showed that the property of the Lys340/Glu354 pair is neutral rather than ionic when His311 is protonated at N_ε. Moreover, a relatively larger ionic character of the Lys340/Glu354 pair when His311 is doubly protonated provides a valuable insight into how the Lys340/Glu354 pair plays a role in shifting the protonated state from N_δ to N_ε in His311 after the acyl-transfer step. Overall, the results demonstrated that His311 is neutral and protonated at N_ε, and the Lys340/Glu354 pair is also neutral in the acyl-enzyme state. Those computational results lead to the conclusion that the decarboxylation reaction is facilitated by an oxyanion hole which is comprised of two catalytic histidines.

In chapter 3.3, the protonation state of catalytic residues in the resting state was revisited because a recent benchmark study[140] showed that the employed semi-empirical method (RM1) in chapter 3.1 tends to overestimate the stabilization of the zwitterionic state. Furthermore, the Lys340/Glu354 pair was considered as purely ionic in chapter 3.1, while it actually has a mixed neutral and ionic character as demonstrated in chapter 3.2. The new investigations employed a larger QM region including the Lys340/Glu354 pair with the BLYP/6-31G** approach, which was proven to be accurate enough for the present purpose by benchmark computations. The new results from the QM/MM MD and FEP computations indicated the catalytic residues to be neutral most probably in the resting state, and this in turn brought up the question how KasA can be activated to initiate the catalytic reaction. On the basis of the results from the MD simulations and FEP computations for the His311Ala mutant in chapter 3.1, we hypothesized that the open conformation of Phe404 would trigger the activation of the catalytic residues by the formation of a strong hydrogen bond. The QM/MM MD simulation proved that the activation of the catalytic residues can indeed be accomplished by the open conformation of Phe404 we suggested, and the corresponding force field based PMF profile also indicated that this conformational change is energetically feasible. The distribution of hydrophilic and hydrophobic residues in the malonyl binding

pocket in conjunction with our computational results further provided a valuable insight into the detailed process how the catalytic residues is activated upon the substrate entering. The long chain fatty acid of the substrate first enters the malonyl binding pocket taking the hydrophobic route of the pocket and causes the conformational change of Phe404 at the entrance of the acyl binding channel. The resulting open conformation of Phe404 in turn promotes the activation, i.e., the proton transfer from Cys171 to His311 via bridging water molecule. The subsequent entry of the phosphopantetheine arm takes the hydrophilic route and pushes away the water molecule from its bridging position. The absent of the bridging water molecule results in the weakening of the hydrogen bond network, thereby making the thiolate of Cys171 less stable. As a result, the catalytic reaction will starts by the attack of Cys171 on the substrate.

Chapter 5 Zusammenfassung

Das Enzym KasA spielt eine entscheidende Rolle in der Biosynthese von Mykolsäuren, den Bausteinen der Zellwände von *Mycobacteriumtuberculosis*. Dessen essentielle Notwendigkeit zeigt sich bei Abwesenheit von KasA in einer Zelllyse (Auflösung von Zellen) bei *Mycobacteriumtuberculosis.*[18] Durch seine Bedeutung für *Mycobacteriumtuberculosis*, dem Erreger von Tuberkulose und damit der zweithäufigsten Todesursache durch Infektionskrankheiten,[1] stellt KasA ein vielversprechendes Ziel für die Entwicklung neuer Medikamente gegen Tuberkulose dar. Durch das Auftreten von extensiv resistenten Stämmen welche die meisten bekannten Antibiotika zur Bekämpfung von Tuberkulose inaktivieren[7] wird es dringend notwendig neue Medikamente gegen Tuberkulose zu entwickeln.

Detailiertes Wissen über den katalytischen Mechanismus des Zielenzyms ist die Grundlage zur systematischen Entwicklung/für das Design wirksamer Inhibitoren. Als erster Schritt zur Entwicklung von neuen Medikamenten gegen Tuberkulose wurde, in dieser Arbeit,

der katalytische Mechanismus von KasA rechnerisch untersucht. In Kapitel 3.1 wird der Protonierungszustand der katalytischen Reste im Ruhezustand untersucht. Vorrausgehend dazu war es notwendig die korrekte Konformation von His345 der Röntgenkristallstruktur zu bestimmen da deren Konformation sich von denen anderer KasA Kristallstrukturen unterscheidet. Die Ergebnisse der Moleküldynamik (MD) Simulationen ergaben, dass die ursprüngliche Konformation der Kristallstruktur nicht korrekt ist. Ausgehend von der korrigierten Konformation wurde der Protonierungszustand der katalytischen Reste untersucht. Für diese Untersuchungen wurden Free Energy Perturbation (FEP) Rechnungen und MD Simulationen verwendet. Die Ergebnisse zeigten, dass der zwitterionische Zustand am wahrscheinlichsten ist. Um diese Aussage mit weiteren handfesten Daten zu untermauern wurden Potential(hyper)flächen (PES) für den Protonentransfer zwischen neutralen und zwitterionischen Zustand mit Hilfe von QM/MM Methoden berechnet. Durch die starke Abhängigkeit der QM/MM Optimierung von der Ausgangsstruktur war es nicht möglich konsistente Ergebnisse für diese Berechnungen zu bekommen. Um dieses Problem zu umgehen wurde ein auf QM/MM basierendes Umbrella Sampling mit Semiempirischen Methoden (RM1) durchgeführt. Die sich daraus ergebende PMF Fläche zeigt das der zwitterionische Zustand stabiler ist als der neutrale Zustand. Zusätzlich, um den molekularen Grund der Acyltransferaktivität des His311Ala Mutanten zu erklären, wurden auch hierfür MD Simulationen und FEP Rechnungen durchgeführt. Die Ergebnisse zeigen das ein ionisiertes katalytisches Cystein massgeblich durch eine Wasserstoffbrückenbindung, ausgehend von der NH Funktion des Phe404, stabilisiert werden kann. Aufgrund dessen bleibt das katalytische Cystein ionisiert auch wenn das His311 nicht vorhanden ist.

In Kapitel 3.2 wurde der Protonierungszustand der entsprechenden Reste im Acyl-Enzym Zustand untersucht. Im Unterschied zu anderen katalytischen Resten ist der Protonierungszustand von His311 ist nicht eindeutig im Acyl-Enzym Zustand und es ergeben sich aus den verschiedenen Protonierungszuständen verschiedene Decarboxylierungsmechanismen. Um den wahrscheinlichsten Protonierungszustand bezüglich der freien Energie zu bestimmen wurden FEP Rechnungen durchgeführt. Die Ergebnisse zeigen, dass der pKa Wert an N_δ beträchtlich durch die Enzymumgebung verringert wird, während dies für N_ε nicht der Fall ist. Der molekulare Grund für diese Störung an N_δ ist die Wasserstoffbindung von Lys340. Wir konnten diesen Effekt durch die Beobachtung des Protonentransfers vom N_δ des doppelt protonierten His311 zum Glu354 in

der folgenden QM/MM MD Simulation bestätigen. Dieses Ergebnis indiziert auch, dass das Lys340/Glu354 Paar, welches in vorherigen Untersuchungen als Salzbrücke betrachtet wurde, nicht rein ionisch ist. Daher wurden die FEP Rechnungen wiederholt um den Einfluss des neutralen Lys340/Glu354 Paares auf die pKa Störung von His311 zu prüfen. Dabei wurde das N_δ des His311 nicht so sehr beeinflusst wie im Fall des ionischen Lys340/Gly354 Paares. Der neu berechnete pKa Wert am N_δ des His311 ist sehr gering (etwa 3) während der Wert für N_ε nahe am intrinsischen pKa Wert liegt. Dies lässt darauf schließen, dass His311 protoniert an N_ε und deprotoniert an N_δ ist. Zusätzlich dazu wurden die PMF Profile für den Protonentransfer zwischen Lys340 und Glu354 mit der QM/MM basierten Umbrella Sampling Methode berechnet. Die Ergebnisse zeigen, dass das Lys340/Glu354 Paar eher neutral als ionisch ist, wenn His311 an N_ε protoniert ist. Ein relativ hoher ionischer Charakter des Lys340/Glu354 Paares, wenn His311 doppelt protoniert ist, gibt einen wertvollen Einblick in die Rolle welche das Lys340/Glu354 Paar beim verschieben des Protonierungszustandes von N_δ zu N_ε im His311 nach dem Acyltransferschritt spielt. Die Ergebnisse zeigen, dass His311 neutral und an N_ε protoniert ist. Ebenso ist das Lys340/Glu354 Paar neutral im Acyl-Enzym Zustand. Diese berechneten Ergebnisse führen zu dem Schluss, dass die Decarboxylierung durch ein Oxyanion Loch erleichtert wird welches aus zwei katalytischen Histidin Resten besteht.

In Kapitel 3.3 wurde der Protonierungszustand der katalytischen Reste im Ruhezustand erneut untersucht da eine aktuelle Benchmarkstudie zeigte,[140] dass die verwendete Semiempirische Methode (RM1) in Kapitel 3.1 dazu tendiert die Stabilisation des zwitterionischen Zustandes zu überschätzen. Auch wurde in Kapitel 3.1 das Lys340/Glu354 Paar als rein ionisch angesehen, während sich in Kapitel 3.2 herausstellte, dass es sich um eine Mischung aus neutralen und ionischen Charakter handelt. Die neuen Untersuchungen beinhalten eine größere QM Region inklusive des Lys340/Glu354 Paares. Der dafür verwendete BLYP/6-31G** Ansatz ist ausreichend akkurat für die aktuelle Fragestellung, was durch Vergleichsrechnungen bewiesen wurde. Die neuen Ergebnisse der QM/MM MD und FEP Rechnungen deuten an, dass die katalytischen Reste im Ruhezustand höchst wahrscheinlich neutral vorliegen. Dies wiederum führt zu der Frage wie KasA aktiviert werden kann um die katalytische Reaktion zu initiieren. Auf der Basis der Ergebnisse der MD Simulationen und FEP Rechnungen für den His311Ala Mutanten in Kapitel 3.1 stellten wir die Hypothese auf, dass die offene Konformation von Phe404 die Aktivierung der

katalytischen Reste durch die (Aus)bildung einer starken Wasserstoffbindung einleitet. Die QM/MM MD Simulation bestätigt dass diese Aktivierung der katalytischen Reste durch die offene Konformation des Phe404 bewerkstelligt werden kann. Das entsprechende auf Kraftfeld basierende PMF Profil zeigt auch, dass dieser Konformationswechsel energetisch realisierbar ist. Die Verteilung der hydrophilen und hydrophoben Reste in der Malonyl Bindungstasche in Verbindung mit unseren berechneten Ergebnissen geben einen Einblick in den detaillierten Prozess der Aktivierung der katalytischen Reste beim Eintritt des Substrates. Die langkettigen Fettsäuren des Substrates treten zuerst in die Malonyl Bindungstasche, belegen den hydrophoben Teil der Tasche und bewirken so die Konformationsänderung von Phe404 am Eingang des Acylbindetunnels. Die resultierende offene Konformation von Phe404 wiederum unterstützt die Aktivierung , d.h. den Protonentransfer vom Cys171 zum His311 über ein verbrückendes Wassermolekül. Der folgende Eintritt des Phosphopantethein Teils belegt den hydrophilen Teil und drückt das Wassermolekül aus der überbrückenden Position. Das Fehlen des überbrückenden Wassermoleküles resultiert in einer Schwächung des Wasserstoffbindungen Netzwerkes, wobei das Thiolat an Cys171 destabilisiert wird. Daraus folgt, dass die katalytische Reaktion mit dem Angriff des Cys171 am dem Substrat beginnt.

Chapter 6 References

(1) WHO | Global Tuberculosis Control 2011 http://www.who.int/tb/publications/global_report/2011/en/ (accessed May 27, 2013).

(2) Rattan, A.; Kalia, A.; Ahmad, N. Multidrug-resistant Mycobacterium Tuberculosis: Molecular Perspectives. *Emerg Infect Dis* **1998**, *4*, 195–209.

(3) Rajbhandary, S. S.; Marks, S. M.; Bock, N. N. Costs of Patients Hospitalized for Multidrug-resistant Tuberculosis. *Int. J. Tuberc. Lung Dis.* **2004**, *8*, 1012–1016.

(4) Ward, H. A.; Marciniuk, D. D.; Hoeppner, V. H.; Jones, W. Treatment Outcome of Multidrug-resistant Tuberculosis Among Vietnamese Immigrants. *Int. J. Tuberc. Lung Dis.* **2005**, *9*, 164–169.

(5) Iseman, M. D. Treatment of Multidrug-resistant Tuberculosis. *N. Engl. J. Med.* **1993**, *329*, 784–791.

(6) Nathanson, E.; Lambregts-van Weezenbeek, C.; Rich, M. L.; Gupta, R.; Bayona, J.; Blondal, K.; Caminero, J. A.; Cegielski, J. P.; Danilovits, M.; Espinal, M. A.; et al.

Multidrug-resistant Tuberculosis Management in Resource-limited Settings. *Emerg Infect Dis* **2006**, *12*, 1389–1397.

(7) Shenoi, S.; Friedland, G. Extensively Drug-Resistant Tuberculosis: A New Face to an Old Pathogen. *Annu. Rev. Med.* **2009**, *60*, 307–320.

(8) Centers for Disease Control and Prevention | Emergence of Mycobacterium tuberculosis with Extensive Resistance to Second-Line Drugs http://www.cdc.gov/mmwr/preview/ mmwrhtml/mm5511a2.htm.

(9) Jain, A.; Mondal, R. Extensively Drug-resistant Tuberculosis: Current Challenges and Threats. *FEMS Immunol. Med. Microbiol.* **2008**, *53*, 145–150.

(10) Bhatt, A.; Molle, V.; Besra, G. S.; Jacobs Jr, W. R.; Kremer, L. The Mycobacterium Tuberculosis FAS-II Condensing Enzymes: Their Role in Mycolic Acid Biosynthesis, Acid-fastness, Pathogenesis and in Future Drug Development. *Mol. Microbiol.* **2007**, *64*, 1442–1454.

(11) Dubnau, E.; Chan, J.; Raynaud, C.; Mohan, V. P.; Lanéelle, M.; Yu, K.; Quémard, A.; Smith, I.; Daffé, M. Oxygenated Mycolic Acids Are Necessary for Virulence of Mycobacterium Tuberculosis in Mice. *Mol. Microbiol.* **2000**, *36*, 630–637.

(12) Glickman, M. S.; Cox, J. S.; Jacobs, W. R., Jr A Novel Mycolic Acid Cyclopropane Synthetase Is Required for Cording, Persistence, and Virulence of Mycobacterium Tuberculosis. *Mol. Cell* **2000**, *5*, 717–727.

(13) Ojha, A.; Anand, M.; Bhatt, A.; Kremer, L.; Jacobs, W. R., Jr; Hatfull, G. F. GroEL1: a Dedicated Chaperone Involved in Mycolic Acid Biosynthesis During Biofilm Formation in Mycobacteria. *Cell* **2005**, *123*, 861–873.

(14) Bhatt, A.; Fujiwara, N.; Bhatt, K.; Gurcha, S. S.; Kremer, L.; Chen, B.; Chan, J.; Porcelli, S. A.; Kobayashi, K.; Besra, G. S.; et al. Deletion of kasB in Mycobacterium Tuberculosis Causes Loss of Acid-fastness and Subclinical Latent Tuberculosis in Immunocompetent Mice. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 5157–5162.

(15) Daffé, M.; Draper, P. The Envelope Layers of Mycobacteria with Reference to Their Pathogenicity. *Adv. Microb. Physiol.* **1998**, *39*, 131–203.

(16) Yuan, Y.; Zhu, Y.; Crane, D. D.; Barry, C. E., 3rd The Effect of Oxygenated Mycolic Acid Composition on Cell Wall Function and Macrophage Growth in Mycobacterium Tuberculosis. *Mol. Microbiol.* **1998**, *29*, 1449–1458.

(17) Glickman, M. S.; Jacobs, W. R. Microbial Pathogenesis of Mycobacterium Tuberculosis: Dawn of a Discipline. *Cell* **2001**, *104*, 477–485.

(18) Bhatt, A.; Kremer, L.; Dai, A. Z.; Sacchettini, J. C.; Jacobs, W. R. Conditional Depletion of KasA, a Key Enzyme of Mycolic Acid Biosynthesis, Leads to Mycobacterial Cell Lysis. *J. Bacteriol.* **2005**, *187*, 7596–7606.

(19) Smith, S.; Witkowski, A.; Joshi, A. K. Structural and Functional Organization of the Animal Fatty Acid Synthase. *Prog. Lipid Res.* **2003**, *42*, 289–317.

(20) Lu, Y.-J.; Zhang, Y.-M.; Rock, C. O. Product Diversity and Regulation of Type II Fatty Acid Synthases. *Biochem. Cell Biol.* **2004**, *82*, 145–155.

(21) White, S. W.; Zheng, J.; Zhang, Y.-M.; Rock, C. O. The Structural Biology of Type II Fatty Acid Biosynthesis. *Annu. Rev. Biochem.* **2005**, *74*, 791–831.

(22) Choi, K. H.; Kremer, L.; Besra, G. S.; Rock, C. O. Identification and Substrate Specificity of Beta -ketoacyl (acyl Carrier Protein) Synthase III (mtFabH) from Mycobacterium Tuberculosis. *J. Biol. Chem.* **2000**, *275*, 28201–28207.

(23) Marrakchi, H.; Ducasse, S.; Labesse, G.; Montrozier, H.; Margeat, E.; Emorine, L.; Charpentier, X.; Daffé, M.; Quémard, A. MabA (FabG1), a Mycobacterium Tuberculosis Protein Involved in the Long-chain Fatty Acid Elongation System FAS-II. *Microbiology (Reading, Engl.)* **2002**, *148*, 951–960.

(24) Quémard, A.; Sacchettini, J. C.; Dessen, A.; Vilcheze, C.; Bittman, R.; Jacobs, W. R., Jr; Blanchard, J. S. Enzymatic Characterization of the Target for Isoniazid in Mycobacterium Tuberculosis. *Biochemistry* **1995**, *34*, 8235–8241.

(25) Schaeffer, M. L.; Agnihotri, G.; Volker, C.; Kallender, H.; Brennan, P. J.; Lonsdale, J. T. Purification and Biochemical Characterization of the Mycobacterium Tuberculosis Beta-ketoacyl-acyl Carrier Protein Synthases KasA and KasB. *J. Biol. Chem.* **2001**, *276*, 47029–47037.

(26) Kremer, L.; Douglas, J. D.; Baulard, A. R.; Morehouse, C.; Guy, M. R.; Alland, D.; Dover, L. G.; Lakey, J. H.; Jacobs, W. R.; Brennan, P. J.; et al. Thiolactomycin and Related Analogues as Novel Anti-mycobacterial Agents Targeting KasA and KasB Condensing Enzymes inMycobacterium Tuberculosis. *J. Biol. Chem.* **2000**, *275*, 16857 –16864.

(27) Slayden, R. A.; Barry, C. E., 3rd The Role of KasA and KasB in the Biosynthesis of Meromycolic Acids and Isoniazid Resistance in Mycobacterium Tuberculosis. *Tuberculosis (Edinb)* **2002**, *82*, 149–160.

(28) Takayama, K.; Wang, C.; Besra, G. S. Pathway to Synthesis and Processing of Mycolic Acids in Mycobacterium Tuberculosis. *Clin. Microbiol. Rev.* **2005**, *18*, 81–101.

(29) Gande, R.; Gibson, K. J. C.; Brown, A. K.; Krumbach, K.; Dover, L. G.; Sahm, H.; Shioyama, S.; Oikawa, T.; Besra, G. S.; Eggeling, L. Acyl-CoA Carboxylases (accD2 and accD3), Together with a Unique Polyketide Synthase (Cg-pks), Are Key to Mycolic Acid Biosynthesis in Corynebacterianeae Such as Corynebacterium Glutamicum and Mycobacterium Tuberculosis. *J. Biol. Chem.* **2004**, *279*, 44847–44857.

(30) Portevin, D.; De Sousa-D'Auria, C.; Houssin, C.; Grimaldi, C.; Chami, M.; Daffé, M.; Guilhot, C. A Polyketide Synthase Catalyzes the Last Condensation Step of Mycolic Acid Biosynthesis in Mycobacteria and Related Organisms. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 314–319.

(31) Gokhale, R. S.; Saxena, P.; Chopra, T.; Mohanty, D. Versatile Polyketide Enzymatic Machinery for the Biosynthesis of Complex Mycobacterial Lipids. *Nat Prod Rep* **2007**, *24*, 267–277.

(32) Lea-Smith, D. J.; Pyke, J. S.; Tull, D.; McConville, M. J.; Coppel, R. L.; Crellin, P. K. The Reductase That Catalyzes Mycolic Motif Synthesis Is Required for Efficient Attachment of Mycolic Acids to Arabinogalactan. *J. Biol. Chem.* **2007**, *282*, 11000–11008.

(33) Price, A. C.; Choi, K. H.; Heath, R. J.; Li, Z.; White, S. W.; Rock, C. O. Inhibition of Beta-ketoacyl-acyl Carrier Protein Synthases by Thiolactomycin and Cerulenin. Structure and Mechanism. *J. Biol. Chem.* **2001**, *276*, 6551–6559.

(34) Kremer, L.; Dover, L. G.; Carrère, S.; Nampoothiri, K. M.; Lesjean, S.; Brown, A. K.; Brennan, P. J.; Minnikin, D. E.; Locht, C.; Besra, G. S. Mycolic Acid Biosynthesis and Enzymic Characterization of the beta-ketoacyl-ACP Synthase A-condensing Enzyme from Mycobacterium Tuberculosis. *Biochem. J.* **2002**, *364*, 423–430.

(35) Olsen, J. G.; Kadziola, A.; von Wettstein-Knowles, P.; Siggaard-Andersen, M.; Larsen, S. Structures of Beta-Ketoacyl-Acyl Carrier Protein Synthase I Complexed with Fatty Acids Elucidate Its Catalytic Machinery. *Structure* **2001**, *9*, 233–243.

(36) Zhang, Y.-M.; Hurlbert, J.; White, S. W.; Rock, C. O. Roles of the Active Site Water, Histidine 303, and Phenylalanine 396 in the Catalytic Mechanism of the Elongation Condensing Enzyme of Streptococcus Pneumoniae. *J. Biol. Chem.* **2006**, *281*, 17390–17399.

(37) Von Wettstein-Knowles, P.; Olsen, J. G.; McGuire, K. A.; Henriksen, A. Fatty Acid Synthesis. Role of Active Site Histidines and Lysine in Cys-His-His-Type Beta-Ketoacyl-Acyl Carrier Protein Synthases. *FEBS J.* **2006**, *273*, 695–710.

(38) Huang, W.; Jia, J.; Edwards, P.; Dehesh, K.; Schneider, G.; Lindqvist, Y. Crystal Structure of [beta]-ketoacyl-acyl Carrier Protein Synthase II from E.coli Reveals the Molecular Architecture of Condensing Enzymes. *EMBO J.* **1998**, *17*, 1183–1191.

(39) Luckner, S. R.; Machutta, C. A.; Tonge, P. J.; Kisker, C. Crystal Structures of Mycobacterium Tuberculosis KasA Show Mode of Action Within Cell Wall Biosynthesis and Its Inhibition by Thiolactomycin. *Structure* **2009**, *17*, 1004–1013.

(40) Wang, J.; Soisson, S. M.; Young, K.; Shoop, W.; Kodali, S.; Galgoci, A.; Painter, R.; Parthasarathy, G.; Tang, Y. S.; Cummings, R.; et al. Platensimycin Is a Selective FabF Inhibitor with Potent Antibiotic Properties. *Nature* **2006**, *441*, 358–361.

(41) Price, A. C.; Rock, C. O.; White, S. W. The 1.3-Angstrom-Resolution Crystal Structure of Beta-Ketoacyl-Acyl Carrier Protein Synthase II from Streptococcus Pneumoniae. *J. Bacteriol.* **2003**, *185*, 4136–4143.

(42) McGuire, K. A.; Siggaard-Andersen, M.; Bangera, M. G.; Olsen, J. G.; von Wettstein-Knowles, P. beta-Ketoacyl-[acyl Carrier Protein] Synthase I of Escherichia Coli: Aspects of the Condensation Mechanism Revealed by Analyses of Mutations in the Active Site Pocket. *Biochemistry* **2001**, *40*, 9836–9845.

(43) Mladenovic, M.; Fink, R. F.; Thiel, W.; Schirmeister, T.; Engels, B. On the Origin of the Stabilization of the Zwitterionic Resting State of Cysteine Proteases: a Theoretical Study. *J. Am. Chem. Soc.* **2008**, *130*, 8696–8705.

(44) Witkowski, A.; Joshi, A. K.; Smith, S. Mechanism of the Beta-Ketoacyl Synthase Reaction Catalyzed by the Animal Fatty Acid Synthase. *Biochemistry* **2002**, *41*, 10877–10887.

(45) Cramer, C. *Essentials of Computational Chemistry: Theories and Models*; Wiley, 2004.

(46) Jensen, F. *Introduction to Computational Chemistry*; Wiley, 2006.

(47) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*; Wiley-VCH Verlag GmbH, 2001.

(48) Thomas, L. H. The Calculation of Atomic Fields. *Mathematical Proceedings of the Cambridge Philosophical Society* **1927**, *23*, 542–548.

(49) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864–B871.

(50) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138.

(51) Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P. RM1: a Reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *J Comput Chem* **2006**, *27*, 1101–1111.

(52) Dewar, M. J. S.; Thiel, W. Ground States of Molecules. 38. The MNDO Method. Approximations and Parameters. *J. Am. Chem. Soc.* **1977**, *99*, 4899–4907.

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

(54) Roux, B. The Calculation of the Potential of Mean Force Using Computer Simulations. *Comp Phys Comm* **1995**, *91*, 275–282.

(55) Souaille, M.; Roux, B. Extension to the Weighted Histogram Analysis Method: Combining Umbrella Sampling with Free Energy Calculations. *Comp Phys Comm* **2001**, *135*, 40–57.

(56) Kamerlin, S. C. L.; Haranczyk, M.; Warshel, A. Progress in Ab Initio QM/MM Free-energy Simulations of Electrostatic Energies in Proteins: Accelerated QM/MM Studies of pKa, Redox Reactions and Solvation Free Energies. *J Phys Chem B* **2009**, *113*, 1253–1272.

(57) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *The Journal of Chemical Physics* **1954**, *22*, 1420–1426.

(58) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-energy Estimation: Umbrella Sampling. *Journal of Computational Physics* **1977**, *23*, 187–199.

(59) Kirkwood, J. G. Statistical Mechanics of Liquid Solutions. *Chem. Rev.* **1936**, *19*, 275–307.

(60) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE Weighted Histogram Analysis Method for Free-energy Calculations on Biomolecules. I. The Method. *J Comp Chem* **1992**, *13*, 1011–1021.

(61) Bosshard, H. R.; Marti, D. N.; Jelesarov, I. Protein Stabilization by Salt Bridges: Concepts, Experimental Approaches and Clarification of Some Misunderstandings. *J. Mol. Recognit.* **2004**, *17*, 1–16.

(62) Riordan, J. F.; McElvany, K. D.; Borders, C. L., Jr Arginyl Residues: Anion Recognition Sites in Enzymes. *Science* **1977**, *195*, 884–886.

(63) Kirsch, J. F.; Eichele, G.; Ford, G. C.; Vincent, M. G.; Jansonius, J. N.; Gehring, H.; Christen, P. Mechanism of Action of Aspartate Aminotransferase Proposed on the Basis of Its Spatial Structure. *J. Mol. Biol.* **1984**, *174*, 497–525.

(64) Singh, J.; Thornton, J. M.; Snarey, M.; Campbell, S. F. The Geometries of Interacting Arginine-carboxyls in Proteins. *FEBS Lett.* **1987**, *224*, 161–171.

(65) Tormo, J.; Blaas, D.; Parry, N. R.; Rowlands, D.; Stuart, D.; Fita, I. Crystal Structure of a Human Rhinovirus Neutralizing Antibody Complexed with a Peptide Derived from Viral Capsid Protein VP2. *EMBO J.* **1994**, *13*, 2247–2256.

(66) Raumann, B. E.; Rould, M. A.; Pabo, C. O.; Sauer, R. T. DNA Recognition by Beta-sheets in the Arc Repressor-operator Crystal Structure. *Nature* **1994**, *367*, 754–757.

(67) Brown, B. M.; Milla, M. E.; Smith, T. L.; Sauer, R. T. Scanning Mutagenesis of the Arc Repressor as a Functional Probe of Operator Recognition. *Nat. Struct. Biol.* **1994**, *1*, 164–168.

(68) Perutz, M. F. Mechanisms of Cooperativity and Allosteric Regulation in Proteins. *Q. Rev. Biophys.* **1989**, *22*, 139–237.

(69) Schirmer, T.; Evans, P. R. Structural Basis of the Allosteric Behaviour of Phosphofructokinase. *Nature* **1990**, *343*, 140–145.

(70) Marqusee, S.; Sauer, R. T. Contributions of a Hydrogen Bond/salt Bridge Network to the Stability of Secondary and Tertiary Structure in Lambda Repressor. *Protein Sci* **1994**, *3*, 2217–2225.

(71) Tissot, A. C.; Vuilleumier, S.; Fersht, A. R. Importance of Two Buried Salt Bridges in the Stability and Folding Pathway of Barnase. *Biochemistry* **1996**, *35*, 6786–6794.

(72) Argos, P. An Investigation of Protein Subunit and Domain Interfaces. *Protein Eng.* **1988**, *2*, 101–113.

(73) Janin, J.; Miller, S.; Chothia, C. Surface, Subunit Interfaces and Interior of Oligomeric Proteins. *J. Mol. Biol.* **1988**, *204*, 155–164.

(74) Ward, W. H.; Jones, D. H.; Fersht, A. R. Effects of Engineering Complementary Charged Residues into the Hydrophobic Subunit Interface of tyrosyl-tRNA Synthetase. Appendix: Kinetic Analysis of Dimeric Enzymes That Reversibly Dissociate into Inactive Subunits. *Biochemistry* **1987**, *26*, 4131–4138.

(75) Rüegg, C.; Ammer, D.; Lerch, K. Comparison of Amino Acid Sequence and Thermostability of Tyrosinase from Three Wild Type Strains of Neurospora Crassa. *J. Biol. Chem.* **1982**, *257*, 6420–6426.

(76) Kelly, C. A.; Nishiyama, M.; Ohnishi, Y.; Beppu, T.; Birktoft, J. J. Determinants of Protein Thermostability Observed in the 1.9-A Crystal Structure of Malate Dehydrogenase from the Thermophilic Bacterium Thermus Flavus. *Biochemistry* **1993**, *32*, 3913–3922.

(77) Fersht, A. R. Conformational Equilibria in -and -chymotrypsin. The Energetics and Importance of the Salt Bridge. *J. Mol. Biol.* **1972**, *64*, 497–509.

(78) Anderson, D. E.; Becktel, W. J.; Dahlquist, F. W. pH-induced Denaturation of Proteins: a Single Salt Bridge Contributes 3-5 Kcal/mol to the Free Energy of Folding of T4 Lysozyme. *Biochemistry* **1990**, *29*, 2403–2408.

(79) Horovitz, A.; Serrano, L.; Avron, B.; Bycroft, M.; Fersht, A. R. Strength and Co-operativity of Contributions of Surface Salt Bridges to Protein Stability. *J. Mol. Biol.* **1990**, *216*, 1031–1044.

(80) Erwin, C. R.; Barnett, B. L.; Oliver, J. D.; Sullivan, J. F. Effects of Engineered Salt Bridges on the Stability of Subtilisin BPN'. *Protein Eng.* **1990**, *4*, 87–97.

(81) Sali, D.; Bycroft, M.; Fersht, A. R. Surface Electrostatic Interactions Contribute Little of Stability of Barnase. *J. Mol. Biol.* **1991**, *220*, 779–788.

(82) Sun, D. P.; Sauer, U.; Nicholson, H.; Matthews, B. W. Contributions of Engineered Surface Salt Bridges to the Stability of T4 Lysozyme Determined by Directed Mutagenesis. *Biochemistry* **1991**, *30*, 7142–7153.

(83) Waldburger, C. D.; Schildbach, J. F.; Sauer, R. T. Are Buried Salt Bridges Important for Protein Stability and Conformational Specificity? *Nat. Struct. Biol.* **1995**, *2*, 122–128.

(84) Dao-pin S; Anderson, D. E.; Baase, W. A.; Dahlquist, F. W.; Matthews, B. W. Structural and Thermodynamic Consequences of Burying a Charged Residue Within the Hydrophobic Core of T4 Lysozyme. *Biochemistry* **1991**, *30*, 11521–11529.

(85) Wimley, W. C.; Gawrisch, K.; Creamer, T. P.; White, S. H. Direct Measurement of Salt-bridge Solvation Energies Using a Peptide Model System: Implications for Protein Stability. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 2985–2990.

(86) Hendsch, Z. S.; Tidor, B. Do Salt Bridges Stabilize Proteins? A Continuum Electrostatic Analysis. *Protein Sci* **1994**, *3*, 211–226.

(87) Zheng, Y.-J.; Ornstein, R. L. What Happens to Salt-Bridges in Nonaqueous Environments: Insights from Quantum Mechanics Calculations. *J. Am. Chem. Soc.* **1996**, *118*, 11237–11243.

(88) Liljefors, T.; Norrby, P.-O. An Ab Initio Study of the Trimethylamine−Formic Acid and the Trimethylammonium Ion−Formate Anion Complexes, Their Monohydrates, and Continuum Solvation. *J. Am. Chem. Soc.* **1997**, *119*, 1052–1058.

(89) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(90) MacKerell; Bashford, D.; Bellott; Dunbrack; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins†. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(91) Mackerell Jr., A. D.; Feig, M.; Brooks III, C. L. Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(92) Foloppe, N.; Sagemark, J.; Nordstrand, K.; Berndt, K. D.; Nilsson, L. Structure, Dynamics and Electrostatics of the Active Site of Glutaredoxin 3 from Escherichia

Coli: Comparison with Functionally Related Proteins. *J. Mol. Biol.* **2001**, *310*, 449–470.

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

(94) Hestenes, M.; Stiefel, E. Methods of Conjugate Gradients for Solving Linear Systems. *Journal of Research of the National Bureau of Standards* **1952**, *49*, 409–436.

(95) Andersen, H. C. Rattle: A "Velocity" Version of the Shake Algorithm for Molecular Dynamics Calculations. *J. Comput. Phys.* **1983**, *52*, 24–34.

(96) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. II. Polar Gases. *J. Chem. Phys.* **1955**, *23*, 1915.

(97) Pitera, J. W.; van Gunsteren, W. F. A Comparison of Non-Bonded Scaling Approaches for Free Energy Calculations. *Molecular Simulation* **2002**, *28*, 45.

(98) Lu, N.; Kofke, D. A. Accuracy of Free-energy Perturbation Calculations in Molecular Simulation. I. Modeling. *J. Chem. Phys.* **2001**, *114*, 7303.

(99) Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; et al. QUASI: A General Purpose Implementation of the QM/MM Approach and Its Application to Problems in Catalysis. *J. Mol. Structure: THEOCHEM* **2003**, *632*, 1–28.

(100) TURBOMOLE V6.1 2009, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com. http://www.cosmologic.de/data/DOK_HTML/node5.html (accessed Apr 27, 2011).

(101) Smith, W.; Forester, T. R. DL_POLY_2.0: A General-Purpose Parallel Molecular Dynamics Simulation Package. *J. Mol. Graphics* **1996**, *14*, 136–141.

(102) Sherwood, P.; de Vries, A. H.; Collins, S. J.; Greatbanks, S. P.; Burton, N. A.; Vincent, M. A.; Hillier, I. H. Computer Simulation of Zeolite Structure and Reactivity Using Embedded Cluster Methods. *Faraday Disc.* **1997**, *106*, 79–92.

(103) Klamt, A.; Schuerrmann, G. 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.

(104) Billeter, S. R.; Turner, A. J.; Thiel, W. Linear Scaling Geometry Optimisation and Transition State Search in Hybrid Delocalised Internal Coordinates. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2177–2186.

(105) Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P. RM1: a Reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *J. Comput. Chem.* **2006**, *27*, 1101–1111.

(106) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins* **2006**, *65*, 712–725.

(107) Grossfield, A. "WHAM: The Weighted Histogram Analysis Method", version 2.0.4, http://membrane.urmc.rochester.edu/content/wham (accessed May 27, 2013).

(108) Case, D.; Darden, T.; Cheatham, T.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Crowley, M.; Walker, R.; Zhang, W.; et al. *Amber 11, University of California, San Francisco*; 2010.

(109) Walker, R. C.; Crowley, M. F.; Case, D. A. The Implementation of a Fast and Accurate QM/MM Potential Method in Amber. *J. Comput. Chem.* **2008**, *29*, 1019–1031.

(110) Bash, P. A.; Field, M. J.; Davenport, R. C.; Petsko, G. A.; Ringe, D.; Karplus, M. Computer Simulation and Analysis of the Reaction Pathway of Triosephosphate Isomerase. *Biochemistry* **1991**, *30*, 5826–5832.

(111) Mulholland, A. J.; Richards, W. G. Acetyl-CoA Enolization in Citrate Synthase: a Quantum Mechanical/molecular Mechanical (QM/MM) Study. *Proteins* **1997**, *27*, 9–25.

(112) Wong, K. F.; Watney, J. B.; Hammes-Schiffer, S. Analysis of Electrostatics and Correlated Motions for Hydride Transfer in Dihydrofolate Reductase. *J. Phys. Chem. B* **2004**, *108*, 12231–12241.

(113) Paasche, A.; Schiller, M.; Schirmeister, T.; Engels, B. Mechanistic Study of the Reaction of Thiol-Containing Enzymes with α,β-Unsaturated Carbonyl Substrates by Computation and Chemoassays. *ChemMedChem* **2010**, *5*, 869–880.

(114) Mulholland, A. J. Chemical Accuracy in QM/MM Calculations on Enzyme-catalysed Reactions. *Chem Cent J* **2007**, *1*, 19.

(115) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247–260.

(116) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J Comput Chem* **2004**, *25*, 1157–1174.

(117) Case, D.; Darden, T.; Cheatham, T.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Crowley, M.; Walker, R.; Zhang, W.; et al. *Amber 11*; University of California, San Francisco, 2010.

(118) Developing Topology and Parameter Files http://www.ks.uiuc.edu/Training/Tutorials/science/forcefield-tutorial/forcefield-html/node6.html (accessed Feb 5, 2013).

(119) Pearlman, D. A. A Comparison of Alternative Approaches to Free Energy Calculations. *J. Phys. Chem.* **1994**, *98*, 1487–1493.

(120) Gao, J.; Kuczera, K.; Tidor, B.; Karplus, M. Hidden Thermodynamics of Mutant Proteins: a Molecular Dynamics Analysis. *Science* **1989**, *244*, 1069–1072.

(121) Lu, N.; Kofke, D. A. Accuracy of Free-energy Perturbation Calculations in Molecular Simulation. II. Heuristics. *The Journal of Chemical Physics* **2001**, *115*, 6866–6875.

(122) Pitera, J. W.; van Gunsteren, W. F. A Comparison of Non-Bonded Scaling Approaches for Free Energy Calculations. *Mol. Simul.* **2002**, *28*, 45–65.

(123) Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. Avoiding Singularities and Numerical Instabilities in Free Energy Calculations Based on Molecular Simulations. *Chem. Phys. Lett.* **1994**, *222*, 529–539.

(124) Zacharias, M.; Straatsma, T. P.; McCammon, J. A. Separation-Shifted Scaling, a New Scaling Method for Lennard-Jones Interactions in Thermodynamic Integration. *J. Chem. Phys.* **1994**, *100*, 9025–9031.

(125) Isborn, C. M.; Götz, A. W.; Clark, M. A.; Walker, R. C.; Martínez, T. J. Electronic Absorption Spectra from MM and Ab Initio QM/MM Molecular Dynamics: Environmental Effects on the Absorption Spectrum of Photoactive Yellow Protein. *J. Chem. Theory Comput.* **2012**, *8*, 5092–5106.

(126) Ufimtsev, I. S.; Martinez, T. J. Quantum Chemistry on Graphical Processing Units. 3. Analytical Energy Gradients, Geometry Optimization, and First Principles Molecular Dynamics. *J. Chem. Theory Comput.* **2009**, *5*, 2619–2628.

(127) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100.

(128) Lee; Yang; Parr Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev., B Condens. Matter* **1988**, *37*, 785–789.

(129) Miehlich, B.; Savin, A.; Stoll, H.; Preuss, H. Results Obtained with the Correlation Energy Density Functionals of Becke and Lee, Yang and Parr. *Chem. Phys. Lett.* **1989**, *157*, 200–206.

(130) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *The Journal of Chemical Physics* **2010**, *132*, 154104–154104–19.

(131) Frisch, M. J.; Pople, J. A.; Binkley, J. S. Self-Consistent Molecular Orbital Methods 25. Supplementary Functions for Gaussian Basis Sets. *J. Chem. Phys.* **1984**, *80*, 3265–3269.

(132) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257–2261.

(133) Ditchfield, R.; Hehre, W.; Pople, J. Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724–728.

(134) Hariharan, P. C.; Pople, J. A. The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies. *Theo. Chem. Acc.* **1973**, *28*, 213–222.

(135) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free-energy Calculations on Biomolecules. I. The Method. *J. Comp. Chem.* **1992**, *13*, 1011–1021.

(136) Ferrenberg; Swendsen New Monte Carlo Technique for Studying Phase Transitions. *Phys. Rev. Lett.* **1988**, *61*, 2635–2638.

(137) Souaille, M.; Roux, B. Extension to the Weighted Histogram Analysis Method: Combining Umbrella Sampling with Free Energy Calculations. *Comp. Phys. Comm.* **2001**, *135*, 40–57.

(138) Lee, W.; Luckner, S. R.; Kisker, C.; Tonge, P. J.; Engels, B. Elucidation of the Protonation States of the Catalytic Residues in mtKasA: Implications for Inhibitor Design. *Biochemistry* **2011**, *50*, 5743–5756.

(139) Moche, M.; Dehesh, K.; Edwards, P.; Lindqvist, Y. The Crystal Structure of [beta]-Ketoacyl-Acyl Carrier Protein Synthase II from Synechocystis Sp. at 1.54 Å Resolution and Its Relationship to Other Condensing Enzymes. *J. Mol. Biol.* **2001**, *305*, 491–503.

(140) Paasche, A.; Schirmeister, T.; Engels, B. Benchmark Study for the Cysteine–Histidine Proton Transfer Reaction in a Protein Environment: Gas Phase, COSMO, QM/MM Approaches. *J. Chem. Theory Comput.* **2013**, *9*, 1765–1777.

(141) Furmanchuk, A.; Isayev, O.; Gorb, L.; Shishkin, O. V.; Hovorun, D. M.; Leszczynski, J. Novel View on the Mechanism of Water-Assisted Proton Transfer in the DNA Bases: Bulk Water Hydration. *Phys. Chem. Chem. Phys.* **2011**, *13*, 4311–4317.

(142) Leenders, E. J. M.; Guidoni, L.; Röthlisberger, U.; Vreede, J.; Bolhuis, P. G.; Meijer, E. J. Protonation of the Chromophore in the Photoactive Yellow Protein. *J. Phys. Chem. B* **2007**, *111*, 3765–3773.

(143) Leung, K.; Rempe, S. B. Ab Initio Molecular Dynamics Study of Glycine Intramolecular Proton Transfer in Water. *J. Chem. Phys.* **2005**, *122*, 184506.

(144) Li, G.-S.; Maigret, B.; Rinaldi, D.; Ruiz-López, M. F. Influence of Environment on Proton-Transfer Mechanisms in Model Triads from Theoretical Calculations. *J. Comp. Chem.* **1998**, *19*, 1675–1688.

(145) Liang, X.; Montoya, A.; Haynes, B. S. Local Site Selectivity and Conformational Structures in the Glycosidic Bond Scission of Cellobiose. *J. Phys. Chem. B* **2011**, *115*, 10682–10691.

(146) Mangold, M.; Rolland, L.; Costanzo, F.; Sprik, M.; Sulpizi, M.; Blumberger, J. Absolute pKa Values and Solvation Structure of Amino Acids from Density Functional Based Molecular Dynamics Simulation. *J. Chem. Theory Comput.* **2011**, *7*, 1951–1961.

(147) Pluhařová, E.; Marsalek, O.; Schmidt, B.; Jungwirth, P. Peptide Salt Bridge Stability: From Gas Phase via Microhydration to Bulk Water Simulations. *J. Chem. Phys.* **2012**, *137*, 185101–185101–8.

(148) Xiao, S.; Wang, L.; Liu, Y.; Lin, X.; Liang, H. Theoretical Investigation of the Proton Transfer Mechanism in Guanine-Cytosine and Adenine-Thymine Base Pairs. *J. Chem. Phys.* **2012**, *137*, 195101–195101–8.

(149) Xie, H.; Jin, L.; Rudić, S.; Simons, J. P.; Gerber, R. B. Computational Studies of Protonated β-d-Galactose and Its Hydrated Complex: Structures, Interactions, Proton Transfer Dynamics, and Spectroscopy. *J. Phys. Chem. B* **2012**, *116*, 4851–4859.

(150) Speakman, J. C. Acid Salts of Carboxylic Acids, Crystals with Some 'Very Short" Hydrogen Bonds. In *Structure and Bonding*; Structure and Bonding; Springer Berlin Heidelberg, 1972; pp. 141–199.

(151) Price, A. C.; Choi, K. H.; Heath, R. J.; Li, Z.; White, S. W.; Rock, C. O. Inhibition of Beta-Ketoacyl-Acyl Carrier Protein Synthases by Thiolactomycin and Cerulenin. Structure and Mechanism. *J. Biol. Chem.* **2001**, *276*, 6551–6559.

(152) Wang, J.; Soisson, S. M.; Young, K.; Shoop, W.; Kodali, S.; Galgoci, A.; Painter, R.; Parthasarathy, G.; Tang, Y. S.; Cummings, R.; et al. Platensimycin Is a Selective FabF Inhibitor with Potent Antibiotic Properties. *Nature* **2006**, *441*, 358–361.

(153) Schaftenaar, G.; Noordik, J. H. Molden: a Pre- and Post-processing Program for Molecular and Electronic Structures. *J. Comput. Aided Mol. Des.* **2000**, *14*, 123–134.

(154) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Montgomery, J.; Vreven, T.; Kudin, K.; Burant, J.; et al. Gaussian 03, Revision C.02. **2003**.

(155) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A Well-behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(156) Mladenovic, M.; Fink, R. F.; Thiel, W.; Schirmeister, T.; Engels, B. On the Origin of the Stabilization of the Zwitterionic Resting State of Cysteine Proteases: A Theoretical Study. *J. Am. Chem. Soc.* **2008**, *130*, 8696–8705.

(157) Zhang, Y.; Kua, J.; McCammon, J. A. Role of the Catalytic Triad and Oxyanion Hole in Acetylcholinesterase Catalysis: An Ab Initio QM/MM Study. *J. Am. Chem. Soc.* **2002**, *124*, 10572–10577.

(158) Ishida, T.; Kato, S. Theoretical Perspectives on the Reaction Mechanism of Serine Proteases: The Reaction Free Energy Profiles of the Acylation Process. *J. Am. Chem. Soc.* **2003**, *125*, 12035–12048.

(159) Schaefer, B.; Kisker, C.; Sotriffer, C. A. Molecular Dynamics of Mycobacterium Tuberculosis KasA: Implications for Inhibitor and Substrate Binding and Consequences for Drug Design. *J Comput Aided Mol Des* **2011**, *25*, 1053–1069.

Acknowledgements

First of all, I would like to express my sincere gratitude to Prof. Bernd Engels for giving me the great opportunity to carry out my doctoral study under his supervision. He was always patient and tried to listen carefully during the discussions we had. I am especially grateful for his trust in me during the doctoral study.

I would like to thank Prof. Christoph Sotriffer for being a referee of my thesis defense.

I would like to thank Prof. Caroline Kisker for the collarboration and the fruitful discussions we had.

My thanks also go to Uschi Rüppel for her great help that she gave me whenever I had problems with German bureaucratic procedure.

I want to thank my office mate, Kay Ansorg, for making a pleasant working environment. Moreover, he always tried to help me to integrate into our group, and I really appreciate that.

I also thank Christof Walter for sharing insights in the area of excited states. He was always willing to give detailed explanations to lots of my questions.

Many thanks to Thomas Schmidt, Christoph Grebner, and Alexander Paasche for all the fruitful discussions.

My thanks also go to all the other group members, Johannes Becker, Zarah Falk, Vera Stehr, Anca Boariu, Anastasia Weickert, Daniel Bellinger, Daniel Weber, Anna Heilos, and Dr. Maxim Tafipolski, and former group members, Prof. Reinhold Fink, Volker Settles, Wenlan Liu, Johannes Pfister and Dr. Prabhat Sahu.

I want to express my sincere gratitude to my parents and my brother in Korea for all the support, encouragement, and all their love I received. I am also grateful to my family-in-law for their support and encouragement.

My greatest thanks go to my wife, Kyungmi, and my son, Dongha, who always supported me through both happy and difficult times during my study.

Finally, I thank God who always leads my way.