Bachelor Thesis

A Molecular Dynamics Study of Proton Transfer in Acetylcholinesterase

by

Rachel Glaves

October 22, 2004

Lehrstuhl für Theoretische Chemie
Ruhr-Universität Bochum

My thanks go to Prof. Dr. Dominik Marx, Department of Theoretical Chemistry, for enabling me to carry out this work in his department and to Axel Kohlmeyer, Holger Langer and Marcel Baer, without whose help this thesis never would happened!

Contents

1 Introduction

The enormous variety of biochemical reactions that comprise life are nearly all mediated by a series of remarkable biological catalysts known as enzymes. Under biologically relevant conditions, uncatalyzed reactions tend to be slow - most biological molecules are quite stable in the neutral pH, mild-temperature, aqueous environment inside cells. Enzymes have extraordinary catalytic power, far greater than that of synthetic or inorganic catalysts. They have a high degree of specificity for their substates, they accelerate chemical reactions tremendously and they function in aqueous solutions under very mild conditions of temperature and pH. Few nonbiological catalysts have all these properties.

Enzymes act in organized sequences to catalyze the hundreds of stepwise reactions by which nutrient molecules are degraded, chemical energy is conserved and transformed and biological macromolecules are made from simple precursors. Through the action of regulatory enzymes, metabolic pathways are highly coordinated to yield a harmonious interplay among the many different activities necessary to sustain life.

The study of enzymes has immense practical importance. In some diseases, especially inheritable genetic disorders, there may be a deficiency or even a total absence of one or more enzymes. For other disease conditions, an excessive activity of an enzyme may be the cause. Many drugs exert their biological effects through interactions with enzymes, that is one reason why it is of paramount importance to investigate exactly which interactions occur between the ligand and the enzymes' active site, and how the enzyme recognises its specific ligand. Once these mechanisms are elucidated, it is possible to produce so called "designer drugs" that act only on one specific enzyme, helping to cure or alieviate enzyme-based diseases.

Molecular dynamics simulation is a successful and well-established method in the computational study of structure and dynamics of biological macromolecules. Due to the large size of most enzymatic systems and the fact that many reactions that they catalyze involve the making and breaking of bonds, the study of enzyme *reactions* remains very challenging for computational methods. Classical and first principles molecular dynamics (FPMD) simulations permit the study of complex, dynamic processes and reactions that occur in biological systems, including protein stability and folding and ion transport in biological systems. Molecular dynamics simulations generate information at the microscopic level, including atomic positions and velocities, and in the case of FPMD also the electronic structure. This information can be converted into macroscopic observables such as pressure, energy, heat capacity, etc, using statistical mechanical analysis.

The aim of this experiment was to investigate the feasability of studying the mechanism of proton transfer in the active site of the enzyme acetylcholinesterase (AChE) with a small FPMD model system. Because of the pivotal role that AChE plays in the nervous system, the enzyme has long been an attractive target for the design of mechanism-based-inhibitors. Clinically, a moderate inhibition of AChE is effective in the treatment of a number of diseases to prolong

the effect of its ligand acetylcholine (ACh) on its receptors. Such treatment is desirable either if there is a reduced concentration of ACh, as in the case of Alzheimer's disease, or if there are fewer receptors, as in the case of myasthenia gravis, a rare autoimmune disease which causes paralysis and death. Therefore studies that help to elucidate every detail of this pivotal enzymes mechanism are a step towards curing many diseases. However, overwhelming inhibition of AChE, particularly by covalent bonding to the active site serine, is invariably lethal. Hence, AChE is a prime target for naturally occuring toxins such as the snake venom fasciculin 2, pesticides such as parathion, and chemical warfare agents such as sarin and VX. These toxins can therefore be used as inhibitors to block the mechanism of AChE and are therefore valuable aids in the quest to elucidate the role and function of acetylcholinesterase.

2 Theoretical background

2.1 Acetylcholinesterase

The enzyme we chose to study is acetylcholinesterase (acetylcholine acetylhydrolase, EC 3.1.1.7; abbrieviated herein AChE), a serine hydrolase that belongs to the esterase family within higher eukaryotes. This family acts on different types of carboxylic esters. AChE's biological role is the termination of the nerve impulse transmissions at cholinergic synapses found at neuromuscular junctions. The skeletal muscle fibers are innervated by large myelinated nerve fibers that originate in the spinal cord. Each nerve fiber then branches and stimulates three to several hundred skeletal muscle fibers. Each nerve ending makes a junction, called the neuromuscular junction, with the muscle fiber near its midpoint and the resulting action potential in the muscle fiber travels in both directions toward the muscle fiber ends. The cationic neurotransmitter acetylcholine (ACh), that participates in the transmission of a given nerve impulse, is synthesized from choline and acetyl-coenzyme A and packaged into small synaptic vesicles in the presynaptic terminal. When an action potential reaches the neuromuscular junction, the large amounts of voltage-gated calcium channels in the presynaptic nerve membrane open and allow large numbers of calcium ions to flow from the synaptic space into the presynaptic nerve terminal. The calcium ions then bind with special protein receptor molecules ("release site") on the inside surface of the presynaptic membrane, causing about 125 vesicles of ACh to fuse with the release sites and to empty their contents into the synaptic space by a process called exocytosis.

The released ACh diffuses across the synaptic space towards the postsynaptic nerve ending (the so called myocytes) and docks onto small ACh protein receptors, which function as acetylcholine-gated ion channels. The binding of two ACh molecules to the receptor cause a conformational change in it which opens the channel. The opened acetylcholine channel has a diameter of about 0.65 nm, which is large enough to allow all the important positive ions, sodium (Na^+), potassium (K^+) and calcium (Ca^{2+}), to move easily through the opening. Conversly, negative ions, such as the ever-present chloride ions (Cl^-), do not pass through because of the strong negative charges in the mouth of the channel. In practice, far more Na^+ ions flow through the channel than others, which creates a local positive potential charge inside the muscle fiber membrane (it normally has a negative potential)-the end plate potential. In turn, this positive potential change initiates an action potential in the postsynaptic muscle membrane and thus causes muscle contraction.

The ACh, once released into the synaptic space, continues to activate the ACh receptors as long as the neurotransmitter persists in the cleft. It follows that the ACh must be degraded within the next few milliseconds before the arrival of the next nerve impulse, otherwise a renewed stimulation of the postsynaptic nerve cell is not possible. This essential task is accomplished in mammels by the 75 kD fast-acting enzyme acetylcholinesterase that is GPI (glycoinositol phospholipid)-anchored to the surface of the postsynaptic muscle membrane inside the synaptic

space. The short period that the ACh remains in the synaptic space – a few milliseconds at most – normally is sufficient to excite the muscle fiber. Then the rapid removal of the ACh by AChE prevents continued muscle re-excitation after the fiber has recovered from the first action potential.

The enzymes' catalytic efficiency is close to the diffusion-controlled limit ($k_{cat}/K_m = 1.5 \cdot 10^8\,\mathrm{M}^{-1}\mathrm{s}^{-1}$ [19]) so that it is a nearly perfect catalyst. The great speed of the enzyme is essential for rapid modulation of synaptic activity, it's among the fastest of enzymes and thus has amazing catalytic power. The turnover number k_{ES} is $> 10^4\,\mathrm{s}^{-1}$ [11], which is among the highest turnover numbers reported for enzyme catalysis. For the physiological substrate ACh k_E is $> 10^8\mathrm{M}^{-1}\mathrm{s}^{-1}$ [11] and therefore, diffusion of ACh to the active site is probably rate determining. This high second-order rate constant is a hallmark of an evolutionarily perfect enzyme. One factor that contributes to the speed of binding of cationic ligands to AChE is electrostatics [12]. However, this electrostatically enhanced collision rate between ligands and the enzyme only act to increase the speed of encounter of the cationic ligands with the enzyme. For ligands to reach the active site, they still must move through the steric bottleneck in the primary gorge, and water molecules in the channel must move out of the way of the ligand as it moves along this narrow pathway (see section 2.3.1).

The products of the ACh degradation reaction, an acetate ion and choline, are transported back into the presynaptic nerve terminal where they are recycled repeatedly to ACh and packaged once again into vesicles.

2.2 Structure of AChE

2.2.1 Primary Structure

In *T. californica* a major form of AChE is a homodimer bound to the plasma membrane through covalently attached phosphatidylinositol (PI). This dimer has a simpler quaternary structure than other forms of AChE and its amino acid sequence and the arrangement of its intrachain disulfide bridges have been determined [13]. The native enzyme contains a single polypeptide of 575 amino acids and has a calculated molecular weight of 65 612 Da [11]. The pattern of disulfide bonds and the consequent protein folding they produce in *T. californica* are different from those of mammelian and bacterial serine proteases. *T.californica* AChE contains three pairs of disulfide bonds.

The primary structure of the catalytic domains of the acetylcholinesterases from such species as *T. californica*, mouse and human are quite well conserved.

2.2.2 Tertiary and Quaternary Structure.

The three-dimensional structure of the 537-amino acid residue AChE from the electric fish *Torpedo californica* has been determined by X-ray analysis to 2.8 Å resolution [13]. The crystallized form was the glycolipid (GPI)-anchored homodimer. (Throughout this thesis, the sequence numbers follow the amino acid abbreviations of *T. californica* AChE). The enzyme monomer has an ellipsoidal shape with dimensions of $\sim$45 Å by 60 Å by 65 Å. The molecule belongs to the class

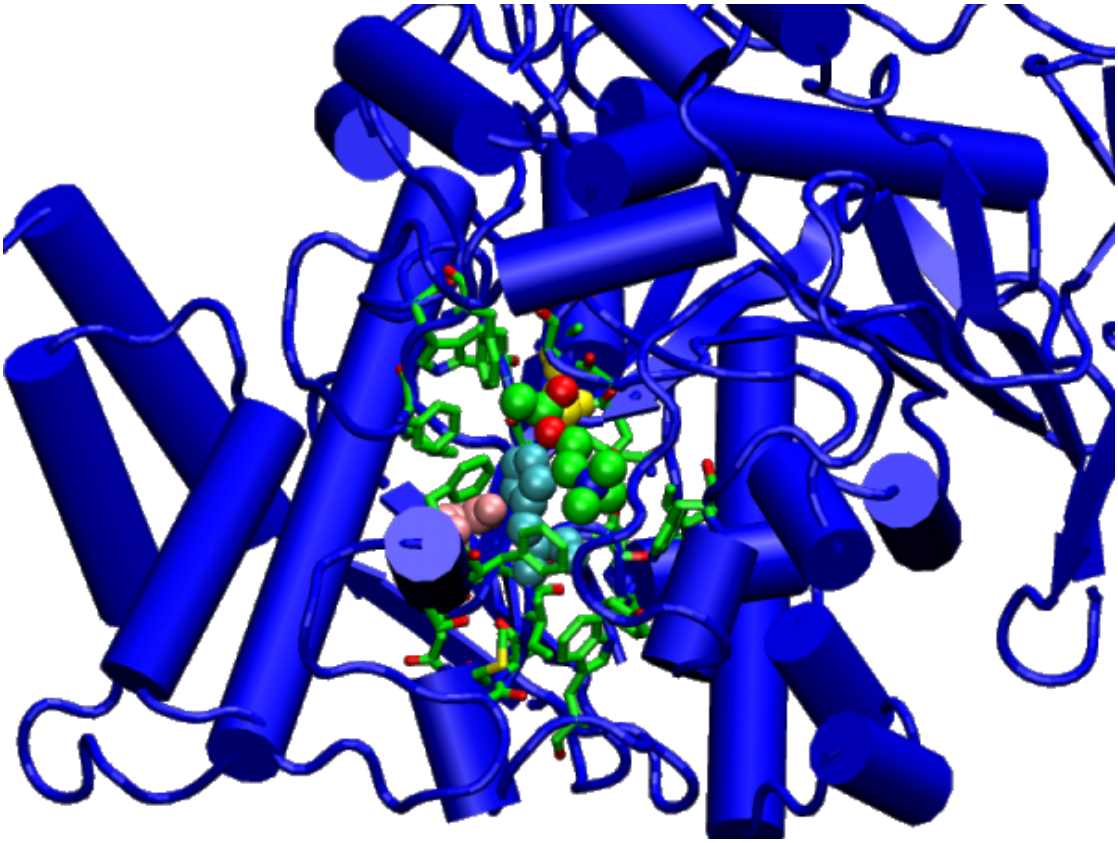

Figure 2.1: The Structure of Acetylcholinesterase looking down the Gorge onto the Active Site.

of α/β proteins and consists of a 12-stranded central mixed β sheet surrounded by 14 α helices. The first and last pairs of strands each form β-hairpin loops that are only loosely hydrogen bonded to the eight central, superhelically twisted strands [13].

The AChE homodimer, whose subunits are related by a crystallographic twofold axis, appears to be held together by a four-helix bundle composed of helices from each subunit. The only interchain disulfide bond involves the carboxyl-terminal Cys^{537}.

AChE is a polymorphic enzyme in vivo with globular catalytic subunits that have masses of 70-80 K [11]. Monomer and disulfide-linked dimer forms may be soluble or attached to the surface of the connective tissue that fills the synaptic space between the presynaptic nerve terminal and the postsynaptic muscle membrane by means of a glycophospholipid (GPI-anchored). The tetramer form may be soluble, lipid-linked to the membrane, or attached to a collagen triple helix [12].

2.2.3 Active Site Structure.

AChE is a serine hydrolase whose catalytic mechanism resembles that of serine proteases such as trypsin. The essential catalytic functional unit of the serine proteases is the catalytic triad,

composed of three precisely orientated active site amino acids (His^{57}, Asp^{102} and Ser^{195} in chymotrypsin and Asp^{32}, His^{64} and Ser^{221} in subtilisin). This triad has been called the charge relay system and it stabilizes chemical catalytic transition states via coupled proton transfers. However, comparison of the primary amino acid sequences from AChE and other serine enzymes show little homology in the active site regions, and no significant homology extends through the rest of the sequences [11]. Therefore the often used serine protease analogy is not supported by comparisons of primary structures.

X-ray structure experiments and primary structure comparisons in *Torpedo californica* have revealed that the catalytic triad consists of the three amino acid residues Ser^{200}, His^{440} and Glu^{327}. The third member of the triad is a glutamic acid residue rather than an aspartic acid residue, only the second instance of a Glu residue in this position among the many serine proteases, lipases and esterases of known structure. It is also unusual because the relation of the triad to the rest of the protein approximates a mirror image to that seen in the serine proteases. The mutation of any of these three residues to alanine leads to a decrease in catalytic activity of at least 3300-fold [21].

The most remarkable feature of the structure of AChE is a deep and narrow gorge, circa 20 Å long, that penetrates halfway into the enzyme and widens out close to its base. The active site, located at the base of this long gorge, consists of two subsites: an "esteratic" subsite containing the catalytic machinery and an "anionic" subsite responsible for binding the quaternary trimethylammonium tail group of ACh (choline-binding pocket). For the substrate to reach the active site from the surface of AChE, molecular dynamic simulations have revealed that the enzyme structural fluctuations play an essential role "breathing", which open the gorge every few picoseconds allow the passage of the substrate [3, 6, 12, 14, 15, 21, 22].

The AChE active site consists of three major domains: a). the esteratic locus (subsite), comprised of the active site serine; b). an anionic locus that is $\sim$4.7 Å from the esteratic serine, where the quaternary ammonium of ACh binds; and c). a hydrophobic region that is contiguous with or near the esteratic and anionic loci and is important in binding aryl substrates and active site ligands.

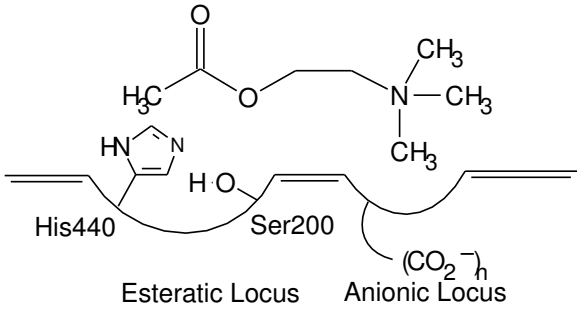

Figure 2.2: Representation of the Active Site of AChE.

A fourth domain in the enzyme binds cationic ligands, such as *d*-tubocurarine [11]. This domain is over 20 Å away from the active site and has been called the peripheral active site. Ligand binding to the peripheral active site frequently causes a conformational change of the active site itself. The four domains of the enzyme act together to produce the complex reaction dynamics

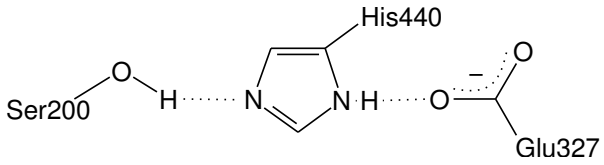

Figure 2.3: Hydrogen Bonds in the Catalytic Triad Setup.

and active site conformational dynamics that are hallmarks of this, and similar, enzymes.

2.2.4 The Esteratic Subsite.

In addition to the nucleophilic element of the active site, serine (Ser^{200}), site-directed mutagenesis has shown that that the esteratic locus contains the imidazole functionality of histidine (His^{440}) [13], which functions as a general acid-base catalyst in the formation and decomposition of tetrahedral intermediates (see section 2.3). Together with a glutamic acid Glu^{327}, the three residues form a planar array that resembles the catalytic triad of chymotrypsin, with the above-mentioned differences. The γ-oxygen atom of Ser^{200}, which can be seen from the surface of the enzyme, is 4 Å above the base of the gorge. Fourteen aromatic residues line a substantial portion ($\sim$40%) of the surface of the gorge [13]. These residues and their flanking sequences are highly conserved in AChEs in different species. The gorge contains only few acidic residues.

The oxyanion hole, formed by the peptidic NH groups of Gly^{118}, Gly^{119} and Ala^{201}, is another important functional unit in the esteratic subsite. The X-ray structure of a transition state analogue complexed with *T. californica* AChE revealed that a three-pronged oxyanion hole exists in the active site of AChE, in contrast to the two-pronged oxyanion hole in most serine proteases [21]. It also revealed that the acetyl headgroup of ACh is held in place by the oxyanion hole through hydrogen bonding of the carbonyl oxygen of ACh with the peptidic NH groups. The catalytic role of the oxyanion hole in serine proteases is therefore generally established to be in stabilizing the high-energy intermediates and the transition states through hydrogen bonding [21].

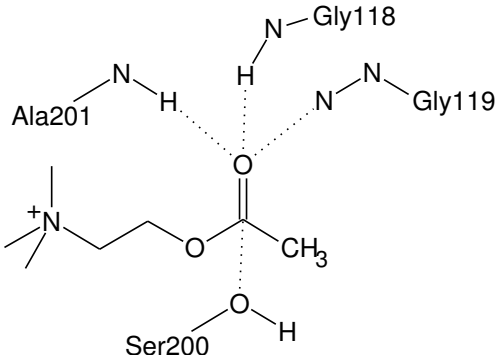

Figure 2.4: The esteratic Binding Site for the Acetyl Headgroup.

2.2.5 The Anionic Subsite.

The anionic locus of the active site is 4.7 Å from the serine hydroxyl of the esteratic locus, and contains multiple negative charges. An anionic subsite is signalled by the interaction of various types of compounds with AChE, including aromatic cations and tetraalkylammonium salts. The subsite binds both quaternary ligands, such as edrophonium, which act as competitive inhibitors, and quaternary oximes, which often serve as effective reactivators of organophosphate-inhibited AChE. Cohen and co-workers [13] have suggested, on the basis of studies in which cationic and uncharged homologs of ACh were used, that the "anionic" subsite is, in fact, uncharged and lipophilic. It contains at most one formal negative charge. This is supported by the presence of aromatic residues in the active site area. Instead the quaternary moiety in ACh appears to bind chiefly through interactions with the π electrons in these aromatic residues in AChE. The gorge is so deep and its aromatic surfaces so extensive that there must be many different ways and places for substrates, agonists and inhibitors to bind to AChE [13].

In the anionic subsite, site-directed mutagenesis indicates that Trp^{86}, Glu^{202} and Phe^{337}[22] play an important role in binding the quaternary trimethylammonium tailgroup. Trp^{86} and Phe^{337} are thought to bind to the cationic moiety mainly through cation-π interactions and hydrophobic interactions, while the interaction between Glu^{202} and the cation moiety is often attributed to electrostatics.

The locus binds the quaternary ammonium tailgroup of ACh and various inhibitors. Experimental studies have established that both the esteratic and anionic loci are responsible for substrate specificity in AChE catalysis [6].

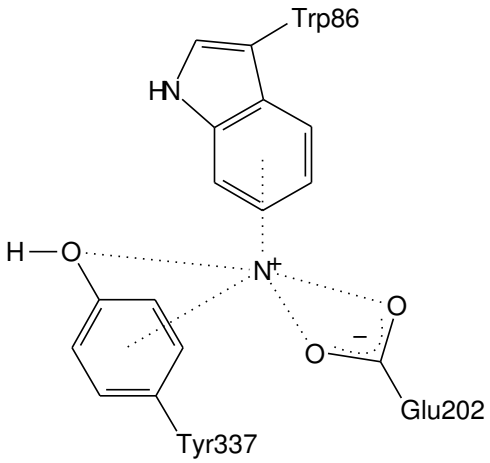

Figure 2.5: The Anionic Binding Site for the Tailgroup.

2.2.6 Hydrophobic Regions.

The anionic and esteratic loci interact with various classes of hydrophobic ligands, for example, aromatic cations and bis-quaternary ammonium ligands. These substrate and ligand interactions are supported by structural studies of AChE that place hydrophobic residues at the active site

[13]. The high aromatic content of the walls and floor of the active site gorge, together with its dimensions, may explain why biochemical studies have revealed a variety of hydrophobic and anionic binding sites distinct from, or overlapping, the active site.

2.3 The Current Picture of the Reaction Mechanism of Acetylcholinesterase

The following section describes the reaction mechanism as it has been deduced using experimental techniques (e.g. point mutations). For some intermediate states, model semi-empirical calculations [20] and single point QM/MM calculations [22] have been performed to try and verify these observations.

AChE catalysis occurs via an acylenzyme mechanism that involves nucleophilic and acid-base elements. The catalytic triad stabilizes the catalytic transition states via coupled proton transfers. It is an acylation-deacylation reaction, which occurs in two stages.

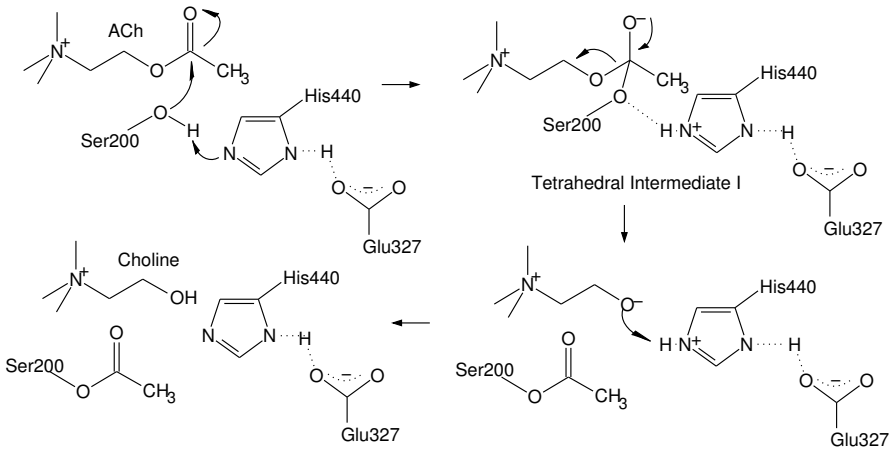

Figure 2.6: The Acylation Reaction.

The first step of the acylation reaction, as shown in figure 2.6, is the simultaneous transfer of the acidic hydroxy proton from the serine residue to the imidazole ring of His^{440} and the nucleophilic attack by the now negatively charged side chain oxygen of Ser^{200} at the carbonyl group of ACh, displacing the choline to form a tetrahedral acyl-AChE intermediate. The acetyl headgroup of ACh is held in place by the oxyanion hole and the anionic subsite stabilizes the ACh tailgroup throughout this step. The contribution of Glu^{327} to the catalytic reaction is thought to be manifold. The first contribution is the electrostatic interaction between the carboxy group of Glu^{327} and the incipient imidazolium cation, which stabilizes the transition state and the tetrahedral intermediate. The second is the so-called "charge-relay" mechanism where the carboxy group of Glu^{327} also serves as the general acid-base catalyst, which is involved in the proton transfer between Glu^{327} and His^{440}. The third and most recently proposed contribution is the "low-barrier hydrogen bond" mechanism, wherein the low-barrier hydrogen bond formed between carboxylate of Glu^{327} and the incipient imidazolium cation of His^{440} provides a special

stabilization energy to lower the reaction barrier [21]. The intermediate then breaks down into choline and acylated serine and, in doing so, restores the original deprotonated status of His^{440}. Thus the His^{440} residue acts as a simple general acid-base catalyst.

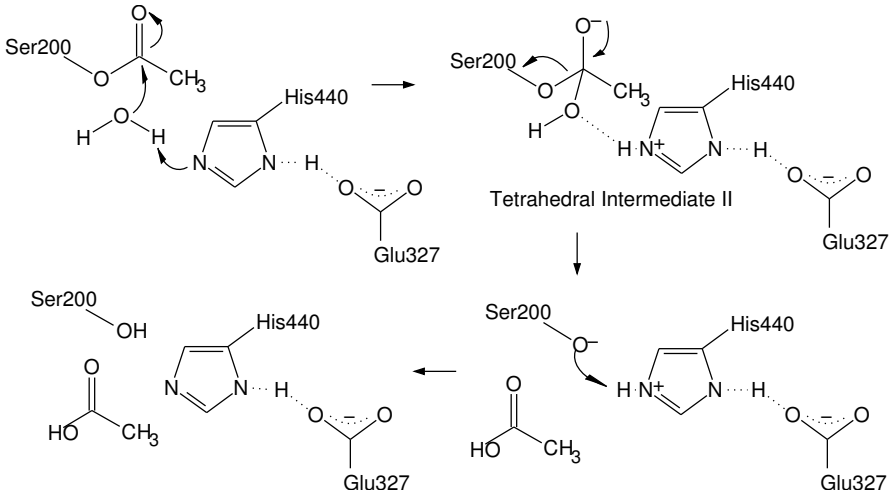

Figure 2.7: The Deacylation Reaction.

Figure 2.7 pictures the deacylation reaction. The previously deprotonated His^{440} nitrogen extracts a proton from a neighboring water molecule and the nucleophilic (negative) water oxygen attacks the carbonyl carbon of the acylated serine. The tetrahedral intermediate formed disintegrates to acetic acid and restores Ser^{200} and His^{440} to their original status.

2.3.1 Conformational fluctuations.

As the active site of AChE is buried deep in the enzyme, gorge breathing motions allow substrate access to the active site dynamic protein fluctuations. Various molecular dynamics simulations [3, 6, 12, 14, 15, 21, 22] of AChE have revealed that the gorge fluctuations are necessary to permit ligands such as ACh to enter the narrow gorge and access the active site. The ligands must move through the steric bottleneck in the channel and water molecules in the channel must move out of the way of the ligand as it moves along this narrow pathway. The simulations have revealed that both fluctuations in the width of the bottleneck in the primary channel or gorge and the transient opening of a second channel (the “back door”) allow water molecules to exchange between the active site and the exterior of the enzyme. The transient existance of additional possible water exchange pathways between the active site region and the exterior through comparatively thin sections of the gorge was also confirmed.

2.4 Simulation Methods.

2.4.1 Molecular Dynamics Simulation.

In the traditional molecular dynamics approach, the history or future of an initial arrangement of atoms is followed by calculating the trajectories of all the particles under the influence of the intermolecular potentials. The motion of small particles such as atoms or molecules is well described by Lagrangian mechanics, with the Lagrangian $\mathcal{L}$ defined as the kinetic energy minus the potential energy:

$$\mathcal{L} = T - V \tag{2.1}$$

Using Newton's second law of motion and the fact that the force can also be expressed as the negative gradient of the potential energy it leads to the set of Newtonian equations of motion for each particle i, with mass m_i and cartesian coordinate r:

$$-\frac{\mathrm{d}V}{\mathrm{d}r_i} = m_i \cdot \frac{\mathrm{d}^2 r_i}{\mathrm{d}t^2} \tag{2.2}$$

Using the knowledge of the force on each atom, it is therefore possible to determine the acceleration of each atom in the system. Integration of the equations of motion then yields a trajectory that describes the evolution of positions, velocities and accelerations of the particles with time. From this trajectory, average values of structural and thermodynamic properties can be determined. Therefore to calculate a trajectory, one only needs the initial positions of the atoms, an initial distribution of velocities and the acceleration, which is determined by the gradient of the potential energy function. The initial positions can be obtained from experimental structures, for example x-ray crystal structures (as is the case in this thesis) or NMR spectroscopy. These calculated trajectories correspond to the conformations that the molecule can sample at the temperature of the simulation. Current empirical force fields (or potential energy functions), like AMBER, GROMOS and CHARMM, have however, certain limitations. One of the most important is, that no drastic changes in electronic structure are allowed; for example bond making and bond breaking cannot be easily modelled. When the chemistry of a system plays an important role in the reaction, that means when there is making and breaking of chemical bonds or changing electronic environments, first principle molecular dynamics (FPMD) schemes come into play.

2.4.2 Car – Parrinello Molecular Dynamics.

FPMD performs a molecular dynamics simulation whilst calculating the forces on the nuclei quantum mechanically (directly from the electron structure), by solving the Hartree-Fock (HF)- or Kohn-Sham (KS)-equations, for example, after every time step. Explicitly treating the electrons in an *ab initio* way in the calculation means that one does not make assumptions about the bonding of the system, allowing one to model chemical reactions.

The Born-Oppenheimer approximation implies that at any instant the state of the electronic system can be well-described by the electronic ground state calculated for the ionic positions at that instant and that it responds instantaneously to changes in ionic positions. For most systems

this is an extremely good approximation given that the mass of the electron is much smaller than that of the atoms. A number of techniques, generally called Born-Oppenheimer methods, therefore calculate the ground state of the orbitals, by minimization of the electronic Kohn-Sham orbitals, see section 2.4.3 below, after each time step. This is extremely time consuming and the computational cost is large, so Car and Parrinello devised a method that does not put the orbitals in their ground state at each time step, but lets them follow a fictitious Newton ion dynamic. It has been shown that the effect of the fictitious dynamic on the motion of the nuclei is very minor and that, provided there is a large enough band gap and a small fictitious mass and time step, one can obtain trajectories that are equivalent to ground state BO – methods. Improved methods of calculating the electronic ground state during BO – MD and faster computers have however reduced the advantages of the CPMD method.

2.4.3 Density-Functional Theory and the Local-Density Approximation.

Density-functional theory (DFT; Kohn and Sham 1965) can be used to describe the effects of correlation and exchange in an electron gas. These electron-electron interactions represent the most difficult problem in calculating the electronic structure or total energy of a chemical structure. The exchange energy is due to the fact that electrons are fermions and their wavefunctions antisymmetrical. This antisymmetry produces a spatial separation between electrons that have the same spin, which reduces the Coulomb energy of the electronic system by a value called the exchange energy. It is straightforward to include exchange in a total-energy calculation,and this is generally referred to as the Hartree-Fock approximation. If electrons that have opposite spins are also separated spatially from one another, the Coulomb energy of the system can be reduced underneath its Hartree-Fock value and the difference between this energy and the Hartree-Fock energy of the system is called the correlation energy, which is extremely hard to calculate. In DFT however, Hohenberg and Kohn proved that the total energy, including that of the many body effects of electrons (exchange and correlation) in the presence of a static external potential (for example, the atomic nuclei) of an electron gas is a unique functional of the electron or charge density. The minimal value of the total-energy functional is the ground state energy of the system of electrons with the ions in positions R_I. The electronic charge density which yields this minimum is then the exact single particle ground state energy. It was then shown by Kohn and Sham (KS) that it is possible to replace the many electron problem by an exactly equivalent set of self-consistent one electron equations. The KS total energy functional can be written as a sum of several terms:

$$E[\{\psi_i\}] = 2\sum_i \int \psi_i \left(-\frac{\hbar^2}{2m}\right) \nabla^2 \psi_i \, \mathrm{d}^3\mathbf{r} + \int V_{ion}(\mathbf{r}) n(\mathbf{r}) \, \mathrm{d}^3\mathbf{r} + \frac{e^2}{2} \int \frac{n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \, \mathrm{d}^3\mathbf{r} \, \mathrm{d}^3\mathbf{r}' + E_{XC}[n(\mathbf{r})] + E_{ion}(\{\mathbf{R_I}\}) \quad (2.3)$$

where E_{ion} is the Coulomb energy associated with the interactions between the nuclei at the positions $\{\mathbf{R_I}\}$, and $E_{XC}[n(\mathbf{r})]$ is the exchange-correlation functional. V_{ion} is the Coulomb potential between the electrons and the nuclei and the third term is the Coulomb interaction

between the electrons, both of which are simply functions of the electronic charge density $n(\mathbf{r})$, defined below:

$$n(\mathbf{r}) = 2\sum_i |\psi_i(\mathbf{r})|^2 \tag{2.4}$$

The set of wave functions ψ_i that minimize the above KS-energy functional (i.e. the wave functions that represent the ground state of the system) are given by the self-consistent solutions to the Kohn-Sham equations (regarded as a Hamiltonian), which are a set of N coupled, three-dimensional, partial differential equations:

$$\left[\frac{-\hbar^2}{2m}\nabla^2 + V_{ion}(\mathbf{r}) + V_{XC}(\mathbf{r})\right]\psi_i(\mathbf{r}) = \varepsilon_i\psi_i(\mathbf{r}) \tag{2.5}$$

where ψ_i is the wave function of the electronic state i, ϵ_i is the Kohn-Sham eigenvalue, and V_H is the Hartree potential given by:

$$V_H(\mathbf{r}) = e^2\int \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\mathrm{d}^3\mathbf{r}' \tag{2.6}$$

The exchange-correlation potential V_{XC}, is defined formally by the functional derivative:

$$V_{XC}(\mathbf{r}) = \frac{\delta E_{XC}[n(\mathbf{r})]}{\delta n(\mathbf{r})} \tag{2.7}$$

The KS equations represent a mapping of the interacting many-electron system onto a system of non-interacting electrons moving in an effective potential due to all the other electrons. If the exchange-correlation functional were known exactly, then taking the functional derivative with respect to the density (see 2.7) would produce a potential that included the effects of exchange and correlation exactly. DFT therefore uses direct minimization or molecular dynamics to solve the eigenvalue problem of the set of KS-equations, once an approximate expression for the exchange-correlation energy is given. The simplest method of desribing the exchange-correlation energy of an electronic system is to use the local-density approximation (LDA; Kohn and Sham, 1965).

In seeking solutions to the system of equations 2.5, it is found that all quantities are represented as functionals of the electronic charge density. The important point that makes this system easier to solve (or more precisely, requires less computational effort) than for example Hartree-Fock equations, is that the effective potential is local. Of course, this is only true if the exchange-correlation energy can be described as a function of the local charge density. One method of doing this is to use the local density approximation, where the exchange-correlation energy of an electronic system is constructed by assuming that the exchange-correlation energy per electron at a point r in the electron gas $\epsilon_{XC}(\mathbf{r})$, is equal to the exchange-correlation energy per electron in a homogeneous electron gas that has the same electron density at the point r. Thus

$$E_{XC}[n(\mathbf{r})] = \int \epsilon_{XC}(\mathbf{r})n(\mathbf{r})\,\mathrm{d}^3\mathbf{r} \tag{2.8}$$

and

$$\frac{\delta E_{XC}[n(\mathbf{r})]}{\delta n(\mathbf{r})} = \frac{\delta[n(\mathbf{r})\epsilon_{XC}(\mathbf{r})]}{\delta n(\mathbf{r})} \tag{2.9}$$

$$\epsilon_{XC}(\mathbf{r}) = \epsilon_{XC}^{hom}[n(\mathbf{r})] \tag{2.10}$$

Where equation 2.10 is the assumption that the exchange-correlation energy is purely local. In LDA, corrections to the exchange-correlation energy due to the inhomogeneities in the electron charge density at a point $\mathbf{r}$ are ignored. Considering the inexact nature of the approximation, it produces remarkably accurate descriptions of the exchange-correlation energy of an electronic system. To improve the inexact nature of the LDA, information about the gradient of the charge density, not just the density itself, at a particular point $\mathbf{r}$ is used to account for the inhomogeneities of the true electron density. This means that one interprets the LDA as the first term of a Taylor expansion of the uniform density to obtain better approximations of the exchange – correlation functional. Functionals of this form, for example PBE [10] and BLYP [2, 7] are known as generalized gradient approximations.

2.4.4 Bloch's Theorem and Plane Waves.

Even though certain observables of the many-body problem can be mapped into equivalent observables in an effective single-particle problem solving the difficult exchange-correlation energy problem, there still remains the formidable task of handling an infinite number of noninteracting electrons moving in the static potential of an infinite number of nuclei or ions. Two difficulties must be overcome: a wave function must be calculated for each of the infinite number of electrons in the system and, since each electronic wave function extends over the entire solid, the basis set required to expand each wave function is infinite. Both problems can be solved by performing calculations on periodic systems, in which a large unit cell containing the configuration in question is repeated periodically throughout space, and applying Bloch's theorem to the electronic wave functions. The ions in a perfect crystal are arranged in a regular periodic way (at 0 K). Therefore the external potential felt by the electrons will also be periodic-the period being the same as the length of the unit cell, 1. That is, the external potential on an electron at r can be expressed as:

$$V(\mathbf{r}) = V(\mathbf{r} + 1) \tag{2.11}$$

This requirement is needed to apply Bloch's theorem, with which it is possible to express the wave function of the infinite crystal in terms of wave functions at reciprocal space vectors of a lattice. Bloch's theorem uses the periodicity of a crystal to reduce the infinite number of one-electron wave functions to be calculated to simply the number of electrons in the unit cell of the crystal (or half that number if the electronic orbitals are assumed to be doubly occupied-that is, spin degenerate). Bloch's theorem states that in a periodic solid each electronic wave function can be written as the product of a cell-periodic part and a wavelike part:

$$\psi(\mathbf{r}) = \exp[i\mathbf{k} \cdot \mathbf{r}] f_i(\mathbf{r}) \tag{2.12}$$

The first term is the wavelike part (see later), the second term is the cell periodic part of the wave function, which can be expressed by expanding it using a basis set consisting of a finite number of plane waves whose wave vectors are reciprocal lattice vectors of the crystal:

$$f_i(\mathbf{r}) = \sum_G c_{i,G} \exp[i\mathbf{G} \cdot \mathbf{r}] \tag{2.13}$$

where $\mathbf{G}$ represents the reciprocal lattice vectors, which are defined by $\mathbf{G} \cdot \mathbf{l} = 2\pi m$ for all $\mathbf{l}$ where $\mathbf{l}$ is a lattice vector of the crystal and m an integer. Therefore each electronic wave function can be written as a sum of plane waves:

$$\psi_i(\mathbf{r}) = \sum_G c_{i,\mathbf{k}+\mathbf{G}} \exp[i(\mathbf{k} + \mathbf{G}) \cdot \mathbf{r}] \tag{2.14}$$

Therefore by the use of Bloch's theorem, the problem of the infinite number of electrons has now been mapped onto the problem of expressing the wave function in terms of an infinite number of reciprocal space vectors within the first Brillouin zone (BZ) of the periodic cell k. This problem is dealt with by sampling the BZ at special sets of k-points. Electronic states are allowed only at a set of k-points determined by the boundary conditions that apply to the bulk solid. The infinite numbers of electrons in the solid are now accounted for by an infinite number of k-points, each with a finite number of occupied electronic states. Therefore one now has the task of calculating a finite number of electronic wave functions at an infinite number of k-points; it would normally require an infinite number of calculations to compute this potential. However, because the wave functions at k-points that are very close together are practically identical, it is possible to represent the electronic wave functions over a close region of k-space by the wave functions at a single k-point. In this case the electronic states at only a finite number of k-points are required to calculate the electronic potential and hence determine the total energy of the solid.

Bloch's theorem thus implies that the wave function of an electron in a periodic potential can be expanded in a plane-wave (PW) basis set, and in principle, an infinite number of plane waves is required by the theory. However, the Fourier coefficients $c_{i,k+G}$ of the wave function decrease with increasing $|k + G|$, so that the PW expansion can be effectively truncated after a finite number of terms, for example, it can be limited to all waves with kinetic energy lower that some particular energy cutoff, E_{cut}. Basically what this means is that the coefficients from plane waves with small kinetic energy are more important than those with large kinetic energies. Introduction of a plane wave energy cutoff reduces the basis set to finite size. It also regretably introduces an error in the total energy of the system, but it can easily be reduced by increasing the size of the basis set by allowing a larger energy cutoff value.

Another advantage of expanding the electronic wave functions in terms of a basis set of plane waves is that the KS-equations take on a particularly simple form. Substitution of the "sum of plane waves" wave function (equation (2.14)) in the KS-equations (equation (2.5)) and integration over r gives the secular equation:

$$\sum_{G'} \left[\frac{\hbar^2}{2m} |\mathbf{k}+\mathbf{G}|^2 \delta_{GG'} + Vion(\mathbf{G}-\mathbf{G}') + V_H(\mathbf{G}-\mathbf{G}') + V_{XC}(\mathbf{G}-\mathbf{G}') \right] c_{i,k+G'} = \epsilon_i c_{i,k+G} \quad (2.15)$$

In this form, the reciprocal space representation of the kinetic energy is diagonal and the various potentials are described in terms of their Fourier transforms ($\{V(G-G')\}$). The usual method to solve the PW expansion of the KS-equations is by diagonalization of the Hamilton matrix whose matrix elements $H_{k+G,k+G'}$ are given by the terms in the brackets in equation 2.15. It follows that the size of the Hamilton matrix is determined by $E_{cut} = (\hbar^2/2m)|\mathbf{k}-\mathbf{G}|^2$ and it is extremely large for systems that contain both valence and core electrons. This severe problem that causes an huge increase in the required computational space is overcome by the use of the pseudopotential approximation. A plane wave basis set imposes no positional preferences within the unit cell, as is the case when using a localised basis set, for example Gaussian functions (Pulay forces). This makes it particularly useful for MD, as the electronic structure can propagate throughout the unit cell along with the atoms. The main disadvantage is that a lot of computational effort is used to deal with the vacuum that fills the unit cell.

2.4.5 Pseudopotentials.

To describe electron-ion interactions, Phillips (1958), Heine and Cohen (1970) developed the pseudopotential approximation. This allows one to replace the strong electron-ion potential with a much weaker "pseudo"-potential that describes all the salient features of a valence electron moving through the solid. This replacement allows the electronic wave functions to be expanded using far fewer plane waves (smaller E_{cut}) than would be needed to expand the wave functions in a full ionic potential. The electronic states of an atom can be separated into the *core states*, which are highly localized and not affected by the chemical environment and the *valence states*, which are extended and responsible for chemical binding.

Since core electrons do not participate in chemical bonding, a good description of the valence wave function inside the ionic cores is, in most cases, unnecessary and therefore there is no lack of crucial information if the inner solution (inside the core radius) is replaced with a smooth, nodeless pseudo-wave function, which also behaves numerically better. The strong nuclear potential is thus replaced by a weaker pseudopotential which acts on a set of pseudo-wave functions rather than the true valence wave function (this causes a screened interaction between the valence electrons and the ionic cores, i.e. nuclei plus core electrons). These pseudoelectrons experience exactly the same potential outside the core region as the original electrons but have a much weaker potential inside the core region.

Figure 2.8 shows these properties for a simple type of pseudopotential. The valence wave functions oscillate rapidly in the region occupied by the core electrons because of the strong ionic potential. These oscillations maintain the orthogonality between the core and valence electrons, which is required by the exclusion principle. The pseudopotential is constructed in such a way that there are no radial nodes in the pseudo wave function in the core region and that the pseudo wave functions and pseudopotentials are identical to the all electron wave function

and potential outside a radius cutoff. The pseudopotential is also constructed such that the scattering properties of the pseudo wave functions are identical to the scattering properties of the ion and core electrons. In general, this is different for each angular momentum component of the valence wave function, therefore the pseudopotential is dependent on the angular momentum (so-called non-local pseudopotentials).

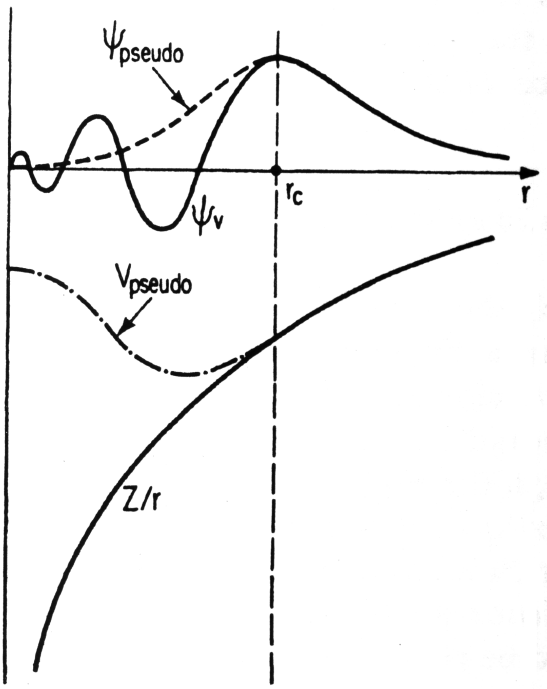

Figure 2.8: An illustration of the full all-electron (AE) wave function and electronic potential (solid lines) plotted against distance, r, from the atomic nucleus. The corresponding pseudo wave function and potential is plotted (dashed lines). Outside a given radius, r_c, all the electron and pseudo-electron values are identical.

The most general form for a pseudopotential is:

$$V_{NL} = \sum_{lm} |lm\rangle V_l \langle lm|, \tag{2.16}$$

where $|lm\rangle$ are the spherical harmonics and V_l is the pseudopotential for the angular momentum l. Acting on the electronic wave function with this operator decomposes the wave function into spherical harmonics, each of which is then multiplied by the relevant pseudopotential V_l.

2.4.6 CPMD Langrangian.

Pseudopotential calculations with a plane wave basis are not very well suited to conventional matrix diagonalization techniques. Due to the drastic increase in the cost of matrix diagonalization as the number of plane wave states increases, conventional matrix diagonalization techniques

are restricted to the order of 1000 plane-wave basis states, and this in turn restricts the number of atoms in the unit cell to the order of 10 atoms. Alternative methods like CPMD can be used directly to calculate the self-consistent Kohn-Sham eigenstates of a system (a so called wave function optimization). These alternative methods minimize the KS-energy functional and lead to the same self-consistent KS eigenstates and eigenvalues as conventional matrix diagonalization, but they are much better suited to performing total-energy pseudopotentials calculations because the computational time and memory requirements scale more slowly with the size of the system, allowing calculations on larger and more complex systems that can be studied using conventional matrix diagonalization techniques.

In the Car-Parrinello method the KS-energy functional $E[c_i]$ is a function of the set of coefficients of the plane wave basis set $[c_i]$. Each coefficient c_i can be regarded as the coordinate of a classical "particle". The essential step was to treat the expansion coefficients of the wave function as dynamical variables and integrate these fictitious wave function dynamics with the classical molecular dynamics by use of a single extended CP Lagrangian that can be written as follows:

$$\mathcal{L}_{\mathcal{CP}} = \sum_i \mu \langle \dot{\psi}_i | \dot{\psi}_i \rangle + \frac{1}{2} \sum_I M_I \dot{R}_I^2 - E[\{\psi_i\}, \mathbf{R_I}, \{\alpha_n\}], \tag{2.17}$$

where μ is a fictitious mass associated with the expansion coefficients of the KS electronic wave function, E is the KS-energy functional, R_I is the position of the ion I, and α_n defines the size and shape of the unit cell. This is analogous to the conventional form of the Lagrangian where the kinetic energy term has been replaced with the fictitious dynamics of the electronic degrees of freedom (electronic wave function ψ_i) and the KS-energy functional has taken the place of the potential energy.

In the dynamics of the orbitals there are now two components: one is their own dynamics, controlled by the fictitious mass μ, and the other arises as a consequence of a dragging force due to the motion of the nuclei, through E_{KS}. The latter fixes the average trajectory of the orbitals, whilst the former superimposes independant oscillations. The mass μ controls both, the energy transfer between orbitals and nuclei and the choice of the integration time step. A compromise is required in order to keep the energy transfer to a reasonably low value whilst preserving a sufficiently large integration time step.

In order to represent an electron density arising from a Slater determinant, the KS electronic orbitals are subject to the constraints of orthonormality:

$$\int \psi_i^*(\mathbf{r}) \psi_j(\mathbf{r}) \, \mathrm{d}^3\mathbf{r} = \delta_{ij} \tag{2.18}$$

These constraints are incorporated in the molecular dynamics Lagrangian by using the method of Lagrange multipliers. The CP- Lagrangian then becomes:

$$\mathcal{L}_{\mathcal{CP}} = \sum_i \mu \langle \dot{\psi}_i | \dot{\psi}_i \rangle - E[\{\psi_i\}, \{\mathbf{R_I}\}, \{\alpha_n\}] + \sum_{ij} \Lambda_{ij} \left[\{\psi_i^*(\mathbf{r}) \psi_j(\mathbf{r}) \mathrm{d}^3 r\} - \delta_{ij} \right] \tag{2.19}$$

The Lagrange multipliers Λ_{jj} ensure that wave functions remain normalized, while the Lagrange multipliers Λ_{ij} $(i \neq j)$ ensure that the wave functions remain orthogonal. The Lagrange

multipliers can be thought of as providing additional forces acting on the wave functions, which ensure that the wave functions remain orthogonal.

We then have a Lagrangian that depends on the nuclear coordinates $[R]$ and on the orbitals $\{\psi(r)\}$. The equations of motion for the electronic states are derived from the Lagrangian equations of motion:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathcal{L}_{CP}}{\partial \dot{R}_I}\right) = -\frac{\partial \mathcal{L}_{CP}}{\partial \mathbf{R}_I} \tag{2.20}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\delta \mathcal{L}_{CP}}{\delta \dot{\psi}_i^*(\mathbf{r})}\right) = -\frac{\delta \mathcal{L}_{CP}}{\delta \psi_i^*(\mathbf{r})} \tag{2.21}$$

The second equation involves functional derivatives because the orbitals are not simple variables but continuous scalar fields. Functional derivation implies to derive with respect to each one of the expansion coefficients, which results in the Car-Parrinello equations of motion:

$$M_I \ddot{R}_I = -\frac{\partial E_{KS}[\{\psi_i(\mathbf{r})\}, R_I]}{\partial R_I} \tag{2.22}$$

$$\mu \ddot{\psi}_i(\mathbf{r}, t) = -\frac{1}{f_i}\frac{\partial E_{KS}[\{\psi_i(\mathbf{r})\}, R_I]}{\partial \psi_i^*(\mathbf{r})} + \sum_{j=1}^{N} \Lambda_{ij}\psi_j(\mathbf{r}, t) \tag{2.23}$$

$$\mu \ddot{\psi}_i = -\hat{h}_{KS}\psi_i(\mathbf{r}, t) + \sum_j \Lambda_{ij}\psi_j(\mathbf{r}, t) \tag{2.24}$$

where $\hat{h}_{KS}$ is the single-particle Kohn-Sham Hamiltonian, and Λ_{ij} are the Lagrange multipliers that ensure the orthonormality of the KS orbitals (wave functions) as they propagate along their molecular dynamic trajectories. The force $-(H\psi_i)$ is the gradient of the KS-energy functional at the point in the Hilbert space that corresponds to the wave function ψ_i. The trajectories for the ionic and electronic degrees of freedom are generated via these equations of motion.

The idea behind the method is that by putting the electrons in the ground state at a fixed set of ionic positions and then allowing the ions to move according to equation 2.22, the electronic orbitals should adiabatically follow the motion of the ions, performong small oscillations around the electronic ground state. The electronic orbitals will have a fictitious kinetic energy associated with their motion and the fictitious mass parameter μ_i. If μ_i is small enough than the motion of the orbitals will be very fast relative to the motion of the ions. Of course, the dynamics of the nuclei in CPMD has only physical meaning if the electronic structure is close to its instantaneous ground-state at each step of the simulation, i.e. if the oscillations around the Born-Oppenheimer surface (ground state) are small. This means that the dynamics of the electronic orbitals has to remain relatively cold. However since the two dynamic susbsystems of ions and electrons are coupled, in principle energy can flow from the relatively hot nuclei subsystem to the colder orbital subsystem, which would lead to deviations in the BO-surface. However, a good choice of μ_i can ensure that the frequency spectra of the electronic orbitals and the ions are well separated from one another, i.e. energetically isolated from one another. An additional way to control the irreversible heat flow from the ionic degrees of freedom to the electronic degrees of freedom is

to attach a so called Nosé-Hoover thermostat to the ionic system [4, 9]. This is an algorithm that supplies energy to the ionic degrees of freedom, to compensate for this energy loss. To control the build-up of energy in the electronic degrees of freedom a separate Nosé-Hoover thermostat is attached to the electronic system [16], set at a much lower temperature that the ionic thermostat.

The conserved quantity in a CPMD is the so called Hamilton energy [5] $E_{cons} = K_e + K_n + E_{KS}$, with $K_e = \mu \sum_i f_i |\dot{\psi}_i(\mathbf{r})|^2 \mathrm{d}r$, the fictitious electronic kinetic energy and $K_n = 1/2 \sum_I M_I \dot{R}_I^2$ the true nuclear kinetic energy. The above condition means $K_e \ll K_n + E_{ks}$.

3 Computational Details and Analysis Methods.

The primary intention of the subsequently described simulations was to devise a minimal model system for studying the role of proton transfers in the AChE catalytic reaction.

The model system consists soley of the three amino acid side chains in the catalytic center, i.e. Ser^{200}, His^{440} and Glu^{327}, and the substrate molecule acetylcholine (ACh). The coordinates of the amino acids were directly extracted from the x-ray crystal structure of *T. californica* (Resolution 4.20 Å; pdb-file reference number: 1C20) and the respective α–carbons saturated with hydrogen atom. The acetylcholine had been manually inserted into the active site, and its structure had been optimized using a molecular modelling calculation. In our simulation of the acylation reaction, the initial reaction "coordinate" chosen was the simultaneous bond formation of the new C-O bond between ACh and Ser^{200} and the simultaneous proton-transfer reaction between Ser^{200} and His^{440}. Figure 3.1 depicts a schematic representation of this start structure.

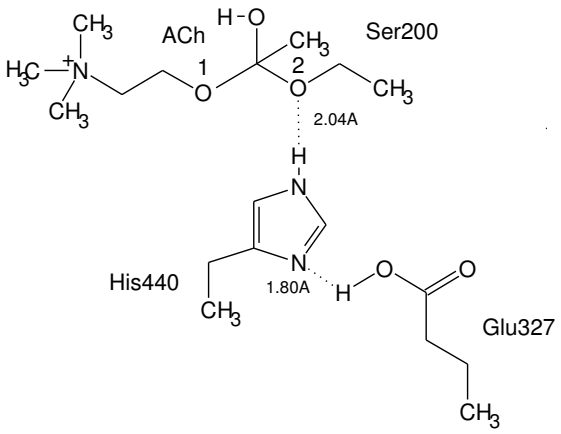

Figure 3.1: Start Geometry of the Acylation Reaction.

To simulate and investigate the effect of solvent in the active site, four water molecules were placed in the active site of AChE, using coordinates from the pdb-file 1C2O. The start geometry is shown in figure 3.2.

The start structure for the simulation of the deacylation reaction is shown in figure 3.3. The choline molecule produced during the acylation simulations was "flipped" around, so that it did not have any effect on the proceeding deacylation reaction. The nearby water molecule was brought nearer to the histidine residue.

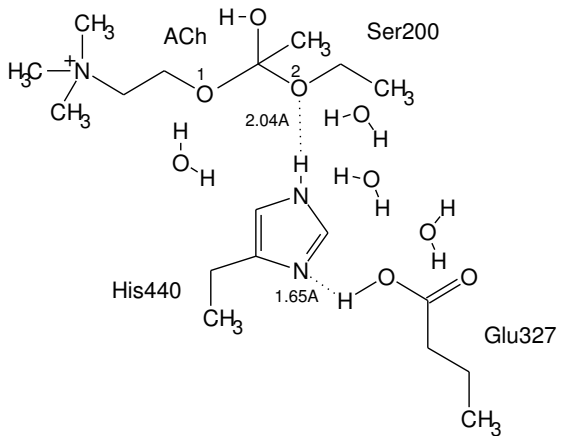

Figure 3.2: Start Geometry plus Solvent Molecules.

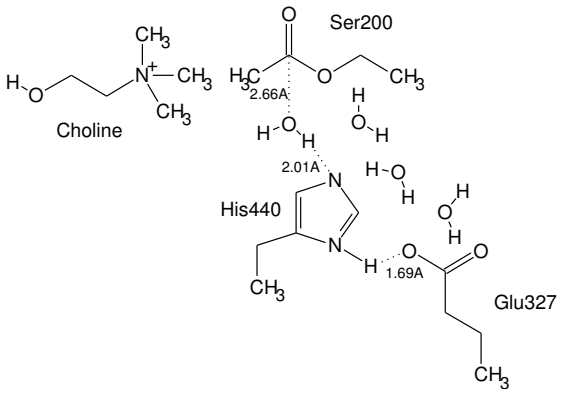

Figure 3.3: Start Geometry of the Deacylation Reaction.

3.1 Technical Details.

These general parameters refer to all wave function optimizations and all molecular dynamics simulations unless otherwise stated in section 3.2.

The system was simulated in the canonical ensemble (nVT), which represents a closed system in which the volume, temperature and the number of particles are kept constant. The dimensions of the simulation box were 12.0 Å x 8.0 Å x 15.0 Å and it had an orthorhombic symmetry. The simulations were carried out under periodic boundary conditions to reduce the computational effort. To be able to run the simulation, some atoms had to be fixed in space with constraints to preserve the rigid backbone of the protein. There were 9 fixed atoms, numbered 1 to 9 in figure 3.4, and 4 fixed torsion angles, the coloured bonds in figure 3.4, unless otherwise specified. The torsion angles were constrained to conserve the spatial arrangement of the amino acids.

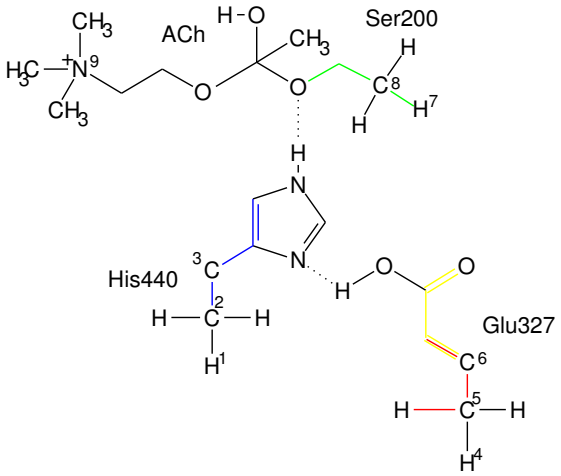

Figure 3.4: Position and Torsion Constraints in the Start Geometry.

All calculations were carried out using the CPMD code [1]. The DFT functional used to calculate the electronic structure in the wave function optimization and the simulations was of type PBE [10] with a density cutoff of 5.0×10^{-6} for the gradient correction. The pseudopotentials used to represent the oxygen, nitrogen, carbon and hydrogen core electrons were of type "Vanderbilt" [17, 18] and a plane wave basis set with an energy cutoff radius of 25 Rydberg was employed.

The wave function optimization was carried out with CPMD using the method of direct inversion with a convergence parameter of 1.0×10^{-7}a.u.

The time step of the MD was 4 a.u. with a maximum of 500 steps (circa 48 fs) for the MD equilibration and 25000 (circa 2.4 ps) steps for the molecular dynamics simulation. The individual simulation lengths are recorded in section 4. The fictitious electron mass, μ, was set at 600 a.u. As the MD equilibration originally starts at 0 K, the system has to be heated up to a certain temperature (300.0 K in our case). For this reason, a temperature scaling algorithm was employed to bring the ions up to 300 K and to scale the temperature up or down whenever it exceeded a value of 300 ± 50.0 K.

In the MD simulations, Nosé-Hoover thermostat chains [4, 8, 9] were coupled on the ions and the electrons (one for each degree of freedom) to control the temperature loss from the ionic system and ensure thermal equilibrium of all the degrees of freedom in the system and the stability of the CP-Hamiltonian. The Nosé-Hoover thermostat for the ions was set at a temperature of 300.0 K and the thermostat for the electrons [16] at 0.006 K.

3.2 The Simulations.

A total of six simulations were carried out, each one with a slightly different start geometry. After the coordinates of the model system had been extracted from the x-ray structure and edited as required, a wave function optimization was carried out in all cases to optimize the electron density of the original structure and calculate the electronic ground state of the structure. After the wave function optimization was completed, we took the optimized coordinates and started a short MD simulation, with temperature scaling, to equilibrate the structure. Once the structure had been equilibrated, another wave function optimization was carried out and the actual MD simulation was started.

All wave function optimizations, except the ones prior to the actual MD simulation, and short MD equilibrations were carried out with the carbonyl oxygen of tetrahedral intermediate I protonated to prevent any reactions from occurring before the system was properly equilibrated.

To equilibrate the water molecules in figure 3.2, a geometry and wave function opimization was carried out with CPMD holding all the atoms fixed with position constraints except the water molecules themselves. They then took on the configuration with the lowest potential energy by aligning themselves up and building hydrogen bonds with each other. The wave function was considered optimized when the energy converged under a value of 1.0×10^{-7}a.u. and the geometry considered optimized under a convergence limit of 1.0×10^{-3}a.u.

Simulation 1: The first simulation consisted of the start geometry shown in figure 3.1 with the hydrogen atom placed on the carboxyl carbon of the tetrahedral intermediate I. The charge of the system was +1.0.

Simulation 2: The hydrogen atom on the carboxyl oxygen of the tetrahedral intermediate in figure 3.1 was then removed in the “atoms section”of the input. The overall charge of the system was changed from +1.0 to 0.0.

Simulation 3: For this simulation, the hydrogen atom was placed back onto the carboxyl carbon of the tetrahedral intermediate and the geometry optimized water molecules were added, as shown in figure 3.2. The charge of the system was altered to +1.0 again.

Simulation 4: The next simulation was carried out using figure 3.2 without the hydrogen on the carboxyl carbon of the tetrahedral intermediate. The charge was altered once again to 0.0.

Simulation 5: To investigate the effects of the relative positions of the amino acids in the catalytic triad to one another, the start geometry in figure 3.2 was used, with the His^{440} residue moved 1.0 Å in x-direction and the Glu^{327} residue moved 0.5 Å in z direction (towards the His^{440} residue) and the proton on the carbonyl oxygen of the tetrahedral intermediate was removed. The charge of the system was 0.0. The simulation broke down after 68 frames (about 66 fs), so we removed two of the torsion constraints, set the time step to 3 a.u. and started the simulation once again. The trajectory files were then combined for analysis.

Simulation 6: To simulate the deacylation step of the AChE reaction, the start geometry shown in figure 3.3 was used. First we did a geometry optimization to optimize the hydrogen positions of the newly flipped around choline rest. All the other atoms were fixed with constraints during the geometry optimization.

As no reaction occurred, we protonated the Glu^{327} residue using a specific growth constraint. The constraint slowly moved the hydrogen atom 0.72 Å away from the imidazolium ring to a distance of 0.97 Å on to the oxygen in the glutamic acid residue and thus made the deacylation step more favourable. The protonation took a total of 800 MD steps (58 fs), each with a timestep of 3 a.u. During the actual MD simulation the hydrogen was then restrained on the Glu^{327} oxygen throughout the simulation, using a different "fixed" constraint. In addition, all the torsion constraints were removed, leaving the molecules free to rotate into the positions of lowest potential energy. The time step of the MD simulation was then set to 3 a.u. and the charge of the system was 0.0.

4 Results

4.1 The Trajectories.

Trajectory 1: The total simulation time was about 1.82 ps. No reaction occurred during the course of this simulation. The atoms that were not fixed by the constraints moved back and forth, as they moved along the trajectory towards the conformation with the lowest potential energy. The start geometry remained the same and the system appeared to be stabile on the simulation time scale.

Trajectory 2: The duration of the simulation was about 1.48 ps. After 24 fs the bond between the carbonyl carbon and oxygen 1 (see figure 3.1) in tetrahedral intermediate I broke. This left an acylated serine residue and an unprotonated choline rest. After the bond broke, the two fragments moved away from each other-the bond length between the carbonyl carbon and oxygen 1 was 1.46 Å to begin with and the end distance was 4.83 Å. The CH_3 group on the serine turns itself away from the ACh-rest and moves away from it.

After 232 fs there was proton transfer between the hydrogen atom on the Glu^{327} and the His^{440} residue and thereafter excessive proton transfer between the histidine and the glutamic acid residues throughout the simulation. The proton ended up on the Glu^{327} residue. The choline rest remained unprotonated throughout the simulation.

Trajectory 3: The length of the simulation was 0.968 ps. Proton transfer occurred between the hydroxyl group from the glutamic acid residue and the nitrogen from the histidine residue after 141 fs. After this transfer, excessive proton transfer took place between the Glu^{327} and the His^{440} residues. The proton ended up on the Glu^{327} residue. Otherwise no reaction happened during the simulation, all the unfixed molecules just equilibrated themselves. The water molecules moved around rather alot and aligned themselves up so that their hydrogen atoms could form hydrogen bonds to the other oxygen atoms.

Trajectory 4: The simulation was stopped after 1.25 ps. After 39 fs the His^{440} residue became protonated by the glutamic acid residue and stayed thus throughout the simulation. The reaction then continued after 167 fs, as the bond between the carbonyl carbon and the serine oxygen 2 (see figure 3.2) broke to give an unprotonated serine residue and ACh. The unprotonated Ser^{200} then removed a proton from the histidine residue almost immediately after 190 fs, restoring it to its original form. The ACh rest and the Ser^{200} residue then moved away from one another.

Trajectory 5: The trajectory was stopped after 0.832 ps. After 117 fs, the His^{440} residue was protonated by the glutamic acid residue and the bond between the carbonyl carbon and oxygen 1

of the tetrahedral intermediate broke almost simultaneously, after 135 fs. A short time later, after 173 fs, the deprotonated choline rest removed the proton from the His^{440} residue and remained protonated throughout the simulation. The two rests then moved further and further apart, as can be seen in figure 5.2.

Trajectory 6: The trajectory was stopped after 1.74 ps. No reaction occurred during the course of the simulation, even after the Glu^{327} residue was forcefully protonated with the growth constraint. The water moved closer to the His^{440} nitrogen, but no reaction or proton transfer took place. The water molecules moved around rather alot and formed hydrogen bonds with each other.

4.2 Simulation Energies.

The energies that are calculated during a molecular dynamics simulation are the most important indication of the reliability of the trajectory produced by the integration of the CP equations of motion. One looks at the course of the energies, especially the fictitious kinetic energy of the electrons and the Hamilton energy, during the simulation to establish the stability of the trajectory. The energies of trajectory 2 are shown in 4.1. The conserved energy in a CPMD simulation is the Hamilton energy (the sum of classical energy and the fictitious kinetic energy of the electrons), which should remain constant throughout the simulation. As one can see in figure 4.1 the Hamilton energy has a slight drift towards higher energies in its course which indicates that the system is not fully preserving the energy of the system. This drift is most likely an artefact originating from using a gradient corrected functional in combination with a low plane wave cutoff.

However, as this was an exploratory simulation, this discrepancy was considered to be of less importance than obtaining trajectories covering a sufficiently large part of phase space.

The Kohn-Sham energy is calculated using DFT and represents the potential energy of the classical system. The classic energy is the total energy of the system, i.e. the sum of the kinetic and potential energy of the ions. The temperature of the system is also recorded, which is calculated as the difference between the classical energy and the KS-energy. The temperature fluctuates around an average energy of 300 K, controlled by the Nosè-Hoover thermostats. The average displacement of the ions during the simulation is also documented.

In figure 4.2, the Hamilton energy, the KS-energy and the classical energy are pictured. The Hamilton energy and the classical energy follow the same path, the difference between them is the fictitious kinetic energy of the electrons, which has a very small value compared with the other energies. Thus the condition that $K_e \ll K_n + E_{ks}$ is met. As we can see the drift in the Hamilton energy and the classical energy does not come from the KS-energy, which stays constant, indicating that the kinetic energy of the system is responsible for the energy drift.

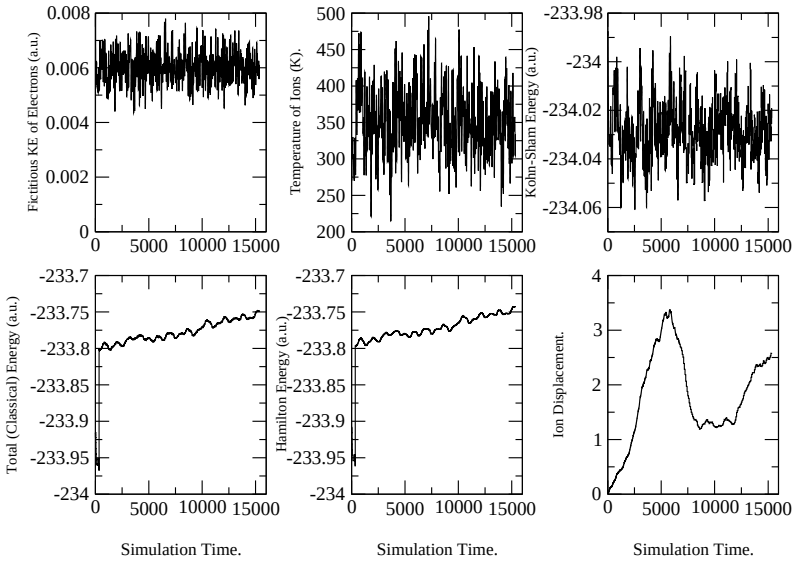

Figure 4.1: The Various Energies calculated in Trajectory 2.

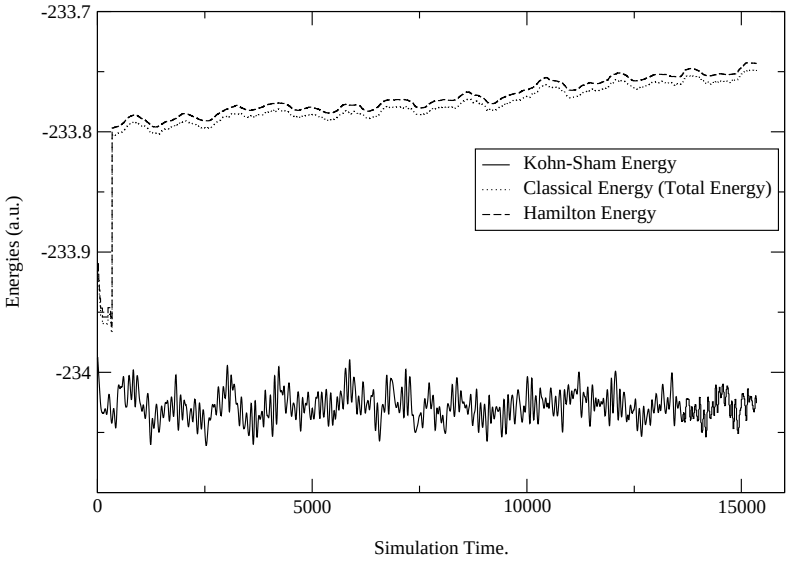

Figure 4.2: A Comparison of the KS-, Hamilton- and Classical Energies from Trajectory 2.

5 Conclusion.

5.1 The Acylation Reaction.

The trajectories 1 and 3 with the protonated carboxyl carbon of tetrahedral intermediate I show that the reaction does not occur when a proton is attached onto the carboxylate group. This result was independant of the presence of the solvent molecules. This useful finding meant that the proton worked as a protecting group, enabling us to carry out the largest part of the molecular dynamics equilibrations on any of the intermediate configurations without a reaction happening already in the preparation phase.

5.1.1 The Effect of Amino Acid Position in the Active Site.

In trajectory 4 the backwards reaction occurred and the bond between the carbonyl carbon and the serine oxygen 2 (see figure 3.2 in the tetrahedral intermediate broke. This simulation shows that the exact positioning of the amino acid residues in the active site may be important for the catalysis reaction, particularly as the correct reaction occurs in simulation 5 where the His^{440} and Glu^{327} residues are positioned somewhat differently. This suggests that the aformentioned breathing motions/fluctuations of the entire protein could alter the positions of the amino acids in the active site because they are contained in the rigid backbone of the protein, and thus steer the course of the reaction using these motions. The opening and closing of the gorge to allow substrate entrance could therefore control the course of the reaction and causing the His^{440} motion to act as a switch for the reaction.

Another contribution to this finding could be the poor resolution of the x-ray structure (4.20Å) used to construct the active site. This result also illustrates that the single steps of the catalytic reaction are reversible.

5.1.2 The Role of Water in Active Site.

In trajectory 2, without the added water molecules, the His^{440} was protonated after the tetrahedral intermediate broke down, whereas in trajectories 4 and 5, with the solvent molecules, the His^{440} became protonated before intermediate I disintegrated. This suggests two different mechanisms for the reaction in the active site of AChE, controlled by solvent presence. These mechanisms will be discussed in section 5.1.3.

As we can see by comparing the time points of the proton transfer between His^{440} and Glu^{327} in the trajectories 2, 3, 4 and 5, protonation occurs much faster with the solvent molecules present than without. This was regardless of which bond in the tetrahedral intermediate broke (4 and 5), or whether the protecting group was present (3). However, this observation must be further verified using many more simulations with multiple start configurations to obtain a

statistical view of the time point of this proton transfer. Therefore we propose that one role of the solvent water molecules is to stabilize the O-C-O-H bond from Glu^{327} by forming hydrogen bonds with its oxygens, which decreases the potential energy of the residue, and thus weakens the O-H bond. This means that His^{440} can remove the Glu^{327} proton much faster than without the solvent, where the Glu^{327} is not stabilized through these electrostatic interactions with water. Basically the water molecules act to lower the activation barrier of proton transfer between Glu^{327} and His^{440}.

The choline anion formed in trajectory 2 does not remove the proton from the histidine residue, even though the His^{440} was protonated by the Glu^{327} beforehand. However, this protonation does occur with the presence of the solvent molecules (see trajectories 4 and 5), thus one could propose that water is essential for this process too. It possibly stabilizes the choline oxygen and "holds" it in the right place for protonation by His^{440}. Alternatively, it could be a side effect of the positive charge of the quaternary nitrogen in a choline in a neighbouring image of the unit cell pulling the carbonyl oxygen away, and thus preventing the protonation step. To avoid such a side effect,a larger simulation cell including the oxyanion hole should be used.

In trajectory 2, the His^{440} proton ended up being transferred back to the glutamic acid residue, restoring the histidine residue to its original status, because the choline anion did not remove the other hydrogen on the other side of the imidazolium ring (as mentioned above). In simulation 5 however, the choline ion was protonated by His^{440}. In both simulations, the two product fragments moved away from each other, due to the strong electrostatic repulsion between the negatively charged choline oxygen and the carbonyl oxygen of the acylated serine. As can be seen in figure 5.1, the unprotonated choline anion in trajectory 2 moves much further away from the acylated serine as does the protonated choline rest in trajectory 5 (see figure 5.2). The choline proton in trajectory 5 obviously shields the electrostatic repulsion between the two oxygens to some extent.

5.1.3 Proton Transfer in the Active Site.

Without the solvent molecules the His^{440} residue was protonated after the tetrahedral intermediate broke down, with the solvent molecules His^{440} became protonated before intermediate I disintegrated. This suggests two different mechanisms for the reaction in the active site of AChE, controlled by solvent presence. As we assume (see [12]) that there are some, albeit a small amount, of gorge water molecules in the active site of AChE, it seems more likely that the second mechanism with the water molecules present will occur.

The trajectories 4 and 5 indicate that the histidine residue is first protonated by Glu^{327}, before the tetrahedral intermediate disintegrates. This bond formed between the nitrogen and the hydrogen on His^{440} lowers the activation energy of the other N-H bond on the opposite side of the imidazolium ring, which induces the bond in the tetrahedral intermediate to break and the choline ion to remove the aformentioned proton, restoring the His^{440} residue back to its normal form. After the first proton is transferred, the potential energy of the residue is lowered, which weakens the next N-H bond in His^{440}. This makes it easier for the bond in the tetrahedral intermediate to break and remove the proton from the weakened N-H bond. (Although water seems to have to be present for this proton transfer-see section 5.1.2). The positive electrostatic potential of the protonated His^{440} could also help to induce the break down of the tetrahedral

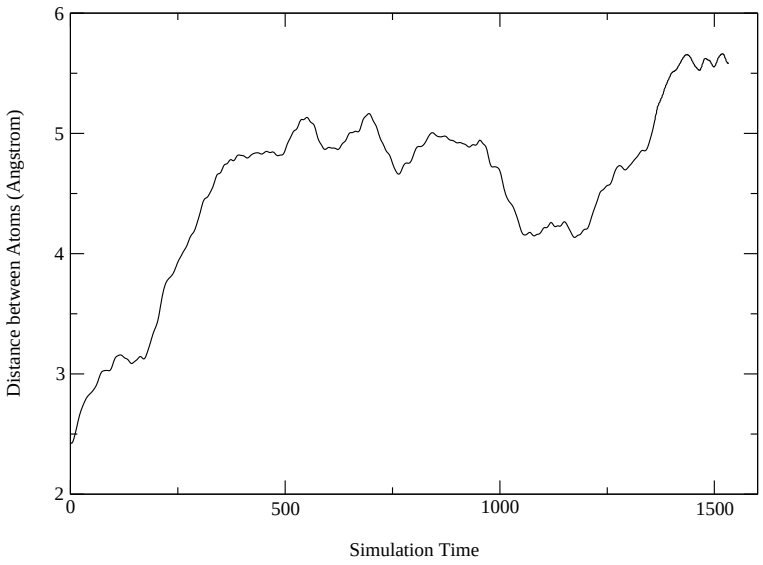

Figure 5.1: Trajectory 2–see text.

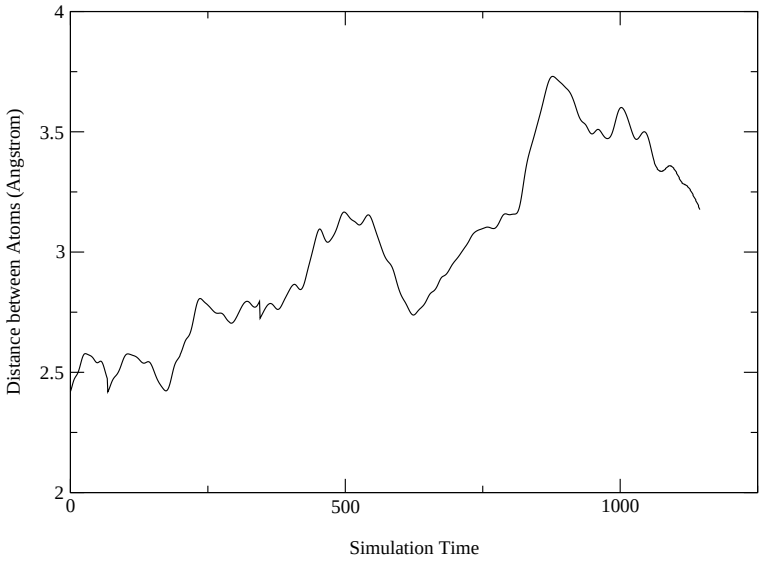

Figure 5.2: Trajectory 5–see text.

intermediate.

This mechanism would confirm the role of Glu^{327} as an acid-base catalyst and suggest that proton transfer between Glu^{327} and His^{440} destabilizes the tetrahedral intermediate I, causing its disintegration (see [21]).

The frequent proton transfer between Glu^{327} and His^{440} in simulations 2 and 3 (see figure 5.3) is a typical indication that the hydrogen bond present between these two residues is a "low-barrier hydrogen bond", which provides a special stabilization energy to lower the reaction barrier of the intermediate (see [21]).

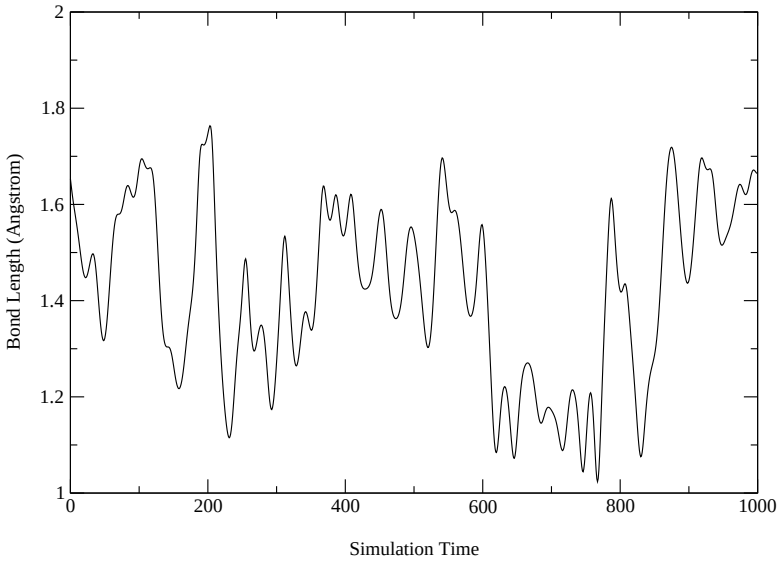

Figure 5.3: Bond Distance between the His^{440} Nitrogen and the Glu^{327} Hydrogen during Trajectory 3.

To conclude, we can say that the simulations carried out suggest that the induced electrostatic interaction of Glu^{327} with His^{440} and the proton exchange between them is essential for the acylation reaction of catalysis in AChE.

5.2 The Deacylation Reaction.

No reaction occurred during the length of trajectory 6, even though the His^{440} proton was forcefully removed to protonate the Glu^{327} residue, which should have started the deacylation reaction. The are many reasons for this failure. The study did not include the enzyme environment, as the entire protein is too large simulate with *ab initio* molecular dynamics, which means that other important electrostatic contributions from the surounding molecules are not present. To avoid this problem, the simulation box should at least be constructed to include the pivotal amino

acids of the oxyanion hole and the esteratic and anionic binding sites. This should also resolve the problem of the quaternary nitrogens' influence on the carbonyl oxygens, caused by the periodic boundary conditions, by compensating its positive charge. However, such an increase in the size of the system of course increases the computational effort and calculation time significantly. Additionally to include the protein backbone motions and thus take a potential steering effect of the His^{440} residue coupled to the low frequency of the motions of the gorge into account, a QM/MM (Quantum Mechanics/Molecular Mechanics) calculation is to be recommended as a follow–up investigation. This enables one to calculate the bulk of the protein outside the active site using classical molecular dynamics and the active site and its immediate surroundings using the more accurate, yet more expensive *ab initio* MD techniques.

However, the most probable reason for the failure lies within the periodic boundary conditions. As the choline rest was just "flipped" over and not completely removed, its positively charged quaternary nitrogen can influence the course of the reaction by attracting the partially negatively charged oxygen from the carbonyl group of the Glu^{327} from "the other side" of the simulation box, stopping the oxygen from getting close enough to deprotonate the His^{440} residue.

In general, the results and postulations discussed above should be investigated further using longer simulations (going into the nanosecond region) that include more of the important amino acid residues in the simulation box to confirm or refute the above findings.

Bibliography

[1] CPMD; VERSION 3.9.1; COPYRIGHT IBM CORP., 1990-2004, COPYRIGHT MPI FÜR FESTKÖRPERFORSCHUNG STUTTGART, 1997-2001; WWW.CPMD.ORG. 3.1

[2] A. D. BECKE. *Phys. Rev. A*, 38, 1988. 2.4.3

[3] JENNIFER M. BUI, KAIHSU TAI, AND J. ANDREW MCCAMMON. ACETYLCHOLINESTERASE: ENHANCED FLUCTUATIONS AND ALTERNATIVE ROUTES TO THE ACTIVE SITE IN THE COMPLEX WITH FASCICULIN-2. *J. Am. Chem. Soc.*, 126(23): 7198–7205, 2004. 2.2.3, 2.3.1

[4] W. G. HOOVER. *Phys. Rev. A*, 31, 1985. 2.4.6, 3.1

[5] JORGE KOHANOFF. *Theory and Simulation of Solids and Liquids*, CHAPTER ELECTRONIC STRUCTURE: PSEUDOPOTENTIALS, PLANE WAVES AND CAR-PARRINELLO. 2000. 2.4.6

[6] JEREMY KUA, YINGKAI ZHANG, AND ANDREW MCCAMMON. STUDYING ENZYME BINDING SPECIFICITY IN ACETYLCHOLINESTERASE USING A COMBINED MOLECULAR DYNAMICS AND MULTIPLE DOCKING APPROACH. *J. Am. Chem. Soc.*, 124:8260–8267, 2002. 2.2.3, 2.2.5, 2.3.1

[7] W.; PAAR R. C. LEE, C.; YANG. *Phys. Rev. B*, 37, 1988. 2.4.3

[8] M. E.; KLEIN M. L.; MARTYNA, G. J.; TUCKERMAN. *Phys. Rev. A*, 97, 1992. 3.1

[9] S. NOSÉ. *Mol. Phys.*, 52, 1984. 2.4.6, 3.1

[10] K.; ERNZERHOF M. PERDEW, J. P.; BURKE. *Phys. Rev. Lett.*, 77, 1996. 2.4.3, 3.1

[11] DANIEL M. QUINN. ACETYLCHOLINESTERASE: ENZYME STRUCTURE, REACTION DYNAMICS, AND VIRTUAL TRANSITION STATES. *Chemical Reviews*, 87:955–979, 1987. 2.1, 2.2.1, 2.2.2, 2.2.3, 2.2.3

[12] TONGYE SHEN, KAIHSU TAI, RICHARD H. HENCHMAN, AND J. ANDREW MCCAMMON. MOLECULAR DYNAMICS OF ACETYLCHOLINESTERASE. *Acc. Chem. Res.*, 35(6): 332–340, 2002. 2.1, 2.2.2, 2.2.3, 2.3.1, 5.1.3

[13] JOEL L. SUSSMAN, MICHAL HAREL, FELIX FROLOW, CHRISTIAN OEFNER, ADRIAN GOLDMAN, LILLY TOKER, AND ISRAEL SILMAN. ATOMIC STRUCTURE OF ACETYLCHOLINESTERASE FROM *Torpedo californica*: A PROTOTYPIC ACETYLCHOLINE-BINDING PROTEIN. *Science*, 253:872–879, AUGUST 1991. 2.2.1, 2.2.2, 2.2.4, 2.2.5, 2.2.6

[14] KAIHSU TAI, TONGYE SHEN, ULF BÖRJESSON, MARIOS PHILIPPOPOULOS, AND J. ANDREW MCCAMMON. ANALYSIS OF A 10-NS MOLECULAR DYNAMICS SIMULATION OF MOUSE ACETYLCHOLINESTERASE. *Biophysical Journal*, 81:715–724, 2001. 2.2.3, 2.3.1

[15] KAIHU TAI, TONGYE SHEN, RICHARD H. HENCHMAN, YVES BOURNE, PASCALE MARCHOT, AND J. ANDREW MCCAMMON. MECHANISM OF ACETYLCHOLINESTERASE INHIBITION BY FASCICULIN: A 5-NS MOLECULAR DYNAMICS SIMULATION. *J. Am. Chem. Soc.*, 124(21):6153–6161, 2002. 2.2.3, 2.3.1

[16] M. TUCKERMAN, M. E.; PARRINELLO. *J. Chem. Phys.*, 101, 1994. 2.4.6, 3.1

[17] C.; LAASONEN K.; CAR R. VANDERBILT, D.; LEE. IMPLEMENTATION OF ULTRASOFT PSEUDOPOTENTIALS IN AB INITIO MOLECULAR DYNAMICS. *Phy. Rev. B*, 43(8):6796 – 6799, 1991. 3.1

[18] D. VANDERBILT. SOFT SELF-CONSISTENT PSEUDOPOTENTIALS IN A GENERALIZED EIGENVALUE FORMALISM. *Phy. Rev. B*, 41(11):7892 – 7895, 1990. 3.1

[19] VOET AND VOET. *Biochemistry*. WILEY & SONS, 2ND EDITION, 1995. 2.1

[20] QINMI WANG, HUALIANG JIANG, JIANZHONG CHEN, KAIXIAN CHEN, AND RUYUN JI. ON THE POSSIBLE REACTION PATHWAY FOR THE ACYLATION OF ACHE-CATALYZED HYDROLYSIS OF ACH: SEMIEMPIRICAL QUANTUM CHEMICAL STUDY. *International Journal of Quantum Chemistry*, 70:515–525, 1998. 2.3

[21] YINGKAI ZHANG, JEREMY KUA, AND J. ANDREW MCCAMMON. ROLE OF THE CATALYTIC TRIAD AND OXYANION HOLE IN ACETYLCHOLINESTERASE CATALYSIS: AN *ab initio* QM/MM STUDY. *J. Am. Chem. Soc.*, 124(35):10572–10577, 2002. 2.2.3, 2.2.4, 2.3, 2.3.1, 5.1.3

[22] YINGKAI ZHANG, JEREMY KUA, AND J. ANDREW MCCAMMON. INFLUENCE OF STRUCTURAL FLUCTUATION ON ENZYME REACTION ENERGY BARRIERS IN COMBINED QUANTUM MECHANICAL/MOLECULAR MECHANICAL STUDIES. *J. Phys. Chem. B*, 107 (18):4459–4463, 2003. 2.2.3, 2.2.5, 2.3, 2.3.1