A Mechanistic Study Of The Enzymatic Excision Mechanism In AP Endonuclease (APE1)

im Fachbereich Physik der Freien Universität Berlin eingereichte

Dissertation

zur Erlangung des akademischen Grades eines Doktors der

Naturwissenschaften (Dr. rer. nat.)

vorgelegt von

Hossein Batebi, Semnan, Iran

Berlin, 2018

Erster Gutachter: Prof. Dr. Petra Imhof
Freie Universität Berlin

Zweiter Gutachter: Prof. Dr. Maria Andrea Mroginski
Technische Universität Berlin

Tag der Disputation: 13.02.2019

Contents

List of Tables

List of Figures

1 Introduction

1.1 DNA structure

Deoxyribonucleic acid (DNA) is a molecule that encodes genetic instructions of all known living organisms and many viruses. Otherwise stated, DNA stores all the information required to grow, develop, maintain and reproduce an organism. DNA and RNA as nucleic acids together with carbohydrates, lipids and proteins form four major types of macromolecules that are crucial for all known forms of organisms. Each DNA is composed of four types of smaller chemical entities called nucleotides: cytosine (C), adenine (A), thymine (T), and guanine (G). Each nucleotide consists of a nucleobase, a (desoxyribose) sugar and a phosphate group. Sugar and phosphate groups are connected by chemical bonds to make up the backbone of DNA. The structural framework of DNA is formed by the sugar phosphate backbone which defines directionality of the molecule. Each nucleotide forms a chemical bond called ester bond with another nucleotide in a chain: the free hydroxyl group on the 3' carbon atom of anucleotide forms an ester bond with the 5' carbon atom of the sugar on the next nucleotide. The phosphate is the 5' end of each nucleotide yet the sugar is the 3' end. During the synthesis of the molecule the sugar-phosphate backbone is defined as expanding in the 5' to 3' direction. To construct double-stranded DNA, two linear sugar-phosphate backbones twist together opposite of each other in a helical shape (see Figure 1). According to the base-pairing rules, guanine (G) binds with cytosine (C) and adenine (A) binds with thymine (T) through hydrogen bonds thereby connecting the two anti-parallel strands of the DNA the arrangement of these nucleotides is called the DNA sequence. The genetic information in long strings of such DNA sequences, the genes, are inherited from their parents during reproduction.

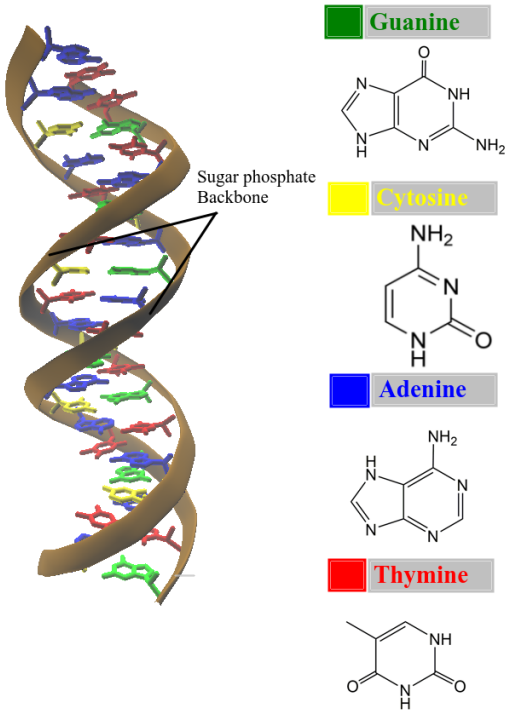

Figure 1: DNA structure

1.2 DNA damages

Integrity and stability of DNA are crucial for life because any alteration in this repository of genetic information would lead to mutations. The backbone of DNA is resistant to cleavage by a wide variety of reagents. DNA, however, is not immune to damage, both environmental factors and metabolic activities can lead to DNA damage. Perhaps the best example of a metabolic agent are free radicals that lead to oxidative damage. Beyond metabolic agents, radiation and tobacco smoke as well-known environmental damage such as ultraviolet (UV) radiation from the sun damages smoking cigarettes can cause mutations in cells. It can be surprising to know that an individual cell can suffer up to one million DNA changes per day [61]. Such damage to DNA can, e.g., alter the genes which are involved in cell growth, and consequently can lead to the development of different types of cancer. In addition, cell death as a consequence of DNA damage may lead to crucial effects for the organism of which the damaged cell is a part [89]. The single base lesion is the most common form of DNA damage. A DNA base can be modified by exogenous DNA damaging agents [46] or it can be oxidized [22, 21] and/or alkylated [171, 153] or hydrolysed [154] during physiologic metabolism. A wide variety of common disorders including atherosclerosis [165], neurodegenerative disease [128], or cancer [23, 20, 127] are related to the accumulation of base damage, mainly oxidized bases.

1.3 Base Excision Repair (BER)

A front-line repair system, called Base excision repair (BER) system, is responsible for maintaining the genome integrity [61]. Based on the total number of nucleotides that are replaced in the damaged strand during the repair process, BER has been classified to two main sub-pathways. In the sub-pathway termed long-patch BER (LP-BER) two or more nucleotides in the damaged strand are replaced whereas in the sub-pathway named single-nucleotide BER (SN-BER) only one nucleotide is replaced in the damage strand [21].The single-nucleotide BER (SN-BER) are initiated by DNA glycosylases [152]. Glycosylases cleave the glycosydic N-C1' bond between the damaged base and the backbone sugar. The product of these processes is DNA with abasic (apurinic/apyrimidinic, AP) sites. Abasic DNA is then processed by AP endonucleases which cleave the phosphodiester bond in the DNA backbone to cut out the abasic site [131, 44]. DNA polymeraseβ fills the gap with a new nucleotide and ligases finally glue broken DNA strands together (see Figure 2). One example where the Base Excision Repair (BER) is at work is the correction of mis-paired bases. DNA glycosylases such as the human thymine DNA glycosylase (TDG) or uracil DNA glycosylase (UDG) recognize G:T or G:U mismatches and remove specifically the mispaired T or U, respectively, leaving sites that are base-deficient, so-called AP sites [5, 107, 4, 109, 38, 106, 162, 142].

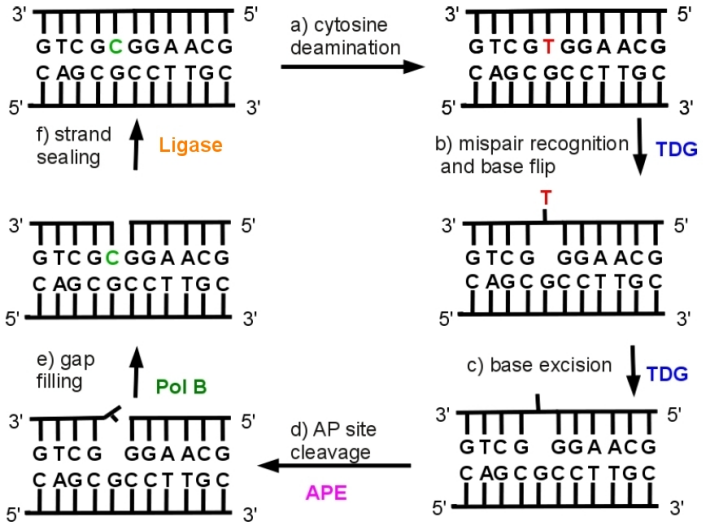

Figure 2: The Base Excision Repair system detects and repairs damaged and mispaired DNA bases: a) Due to erroneous deamination a (methyl-) cytosine is transformed to a thymine. b) The enzyme Thymine DNA glycosylase (TDG) recognizes the mispaired base and flips it out of the DNA double helix c) TDG excises the thymine and generates an abasic (AP) site. d) AP endonuclease (APE) cleaves at the 5'-end of the abasic site. The repair is performed by e) gap filling by polymerase β (PolB) and subsequent f) ligation by a DNA ligase enzyme.

1.4 Human AP endonuclease1 (APE1)

Abasic (apurinic-apyrimidic) AP sites are both mutagenic and cytotoxic [177] in such that, if remained unrepaired within the DNA, they cause an overall loss of genomic integrity, increased mutations and stalled replication [151]. The major abasic endonuclease of human cells, APE1, recognizes and removes abasic sites in damaged DNA [85, 169, 170]. Human AP endonuclease1 (APE1) is an Mg^{2+}-dependent enzyme, imperative for DNA repair via the BER pathway [16]. APE1 hydrolyzes the phosphodiester backbone of the DNA 5' to the AP site and generates 3'-OH 5'- deoxyribose phosphate (dRP) that has to be removed before gap filling reactions by a DNA polymerase and subsequent ligation of the product DNA. As a metalloenzyme [146], Magnesium, is essential for the enzymatic phosphodiester cleavage of APE1 at the abasic site and it influences the enzyme's binding interaction with DNA. In addition to AP endonuclease (APE1) in metazoan eukaryotes, the bacterial endonuclease IV can remove the AP sites with the same enzyme activities and different metal dependence and structural fold. Unlike Mg^{2+}-dependent AP endonucleases APE-1 and Exo III, Endo IV is a Zn^{2+}-dependent endonuclease which has either two Zn^{2+}-ions and one Mn^{2+}-ion or three Zn^{2+}-ions [41]. In *Schizosaccharomyces pombe*, the AP site can be removed with the major role of Endonuclease III (Nth1) instead of APE1 starring with cleavage of AP site by β-elimination [87].

1.4.1 APE1 phosphate hydrolysis

Cleavage of the phosphodiester backbone, uncatalyzed in water, as enzymatic reaction, or RNA-catalyzed attracted many scientists' attention during the last decade [66, 67, 65, 124, 32, 150, 78, 108, 115, 63, 123, 48, 64, 92, 120, 60] as phosphodiester hydrolysis is the fundamental reaction of endonuclease enzymes cleaving the backbone of DNA and RNA, usually at a recognition site such as a specific sequence (in e.g. restriction enzymes [114]). Moreover, phosphodiester cleavage can also occur in ribozymes, i.e. catalytic RNA, acting as ribonucleases [60, 19, 28].

The two extremes of a phosphate hydrolysis reaction are through an associative or dissociative mechanism. In an associative mechanism the first step is the nucleophilic attack that results in the formation of a pentacovalent molecular geometry. The second step is then the actual cleavage of the scissile P–O bond. In contrast, the first step of the alternative, dissociative mechanism is

the P–O bond scission, leaving a trigonal pyramidal metaphosphate intermediate that is processed to the product via nucleophilic attack. Several intermediate mechanisms, such as a concerted bond breaking/bond formation are, however, conceivable. Depending on the nature and position of active site residues surrounding the scissile phosphate group, nucleophile activation or leaving group departure, and thereby a more associative or more dissociative mechanism, may be favored. Available high-resolution X-ray structures of many endonuclease enzymes in complex with target DNA or RNA or substrate analogues [51, 47, 130, 168] and various biochemical studies have shown that metal cofactors are essential for the enzymatic phosphodiester cleavage [77, 159, 173, 82]. The metal ions not only play a direct role in the chemical step but also influence the binding interaction of the enzyme with the DNA or RNA [98, 135].

1.4.2 APE1 structural details

The structure of the Ape1–DNA complex has been solved with varying numbers and types of metal bound. The 2.95 Å and 2.65 Å resolution crystal structure of DNA-bound Ape1 product (PDB ID: 1DE8) and substrate complex without metal ions (PDB ID: 1DEW), respectively, were published in 2000 [31]. More recently published structures contain either one or two metal ions, giving rise to different suggested mechanisms [130]. For example, a catalytic mechanism proposing two divalent metal ions in the active site of Ape1 was published by Rupp and coworkers in 2001 [130]. Based on molecular dynamics simulations a shuttle mechanism in which the magnesium ion moves from one binding site to another was proposed [111]. To gain further structural and mechanistic information a plethora of experimental and theoretical studies were undertaken [110, 160, 15, 134, 37, 133, 111, 130]. Still, the exact position and number of magnesium ions remain ambiguous. In addition to the magnesium ion(s), the role of catalytically important amino residues such as Asp70, Glu96, Tyr171, Asp210, Asn212, Asp283, Asp308 and His309 have been debated in the literature [90, 40, 138, 91]. For instance, mutational studies of key amino acid residues account for the critical impact of them on the enzyme's DNA-binding capacity and/or the catalytic cleavage reaction [149, 93, 148, 91, 160]. Mutation studies on these amino acid residues showed Y171F, N212A, D212N reduction by 1200-fold, 7000 and 10000-fold, respectively, in the processing rate of DNA duplexes containing a substrate-analog [160]. Several mutational studies on Asn212 demonstrated the critical role of this amino acid residue in abasic DNA recognition by Ape1 and in the incision step [91, 138]. N212A does not have a notable impact on the ability of Ape1 to recognize abasic DNA but it reduces the rate of the incision step considerably [91]. In contrast, N212D reduces the ability of Ape1 in abasic DNA recognition, although this mutant incises the abasic DNA more efficiently than N212A [91]. Different hypotheses for the role of Asn212 in the incision mechanism have been proposed. A mechanism in which the Mg^{2+}-ion is bound to a site from which the nucleophilic water molecule is likely to attack [110, 111, 130] includes Asn212 as responsible for Mg^{2+}-ion coordination and formation of hydrogen bonds with the abasic DNA. Motivated by the mechanistic ambiguity that arose from the lack of Ape1 active-site structural details, several structures of Ape1 were published in 2015 [168]. These high-resolution crystal structures include Ape1 bound to the abasic phosphorothioate substrate without metal-ion (PDB ID: 5DFI), an inactive phosphorothioate substrate complex with Mn^{2+} (PDB ID: 5DG0) and a product complex (PDB ID: 5DFF) which diffracted to 1.63 Å, 1.80 Å, and 1.57 Å, respectively.

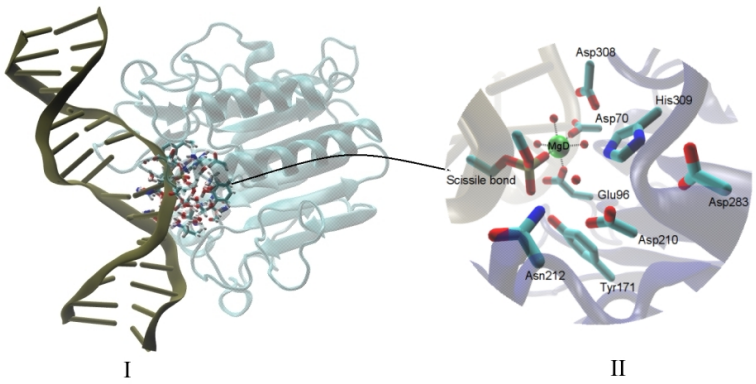

Figure 3: I) Snapshot of APE1-DNA substrate complex bound to a magnesium ion, II) Important active-site residues in the substrate complex of Ape1-DNA.

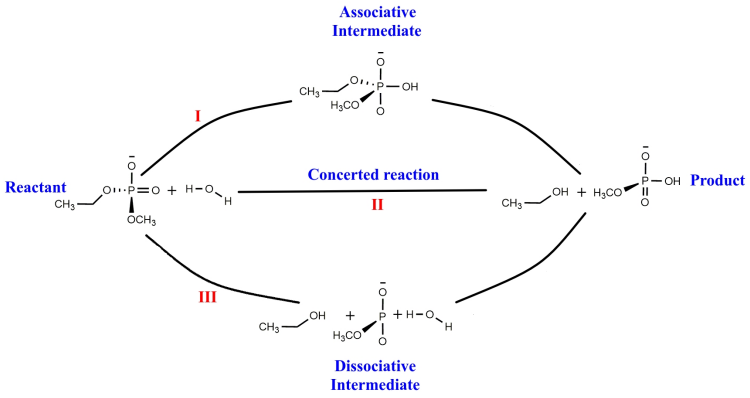

Figure 4: General scheme of a phosphate hydrolysis reaction

1.5 Approaches and motivations

Current high resolution crystal structures are giving the opportunity to investigate structural and functional details of an enzyme at atomic resolution. In order to investigate the catalytic power of an enzyme a reasonable way is comparing the reaction rate in the enzyme environment and the uncatalyzed solution as a reference. Investigating the reaction rate and type of mechanism can be done via free energy calculations along the reaction pathway. It is sensible to determine the different reaction mechanism pathways in the enzyme environment or in solution going from reaction state to product state via passing energy barriers (at transition states). The puzzling part is how to explain lower energy barrier in enzyme environment in comparison with the reaction in the solution. More precisely, the most critical question is: which residues, often acids and bases, are important in the enzyme environment to accelerate the reaction?

It is almost impossible to reach an exact conclusion about the precise role of forces which contribute to a catalytic effect, however, some mutational studies on APE1 pointed toward the importance of electrostatic forces. According to the definition, "electrostatic catalysis" includes the solvation by bound water molecules, permanent dipoles, polarizability (induced dipoles), protein charges [158] . It should be noted that it does not include orientational entropy or any dynamical effects, neither Van der Waals effects and charge-transfer covalent interactions. Yet, all these effects may work in concert in an enzymatic mechanism.

With recent growing in computational resources, simulations of biological and chemical processes at different levels of time-scales and accuracy may help to asses the different driving forces and their contribution in the catalytic effect of an enzyme.

In APE1 enzyme, although experiments have provided us with structural information, theoret-

ical approaches are required for an understanding of the mechanism at an atomic level of detail. In fact, electronic level insight is not present in previous studies and we still lack a detailed understanding of the phosphodiester backbone cleavage by the APE1 enzyme. To this end, the enzymatic mechanism by APE is addressed in this theses as outlined below:

First, we follow a "reductionist approach" wherein the most frequent components of endonuclease active sites are gathered and possible mechanisms are evaluated. The goal of implementing the "reductionist approach" is to elucidate the catalytic effect of the building blocks, metal ions and amino acid residues, in a minimalistic (cluster) model that can be regarded as a universal, but necessarily simplistic, endonuclease. The cluster models include either one or two metal ions and also vary in the number and position of amino acid residues, the number of water molecules involved in the metal ion coordination, and in the protonation state of a catalytically important histidine residue. Histidine residues may catalyze the reaction via activating the nucleophile through deprotonation or facilitating leaving group departure through protonation as has been discussed widely in the context of phosphodiester backbone cleavage [100, 74]. It is therefore plausible to consider the impact of different protonation states of histidine in the reaction pathways. Aspartate and glutamate are common amino acid residues in active sites of nucleases that are important in positioning the metal cofactors but may also act as general base/general acid. These rather small models can mimic an enzymatic active site in a protein only to a limited extend. For instance, the heterogeneous electrostatic environment of protein residues farther from the active site is neglected, and the protein's flexibility cannot be taken into account.

As the second step of this thesis the Ape1-DNA substrate complex was modeled based on the crystal structure of the phosphorothioate substrate complex with Mn^{2+} as metal cofactor (PDB ID: 5DG0)[168]. The sulfur atom in the thiophosphate group was replaced by an oxygen atom. The Mn^{2+}-ion was substituted with one or two magnesium ions, respectively, placed at different metal binding sites as proposed by recent crystal structures. The position and number of the metal ion(s) as well as important active-site residues in the substrate complex of Ape1-DNA are analyzed in detail so as to investigate the possible impact of metal ion binding on the stabilization of the active site in the Ape1-DNA complex in a reaction-competent conformation and hence enzymatic activity. APE1 active site structural details provide us with a reasonable statistics and information to postulate a multitude of reaction mechanisms [114]

The final step describes the investigation of the enzymatic reaction pathways of phosphodiester backbone cleavage by APE1 enzyme calculated using a combined quantum mechanical/molecular mechanical (QM/MM) approach, treating the active site quantum mechanically and the remaining enzyme classically. This combined approach afforded us an accurate description of the groups directly involved in the reaction while treating the whole enzyme in atomic detail [95]. Contrary to simple implicit descriptions of the protein (e. g. constant dielectric), the QM/MM approach permits the contribution of the enzyme to catalysis to be investigated and the contribution of individual residues to be quantified.

2 Methods

2.1 Quantum mechanical Methods

2.1.1 Many-body systems

Over the past several decades, quantum chemistry calculations have become an invaluable tool for the investigation of molecular properties on an electronic structure level This is to a large extend due to the development in approximation methods that explain correlation exchanges and interactions with sufficient accuracy while remaining computationally tractable. Quantum chemistry has offered the terminological and conceptual framework required to understand a wide variety of questions, including interpretation of chemical reactions, supra-molecular structure, spectroscopic properties, etc. The Schrödinger equation can be counted as the basic equation of physics to describe quantum mechanical behavior:

$$\hat{\mathrm{H}}\Psi = \mathrm{E}\Psi$$

Where $\hat{\mathrm{H}}$ is the Hamiltonian operator, E stands for the total energy of the system and the wave function of which is a function of the positions of electrons and nuclei in the system as described by Ψ. The time-dependent Schrödinger equation for a single particle of mass m located at position $\mathbf{r}$ at each time t is:

$$\{-\frac{\hbar^2}{2m}\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)+V\}\Psi(\mathbf{r},t)=i\hbar\frac{\partial\Psi(\mathbf{r},t)}{\partial t} \tag{1}$$

Where V is the external field (later on denoted as the electrostatic potential caused by the motion of the molecule's nuclei) and the term Ψ is the single particle wave function, $\hat{\mathrm{H}}$ for the time independent schrödinger equation described in curly brackets {} in eq. 1. Assuming the external potential as time-independent, using the separation of variables may result to $\Psi(r,t) = \psi(r)T(t)$ as such the time-independent Schrödinger equation can be written as following:

$$\left\{-\frac{\hbar^2}{2m}\nabla^2+V\right\}\Psi(\mathbf{r})=E\Psi(\mathbf{r}) \tag{2}$$

Therefore, the Schrödinger equation is reduced to a set of partial eigenvalue equations [3] and it can be solved exactly only for a limited number of problems, e.g. hydrogen atom. The lack of an exact solution to the Schrödinger equation for systems with more than one electron (i.e. including three or further interacting particles) supports approximate solutions that suggested more than one functional form of the wave function.

2.1.2 Born-Oppenheimer Approximation

As outlined above, the exact Schrödinger equation can not be solved for molecular systems. However, if the motion of the nuclei is decoupled from the motion of electrons, the exact Schrödinger equation for simplest molecular types like $H2^+$ can be solved using the Born-Oppenheimer approximation. Due to greater masses of the nuclei in comparison with that of the electron, it sounds reasonable to say that anelectron can adjust instantly to any changes in the positions of nuclei. Hence, the electronic wave function only relies on the positions of nuclei and not their momenta. According to the Born-Oppenheimer approximation, the total wave function for the molecule can be written as :

$$\Psi_{tot}(nuclei, electrons) = \Psi(nuclei)\Psi(electrons)$$

Subsequently, total Energy can be written as following where Ψ(electrons) and hence also E(electrons) parametric-ally depend on the position of nuclei:

$$E_{tot} = E(nuclei) + E(electrons).$$

Given that electronic energy includes electron-electron repulsion besides potential and kinetic energy of electrons stirring in the electrostatic field of the nuclei. It is important to note that when

the Born-Oppenheimer approximation is implemented, electronic motions are mainly studied and the nuclei are assessed to be fixed. Therefore, the Schrödinger equation is solved for electrons in the field of the nuclei, for each arrangement of the nuclei.

2.1.3 Spin

Spin of the electron in the single-electron wave function can be understood as a product of the spatial function ϕ that characterizes the electron density in the space and spin function χ which accounts for the electron spin to take the values 0 or 1. An acceptable functional form of the wave function for the polyelectronic systems is defined by a Slater determinant, it explains the anti-symmetry property of the wave function and can be written as:

$$\Psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(1) & \chi_2(1) & \dots & \chi_N(1) \\ \chi_1(2) & \chi_2(2) & \dots & \chi_N(2) \\ . & . & & . \\ . & . & & . \\ . & . & & . \\ \chi_1(N) & \chi_2(N) & \dots & \chi_N(N) \end{vmatrix}$$

In this determinant, $\chi_i(N)$ are the N electron spin orbitals and the pre-factor that results from the indistinguishably of the electron accounts for the normalization.

2.1.4 Calculating the Energy

The spin orbital for a molecule can be postulated as a linear combination of atomic orbitals (LCAO approach):

$$\psi_i = \Sigma_{\mu=1}^{K} c_{\mu i} \phi_\mu \tag{3}$$

Where ψ_i indicates the spatial part of the molecular orbital i and ϕ_μ is an atomic orbital. The energy can be obtained from the equation

$$E = \frac{\int \Psi H \Psi d\tau}{\int \Psi \Psi d\tau} \tag{4}$$

considering the Hamiltonian of the N-electron system is written in the general form and atomic units as:

$$H = \left(-\frac{1}{2} \sum_{i=1}^{N} \nabla_i^2 - \frac{1}{r_{1A}} - \frac{1}{r_{1B}} ... + \frac{1}{r_{12}} + \frac{1}{r_{13}} + ... \right) \tag{5}$$

Wherein the labels have the convention that r_{iA} denotes the distance between the electron i from the nucleus A.

Energy equations thus contain a large number of coupled terms, expressed as integrals (see below) which make it impossible to be solved exactly. An advantageous approach to avoid this problem is to express the Energy terms distributed into three types of interactions which exist in the total energy expression as follows: The kinetic and potential energy of N electrons in N molecular orbitals and in the field of nucleus A

$$E_{total}^{core} = \Sigma_{i=1}^{N} \int d\tau_1 \chi_i(1) \left(-\frac{1}{2} \nabla_i^2 - \Sigma_{A=1}^{M} \frac{Z_A}{r_{1A}} \right) \chi_i(1) = \Sigma_{i=1}^{N} H_{ii}^{core} \tag{6}$$

where H_{ii}^{core} is this energy related to the molecular orbital χ_i for M nuclei, the total electrostatic energy of the system due to the Coulomb interactions

$$E_{total}^{coulomb} = \Sigma_{i=1}^{N} \Sigma_{j=i+1}^{N} \int d\tau_1 d\tau_2 \chi_i(1) \chi_i(1) \frac{1}{r_{12}} \chi_j(2) \chi_j(2) = \Sigma_{i=1}^{N} \Sigma_{j=i+1}^{N} J_{ij} \tag{7}$$

in which J_{ij} is the energy due to the the Coulomb interaction between every electron pairs in different spin orbitals i and j, and the total exchange interaction energy between each spin in χ_i and all other N-1 electrons in the system

$$E_{total}^{exchange} = \Sigma_{i=1}^{N}\Sigma_{j'=i+1}^{N}\int\int d\tau_1 d\tau_2 \chi_i(1)\chi_j(2)\left(\frac{1}{r_{12}}\right)\chi_i(2)\chi_j(1) = \Sigma_{j=1}^{N}\Sigma_{j'=i+1}^{N}K_{ij} \quad (8)$$

that K_{ij} is the energy resulted from the exchange interaction between every electron pairs in different spin orbitals i and j.
The third contribution is originated from the fact that the motion of the pairs of electrons with parallel spins are correlated.

2.1.5 Calculating the molecular orbitals

As noted before, it is not possible to provide the correct solution of theSchrödinger equation for systems with more than one electron. However, approximate solutions can be obtained. The Variation Theorem can help to determine which specific solution to the wave function is better than other solutions. Due to the approximations that always ignore some level of complexities (and therefore some interactions) in the problem, the energy calculated from the approximated solution is always larger than the real solution. Hence, the proper wave function can be considered as the one which minimizes the energy of the system. This can be expressed as $\delta E = 0$, which together with the constraint that requires the molecular orbitals to be orthonormal, are solved using the method of Lagrange Multipliers.
The qualitative explanation of this approach is as follows:
As accepted for the orbital picture in the system, the single electrons are considered to be associated to specific spin orbitals. In the many-body problem, the motion of all the electrons have to be determined simultaneously which arises from the fact that the electronic motions are coupled. Referring to the three distinct energy contributions expressed in the previous section, it is convenient to introduce the associated operators as follows:

The core Hamiltonian operator:

$$\hat{H}^{core}(1) = -\frac{1}{2}\nabla_1^2 - \Sigma_{A=1}^{M}\frac{Z_A}{r_{1A}} \quad (9)$$

To describe the motion of an individual electron in the field of an isolated nucleus, The Coulomb operator can be defined

$$\hat{J}_j(1) = \int d\tau_2 \chi_j(2)\frac{1}{r_{12}}\chi_j(2) \quad (10)$$

subsequently, denoting the average potential caused by an electron in one spin orbital χ_j
The exchange operator will lead to

$$\hat{K}_j(1)\chi_i(1) = \left[\int d\tau_2 \chi_j(2)\frac{1}{r_{12}}\chi_i(2)\right]\chi_j(1) \quad (11)$$

which is represented as acting on the spin orbital χ_i.

In a compact notation, the so-called ***Hartree-Fock equations*** can be written as

$$f_i\chi_i = \Sigma_j \varepsilon_{ij}\chi_j \quad (12)$$

wherein $\hat{f}_i$ is called the *Fock operator*, the effective single-electron Hamiltonian in the many-body system, and can be defined as

$$\hat{f}_i(1) = \hat{H}^{core}(1) + \Sigma_{j=1}^{N/2}\left\{2\hat{J}_j(1) - \hat{K}_j(1)\right\} \quad (13)$$

in the closed-shell systems.]
Consequently, the standard eigenvalue form of the Hartree-Fock equations is

$$f_i\chi_i = \varepsilon_i\chi_i \tag{14}$$

The general method for solving these equations is the so-called *self-consistent field (SCF)* approach. Since a single-determinant approximation does not include Coulomb correlation, the exact wave functions can not be exposed by a single determinant. All in all, neglecting to provide an acceptable description of the electron correlations can be counted as a big disadvantage of the Hartree-Fock method. To tackle electron correlation issue several post-Hartree–Fock methods have been proposed.

2.1.6 Density Functional Theory (DFT)

DFT is an approximation method for contributing the electronic structure of atoms and molecules in the calculations of the system. As stated for the Hartree-Fock approach, a slater determinant is represented to calculate the full N-electron wave function in a molecule. Similarly, the one-electron wave functions are assumed in DFT method, yet only the total energy of the electrons as well as the total electron density are calculated. According to this theory, the electron density can be used to describe the properties of the system (including the ground-state energy) uniquely, which has been proved by Hohenberg and Kohn in 1964 [118]. It can be stated that energy is a unique functional of $\rho(\mathbf{r})$ and is written as the sum of two major terms as the following:

$$E[\rho(\mathbf{r})] = \int V_{ext}(r)\rho(\mathbf{r})dr + F[\rho(\mathbf{r})] \tag{15}$$

The first term represents the interaction between the electronic density and the external potential. The second term comprises the total Kinetic energy as well as the contributions from the interaction of all electrons.
The number of electrons in the system is fixed using the variational method and the constraint condition.

$$N = \int \rho(r)dr, \tag{16}$$

The ground state energy can be obtained by the method of Lagrangian multipliers as

$$\frac{\delta}{\delta\rho(r)}[E[\rho(r)] - \mu\int \rho(r)dr] \tag{17}$$

where $-\mu$ is the Lagrangian multiplier and can be rewritten as follows:

$$\left(\frac{\delta E[\rho(r)]}{\delta\rho(r)}\right)_{V_{ext}} = \mu. \tag{18}$$

The problem with this method is that the functional $F[\rho(r)]$ is unknown. A practical way to deal with this problem was proposed by Kohn and Sham (1965) [86], such that $F[\rho(r)]$ by an approximation can be written as a sum of the following terms:

$$F[\rho(r)] = E_{KE}(\rho(r)) + E_H[\rho(r)] + E_{XC}[\rho(r)], \tag{19}$$

The term $E_{KE}(\rho(r))$ was introduced to denote the kinetic energy of non-interacting electrons, $E_H[\rho(r)]$ defined as the Hartree electrostatic energy which includes the Coulomb energy between pairs of electrons. The term $E_{XC}[\rho(r)]$ comprises correlation and exchange contributions as well as the contribution to the real kinetic energy that is not calculated in the $E_{KE}(\rho(r))$. Lastly, the Kohn-Sham scheme for the full energy of the N-electron system can be written as:

$$\begin{aligned} E[\rho(r)] =& \sum_{i=1}^{N}\int \psi_i(r)(-\frac{\nabla^2}{2})\psi_i(r)dr + \frac{1}{2}\iint \frac{\rho(r_1)\rho(r_2)}{|r_1 - r_2|}dr_1dr_2 \\ &+ E_{XC}[\rho(r)] - \sum_{A=1}^{M}\int \frac{Z_A}{|r - R_A|}\rho(r)dr \end{aligned} \tag{20}$$

The single-electron Kohn-Sham equations are then obtained by defining the electronic density of the system as $\rho(r) = \sum_{i=1}^{N} |\psi_i(r)|^2$ and using the variational method as follows:

$$\left\{ -\frac{\nabla^2}{2} - (\sum_{A=1}^{M} \frac{Z_A}{r_{1A}}) + \int \frac{\rho(r_2)}{r_{12}} dr_2 + V_{XC}[r_1] \right\} \psi_i(r_1) = \varepsilon_i \psi_i(r_1) \tag{21}$$

Therein ε_i is the i-th orbital energy and $V_{XC}[r_1]$ is the exchange correlation functional that can be written in terms of the exchange-correlation energy as follows:

$$V_{XC}[r] = (\frac{\delta E_{XC}[\rho(r)]}{\delta \rho(r)}). \tag{22}$$

The Kohn-Sham equations are solved employing a *self-consistent approach* in a way that at each iteration, an improved value for the density is created and inserted into the equations again in expectation that calculations converge.

The most important challenge of the Density Functional Theory method is, how appropriate the approximations to the *exchange-correlation functional* are chosen. The success of DFT relies on the fact that even the simplest approximations can turn over solid results. Some approximations will be explained in the following sections.

2.1.7 Local Spin Density Approximation (LSDA)

This approximation, which is also termed as **LDA**, is a form of the *uniform electron gas model* and presumes that the electron density is constant in space. More precisely, the space is assumed to be inhomogeneous in the distribution of the electron density $\rho(r)$ while it is composed of volume elements that contain a constant electron density (of a point $\mathbf{r}$) at all points of each individual volume element. With this assumption, the equation 15 for the exchange-correlation energy can be rewritten as:

$$E_{XC}[\rho(r)] = \int \rho(r) \varepsilon_{XC}(\rho(r)) dr \tag{23}$$

Therefore, the exchange-correlation functional will be as follows:

$$V_{XC}[r] = \rho(r) \frac{d\varepsilon_{XC}(\rho(r))}{d\rho(r)} + \varepsilon_{XC}(\rho(r)) \tag{24}$$

Where $\varepsilon_{XC}(\rho(r))$ represents the exchange-correlation energy per electron in the uniform electron gas. These functionals are called *'local'*. The values of $\varepsilon_{XC}(\rho(r))$ for the uniform electron gas can be obtained with a good precision, for instance using the quantum Monte Carlo methods [42]. Sometimes it is more practical to express the exchange-correlation energy per electron as an analytic function of the $\rho(r)$ and also to realize it as two separate parts of exchange and correlation terms.

2.1.8 Gradient-corrected functionals

In spite of the very good performance of the LDA, it is still demanded to develop better approximations for different problems. Instead of the constant density, gradient-corrected, *'non-local'* functionals are an extension to the LDA method and consider the gradient of the density in a volume element in space. For these functionals, the exchange and correlation contributions are divided. A well-known gradient correction form to the exchange functional is introduced by Becke (1988, 1992) [7, 6] which is as follows:

$$E_X[\rho(r)] = E_X^{LSDA}[\rho(r)] - b \sum_{\sigma=\alpha,\beta} \int \rho_\sigma^{4/3} \frac{x_\sigma^2}{(1 + gbx_\sigma sinh^{-1} x_\sigma)} dr, \tag{25}$$

$x_\sigma = \frac{|\nabla\rho_\sigma|}{\rho\sigma^{4/3}}$ is a dimensionless parameter and $E_X^{LSDA}[\rho(r)]$ is the exchange energy in its Slater functional form.

A popular correlation functional form is also proposed by Lee, Yang and Parr [6] which can be expressed as follows:

$$E_C(\rho(r)) = -a\int \frac{1}{1+d\rho^{-1/3}}\left\{r + b\rho^{-2/3}[C_F\rho^{5/3} - 2t_W + (\frac{1}{9}t_W + \frac{1}{18}\nabla^2\rho)e^{-cr^{-1/3}}]\right\}dr \quad (26)$$

where $a = 0.049, b = 0.132, c = 0.2533$ and $d = 0.349$ are constants and

$$t_W = \sum_{i=1}^{N}\frac{|\nabla\rho_i(r)|^2}{\rho_i(r)} - \frac{1}{8}\nabla^2\rho \ ; \qquad C_F = \frac{3}{10}(3\pi^2)^{2/3}.$$

The Becke gradient-corrected exchange functional of the LSDA method and the Lee-Yang-Par correlation functional may be combined which are called the ***'BLYP' functionals*** and are widely used.

2.1.9 Hybrid Hartree-Fock/Density Functional Methods

As it was stated before, the Hartree-Fock method affords an exact solution to the exchange functional. However, the calculation of the correlation effects in this method causes too much complexity. Therefore, it is worth to consider a combination of the approximated correlation functionals with the Hartree-Fock exchange functional done by combining the exact exchange energy from the Kohn-Sham Slater determinant and the correlation part of the local density approximation suggested by Becke (1993) [9] in the following pattern:

$$E_{XC} = \int_0^1 v_{XC}^{\lambda} d\lambda \quad (27)$$

The λ is a coupling parameter ranging from 0 for the system without Coulomb repulsion v_{XC} between the electrons (Kohn-Sham non-interacting states), to 1 for the real system with full interactions. The integral of the exchange-correlation energy can not be solved analytically thus needs to be approximated.
Finally, the exchange-correlation energy can be written as:

$$E_{XC} = E_{XC}^{LSDA} + a_0(E_X^{exact} - E_X^{LSDA}) + a_X\Delta E_X^{GC} + a_C\Delta E_C^{GC} \quad (28)$$

With a_0, a_X and a_C being the empirical coefficients achieved from the experimental data.

One other popular combination of the exchange and correlation functionals is obtained by the Becke's gradient-corrected exchange functional used with Lee-Yang-Par correlation functional and the standard local correlation functional of Vosko, Wilk and Nusair (VWN) called **B3LYP density functional**. It reads as follow:

$$E_{XC}^{B3LYP} = (1-a_0)E_X^{LSDA} + a_0E_X^{HF} + a_X\Delta E_X^{B88} + a_CE_C^{LYP} + (1-a_C)E_C^{VWN}. \quad (29)$$

2.1.10 Basis functions

In electronic structure methods, such as DFT, MP2 and Hartree-Fock methods , the Kohn-Sham orbitals are often suitable to be indicated as a linear combination of the single-atom ***basis functions*** for dealing with the molecular systems, :

$$\psi_i(\mathbf{r}) = \Sigma_{\nu=1}^{K} c_{\nu i}\phi_\nu. \quad (30)$$

The basis functions ϕ_ν may take several functional forms such as *Gaussian* and *Slater type orbitals*. The Slater type orbitals are numerical basis functions which are obtained by solving the isolated-atom Kohn-Sham equations.

2.2 Molecular Mechanics methods

In Molecular mechanics methods as molecular modeling methods the energy of the system is calculated based on the motion of the nuclei whilst the motion of the electrons are neglected. This idea arose from the Born-Oppenheimer approximation which supports neglecting the motion of electrons based on the considerable difference between the mass of the electrons and the nucleus. In general, in systems where the electronic structure is not crucial, MM methods are a common choice to be used. Furthermore, computational costs for systems with a large number of atoms dictate the use of MM methods over quantum mechanical calculations to mimic molecular and atomic interactions by a simpler model.

2.2.1 Force Fields

Empirical potential functions (Force Fields) are modeled to predict special properties of the system with providing the chance of subsequently derived associated parameters. One of the main features of the force field is transferability which means; a set of parameters developed for a molecule can be used for many related molecules. A convenient way of modeling the force field is to represent it as a sum of four contributions which are potential energy terms that arise from the deviation of the bonds and angles from their equilibrium configurations as well as the interaction between non-bonded components. With this representation, some internal coordinates can be defined to reflect changes in their terms. The force field can be written in the following form:

$$U(r^N) = \sum_{bonds} \frac{k_b}{2}(l_i - l_{i,0})^2 + \sum_{angles} \frac{k_\theta}{2}(\theta_i - \theta_{i,0})^2 \tag{31}$$
$$+ \sum_{torsions} \frac{V_n}{2}(1 + \cos(n\omega - \gamma)) + \sum_{i=1}^{N}\sum_{j=i+1}^{N}(4\varepsilon_{ij}[(\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^{6}] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}})$$

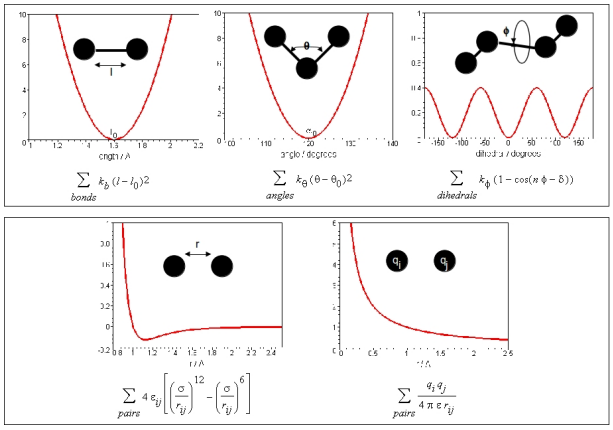

Figure 5: Bonded (top) and non-bonded (bottom) terms in an empirical force field.

Wherein l_i, θ_i,γ_i stands for bonds, angles and dihedral phase, $l_{i,0}$, $\theta_{i,0}$ indicates the minimum equilibrium bond length and bond angle, respectively. q_i q_j represent partial charges for atom i and j while r_{ij} is the distance between those atoms. ϵ_{ij} and $\sigma_{ij}$are the interaction energy and distance at minimum of the Lennard-Jones potential. The term $U(r^N)$ is the potential energy and the contributions to the function are potential energy terms due to the interaction between pairs of atoms in a bond, three atoms in an angle, four atoms in a dihedral and non-bonded terms which consist of electrostatic interaction modeled by a Coulomb potential and Van-der-Waals interactions modeled by a Lennard-Jones potential. Figure 5 shows the various terms in the described force field. The terms associated with the bonds and angles are modeled by harmonic potentials and the terms related to the torsion of bonds are sculptured by a periodic potential. The bonded atoms

mostly interchange a firm force which evokes the high energy deviation compared to the normal values, this strength is embodied in the *force constant's* magnitude.

2.3 Minimization methods

In molecular modeling, the minimum energy positions on the *Potential Energy Surface* (PES) represent the stable configurations of the system. Hence, minimization algorithms are demanded to specify the minimized geometry of each system.
Minimization algorithms fall into two categories: Algorithms for taking the *derivatives of the energy* as a function of the coordinates and *non-derivative* minimization algorithms.
Non-derivative minimization methods include several approaches such as the *simplex method* and the *sequential Univariate method.*

Derivatives of energy are obtained either analytically or numerically. The numerical derivative is calculated by dividing the change in the energy of the system with respect to a small change in the coordinate x_i, $\frac{\delta E}{\delta x_i}$, and returns the derivative at the middle distance point.
The potential energy function can be expanded in Taylor series as follows:

$$U(x) = U(x_k) + (x - x_k)U'(x_k) + (x - x_k)^2U"(x_k)/2 + ... \tag{32}$$

The derivative methods are distinguished by the highest order of the accepted derivative. In this sense, first-order methods apply the first derivative and second-order methods use up to second-order derivative.

First-order derivative minimization

There are some commonly used minimization algorithms of the first-order type derivative like *steepest descent* and *conjugate gradient* methods. In these methods, at each iteration the system is taken closer to the minimum point yet the change in the coordinates occur slowly.

The Steepest Descents methods

In this method, the direction of the motion is similar to the net force. The location of the minimum point can be determined by either making a line search or taking an arbitrary size step parallel to the net force.

Conjugate Gradients Minimization

In the steepest descents method, the gradient of motion and the following steps are orthogonal, while in the conjugate gradients approach, the directions at each point are conjugate and gradients are orthogonal. As a result, in this method, several directions are created. As a rule, for a quadratic function of M variables, M steps are needed to locate the minimum. The direction of motion is defined as

$$\mathbf{v}_k = -\mathbf{g}_k + \gamma_k \mathbf{v}_{k-1} \tag{33}$$

where γ_k is a scalar constant:

$$\gamma_k = \frac{\mathbf{g}_k \cdot \mathbf{g}_k}{\mathbf{g}_{k-1} \cdot \mathbf{g}_{k-1}} \tag{34}$$

The Eq.33 can not be used for the first step, albeit from the second step on. Therefore the first step in this method is in the direction of the gradient resembling the first step in the steepest descent.

2.4 Conjugate Peak Refinement (CPR)

In macromolecules or any other systems with a large number of degrees of freedom, finding local minima related to a stable molecular conformation is the first step in sketching the energy surface of the system. The next step is finding the saddle point between located minima to understand the adiabatic reaction path connecting them. An algorithm which only needs to compute gradient and evaluate the energy function called Conjugate Peak Refinement (CPR) [141] has been developed to determine multi-dimensional reaction coordinates between two defined conformers. This method can be useful for studying conformational changes in macromolecules, complex isomerization reactions, biochemical reactions, etc. On the potential energy landscape (PES), the CPR method attempts to construct a path between two minima. Chain-of-states especially near to the saddle points are connected together in the CPR algorithm to get closer to the exact first order saddle point. In the CPR algorithm we can assume that a quadratic function provides a well approximated energy landscape, therefore, if an initial search direction (S_i) and starting point (X_i) would be known, the saddle point can be located near (X_i). The method of Sinclair & Fletcher [72] is the theoretical basis of CPR algorithm for finding an initial search direction (S_i) and starting point (X_i). Sinclair & Fletcher emphasized on the fact that the Hessian has only one negative eigenvalue at a saddle point. In Sinclair & Fletcher [72] method, finding saddle point is assured when the function is quadratic.

2.5 Molecular Dynamics Simulation

In the energy minimization process, a few minimum configurations of the system are created. In case of a small number of atoms, the partition function can be calculated using the set of minimum configurations. Consequently, the thermodynamics properties can be premeditated. However, in more complex systems such as solutions and solids, the exceedingly large number of minima on the energy surface is a big hindrance to the determination of properties of the system directly from analyzing the energy surface.
Molecular Dynamics simulation, is a deterministic method during which small copies of the whole system are generated, representing the macroscopic system in different configurations. The real dynamics of the system over the time is calculated by predicting the future state of the system from its current state. Hence, it is feasible to ascertain the time-dependent behavior of the thermodynamic properties during Molecular Dynamics simulation which is done by integrating Newton's equation of motion 3538.

2.5.1 Ensemble average

Experimentally, values of the properties are estimated by taking a *time average* over the measurement time

$$A_{ave} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0} A(p^N(t), r^N(t)) dt, \tag{35}$$

while in MD simulations, an ***Ensemble Average*** can be calculated from the time average Ensemble average is the expectation value of the property as an average over all the representative replica of the system

$$\langle A \rangle_{ensemble} = \int \int dp^N dr^N A(p^N, r^N) \rho(p^N, r^N) \tag{36}$$

. According to the ergodic hypothesis, the ensemble average and the time average of the property are equal. $A(p^N, r^N)$ as an observable is a function of the momenta, p, and the positions, r, of the system and $\rho(p^N, r^N)$ is the probability density of the ensemble which has a different shape for different ensembles and partition functions.
The form of the probability density function for a ***canonical ensemble*** (NVT ensemble) is

$$\rho_{N,V,T}(p^N, r^N) = \frac{exp[-E(p^N, r^N)]/k_B T]}{Q} \tag{37}$$

therein Q denotes the partition function of the system, k_B is the Boltzmann constant and T is the temperature. The partition function of the NVT ensemble is:

$$Q_{NVT} = \frac{1}{N!}\frac{1}{h^{3N}}\int\int dp^N dr^N exp\left[-\frac{E(p^N, r^N)}{k_B T}\right] \tag{38}$$

2.5.2 Integration algorithms

The motion of a point mass (as a representative of the nucleus or the whole atom) is determined by Newton's equation of motion,

$$F = ma \Rightarrow \frac{d^2 r_i}{dt^2} = \frac{F_{r_i}}{m_i} \tag{39}$$

where F is the force applied on the nucleus i and r_i is the position of it. On the other hand, the force can be obtained by:

$$F(r_i) = -\nabla U(r_i) \tag{40}$$

Accordingly, by integrating the equation of motion, every next step of one atom is taken by the current position of all other atoms in the system. As a result, a trajectory of atoms positions is created from which all other properties such as velocities and momenta can be captured.

There are several algorithms for integrating the Newton's equation of motion. It is assumed that the position and the other properties of the particle can be expanded in Taylor series as follows:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \delta t\mathbf{v}(t) + \frac{1}{2}\delta t^2\mathbf{a}(t) + \frac{1}{6}\delta t^3\mathbf{b}(t) + \frac{1}{24}\delta t^4\mathbf{c}(t) + ... \tag{41}$$

$$\mathbf{v}(t+\delta t) = \mathbf{v}(t) + \delta t\mathbf{a}(t) + \frac{1}{2}\delta t^2\mathbf{b}(t) + \frac{1}{6}\delta t^3\mathbf{c}(t) + ... \tag{42}$$

$$\mathbf{a}(t+\delta t) = \mathbf{a}(t) + \delta t\mathbf{b}(t) + \frac{1}{2}\delta t^2\mathbf{c}(t) + ... \tag{43}$$

$$\mathbf{b}(t+\delta t) = \mathbf{b}(t) + \delta t\mathbf{c}(t) + ... \tag{44}$$

Where $\mathbf{v}$, $\mathbf{a}$, $\mathbf{b}$ denote the velocity, acceleration and the third derivative respectively.
A widely used algorithm is the so-called ***Verlet algorithm*** (1967) [163] which has the following form

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \delta t\mathbf{v}(t) + \frac{1}{2}\delta t^2\mathbf{a}(t) + ... \tag{45}$$

$$\mathbf{r}(t-\delta t) = \mathbf{r}(t) - \delta t\mathbf{v}(t) + \frac{1}{2}\delta t^2\mathbf{a}(t) - ... \tag{46}$$

after combing them gives the result:

$$\mathbf{r}(t+\delta t) = 2\mathbf{r}(t) - \mathbf{r}(t-\delta t) + \delta t^2\mathbf{a}(t) \tag{47}$$

It means the new positions are calculated using the current positions, accelerations and the positions from the previous step.
The velocities can be calculated accordingly via:

$$\mathbf{v}(t) = \frac{[\mathbf{r}(t+\delta t) - \mathbf{r}(t-\delta t)]}{2\delta t} \tag{48}$$

2.6 Hybrid quantum mechanics/molecular mechanics (QM/MM) simulations

Because of the enormous computational cost, complex biomolecular systems such as proteins, enzymes or DNA/RNA are too large to be fully treated with Quantum mechanics level of theory. Guided by the scarcity of computational sources, systems can be partitioned into two or more regions. In hybrid quantum mechanics/molecular mechanics (QM/MM) simulations, the region wherein chemical processes take place is treated with an appropriate quantum mechanical level of theory (QM) while the remainder which is commonly larger than the QM part is treated with a molecular mechanics force field (MM). In a hybrid QM/MM approach, considering the system as a whole with the QM-treated region embedded in the MM-treated environment can lead to the realistic result with reasonable computational costs [180, 179, 116, 161].

2.6.1 QM-MM Partitioning

The whole system is divided into three regions including an inner part, treated quantum mechanically, surrounded by an outer sphere, handled by a force field. Sometimes in systems with a large number of atoms like bio-molecular structures only atoms around 10˜15 Å of the QM part (usually the active site) are free to move while atoms in the remainder are kept fixed at their initial positions. A sketch of the system including protein and DNA solvated in water and ions is shown in Figure 6.

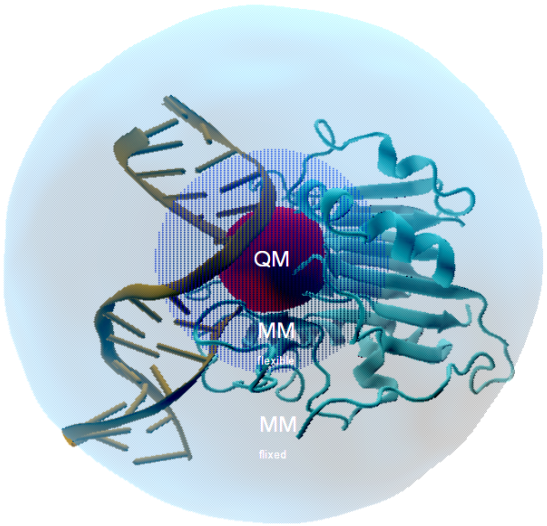

Figure 6: Partitioning of the entire system Into Outer part ($MM_{Flexible}$ + MM_{fixed}) and inner part (QM)

According to the type of QM/MM scheme, different approaches are implemented to treat the boundary region between QM and MM part. It is very crucial to treat the boundary between subsystems properly, specially, if it "cuts" covalent bonds. In addition, coupling terms need to be considered, hence the boundary region may include specific atoms that need to be considered in QM and MM calculation. For instance, link atoms (addition atom) may be used to cap the QM part. Based on the region to be considered as the active region two types of fixed and adaptive partitioning have been introduced. In systems wherein the active region is fixed from starting point till the end of simulation, fixed QM/MM portioning is a valid nominate to be used. Contrarily, in

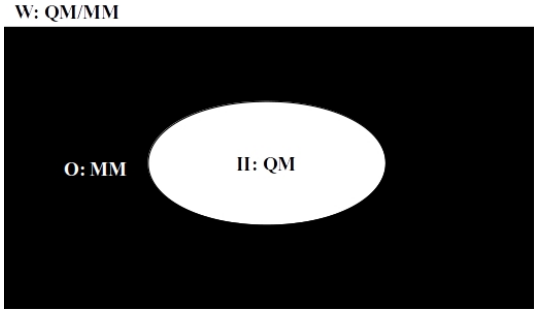

Figure 7: W: (QM/MM) entire system, O: (MM) outer part, II: QM Inner part.

processes with shifting active region(s), adaptive partitioning has to be implemented.

2.6.2 Additive and Subtractive QM/MM Schemes

In subtractive QM/MM schemes, in addition to QM calculation on the inner subsystem, MM calculation on inner subsystem and MM calculation on the entire system should be carried out. The QM/MM energy of the entire system can be calculated as follow:

$$E_{QM/MM}^{Subtractive}(Entire\_system) = E_{MM}(W) + E_{QM}(II) - E_{MM}(II) \tag{49}$$

In order to avoid double counting in equation 49, $E_{MM}$was subtracted from the summation of $E_{MM}(II)+E_{QM}(II)$. Depending on the partitioning approach, the equation 49 can be rearranged. For instance, if we use link atoms the equation would be as following:

$$E_{QM/MM}^{Subtractive}(W) = E_{MM}(W) + E_{QM}(II + \text{L}) - E_{MM}(II + \text{L}) \tag{50}$$

In equation 50, L stands for link atoms, therein calculations are carried out on the inner, capped, system. Abstractly speaking, in the subtractive QM/MM scheme, while the whole system is treated with an MM approach, a definite region which has been cut out from the entire system would be treated quantum mechanically. The absence of justifications on standard process of QM or MM calculation or necessity of explicit coupling terms make subtractive scheme quite simple which can be considered as the main advantage of the subtractive approach. On the other hand, thecalculation of electrostatic interactions is complicated, in particular with thinking that coupling between subsystems (QM and MM) is treated molecular mechanically. This issue is usually represented by interaction between fixed atomic charges in the QM and MM regions as the Coulomb interaction. In addition to the mentioned drawback, providing a complete set of MM parameters for the inner parts may be complicated.

Another widely used QM/MM approach is an additive scheme which can be written as follows:

$$E_{QM/MM}^{Additive}(W) = E_{MM}(O) + E_{QM}(II + \text{L}) + E_{QM/MM}(II, O) \tag{51}$$

Interaction terms between the two subsystems (QM and MM) appear as an explicit term presented by $E_{QM/MM}(II, O)$ in equation 51. Besides what has been assessed in the MM force field (31), electrostatic interactions between QM and MM atoms, Van der Waals and bonded interactions are considered in the additive scheme.

$$E_{QM/MM}(II, O) = E_{QM/MM}^{Elec} + E_{QM/MM}^{VdW} + E_{QM/MM}^{bonded} \tag{52}$$

In the following parts several terms which take part in $E_{QM/MM}(II, O)$ including non-bonded, bonded and electrostatic interactions will be discussed.

2.6.3 The Electrostatic QM–MM Interaction

Several approaches offer different ups and downs including mechanical embedding, electrostatic embedding and polarized embedding have been developed to define electrostatic coupling between the fixed charges used in the MM region and the QM charge density. However, polarized embedding may change the MM charge**.** In mechanical embedding (model A), the same treatment with MM-MM electrostatic interaction is implemented in QM-MM electrostatics. It means, the fixed charge of the MM force field is used in the QM region as well. The mechanical embedding approach has several restrictions and disadvantage, for instance, it is chemically reasonable to observe different charges during a reaction according to the modification in charge distribution of the QM region. In addition, the outer region is not polarized as it does not interact directly with the QM charge density. Sometimes, the fixed charges extracted from the MM force field can not reproduce reliable MM point charges for the inner region. Apart from these disadvantages, the mechanical embedding approach is very straight forward and computationally affordable.

In the case of electrostatic embedding, running a QM calculation beside an MM charge model can solve many problems which mechanical embedding has been faced with . For example, MM point charges can be written as one electron term in QM Hamiltonian as it has been shown as following:

$$\hat{H}^{Elec}_{QM/MM} = -\sum_{i}^{N}\sum_{J\epsilon O}^{L}\frac{q_J}{\mid r_i - R_J \mid} + \sum_{a\epsilon II+}^{M}\sum_{J\epsilon O}^{L}\frac{q_j Q_a}{\mid R_a - R_J \mid} \tag{53}$$

The indices, J, a, and i represent L point charges while the term M stands for QM nuclei and N for electrons. The symbols, Q_a stands for nuclear charges of QM atoms at R_a, q_J is the MM point charges placed at R_J and r_i introduces electron positions. In the electrostatic embedding approach, inner region would be automatically polarized as the electronic structure of the inner part follows the alteration in the charge distribution of the outer region. On account of QM-level treatment of electrostatic interaction the electrostatic embedding scheme is more accurate compared to the mechanical embedding scheme, however it is obvious that it isnot as computationally efficient and straightforward as mechanical embedding. In addition to all strengths which support the idea to use electrostatic embedding scheme, it has some issue. For instance, when the QM/MM boundary runs over a covalent bond, fixed MM charges which are near to the QM electron density can cause over-polarization. It is worth noting that in biomolecular systems, electrostatic embedding is the most prevalent embedding scheme which can generate reasonable results. As already mentioned, electrostatic embedding is widely known for the interaction of polarized QM density with rigid MM charge environment.

However, the lack of a model that describes the interaction of the QM density with flexible MM charges has led to the development of two polarized embedding models (model 1 and model 2). In model 1, the QM electric field polarizes the polarizable charge model in the MM region but this MM region does not send their polarization back to the QM region. Covering polarizable MM model into the QM Hamiltonian in a fully self-consistent formulation causes mutual polarization (model 2). These polarized embedding schemes (model 1 and model 2) are not still widely-applied in biomolecular applications. Moreover, polarizable MM modelsare not well developed to be used in biomolecular simulations

2.6.4 VdW and Bonded QM/MM interactions

Among discussed coupling terms in equation 52, in addition to electrostatic interactions, Van der Waals and bonded interactions participate in QM-MM coupling term. Fortunately, they can be assessed more feasible, in comparision to electrostatic embedding in both additive and subtractive QM/MM scheme. More precisely, with respect to QM-MM van der Waals coupling, additive and subtractive schemes are equal. A Lennard-Jones potential is a common choice to describe van der Waals interactions. Therefore we need proper parameters for QM atoms. Although, the character of QM atoms can be changed as a result of different situations, such as a reaction, in most cases parameters for QM atom can be extracted from well known MM force field.

2.6.5 Overview of Boundary Schemes

Depending on the system we need to define the boundary between outer and inner regions, for instance in a system including a small peptide solvated in water, the boundary does not cross over a covalent bond since a small peptide can be completely treated by QM level of theory while the remainder is treated molecular mechanically. In some cases, it is certainly necessary to cut the covalent bond for passing the QM/MM boundary, therefore, it is unavoidable to be careful about some possible issues, for instance, overpolarisation induced by MM charges close to the edge to the QM density should be avoided. Selecting a proper scheme can be based on some criteria like minimized artifacts. In general, the goal is defining a consistent explanation for QM/MM interactions at boundary region between the two subsystems. In order to ease the discussion we define some labels: The first atom in the shell of a subsystem which is treated by MM and QM are called M_1 and Q_1 the atoms bonded to M_1 and Q_1 are labeled as M_2 and Q_2 respectively.. During the QM-MM division the free valency at Q_1 is created. In a link atom scheme this free valency at Q_1 can be filled by an additional atom (L) to cap the QM part. This additional link atom can be problematic since it is dissimilar to the group which it exchanges. In addition, each link atom generates three artificial structural degrees of freedom. Fortunately, this problems can be fixed by some efforts. For example, artificial degrees of freedom can be removed by the application of constraints. By forcing the link atom as a function of position of M_1 and Q_1 the link atom L seats between M_1 and Q_1. Furthermore, a scaling factor can be defined to relate the distance Q_1-L to the distance Q_1. It should be noted that the polar bond should not be selected as bond to be cut and the selected bond should not participate in conjugative interactions. It is also recommended that the selected bond to be cut has to be three bonds away from the defined boundary since the dihedral terms expand up to two bonds toward inner region. Additionally, cutting over an MM charge group has to be neglected to avoid accumulation of artificial net charge beside QM density. All in all, a reasonable place to cut is C-C bond whereas an amide bond would be a complicated candidate with partial double bond character.

2.7 Choice of quantum mechanic method in QM/MM scheme

Both accuracy and computational costs can be counted as leading criteria in choosing a decent quantum mechanic method in QM/MM calculations. The first-principle quantum methods are not always the best choice to be chosen for solving all problems. For instance, DFT methods are restricted to a limited number of atoms ($<$100) due to elevated computational costs. Recently, approximate quantum chemistry methods have shown great efficiency and reasonable accuracy to handle a plethora of problems pertaining to chemistry, biology and physics. In addition to semi-empirical molecular orbital methods like AM1, MNDO, PM3, the self consistent charge density-functional tight-binding (SCC-DFTB) method derived from density functional theory (DFT) has been developed as alternative approximate approach to DFT and ab initio methods. Despite the lower accuracy of approximate methods in comparison with DFT and ab intio methods their computational speed is invincible as they can be 3 orders of magnitude faster on average than DFT and ab initio methods like Hartree-Fock (HF).

2.7.1 Density-functional tight-binding (DFTB)

The main core of all later extensions of DFTB is the non-self consistent version of DFTB. In common with other tight-binding method, a stationary approximation to DFT compiled with Harris functional approach can be inferred as the DFTB energy. Unfortunately, the molecular systems with intermediate charge transfer can not be treated by non-self consistent version of DFTB, yet it can be served for large and small intramolecular charge transfer.

2.7.2 Self-consistent-charge density-functional tight-binding (SCC-DFTB)

SCC-DFTB has been developed on the basis of a second-order expansion of the KohnSham total energy with reference, $\delta\rho$, to charge density variation relative to reference density ρ_0 . The equation for the total energy can be described as follows:

$$E = \sum_{i}^{occ} < \phi_i | \hat{H}^0 | \phi_i > + \frac{1}{2} \int \int \left(\frac{1}{| \overrightarrow{r} - \overrightarrow{r}' |} + \left. \frac{\delta^2 E_{ex}}{\delta \rho \delta \rho'} \right|_{\rho_0} \right) \times \delta\rho \left(\overrightarrow{r} \right) \delta\rho' \left(\overrightarrow{r'} \right) \overrightarrow{dr} \overrightarrow{dr'}$$

$$+ \left\{ -\frac{1}{2} \int \int \frac{\rho_0(\overrightarrow{r'}) \rho_0(\overrightarrow{r})}{| \overrightarrow{r} - \overrightarrow{r}' |} \overrightarrow{dr} \overrightarrow{dr'} + E_{xc} \left[\rho_0 \left(\overrightarrow{r} \right) \right] - \int V_{xc} \left[\rho_0 \right] \rho_0 \left(\overrightarrow{r} \right) \overrightarrow{dr} \right\} + E_{core} \tag{54}$$

E_{xc} and V_{xc} stand for exchange-correlation energy functional and potential, respectively. The Kohn-Sham orbitals are ϕ_i and the term reference density is represented by ρ_0. The effective Kohn-Sham Hamiltonian $\hat{H}^0$ can be understood as the only dependent of ρ_0.

$$\hat{H}^0 = -\frac{1}{2} \nabla^2 - \sum_k \frac{Z_k}{\left| \overrightarrow{R} - \overrightarrow{r} \right|} + \int \frac{\rho_0(r')}{\left| \overrightarrow{r} - \overrightarrow{r}' \right|} \overrightarrow{dr'} + V_{xc} \left[\rho_0(\overrightarrow{r}) \right] \tag{55}$$

wherein Z_k stands for the charge of the k_{th} nucleus. In equation 54, the second term stands for the participation of the density variations to the total energy. The third term represent the correction for the double counting terms arise from exchange-correlation and Coulombic participations contained in the H^ 0 matrix elements. The last term in equation 54 including E_{core} correspond to the core-core repulsion energy which plays a critical role in the parametrisation. In equation 55 Z_k is adjusted to treat the nuclear charge and the valance electron explicitly. A minimal basis of localized pseudo-atomic Slater orbitals χ_μ includes Hamiltonian matrix elements $<\phi_i|\hat{H}^0|\phi_i>$ can be shown as:

$$\phi_i = \sum_\mu C_\mu^i \chi_\mu \tag{56}$$

Accordingly, the Hamiltonian matrix, $H^0_{\mu\nu} \equiv <\chi_\mu|\hat{H}^0|\chi_\nu>$, can be understood as follows:

$$H^0_{\mu\nu} = < \chi_\mu | \hat{T} + V_{eff} \left[\rho_0^A + \rho_0^B \right] | \chi_\nu >, \mu \epsilon A, \nu \epsilon B \tag{57}$$

In equation 57, A and B define atoms where Overlap matrix element $S_{\mu\nu}$ and the Hamiltonian matrix $H^0_{\mu\nu}$ can be calculated from the distance between atom A and B. In addition, V_{eff} stands for effective Kohn-Sham potential and ρ_0^A corresponds to the electron density of the atom A. The approximation of the sum of a set of pairwise atom-atom potentials can be represented by the term E_{rep}. These pairwise atom-atoms potential can be fitted into a reference systems. For instance, H_3C-CH_3, $H_2C{=}CH_2$, HC-CH can be references for consideration of C-C interactions. It is assumed that these references provide compatibility to be transferred for interactions calculations in more complicated systems.

$$E_{rep} = \sum_{A,B} U^{A-B}(R_{AB}) \tag{58}$$

In equation 58, for the selected reference systems at a series of inter-atomic distances, the pairwise potential is fitted with polynomials functions. Precisely, it is implemented as the difference between the SCC-DFTB electronic energy and DFT total energy. The DFT method which is usually chosen as reference for fitting is the B3LYP method with an intermediate-size basis-set. A superposition of atomic contributions, $\delta\rho^A$corresponds to the charge density variation, $\delta\rho$, adjusted with the reference density can be written as:

$$\delta\rho = \sum_A \delta\rho^A \tag{59}$$

The superposition of atomic contributions, $\delta\rho^A$, are assessed by the charge changes at the atoms A, $\Delta q^A = q^A - q_0^A$. The Mulliken population is represented by the quantity q^A while q_0^A is the number of valence electrons of the atom A. A function γAB can be established to estimate the

second derivative of the total energy regarding the charge density variations, $(\delta^2 \mathrm{E}/\delta\rho\delta\rho')|_{\rho_0}$. According to the analysis of Coulomb interaction between two spherical charge distributions centered on the two nuclei, a functional form which represented γAB was suggested as follows:

$$\gamma AB = \iint \frac{s(\tau_{A,}\overrightarrow{R}_A)s(\tau_{B,}\overrightarrow{R}_B)}{|\overrightarrow{r} - \overrightarrow{r}'|}\overrightarrow{dr}\overrightarrow{dr'} =$$

$$\frac{1}{R_{AB}} - \left[e^{-\tau_A R_{AB}}K(\tau_A, \tau_B, R_{AB}) + e^{-\tau_B R_{AB}}K(\tau_B, \tau_A, R_{AB})\right] \tag{60}$$

Therein, $s(\tau_{A,}\overrightarrow{R}_A)$ stands for 1s Slater orbital with exponent τ_A centered on $\overrightarrow{R}_A$. The term $K(\tau_B, \tau_A, R_{AB})$ is defined as:

$$K(\tau_B, \tau_A, R_{AB}) = \frac{\tau_B^4 \tau_A}{2(\tau_A^2 - \tau_B^2)^2} - \frac{\tau_B^6 - 3\tau_B^4\tau_A^2}{2(\tau_A^2 - \tau_B^2)R_{AB}} \tag{61}$$

The function γAB reaches to γAA at the limit of $\mathrm{R}_{AB} \rightarrow 0$ and γAB attains the asymptotic form $(1/\mathrm{R}_{AB})$ at long distance. Another condition connects the width of the atomic charge density inversely to the chemical hardness (U_A) as:

$$\tau_A = \frac{16}{5}U_A \tag{62}$$

The total energy of SCC-DFTB compiled with the approximations noted above can be written as follows:

$$E^{SCC-DFTB} = \sum_i^{occ} < \phi_i|\hat{H}^0|\phi_i > + \frac{1}{2}\sum_{A,B}\gamma AB\Delta q^A \Delta q^B + E_{rep} \tag{63}$$

In molecular dynamics simulations and geometry optimization, obtaining analytic first derivatives of the SCC-DFTB total energy is imperative. They can be expressed as:

$$\frac{\partial E^{SCC-DFTB}}{\partial \overrightarrow{R}_A} = \sum_i n_i \sum_{\mu\nu} c_\mu^i c_\nu^i \times$$

$$\left[\frac{\partial H^0_{\mu\nu}}{\partial \overrightarrow{R}_A} - \left(\epsilon_i - \frac{1}{2}\sum_c \left(\gamma_{AC} + \gamma_{BC}\right)\Delta q^C\right)\frac{\partial S_{\mu\nu}}{\partial \overrightarrow{R}_A}\right] +$$

$$\Delta q^A \sum_C \frac{\partial \gamma_{AC}}{\partial \overrightarrow{R}_A}\Delta q^C + \frac{\partial E_{rep}}{\partial \overrightarrow{R}_A}\mu\epsilon A; \nu\epsilon B \tag{64}$$

The term $\mathrm{S}_{\mu\nu}$ stands for the tabulated overlap matrix element while ϵ_i and n_i correspond to eigenvalue and occupation number for the i_{th} molecular orbital respectively. The tabulated matrix elements in equation 64 and two-center approximation mentioned in equation 57 can be regarded as two reasons which make DFTB calculations faster than DFT.

2.7.3 QM/MM Implementation of DFTB

The QM/MM scheme of the Charmm program [24, 26] is the default implementation of QM/MM calculation for DFTB. It means, the QM part is treated with DFTB through wherein MM atoms interact with QM atoms over the van der Waals and electrostatic terms. Accordingly, the total energy can be written as:

$$E^{tot} = < \Psi \left| \hat{H}^{QM} + \hat{H}_{el}^{QM/MM} \right| \Psi + E_{van}^{QM/MM} + E^{MM} \tag{65}$$

Therein, the electrostatic interaction from the MM part $< \Psi \left| \hat{H}^{QM} + \hat{H}_{el}^{QM/MM} \right| \Psi >$ extracted from equation 63 is represented by the first term. The second term shows the van der Waals interactions between QM and MM atoms. The empirical energy of the MM atoms can be seen in

the last term. The electron operator is chosen to describe the Coulombic interaction between QM nuclei and electrons and the MM partial charges

$$\hat{H}_{el}^{QM/MM} = \sum_{A\epsilon MM}\sum_{B\epsilon QM}\frac{Q_A Z_A}{\left|\overrightarrow{R}_A - \overrightarrow{R}_B\right|} - \sum_{A\epsilon MM}\sum_{i=1}^{N_{el}}\frac{Q_A}{\left|\overrightarrow{R}_A - \overrightarrow{r}_i\right|} \tag{66}$$

Wherein, the term N_{el} is the number of QM electrons, Q_A stands for partial charge on MM atom A, the charge of QM atom B is represented by Z_B. In DFT and ab initio implementations of QM in QM/MM calculation the operator in equation 66 is treated analytically. The three-center integrals of the last terms have to be calculated which is computational expensive. The Coulomb interaction between the MM partial charge and Mulliken charge of the QM atoms given by SCC-DFTB can be used to reduce the computational cost of QM/MM electrostatic interaction

$$\hat{H}_{el}^{QM/MM} \approx \sum_{A\epsilon MM}\sum_{B\epsilon QM}\frac{Q_A \Delta q^B}{\left|\overrightarrow{R}_A - \overrightarrow{R}_B\right|} \tag{67}$$

Through transferring equation 67 into equation 63 eigenvalue equation can be obtained

$$\sum_{\nu}(H_{\mu\nu} - \epsilon_i S_{\mu\nu})C_\nu^i = 0 \tag{68}$$

Therein the matrix elements can be written as:

$$H_{\mu\nu} = H_{\mu\nu}^0 + \frac{1}{2}S_{\mu\nu}\sum_{B\epsilon QM}(\gamma_{CB} + \gamma_{DB})\Delta q^B +$$

$$\frac{1}{2}S_{\mu\nu}\sum_{A\epsilon MM}\left(\frac{Q_A \Delta q^C}{\left|\overrightarrow{R}_C - \overrightarrow{R}_A\right|} + \frac{Q_A \Delta q^D}{\left|\overrightarrow{R}_D - \overrightarrow{R}_A\right|}\right)\mu\epsilon C; \nu\epsilon D$$

The terms B and A indicate MM atoms while C and D are QM atoms.

2.8 QM/MM Molecular-Dynamics simulation

Investigating reaction free energies in solution or solvation free energies have been most common reasons to do full QM/MM including freely moving QM atoms in explicit-solvent. Particularly, along with Molecular Dynamics simulations for sampling, free energy treatment such as thermodynamic integration, umbrella sampling are needed to calculate e.g. reaction free energies in solution. Due to the intense cost of first-principles QM methods, like DFT, semi empirical methods have been a popular choice for biomolecular systems. In MD studies, through using Born-Oppernheimer MD, the QM forces and energies normally are given by SCF calculation in each step. There are other options such as Car-Parrinello MD (CP-MD) which include the electronsexplicitly as active degrees of freedom, by fictitious dynamical variables within a Lagrangian scheme.

2.9 Free energy calculations

The free energy, and changes therein, is an important quantity to characterize many processes such as chemical reactions. Calculation of free energy differences as one of the essential tasks in studying chemical or biochemical processes is quite challenging. Free energy measures the available phase space as it includes entropy. However, broad sampling is needed for systems including more than a few atoms. The canonical partition function ϱ of a system can be calculated by an integral over the whole phase space which is the summation of momentum and configuration space . If we assume that the potential energy E is independent of the momentum, an integral over momentum space can be neglected as it is constant to ϱ. Consequently, ϱ can be acquired as

$$\varrho = \int exp\left[-\beta E(r)\right] d^N r \tag{69}$$

wherein, N is the number of the degree of freedom of the system, β is $\frac{1}{k_B T}$, the term k_B is Boltzmann's constant and T is the absolute temperature. The partition function ϱ is related to the Helmholtz free energy with A = -1/β ln ϱ. In a canonical ensemble (N, V, T), number of particles, volume and temperature are constant. Swapping constant volume with constant pressure can lead to (N, P, T) ensemble. The Gibbs free energy (G), can be calculated out of the partition function of (N, P, T) ensemble with the same deviations and formalisms to Helmholtz free energy (A) from (N, V, T) ensemble [68, 76].

Along the reaction coordinate ξ, the probability distribution of the system can be calculated as follows:

$$\varrho\left(\xi\right) = \frac{\int \delta\left[\xi(r) - \xi\right] exp\left[(-\beta E) d^N r\right]}{\int exp\left[(-\beta E) d^N r\right]} \tag{70}$$

The Helmholtz free energy along the reaction coordinate (also called PMF) can be calculated via $A(\xi)$ = -1/βln $\varrho(\xi)$.

Assuming that the system is ergodic, the $\varrho(\xi)$ is

$$\rho(\xi) = \lim_{t \to \infty} \frac{1}{t} \int_0^t \rho\left[\xi(t')\right] dt' \tag{71}$$

In ergodic systems with infinite sampling, the ensemble average$\varrho(\xi)$ is equal to the time average $\rho(\xi)$. In equation 71 ρ is the number of incidence of ξ and t stands for the time. By checking $\rho(\xi)$, A(ξ) can be directly acquired out of the MD simulations. In simulations, samplings are confined with finite run time, as a consequence, direct sampling is tough for events with an energy barrier significantly larger than $k_B T$. In general, regions with lower energy around a minimum are sampled better than regions with higher energy. To acquire a realistic PMF A(ξ) , higher energy points (rare events) are required to be included, hence, several different approaches have been developed. To name only two

1. Methods such as metadynamics [2] where a penalty is added during free energy calculations to well-visited regions of the conformational space .

2. Methods in which the system is forced (or restrained) to be in a certain region of conformational space, usually along only one or two degrees of freedom (reaction coordinates)

In methods of type 2 the reaction path is divided into windows associated with the center value of the bias in ξ. All windows are sampled separately and a certain amount of the reaction coordinate ξ is covered by each window. A global free energy profile A(ξ) can be obtained by combining all individuality calculated windows.

2.10 Umbrella Sampling

Umbrella sampling, biased molecular dynamics, is one of the widely used methods to calculate the free energy along a reaction pathway in biomolecular systems. As noted above, to guarantee sampling in all relevant regions, the reaction coordinate ξ is split into several windows. Normally, a bias function, often a simple harmonic bias of strength K, is implemented in each window to restrain the system in this window i around the reference point ξ_i^{ref} .

$$\omega_i(\xi) = K/2(\xi - \xi_i^{ref})^2 \tag{72}$$

In equation 72, choosing a proper amount for K is crucial as the selected amount should be huge enough to allow the system to pass the barrier. However, it should not be too large to induce too narrow distributions. In addition, if the K is too huge or the time step is too big, the high energy points along the reaction pathway are over-counted in the free energy profile A(ξ) [140]. To analyze the Umbrella Sampling Simulations, and obtain free energy curves, an additional technique such as umbrella integration [69] or Weighted Histogram Analysis Method (WHAM) is required [143, 147]. Distributions between two windows next to each other in reaction coordinate ξ should overlap enough to adopt reasonably with WHAM.

2.11 Weighted Histogram Analysis Method (WHAM)

In WHAM, the global distribution $P^u(\xi)$ can be calculated as follows:

$$P^u(\xi) = \sum_{i}^{windows} p_i(\xi) P_i^u(\xi)$$

To minimize the statistical error of $P^u(\xi)$ as biased distribution in windows i, the weights p_i are selected such that

$$\frac{\partial \sigma^2(P^u)}{\partial p_i} = 0$$

With $\sum p_i = 1$ it follows that

$$p_i = \frac{a_i}{\sum_j a_j}, a_i(\xi) = N_i exp\left[\beta \omega_i(\xi) + \beta F_i\right]$$

The term N_i stands for the total number of steps sampled for window i and the term a is the representative of each window. Consequently, the F_i which is the free energy offset from window i can be calculated as

$$exp(-\beta F_i) = \int P^u(\xi) exp\left[-\beta \omega_i(\xi)\right] d\xi$$

It should be noted that WHAM equation is solved iteratively, since F_i which can be calculated from $P_i(\xi)$ is necessary to solve ${P_i}^u(\xi)$

3 Model setup

3.1 Cluster models

We have set up cluster models with one and two magnesium ions and variations of amino acid residues. For the one-metal models, a first type includes a protonated histidine side chain (1Mg-Hsp), and a second type contains a neutral histidine side chain (1Mg-Hsd) (see Figure 8). The two-metal models, termed 2Mg, have been setup with neutral histidine (see Figure 9). The single magnesium ion is coordinated by one glutamate and one aspartate residue, one oxygen atom of the phosphate group and a water molecule. Motivated by several high resolution crystal structures of the active sites of endonuclease enzymes [51, 47, 168, 133], the magnesium ion is positioned either on the "attack site", binding to O2P, and termed MgA, or on the "departure site", coordinated by O1P and labeled as MgD (see Figure 8). The cluster models are inspired by actual enzymatic active sites, such as those shown in Figure 13. They are, however, not based on any particular crystal structure (pdb file) but created artificially.In the two-metal models, the glutamate residue bridges the two magnesium ions, here also called MgA and MgD for "attack-site" and "departure site", respectively. Both ions are coordinated by the same oxygen atom of the phosphate group (O1P). Further ligands are one aspartate residue, and two water molecules for each metal ion in models 2Mg-Hsd-Glu-2Asp and 2Mg-Hsd-Glu-2Asp-1Asp, respectively. In model 2Mg-Hsd-Glu-3Asp the second water ligand on MgD has been replaced by an aspartate residue and in model 2Mg-Hsd-Glu-1Asp the bidentate aspartate ligand on MgD has been replaced by two water molecules (see Figure 9). The 1Mg-Hsd models and the 2Mg model have been setup in an additional variant, extended by one more aspartate residue. In the one-metal case, this aspartate residue has been placed in a position hydrogen-bonded to the histidine residue. In the two-metal case, the extra aspartate residue is either positioned in hydrogen-bond distance to the histidine residue (2Mg-Hsd-Glu-2Asp-1Asp), similar to the one-metal scenario, or it replaces a metal-coordinating water molecule (2Mg-Hsd-Glu-3ASP). Table 1 lists all models and mechanisms computed in this section. The nomenclature for the mechanisms is as follows: 1Mg or 2Mg for models with one or two magnesium ions, respectively. In the one-metal models 1MgA refers to a magnesium ion located at the attack site whereas 1MgD refers to a magnesium ion located at the departure site. The further labels -p, -d, -h refer to a mechanisms with protonated histidine, neutral histidine and direct nucleophilic attack, and neutral histidine-assisted nucleophilic attack, respectively. In the two-metal case, mechanisms for models with three, two, or one aspartate residues, corresponding to the labels -d3, -d2, and, -d1, respectively have been computed. The label -h again refers to a histidine-assisted mechanism (in a model with three aspartate residues). States along a mechanism are labeled reactant_XX, TS1_XX, Int2_XX, product_XX, etc, for the reactant, first transition state, second intermediate, and product state of mechanism XX, respectively.

Table 1: List of models and mechanism analyzed in this study. For the reactant states of the one-metal models (1Mg) see Figure 8 and for the two-metal models (2Mg) see Figure 9, respectively.

Model	Mechanism	Figure
1MgA-Hsp-Glu-1Asp-1Asp	1MgA-p	14
1MgD-Hsp-Glu-1Asp-1Asp	1MgD-p	15
1MgA-Hsd-Glu-1Asp-1Asp	1MgA-d	18
	1MgA-h	20
1MgD-Hsd-Glu-1Asp-1Asp	1MgD-d	19
	1MgD-h	21
2Mg-Hsd-Glu-3Asp	2Mg-d3	22
2Mg-Hsd-Glu-2Asp	2Mg-d2	23
2Mg-Hsd-Glu-1Asp	2Mg-d1	24
2Mg-Hsd-Glu-2Asp-1Asp	2Mg-h	25

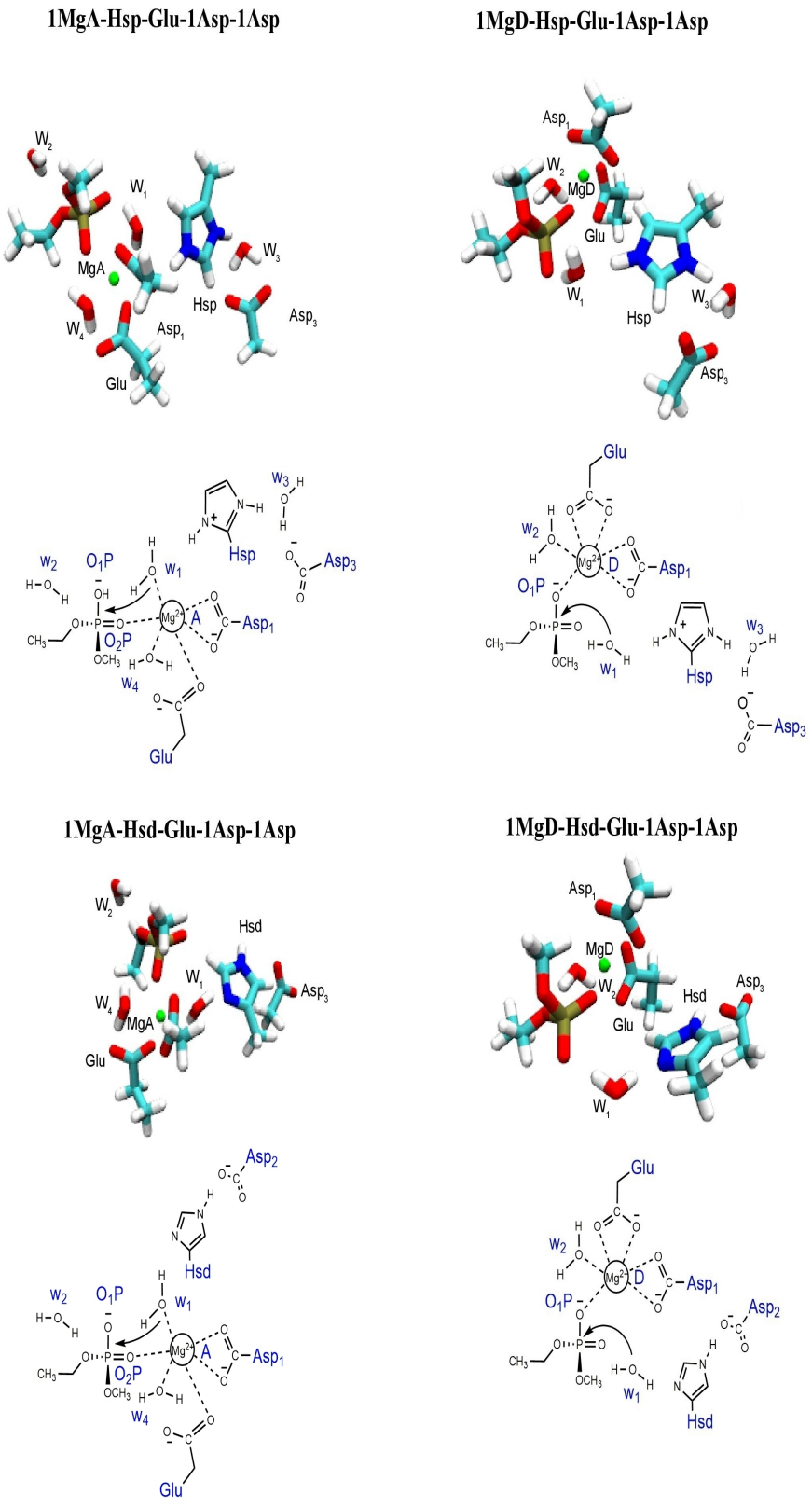

Figure 8: Reactant states for one-metal models. The scheme indicates the attack by the nucleophilic water molecule.

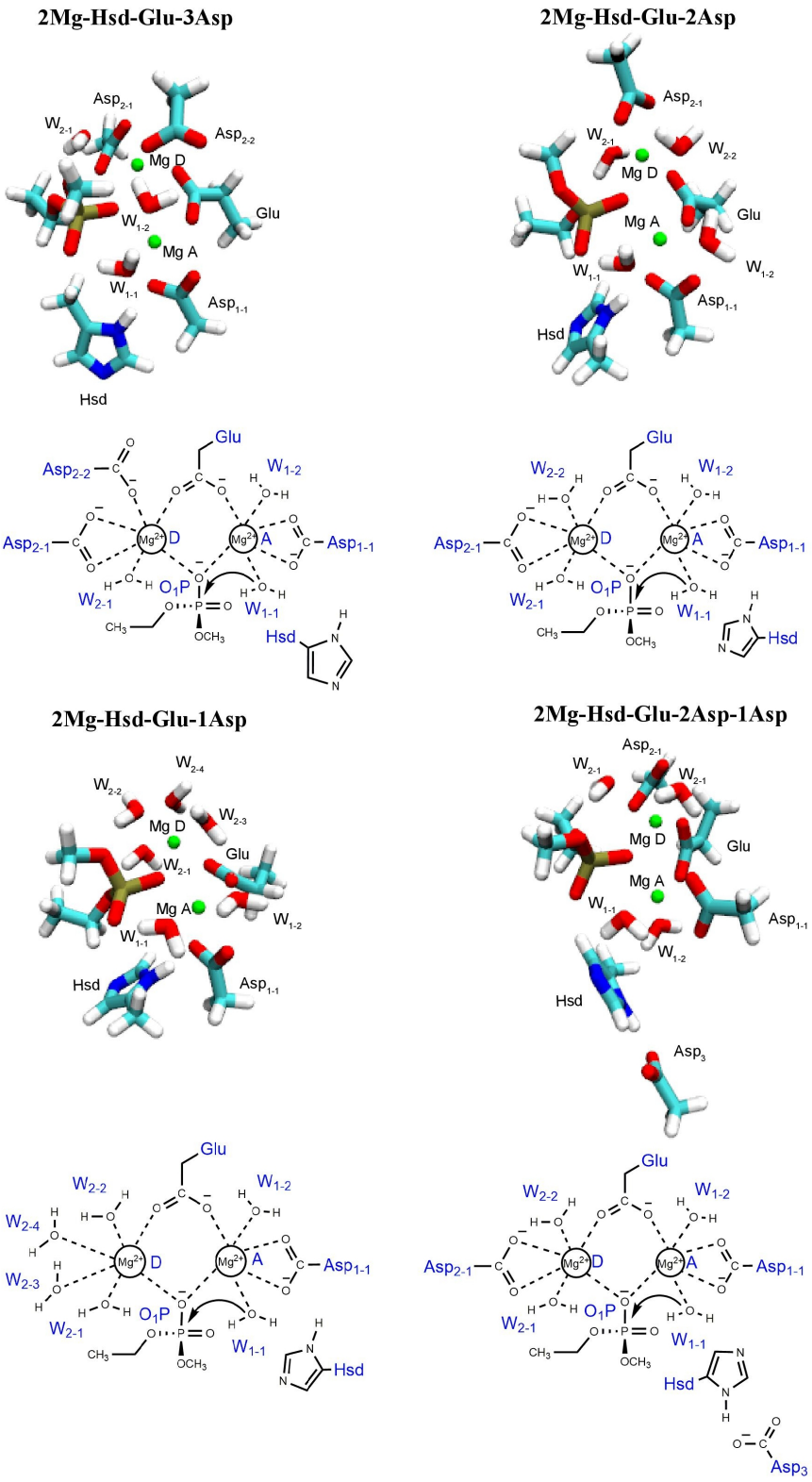

Figure 9: Reactant states for two-metal models. The scheme indicates the attack by the nucleophilic water molecule.

In order to model an architecture that mimics the one in proteins, all amino acid residues were modeled keeping the positions of those C-atoms fixed that represent Cβ atoms, hence serving as anchor atoms. All minima and transition state structures have been initially calculated using CHARMM [25] interfaced to the semi-empirical DFTB3/3OB [132] as quantum method, using specified parameter for phosphorus and magnesium [96]. Minima for reactant, product, and intermediate states were obtained with 10^{-7} a.u. as SCF convergence criterion and 10^{-4} a.u./Å as criterion for the geometry optimization. As outlined in the introduction, two extreme types of mechanisms, dissociative or associative, are possible in the cluster models. Two dimensional scans of the DFTB potential energy surface (PES) using the P–OH2 distance (nucleophilic attack) and the P–O3' distance (leaving group departure) as coordinates, have been carried out to explore whether an associative or a dissociative mechanism is more likely. The PES scans of different one-metal and two-metal models suggest an associative mechanism to be the most favorable (see Figures 32 to 29). Moreover, no dissociative intermediate could be trapped on the B3-LYP of theory.We then inspected the respective associative mechanisms more closely by modeling and optimizing various intermediate states of nucleophilic attack, leaving group departure, and several proton transfer steps preceding or accompanying the P–O bond making and breaking. Minimum energy pathways between those minima have then been computed using the conjugate peak refinement (CPR) method [141] as implemented in CHARMM. Subsequently, selected minima and transitions states were refined applying the B3-LYP hybrid functional [29, 8] and the 6-31G(d,p) basis set [122] as implemented in the Gaussian 09 program package [102]. The geometry convergence criteria was tightened to 1×10^{-8}a.u./Å. The nature of the optimized stationary points (minima or transition states) was verified by normal mode analysis. Additionally, intrinsic reaction coordinates have been followed, starting from the transition structures in both forward and backward direction, so as to validate the minima connected by the respective transition.

All optimizations have been carried out in vacuum and in implicit water, using the polarization continuum method as implemented in the Gaussian 09 with a dielectric constant of $\epsilon = 80$. Partial atomic charges were calculated on the basis of natural bond orbital (NBO) analyses [55].

3.2 Molecular Dynamics Simulations

The Ape1-DNA substrate complex was modeled based on the crystal structure of the phosphorothioate substrate complex with Mn^{2+} as metal cofactor (PDB ID: 5DG0) [168]. The sulfur atom in the thiophosphate group was replaced by an oxygen atom. The Mn^{2+}-ion was substituted with one or two magnesium ions, respectively, placed at different metal binding sites proposed by recent crystal structures. In one model, termed MgF, the Mn^{2+}-ion is simply replaced by a Mg^{2+}-ion at the same position as in the structure 5DG0 [168], that is coordinated by Asp70, Glu96, Asp308 and three water molecules (see Figure 10a). Crystal structures of the catalytically active Ape1-DNA product complex and previous simulations studies based on reactant state models thereof report a Mg^{2+}-ion position that is shifted a bit towards a non-bridging oxygen atom (O1P) of the AP-site [133, 31]. The models with a Mg^{2+}-ion placed at this position on the "departure site" are termed MgD in this work. In these models the direct contact with Asp308 is lost and MgD, coordinates three water molecules, one oxygen atom from the carboxyl group of Asp70, one of Glu96 and one non-bridging oxygen atom (O1P) of the AP-site (see Figure 10b). Another model was set up with Mg^{2+}-ion binding on the "attack site" (MgA), that is coordinated by one non-bridging oxygen atom (O2P) of the AP-site, the side chain oxygen atom of Asn212, the carboxyl group of Asp210, one oxygen atom of Glu96 as suggested by the crystal structure of the Ape1-DNA complex with two metal ions (Pb^{2+} in the crystal structure 1HD7) [130] and previous theoretical studies [110, 111] (see Figure 10c)). After minimization and equilibration, the single Mg^{2+}-ion located at site A in the MgA models is, however, not coordinated by Glu96 but by two water molecules. In order to investigate the effect of the second divalent metal ion on the active site conformation (and hence enzymatic activity) of Ape1, the second Mg^{2+}-ion was positioned at the attack site "A" in the active site of the high-resolution crystal structure of Ape1-DNA substrate (5DG0) while the other Mg^{2+}-ion was placed at site D (see Figure 10d)). Hydrogen atoms were added to both the DNA and the protein. The majority of the residues were kept in the standard protonation state in all MD runs. Exceptional protonation states were used for active site residues His309 and Asp210. The protonation state of His309 was varied explicitly and MD simulations were conducted with

protonated His309 (Hsp) and the two variants of neutral His309 (Hsd: (un-protonated Nϵ and protonated Nδ) and Hse: (protonated Nϵ and un-protonated Nδ)). The standard protonation state of Asp210 is deprotonated, however, this residue has a potential functional importance and therefore simulations with protonated Asp210 were performed as well. To further investigate the role of residues D210 and N212, we also performed simulations of the D210N and N212A mutants of the APE-DNA substrate complex. Table 2 summarizes which models have been simulated in this work.

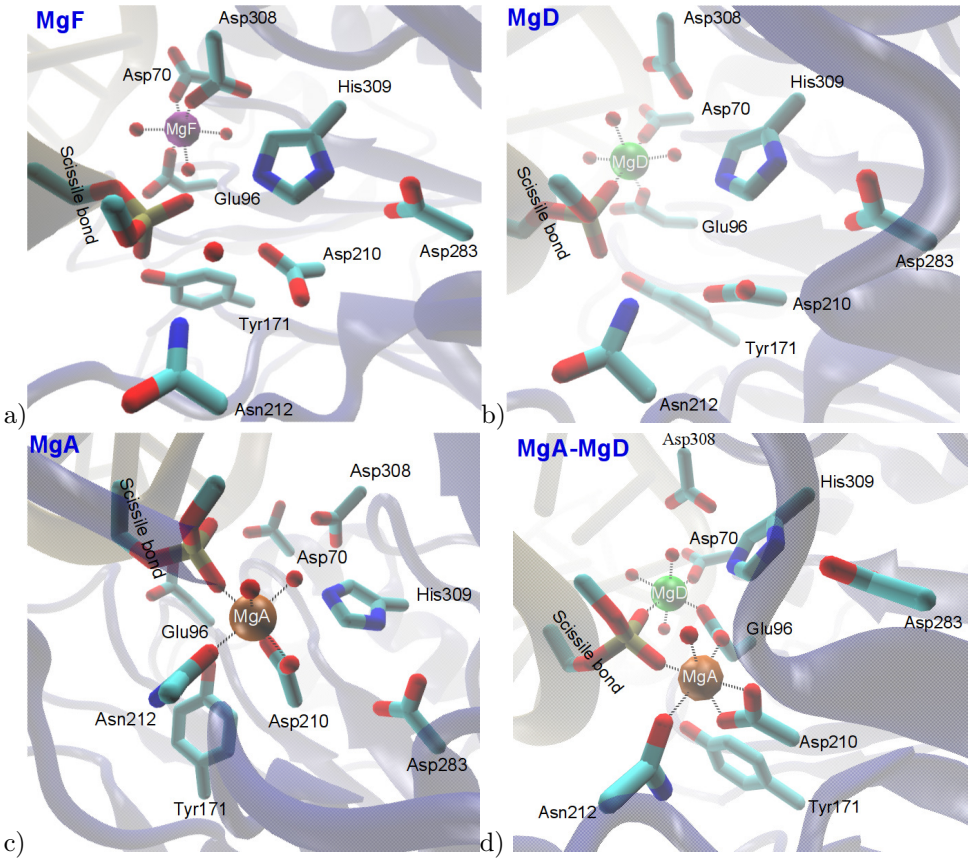

Figure 10: Key Amino acid residues in the active site of the Ape1-DNA substrate complex with a) one magnesium ion at metal binding site F, b) one magnesium ion at metal binding site D, c) one magnesium ion at metal binding site A, d) two magnesium ions at site A and D.

Table 2: List of APE-DNA models simulated in this work

Mg^{2+}-ion position(s)	Asp210 charged			Asp 210 protonated		
	Hsp309	Hsd309	Hse309	Hsp309	Hsd309	Hse309
MgF	3·100ns	3·100ns	3·100ns	1·100ns	1·100ns	1·100ns
MgD	3·100ns	3·100ns	3·100ns	1·100ns	1·100ns	1·100ns
MgA	3·100ns	3·100ns	3·100ns	1·100ns	1·100ns	1·100ns
MgA-MgD	3·100ns	3·100ns	3·100ns	1·100ns	1·100ns	1·100ns

Mg^{2+}-ion position(s)	D210N		N212A
	Hsp309	Hsd309	Hsp309
MgF	1·100ns	1·100ns	1·100ns
MgD	1·100ns	1·100ns	1·100ns
MgA	1·100ns	1·100ns	1·100ns
MgA-MgD	1·100ns	1·100ns	1·100ns

The systems were solvated in a cubic box of 90 Å side length with TIP3P [172] water and sufficient Na^+/Cl^- ions were placed randomly within a minimum distance of 10.5 Å from the solute and 5 Å between ions to ensure a zero net charge for the solute–solvent–counter-ion complex. The systems contained ∼80,000 atoms.

The final complexes were relaxed by 5,000 steps of energy minimization, followed by a 30 ps MD simulation (time step 1 fs) to heat up the system to 300 K by velocity scaling. Next, a 200 ps (time step 2 fs) equilibration was performed for an NPT ensemble. Periodic boundary conditions were implemented with the particle-mesh Ewald method [39] for electrostatic interactions with a cut-off distance of 14 Å. Lennard–Jones interactions were truncated at 14 Å. The SHAKE algorithm was applied to constrain all bonds involving hydrogen atoms. The terminal base pairs of the DNA fragments were restrained harmonically with a force constant of 20 kcal/mol to keep the centers of mass of the respective donor and acceptor atoms involved in Watson-Crick hydrogen bonds at a distance of ∼3 Å .Three independent, 100-ns-long molecular dynamics (MD) simulations were performed for each system setup, started with different initial velocities. For the D210N mutant and the systems with protonated Asp210 only one 100ns long simulation has been performed. During all simulations, pressure (1 bar) and temperature (300 K) were controlled using a Langevin barostat and Nosé—Hoover Langevin thermostat. The simulations were performed with the program package NAMD 2.10 [71] using the CHARMM27 force field [83]. The first 20 ns of each run were disregarded in all analyses as the measured root-mean-square deviation (RMSD) (see Figures 69 to 72) from the native structure of the Ape1-DNA for the systems studied revealed fast growth in conformational changes from 0 to 20 ns. Consequently, this first part of each simulation is regarded as initial equilibration. Analyses have been performed for each of the three runs individually and then averaged. Errors have been estimated as the standard deviation from these mean values, regarding the respective three individual trajectories per model as independent samples.

3.2.1 QM/MM

Based on our studies on the active-site of Ape1-DNA substrate complex [56] [57] and mutational studies in the literature [149, 93, 148, 91, 160] showing a model of Ape1-DNA substrate complex with one Mg^{2+} ion and protonated His309 (MgD_Hsp) as the most probable reactant complex. We have therefore selected a snapshot from the classical MD simulation of this Ape1-DNA model to serve as starting point for our investigations of the enzymatic reaction mechanism. The QM region consists of 127 atoms, precisely, the phosphate backbone 5', 3', non-bridging oxygen atoms of the 5´-phosphate, the nucleophilic water molecule, the side chains of Asp210, Asp70, Glu96, His309, Tyr171, Asn212, Lys98, and Asn68, one Mg^{2+} ion, three metal ligated water molecules, and water molecules around 4 Å of the 5'-phosphate. The selected residues have been shown by mutation studies to have an impact on the chemical reaction of this enzyme. The remainder of the system, including water molecules, was treated with molecular mechanics.

Semi-empirical DFTB3/3OB [132] using specific parameters for phosphorus and magnesium was employed to represent the QM region [96] which has proven a good approximation to computationally more expensive DFT calculations for phosphate hydrolysis reactions [132, 96, 52]. The molecular mechanical part of the system was represented by the CHARMM27 [83] force field and the TIP3P [172, 121] water model. The QM/MM boundary was treated with the link atom approach [101]. All atoms which are within 10 Å from the 5'-phosphate (∼1000) were free to move in the QM/MM calculations, the other atoms were kept fixed at their initial positions. Minima for reactant, product and intermediate states were obtained with 10^{-7} kcal/mol as SCF convergence criterion and 10^{-4} a.u. /Å as criterion for the geometry optimization. Minimum energy pathways between those minima have then been computed using the conjugate peak refinement (CPR) [141] method as implemented in CHARMM. For selected mutants, D210A, N212A, Y171F, and H309N, 400 ps long QM/MM molecular dynamics simulations (see section 3.2.1) have been carried out. Subsequently, a minimized structure of the last snapshot from each of these molecular dynamic simulation has been selected as the initial structure for minimum energy pathway calculations of the phosphodiester hydrolysis reaction.To obtain a free energy profile for each step along the reaction pathway, we have performed umbrella sampling molecular dynamics simulations. To this end, first for each minimum (reactant, intermediate, and product state), we have performed a QM/MM MD simulation, followed by subsequently biasing the transition towards the next (or previous)

state. In these MD simulations, As a reaction coordinate, we chose a combination of distances that best reflects the canonical reaction coordinate obtained in the minimum energy path calculations with CPR [141] of that step. The reaction coordinate was binned into several "windows" of bin size 0.08 Å. The system was gradually heated to 300K and equilibrated for 10 ps. Then, for each window, three independent, 100-ps-long molecular dynamics (MD) simulations were performed, started with different initial velocities, in the NVT ensemble. Temperature was controlled by a Nosé—Hoover Langevin thermostat [27]. The integration step was 1 fs. The system was restrained around the reaction coordinate value of the respective window by a harmonic potential with force constant of 1000 kcal·mol^{-1}·Å^{-2} in the umbrella sampling simulations. Weighted histogram analysis method (WHAM) [143, 147] has been employed to connect the probability distributions of the separate windows. Around 77 windows are sampled to cover leaving group departure (P-O3') and nucleophilic attack of activated water (P-OH). Specifically, 14 windows for the first proton transfer (Lys98 (NZ-OD) ASP70), 33 windows for leaving group departure (P-O3') and 30 windows for nucleophilic attack of activated water (P-OH). The defined coordinates along the reaction pathway from reactant to product state are presented in detail in Table 3 and graphically depicted in Figure 11. While the reaction develops from Reactant toward Int3 the pathway is called "forward" and vice versa (from Int3 toward reactant) it is named "backward". All QM/MM calculations have been performed using CHARMM [25] interfaced to SCC-DFTB [132][96]. The density functional tight binding method, DFTB3, specified for phosphorus [132, 96, 57] has been implemented to calculate Mulliken charges for important amino acid residues and atoms in the active-site of the APE1-DNA complex during the phosphodiester hydrolysis. In order to cover the whole reaction pathway, one snap-shot has been extracted from each sampled window. For instance, the reaction path between reactant state and Int3 was sampled through 14 windows (see Table 4) and Mulliken charges have been calculated for 14 snapshots. Afterwards, an average of 14 snapshots along with the standard deviation for error estimation has been calculated.

Table 3: Reaction coordinates defining the phosphodiester hydrolysis pathway in the APE enzyme from reactant to intermediate state Int3.

Reaction coordinate	Number of windows	Description
[(K98_**NZ-HZ2**_K98) - (K98_**HZ2-OD2**_D70)]	14	First proton transfer: R-Int_1
[(D70_**OD2-HZ2**_K98) - (K98_**HZ2-OH**_W2)]+ 0.75*[(K98_**HZ2-OH**_W2) - (W2_**H2-OH**_W2)]+ 0.7*[(W2_**H2-OH**_W2) - (W2_**H2-OH**_W3)]+ 0.7*[(W3_**H2-OH**_W3) - (W2_**H2-O3'**_CYT7)]+ 0.2*[(CYT7_**O3'-P**_THY8)]	33	Leaving group departure: Int_1-Int_2
0.5*[(D210_**OD2-H1**_W0) - (W0_**OH-H1**_W0)]+ [(W0_**OH-P**_THY8) - (CYT7_**O3'-P**_THY8)]	30	Proton transfer from nucleophile to D210 and nucleophilic attack: Int_2-Int_3

Table 4: Reaction coordinates defining the phosphodiester cleavage pathway in the APE enzyme from intermediate state Int3 to product state

Reaction coordinate	Number of windows	Description
0.5*[(D210_**OD2-H1**_W0) - (W0_**H1-O2P**_THY8)]+ [(W0_**OH2-H2**_W0) - (W0_**H2-OD1**_D210)]	18	Proton transfer: Int3-Int4
0.2*[(D210_**OD2-H2**_W0) - (W0_**H2-OH**_W0)]+ 0.2*[(THY8_**O2P-H1**_W0) - (W0_**H1-OH**_Y171)]+ [(Y171_**OH-HH**_Y171) - (Y171_**HH-OE2**_E96)]	20	Proton transfer: Int4-Int5
[(Y171_**HH-OE2**_E96) - (Y171_**HH-OD1**_D70)]	16	Proton transfer: Int5-Int6
[(Y171_**HH-OD1**_D70) - (Y171_**HH-NZ**_K98)]	13	Proton transfer: Int6-Product

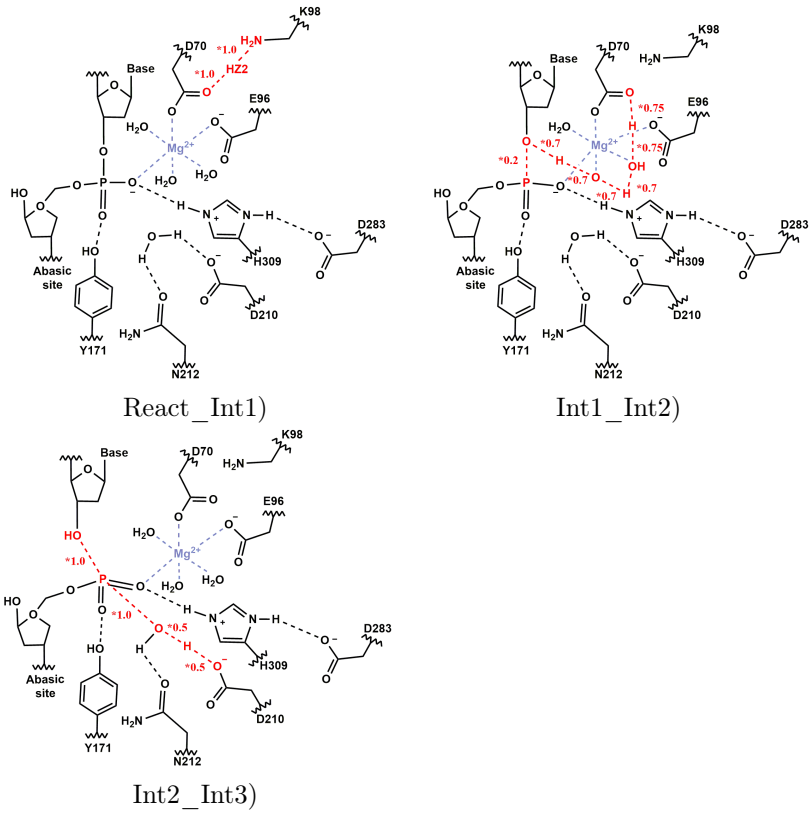

Figure 11: Reaction coordinates (shown in red color) defining the phosphodiester hydrolysis pathway in the APE enzyme from reactant to Int3

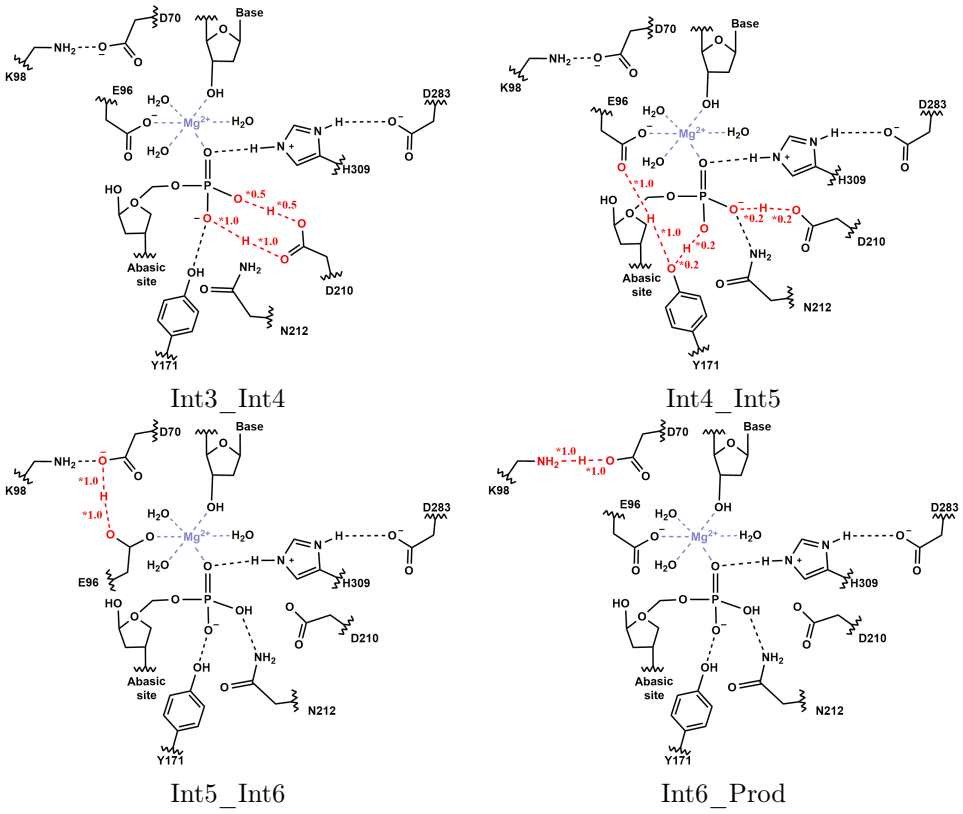

Figure 12: Reaction coordinates (shown in red color) defining the phosphodiester cleavage pathway in the APE enzyme from intermediate state Int3 to product state

4 Role of magnesium ions and histidine protonation in phosphodiester hydrolysis in active-site models

Parts of this chapter are published in "*Batebi, H. and Imhof, P., Phosphodiester hydrolysis computed for cluster models of enzymatic active sites. Theor. Chem. Acc. 135 (2016) 262*, DOI"

Phosphodiester hydrolysis is the fundamental reaction of endonuclease enzymes cleaving the backbone of DNA and RNA, usually at a recognition site such as a specific sequence (in e.g. restriction enzymes [114]), at an abasic site (such as the apurinic/apyrimidinic (AP) endonuclease in the Base excision repair system)[16], or escherichia coli endonuclease IV (endoIV-DNA) [41] or endonuclease III homologue 1 (NTH1) that initiates the repair by β-elimination [87]. Moreover, phosphodiester cleavage can also occur in ribozymes, i.e. catalytic RNA, acting as ribonucleases [60, 19, 28].

In order to elucidate the putative role that Magnesium cofactors and other residues which are typically found in active sites of endonuclease enzymes play in the cleavage of a phosphate backbone, we have performed quantum mechanical calculations of reaction pathways based on different minimalistic (cluster) models that can be regarded as a universal, but necessarily simplistic, endonuclease, mimicking such active sites of endonuclease enzymes (see Figure 13). In this "reductionist approach", the cluster models include either one or two metal ions and also vary in the number and position of amino acid residues, the number of water molecules involved in the metal ion coordination, and in the protonation state of a catalytically important histidine residue. Histidine residues may catalyze the reaction via activating the nucleophile through deprotonation or facilitating leaving group departure through protonation as has been discussed widely in the context of phosphodiester backbone cleavage [100, 74, 45] and it is therefore plausible to consider the impact of different protonation states of histidine in the reaction pathways. Aspartate and glutamate are common amino acid residues in active sites of nucleases that are important in positioning the metal cofactors but may also act as general base/general acid, Figure 13.

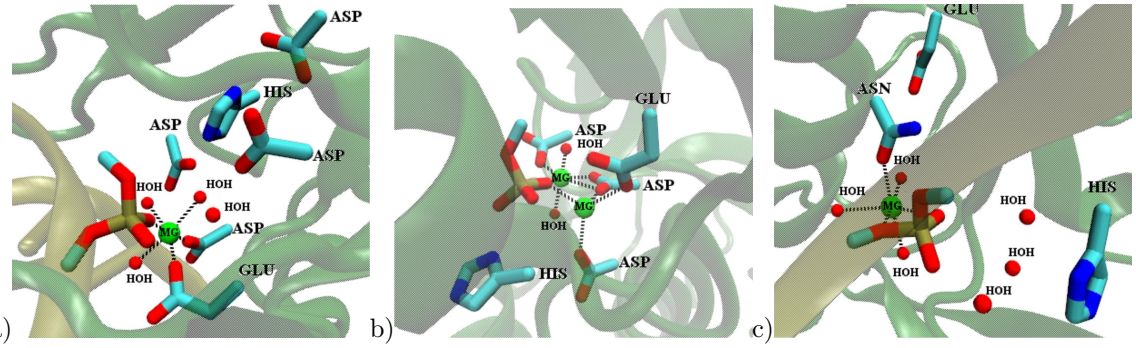

Figure 13: Active-site residues in nuclease enzymes a) human APE1 endonuclease-DNA substrate complex with one magnesium ion b) nuclease domain of 3'hExo, bound to rAMP with two magnesium ions c) intron-encoded homing endonuclease I-PPOI (H98A)/DNA homing site complex with one magnesium ion. Only the side chains of the active site amino acids and the phosphate group are shown in licorice representations. The metal ions and water molecules are depicted as green or red spheres, respectively. The protein environment and the nucleic acids are indicated by a cartoon representation.

In this section we report calculations of phosphate hydrolysis mechanisms in cluster models of different size and composition. We find associative mechanisms to be favorable in which the direct attack of the nucleophilic water molecule with proton transfer to the phosphate group is facilitated by a second metal ion or a positively charged histidine residue whereas no catalytic effect on leaving group departure is observed. The energetically most favorable mechanism is initiated by proton transfer from the nucleophilic water molecule to a histidine-aspartate cluster. The nucleophilic attack is stabilized by the metal ions in such a way that formation of a pentacovalent intermediate

is feasible while still allowing subsequent P–O bond dissociation. This balance in stabilization and weakening of P–O bonds is achieved by a delicate structural arrangement of metal ions and active site residues.

4.1 One-metal models

4.1.1 One magnesium ion and a protonated histidine: Mechanisms 1MgA-p and 1MgD-p

For the 1MgA-p mechanism computed for a model with one magnesium atom located at the attack site, protonated histidine, and one extra aspartate residue (see model 1MgA-Hsp-Glu-1Asp-1Asp in Figure 8) the barrier for nucleophilic attack (TS1_1MgA-p) is 34.9 kcal/mol in vacuum (34.8 kcal/mol in solvent) (see Figure 14). Upon this attack a proton is transferred from the water molecule to the O2P atom of the phosphate group (see Figure 16) leading to a pentacovalent intermediate, Int1_1MgA-p, with an energy about as high as the transition state leading to it (34.7 kcal/mol and 33.7 kcal/mol in vacuum and solvent, respectively). In the second step of the mechanism the HO2–P rotatessuch that the hydrogen atom points towards the O3' atom of the leaving group. The corresponding transition state, TS2_1MgA-p, has a similar energy as the nucleophilic attack in this mechanism (34.9 kcal/mol in vacuum and 34.7 kcal/mol in solvent). Leaving group departure, coupled to a proton transfer from the phosphate O2P atom to the O3' atom, goes through the transition state with the highest energy along this 1MgA-p mechanism (40.4 kcal/mol in vacuum and 39.9 kcal/mol in solvent, respectively).

Placing the magnesium ion at the departure site (see model 1MgD-Hsp-Glu-1Asp-1Asp in Figure 8), results in a mechanism 1MgD-p with a reduced barrier of 29.0 kcal/mol ($\sim$28kcal/mol in solvent, Figure 15) for the nucleophilic attack. Also the pentacovalent intermediate in this mechanism, Int1_1MgD-p has a lower energy than in the 1MgA-p mechanism with the magnesium ion located at the attack site: 23.7 kcal/mol in vacuum and 24.kcal/mol in solvent, respectively.

It should be noted that the magnesium ion is coordinated by the same type and number of ligands (water molecules and amino acid residues) in both positions, i.e. on the attack and on the departure site. One difference is, however, the loss of a favorable hydrogen-bonded contact of the nucleophilic water molecule to residue Asp1 in the reactant state of mechanism 1MgA-p upon attack which is a possible explanation for the increase in activation energy compared to the 1MgD-p mechanism.

The rotation of the HO2–P group has a transition state, TS2_1MgD-p, of relative energy of 35.0 kcal/mol (34.3 kcal/mol in solvent) which is the highest point along this mechanism (cf. Figure 15). Leaving group departure in this mechanism has a transition state, TS2_1MgD-p, of similar energy (30.0 kcal/mol) as the one of the nucleophilic attack, TS1_1MgD-p.

The barriers calculated for the 1MgD-p mechanism are somewhat lower than those calculated for uncatalyzed phosphodiester hydrolysis in which only the phosphate group and the nucleophilic water molecules are considered [120], or those calculated in the presence of a water-ligated magnesium ion [119]. The small catalytic effect by the additional residues in the present model (glutamate and aspartate ligating the magnesium ion, and a histidine residue) can be explained by a positioning of the nucleophile. The substantially higher barrier (40.5 kcal/mol) for P-O bond scission with a proton transfer from a metal-ligated water molecule to the O3'-atom of the leaving group renders also the catalytic effect of the magnesium ion to be structural rather than by direct participation on the reaction.

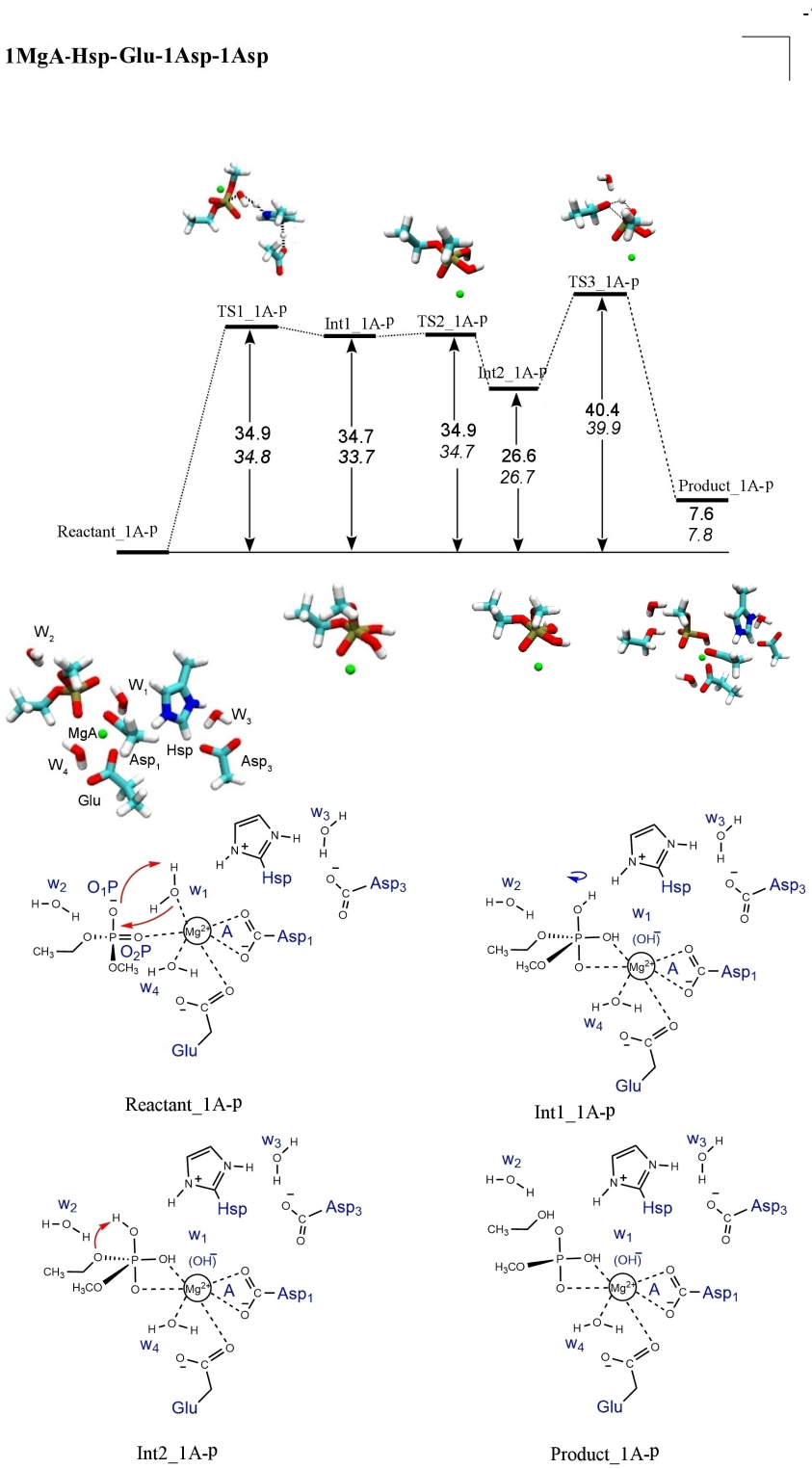

Figure 14: Mechanism, 1MgA-p, for model with one magnesium ion located a the attack site, A, and protonated histidine (see 1MgA-Hsp-Glu-1Asp-1Asp in Figure 8), calculated by B3LYP/6-31G(d,p) in vacuum (upright numbers) and solvent (in italics). All energies (in kcal/mol) are given relative to the reactant state.

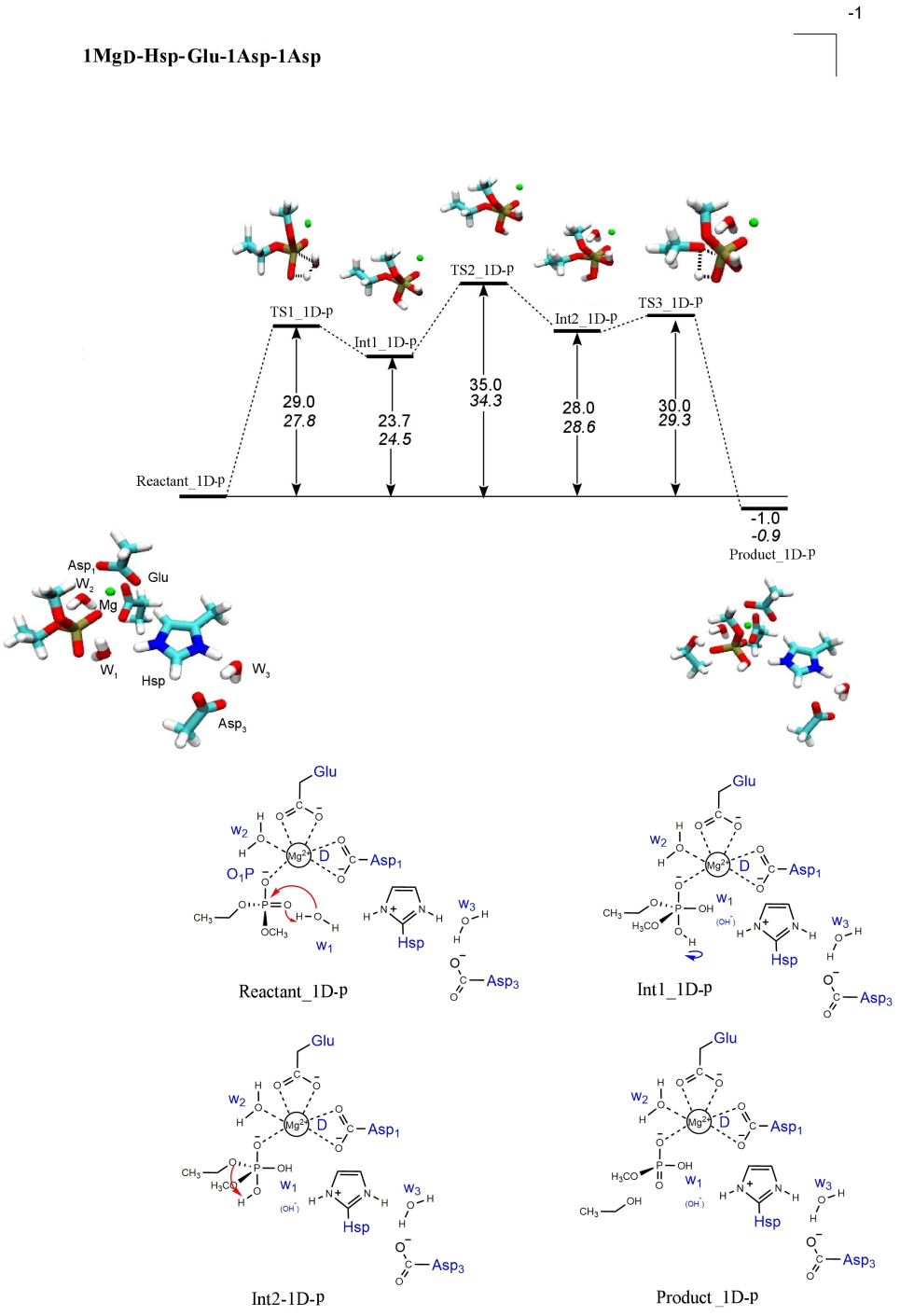

Figure 15: Mechanism,1MgD-p, for model with one magnesium ion located at the departure site, D, and protonated histidine (see 1MgD-Hsp-Glu-1Asp-1Asp in Figure 8), calculated by B3LYP/6-31G(d,p) in vacuum (upright numbers) and solvent (in italics). All energies (in kcal/mol) are given relative to the reactant state.

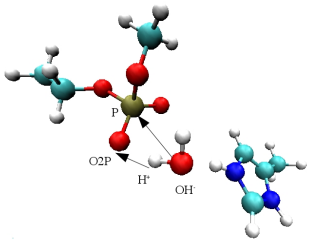

Figure 16: Attack of a water molecule on the phosphorus atom of a phosphodiester.

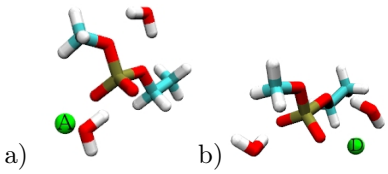

Figure 17: Different position of the magnesium ion in a) model with one magnesium ion at the attack site, A, and b) model with one magnesium ion at the departure-site, D.

4.1.2 One magnesium ion and neutral histidine: Mechanisms 1MgA-d and 1MgD-d

To check whether the positive charge of the protonated histidine plays an important role in stabilizing the transition state for nucleophilic attack we computed a similar associative mechanism of a cluster model with one magnesium ion and a neutral histidine, again with the magnesium ion located on either the attack site or at the departure site (see models 1MgA-Hsd-Glu-1Asp-1Asp, and 1MgD-Hsd-Glu-1Asp-1Asp in Figure 8, respectively).

The transition state for nucleophilic attack with proton transfer from the attacking water molecule to the phosphate group has an energy of 29.6 kcal/mol in vacuum (38.6 kcal/mol in solvent) if the magnesium ion is located on the attack (Figure 18, TS1_1MgA-d) site and 35.1 kcal/mol in vacuum (39.4 kcal/mol in solvent) if the metal ion is located on the departure site (Figure 19, TS1_1MgD-d). Comparison of these energies with those obtained for the 1MgA-p and 1MgD-p mechanism for models with protonated histidine suggest a favorable role of protonated histidine for the nucleophilic attack if no metal ion is located on the attack site. In that case the barrier for nucleophilic attack is similar to the one computed for the 1MgA-p mechanism, indicating that the presence of a positive charge on the attack site is favorable for this first step. Because of the high initial barrier no further steps have been computed for the 1MgD-d mechanism. The rate-determining step in the 1MgA-d mechanism is the cleavage of the scissile bond with a transition state energy of 39.2 kcal/mol (39.5 kcal/mol in solvent).

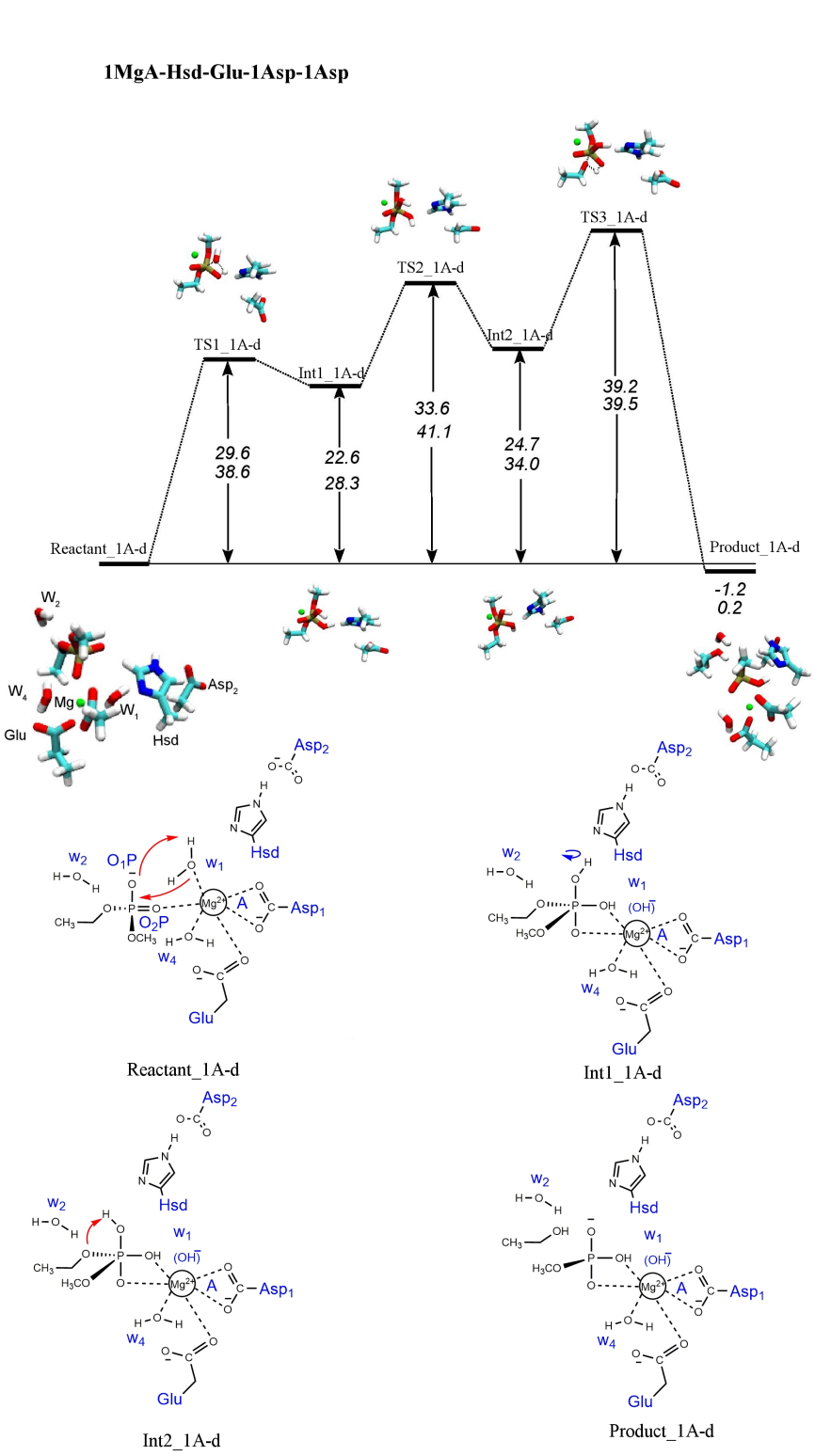

Figure 18: Direct mechanism, 1MgA-d, in the model with one magnesium located at the attack site, neutral histidine, and two aspartate residues (see 1MgA-Hsd-Glu-1Asp-1Asp in Figure 8), calculated by B3LYP/6-31G(d,p) in vacuum (upright numbers) and solvent (in italics). Energies (in kcal/mol) are given relative to the reactant state.

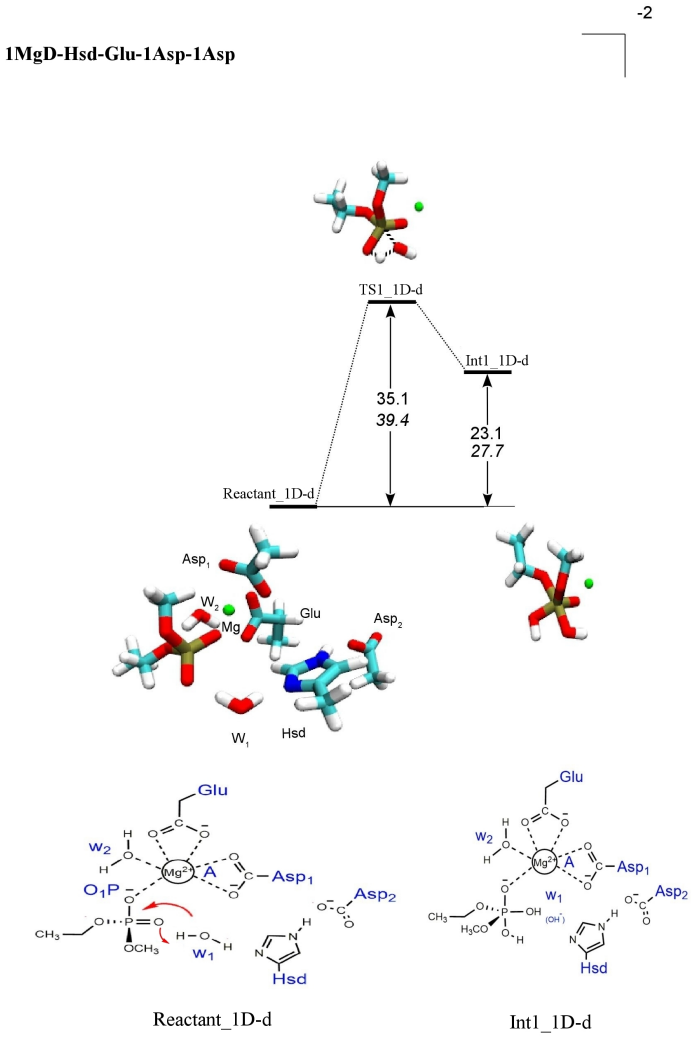

Figure 19: Direct nucleophilic attack, 1MgD-d, in the model with one magnesium ion located at the departure site, neutral histidine, and two aspartate residues, (see 1MgD-Hsd-Glu-1Asp-1Asp in Figure 8), calculated by B3LYP/6-31G(d,p) in vacuum (upright numbers) and solvent (in italics). Energies (in kcal/mol) are given relative to the reactant state.

4.1.3 One magnesium ion and neutral histidine: Mechanisms 1MgA-h and 1MgD-h

Another possibility to activate the water molecule for nucleophilic attack is the transfer of a proton to the Nδ atom of the unprotonated histidine residue. In the corresponding models (see 1MgA-Hsd-Glu-1Asp-1Asp and 1MgD-Hsd-Glu-1Asp-1Asp in Figure 8), the aspartate mimic is positioned close to the histidine, forming a hydrogen bond to the Nϵ atom. This conformation resembles a catalytic triad [103, 40] (if the nucleophile is included) that allows shuttling of a proton between histidine and aspartate, thus facilitating acceptance of a proton from the nucleophilic water molecule. Indeed, a significant reduction in the energy barrier can be observed in such a histidine-assisted nucleophilic attack, dropping to 23.5 kcal/mol (25.4 kcal/mol in solvent) in the 1MgA-h mechanism (Figure 20, TS1_1MgA-h) and to 27.4 kcal/mol (26.2 kcal/mol, in solvent) in the 1MgD-h mechanism (Figure 21, TS1_1MgD-h). In the latter model, this is the rate determining step in this mechanism (Figure 21). In the resulting intermediate, Int1_1MgD-h, the scissile bond is already cleaved, thus proton transfer, nucleophilic attack and leaving group departure all take place in one step. The remaining reprotonation steps have negligible barriers. In contrast, in the mechanism calculated for a model with the metal ion located on the attack site, 1MgA-h, nucleophilic attack and leaving group departure are separate steps, indicating some stabilization of the pentacovalent intermediate, Int1_1MgA-h, by the metal ion. In fact, leaving group departure is preceded by proton transfer from the histidine to the phosphate group (TS2_1MgA-h with 28.3 kcal/mol barrier in vacuum and 31.1 kcal/mol barrier in solvent, respectively). The final cleavage of the scissile bond, assisted by proton transfer from a magnesium-bound water molecule, has a transition state energy of 39.2 kcal/mol (39.5 kcal/mol in solvent) and is the rate determining step like in the 1MgA-d mechanism. These rate-determining barriers for leaving group departure are close to those of uncatalyzed dimethyl-phosphate hydrolysis [120]. This is reasonable since the magnesium ion situated on the "attack site" does not have a proper catalytic effect on leaving group departure.

Table 5: Energies of stationary points along the most favorable phosphate hydrolysis mechanisms computed for the one-metal models. All energies (in kcal/mol) have been calculated with (B3LYP, 6-31G(d,p)) and are given relative to the respective reactant state.

Mechanism	TS1	Int1	TS2	Int2	TS3	Int3	TS4	Product
1MgA-p vacuum	34.9	34.7	34.9	26.6	40.4			7.6
1MgA-p solvent	34.8	33.7	34.7	26.7	39.9			7.8
1MgD-p vacuum	29.1	23.7	35.0	28.0	30.0			-1.0
1MgD-p solvent	31.3	26.7	36.2	29.0	31.5			-0.7
1MgA-d vacuum	29.6	22.6	33.6	24.7	39.2			-1.2
1MgA-d solvent	38.6	28.3	41.1	34.0	39.5			0.2
1MgA-h vacuum	23.5	23.3	28.3	22.6	33.6	24.7	39.2	-1.2
1MgA-h solvent	25.4	25.0	31.1	28.3	41.1	34.0	39.5	0.2
1MgD-h vacuum	27.4	-2.61	22.8	-4.7	-4.3			-6.1
1MgD-h solvent	26.2	-1.4	14.22	-4.6	-2.3			-2.6

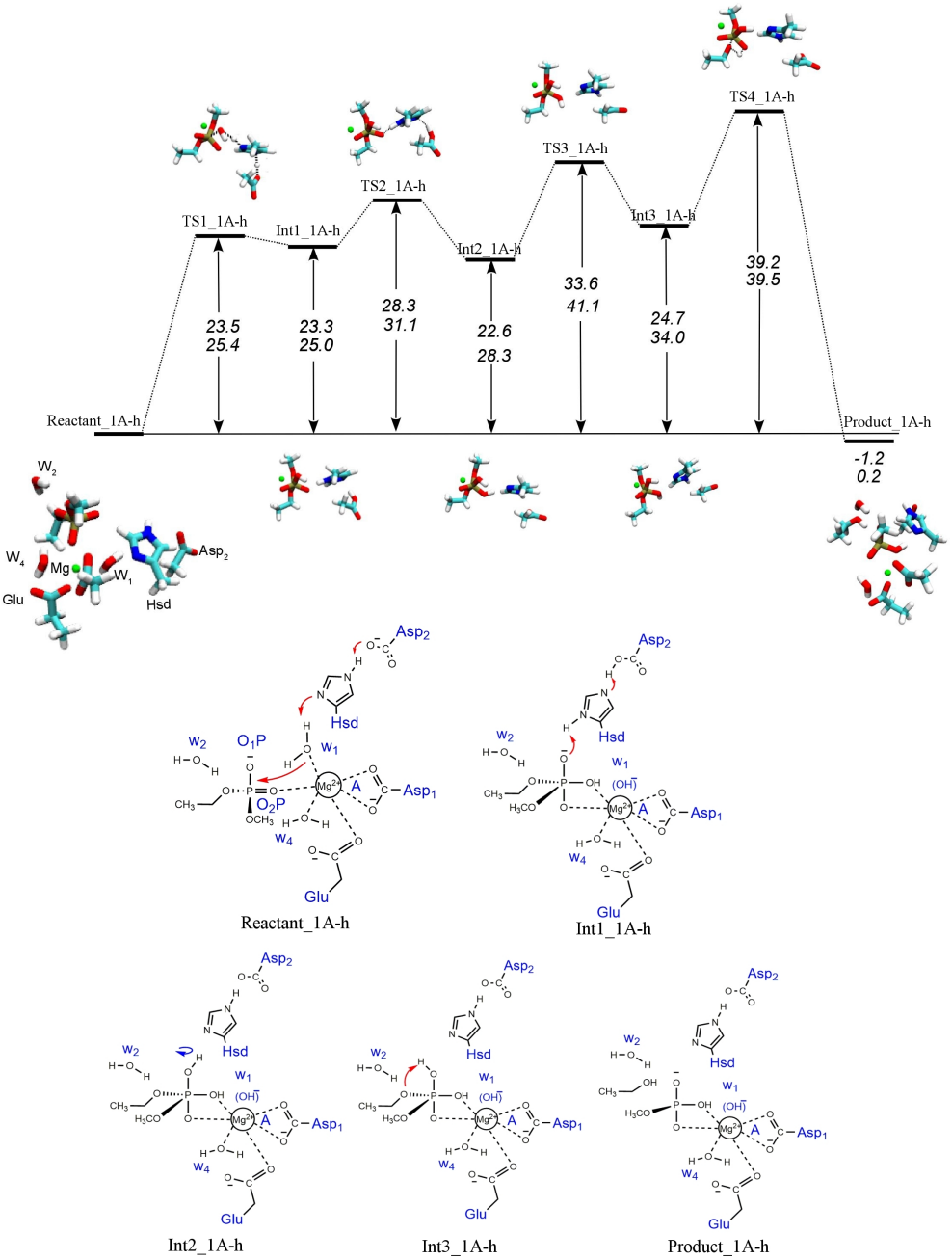

Figure 20: Histidine-assisted mechanism (1A-h) for model with one magnesium ion located at the attack site, neutral histidine and two aspartate residues, 1MgA-Hsd-Glu-1Asp-1Asp, calculated by B3LYP/6-31G(d,p) in vacuum (upright numbers) and solvent (in italics). All energies (in kcal/mol) are given relative to the reactant state.

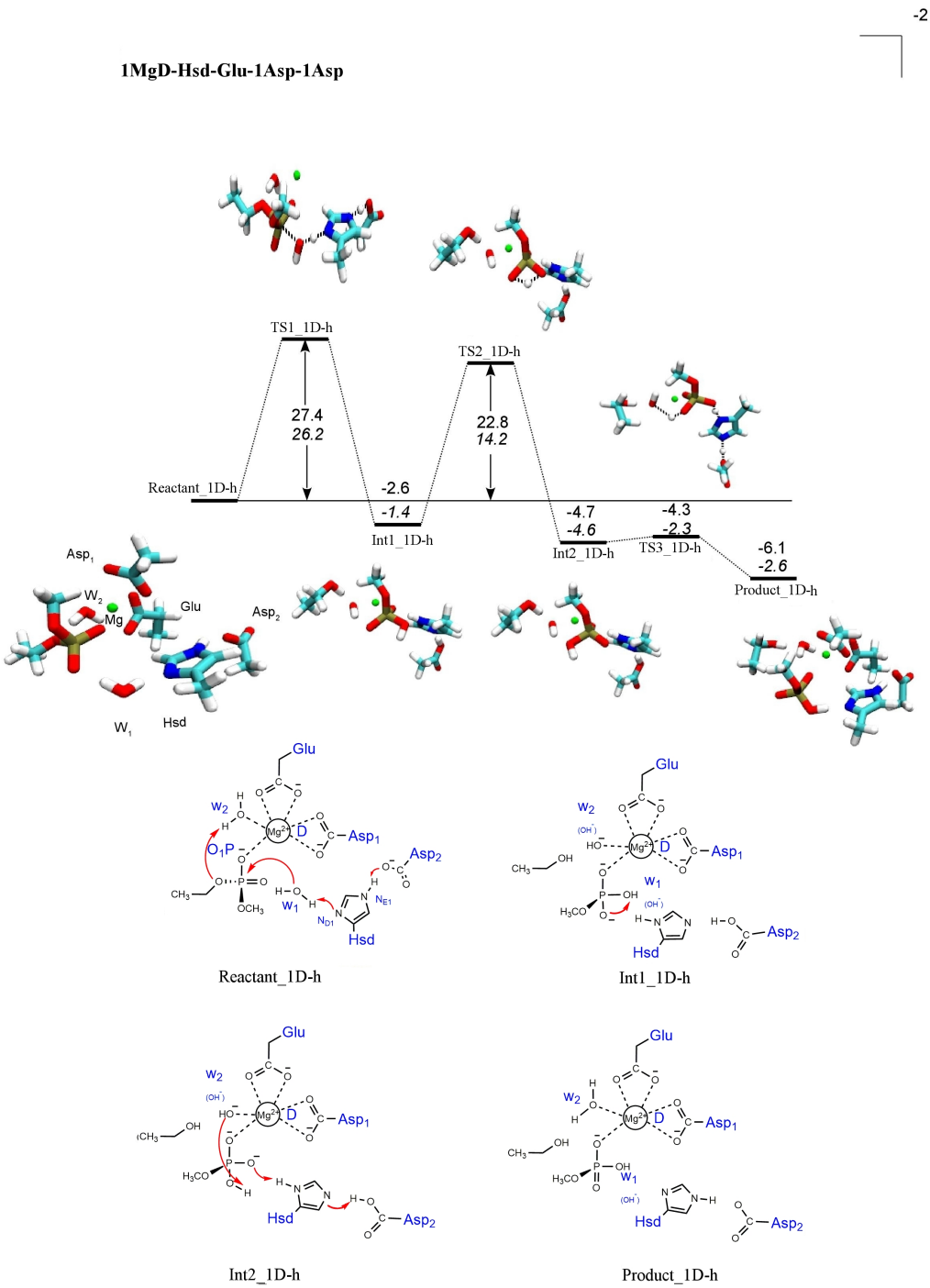

Figure 21: Histidine-assisted mechanism (1D-h) for model with one magnesium ion located at the departure site, neutral histidine and two aspartate residues, 1MgD-Hsd-Glu-1Asp-1Asp, calculated by B3LYP/6-31G(d,p) in vacuum (upright numbers) and solvent (in italics). All energies (in kcal/mol) are given relative to the reactant state..

4.2 Two-metal models

According to the potential energy scans (Figure30 and 32), also the two-magnesium models exhibit associative mechanisms for phosphate hydrolysis. Also for these models, no dissociative intermediate could be trapped on the B3LYP/6-31G(d,p) level of theory.

In the two-magnesium models, one of the two metal ions, MgA is situated on the "attack site", that is the site on which the nucleophile is located, and the second metal ion, MgD is located on the "departure site", i.e. closer to the O3' atom of the leaving group. Each metal ion is coordinated by an aspartate residue and one water molecule, a shared oxygen atom of the phosphate group, one of the two oxygen atoms of a glutamate residue, and double-coordinated by one aspartate residue (see Figure 9).

4.2.1 Two magnesium ions, neutral histidine and one, two, or three aspartate residues: Mechanisms 2Mg-d0, 2Mg-d1, and 2Mg-d2

To test whether an increase in positive charge on the "departure site" has an activating effect on the leaving group, we have setup models in which the second metal ion is coordinated by two, one, or no aspartate residue (see models 2Mg-Hsd-Glu-3Asp, 2Mg-Hsd-Glu-2Asp, and 2Mg-Hsd-Glu-1Asp in Figure 9). The corresponding two mechanisms are termed 2Mg-d2, 2Mg-d1, and 2Mg-d0, respectively. For these direct mechanisms in the two-metal models the energy barriers for nucleophilic attack (26.0 kcal/mol in vacuum and 26.6 kcal/mol in solvent, 22.5 kcal/mol in vacuum and 22.4 kcal/mol in solvent, and 29.2 kcal/mol in vacuum and 29.1 kcal/mol in solvent for TS1_2Mg-d2, TS1_2Mg-d1, and TS1_2Mg-d0, respectively) do not show any trend with respect to the total charge of the system with hardly any difference between vacuum and solvent models. Moreover, these energy barriers are similar to the barrier computed for the histidine-activated nucleophilic attack in the one-metal mechanism 1MgD-h (TS1_1D-h 27.4 kcal/mol in vacuum and 26.2 kcal/mol in solvent. respectively) and the one-metal 1MgD-p mechanism with protonated histidine (TS1_1MgD-p 29.0 kcal/mol in vacuum and 27.8 kcal/mol in solvent, respectively). In the latter, as well as in this two-metal model, one proton of the nucleophilic water molecule is directly transferred to an oxygen atom of the phosphate group and the remaining OH^- ion attacks to the phosphorous atom (Figure 22 and 15). In all these cases, generation of an OH^- ion can be regarded as favored by the nearby residues: the positive charge of the histidine or the magnesium ion "push" the proton from the nucleophile towards the phosphate and/or stabilize the negatively charged hydroxide ion. In the histidine-assisted mechanism, the proton is ultimately accepted by the aspartate residue at the end of the triad. An impact of the charge difference due to another aspartate residue on phosphate hydrolysis can, however, be seen in the reaction energies in vacuum which are 3.2 kcal/mol in the 2Mg-d2 mechanism as opposed to the exothermic reaction in the 2Mg-d1 mechanism with -6.0 kcal/mol relative product energy. This effect is not observed in solvent where the reaction energies, -3.3 kcal/mol and -2.2 kcal/mol in the 2Mg-d2 and 2Mg-d1 mechanism, respectively, are rather similar, indicating that the additional negative charge introduced by an aspartate residue, is screened in solvent. The computed direct reaction pathways for the two-metal models, 2Mg-d2, 2Mg-d1, and 2Mg-d0, show no significant difference in the energy barrier for departure of the leaving group which is the rate-determining step in all these mechanisms (with 37.6 kcal/mol, 32.0 kcal/mol and 32.1 kcal/mol in vacuum, and 34.3 kcal/mol, 34.9 kcal/mol, and 34.1 kcal/mol in solvent, respectively, and see also Table 6) In the 2Mg-d3 mechanism the transition state energy for this step is even higher than the respective transition state energies in all the one-metal mechanisms calculated in this work. Therefore, the second magnesium ion near the O3' likely does not play a catalytic role in leaving group departure. Dissociation of the scissile bond in the pentacovalent intermediate is accompanied by a proton transfer from the O2P atom of the phosphate group to the leaving group. This proton can alternatively be transferred from a metal-ligated water molecule. In all models studied here, the energy barriers for the latter option, leaving group departure with accepting a proton from a metal-ligated water molecule, (47.5 kcal/mol, 37.8 kcal/mol and 38.5 kcal/mol for 2Mg-d2, 2Mg-d1, and 2Mg-d0, respectively) are higher than those for leaving group departure with a proton transfer from the O2P atom. This preference for the proton transfer from the O2P atom can be explained by the metal-bound phosphate group being a stronger Bronsted acid, i.e. a better proton donor than a metal-activated water molecule.

2Mg-Hsd-Glu-3Asp

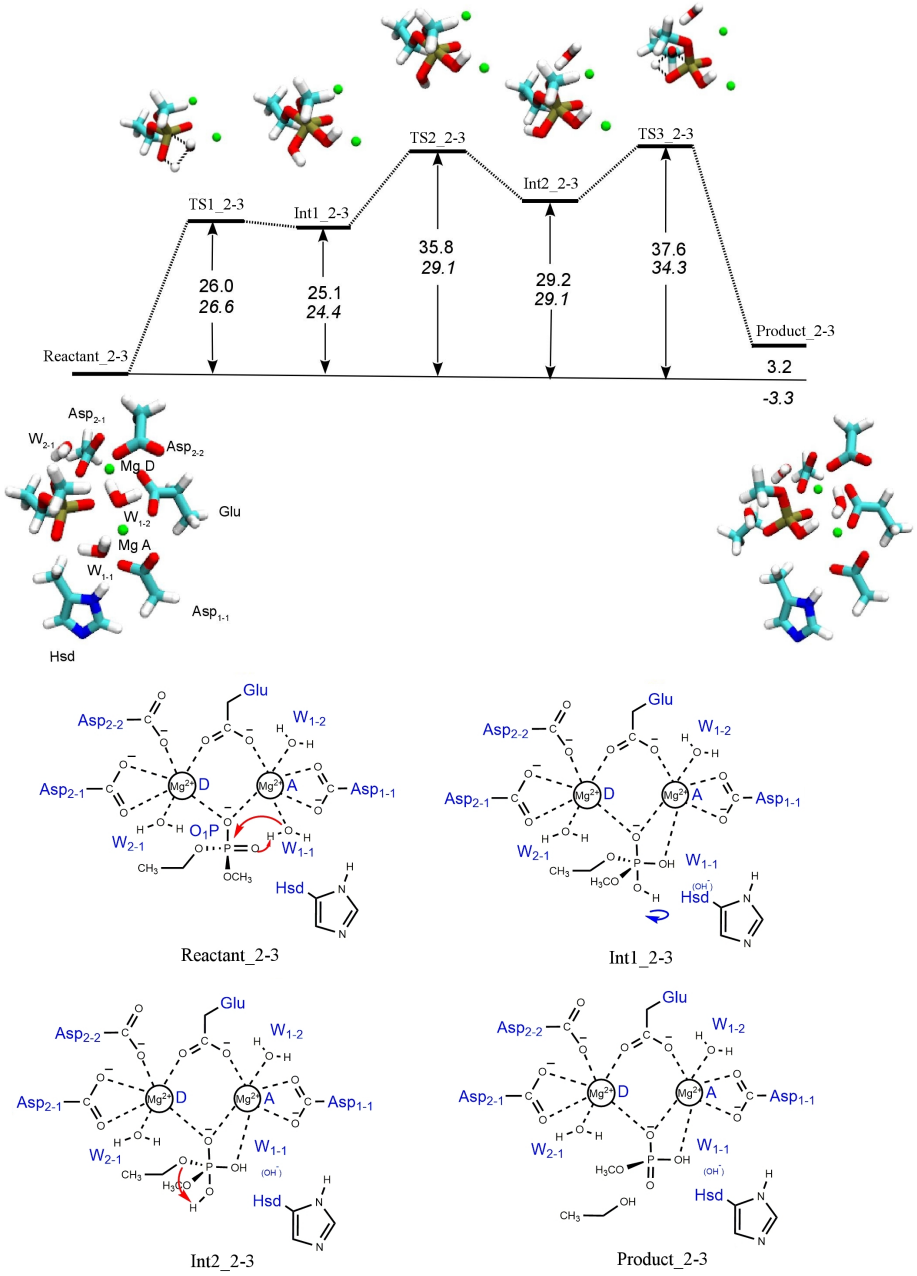

Figure 22: Mechanism 2Mg-d2 for model with two magnesium ions, neutral histidine, and three aspartate residues (see 2Mg-Hsd-Glu-3Asp in Figure 9), calculated by B3LYP/6-31G(d,p) in vacuum (upright numbers) and solvent (in italics). Energies (in kcal/mol) are given relative to the reactant state.

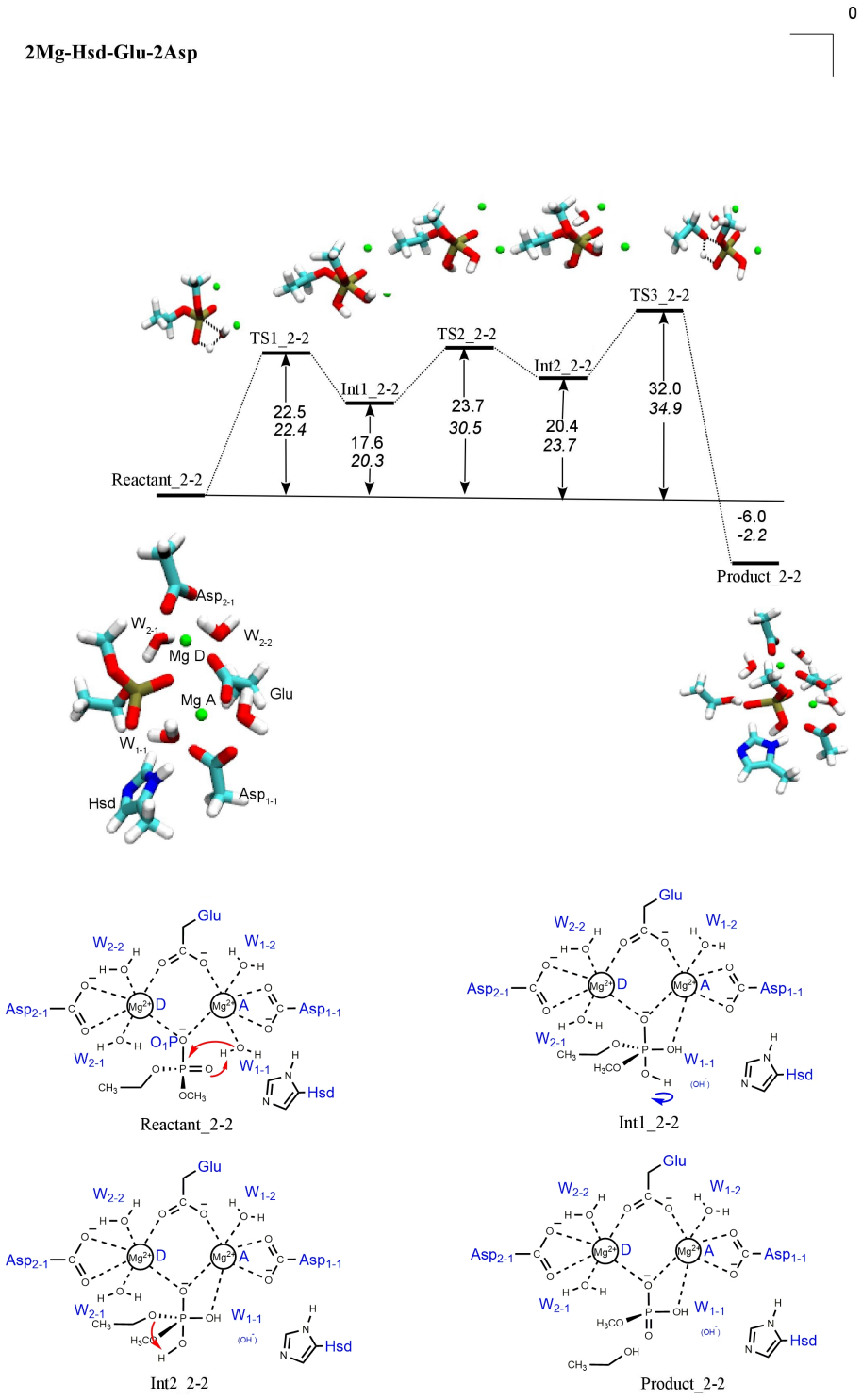

Figure 23: Mechanism 2Mg-d1 for model with two magnesium ions, neutral histidine, and two aspartate residues (see 2Mg-Hsd-Glu-2Asp in Figure 9), calculated by B3LYP/6-31G(d,p) in vacuum (upright numbers) and solvent (in italics). All energies are in kcal/mol relative to the reactant state.

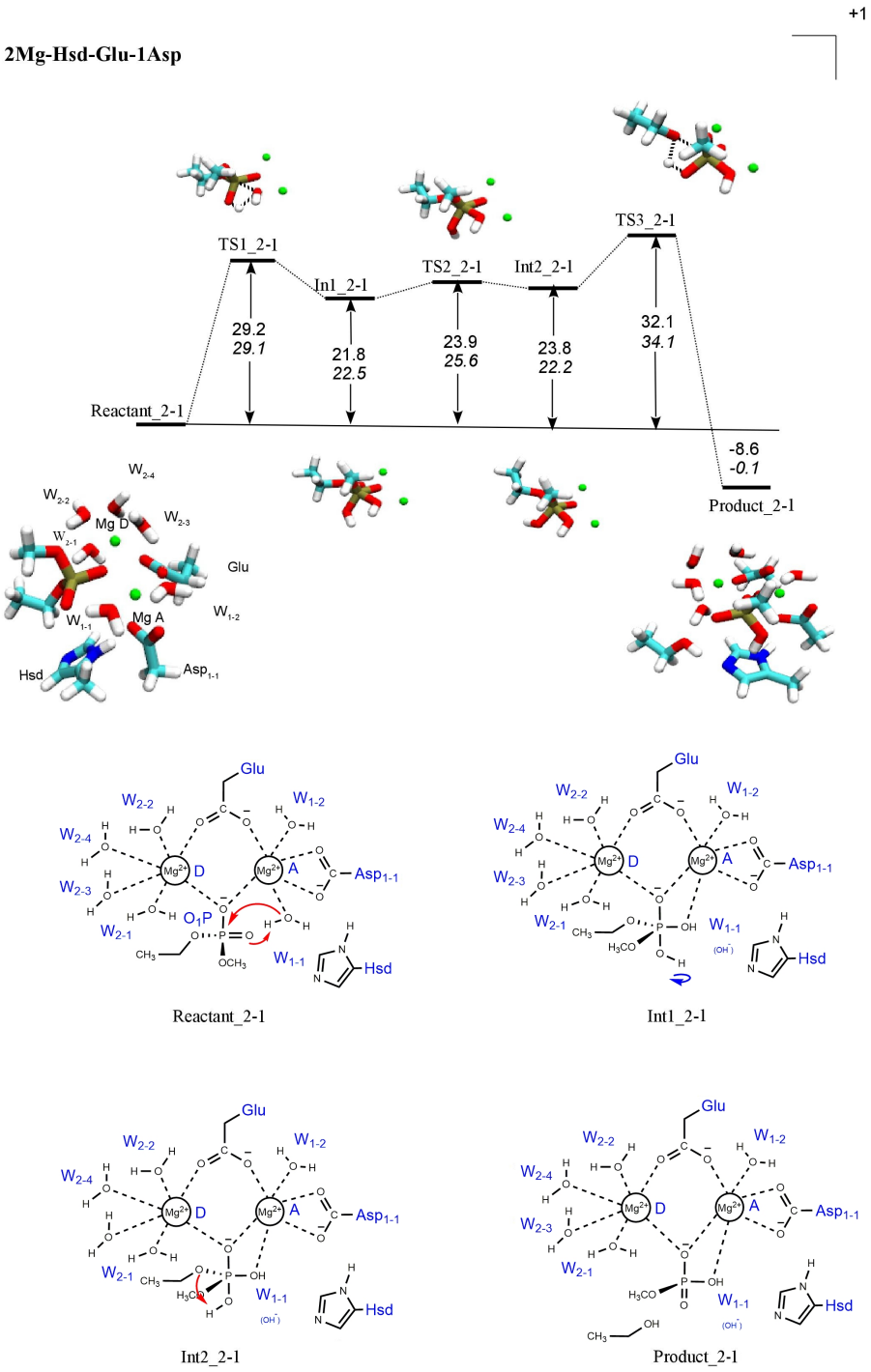

Figure 24: Mechanism 2Mg-d0 for model with two magnesium ions, neutral histidine, and one aspartate residue (see 2Mg-Hsd-Glu-1Asp in Figure 9), calculated by B3LYP/6-31G(d,p) in vacuum (upright numbers) and solvent (in italics). All energies are in kcal/mol relative to the reactant state

4.2.2 Two magnesium ions, neutral histidine and three aspartate residuesMechanism 2Mg-h

In order to calculate a histidine-assisted mechanism similar to those obtained for the one-metal models, 1MgA-h and 1MgD-h, (Figures 20 and 21), we placed the third aspartate residue in hydrogen-bond distance to the histidine (see model 2Mg-Hsd-Glu-2Asp-1Asp in Figure 9). Direct attack of the water molecule with proton transfer to the phosphate group in this model has a barrier of 29.5 kcal/mol (25.4 kcal/mol in solvent) and is thus comparable to the direct nucleophilic attack in the other two-metal models..

In contrast, the mechanism with shuttling a proton from the nucleophilic water molecule to the Nϵ atom of the histidine, 2Mg-h, has an energy barrier for nucleophilic attack that is reduced to 9.7 kcal/mol (16.9 kcal/mol in solvent). The second transition state for dissociation of the scissile bond is 12.1 kcal/mol (18.6 kcal/mol in solvent) and thus also considerably lowered compared to the leaving group departure steps in the other two-metal models as well as lower than in the histidine-activated mechanism of the one-magnesium models (1MgA-h and 1MgD-h). The rate determining step in the 2Mg-h mechanism is the final proton transfer from the Nϵ atom of the histidine residue to the dissociated alcohol. The transition energy for this proton transfer step is 25.8 kcal/mol (29.2 kcal/mol in solvent). This rather high energy barrier is still the lowest rate-determining barrier among all mechanisms and vacuum models studied in this work. In solvent, the barrier for this reprotonation step is comparable or slightly higher than the ones computed for nucleophilic attack of the other models. The barriers for the nucleophilic attack and dissociation of the scissile bond, however, are significantly lower than in all other models and mechanisms proposed in this work.

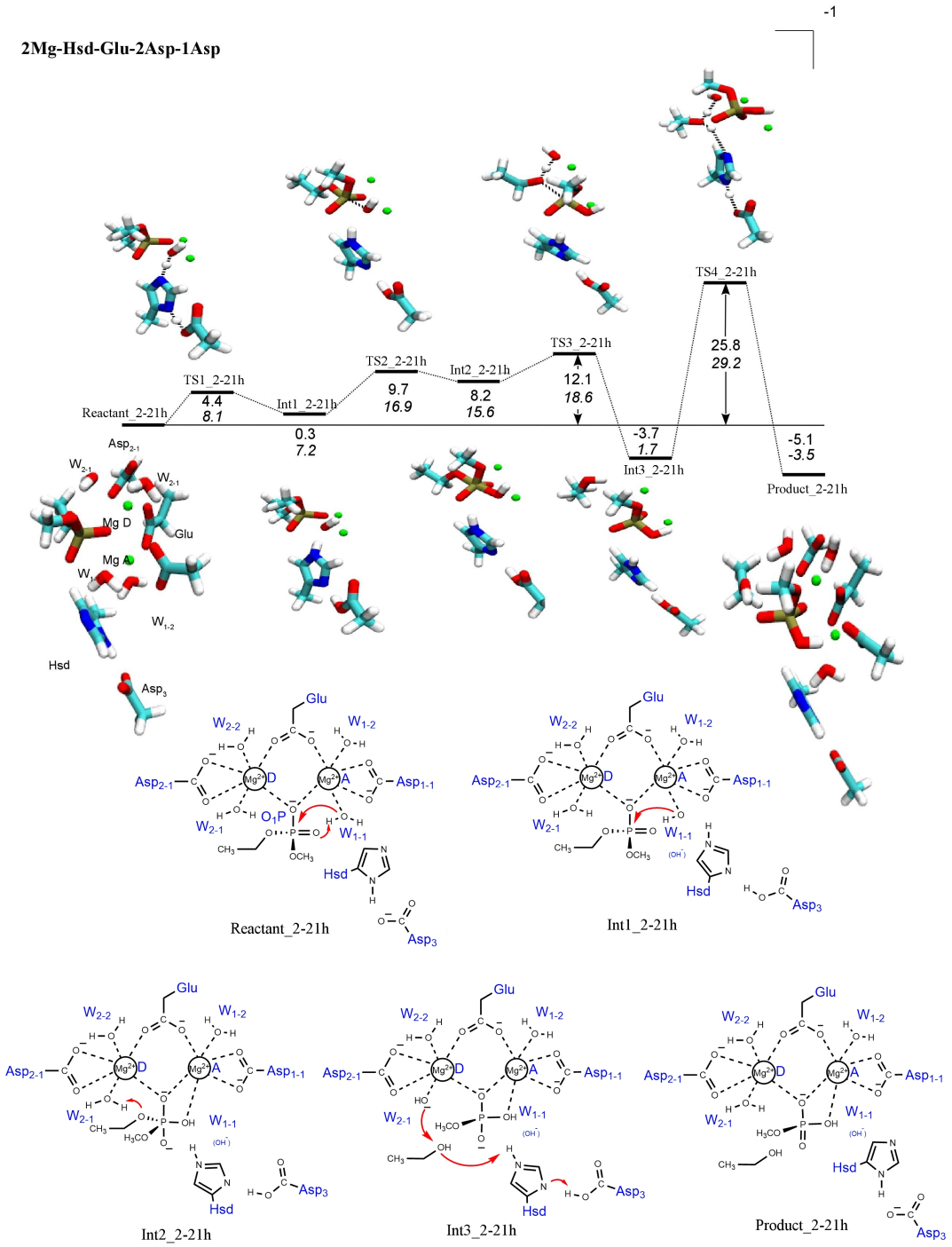

Figure 25: Histidine-assisted mechanism, 2Mg-h, for model with two magnesium ions, neutral histidine, and three aspartate residues (see 2Mg-Hsd-Glu-2Asp_1Asp in Fig. 9) calculated by B3LYP/6-31G** in vacuum (upright numbers) and solvent (in italics). All values are in kcal/mol relative to the reactant state.

Table 6: Energies of stationary points along the phosphate hydrolysis mechanisms computed for the two-metal models (2Mg). All energies are given in kcal/mol, relative to the respective reactant state.

Model	TS1	Int1	TS2	Int2	TS3	Int3	TS4	Product
2Mg-d2 vacuum	26.0	25.1	35.8	29.2	37.6			3.2
2Mg-d2 solvent	26.6	24.4	29.1	29.1	34.3			-3.3
2Mg-d1 vacuum	22.5	17.6	23.7	18.3	32.0			-6.0
2Mg-d1 solvent	22.4	20.3	30.5	23.7	34.9			-2.2
2Mg-d0 vacuum	29.2	21.8	23.9	23.8	32.1			-8.6
2Mg-d0 solvent	29.1	22.5	25.6	22.2	34.1			-0.1
2Mg-h vacuum	4.4	0.3	9.7	8.2	12.1	-3.7	25.8	-5.2
2Mg-h solvent	8.1	7.2	16.9	15.6	18.6	1.7	29.2	-3.5

4.3 Potential energy surface scans

According to Figure 26 there are two intermediates, (structures (I) and (II)), which are favorable in terms of energy. However, in both of them a metal ligated water forms a bond with the phosphate group after donating a proton to the leaving group whereas the nucleophilic water has not attacked. The main differences between structure (I) and (II) is the position of the leaving group which is positioned near the glutamate residue in (II). Moreover, the distances between the phosphate group and the nucleophilic water are also different. These structures are not final products and energy barriers between reactant and these intermediates are more than 40 kcal/mol.

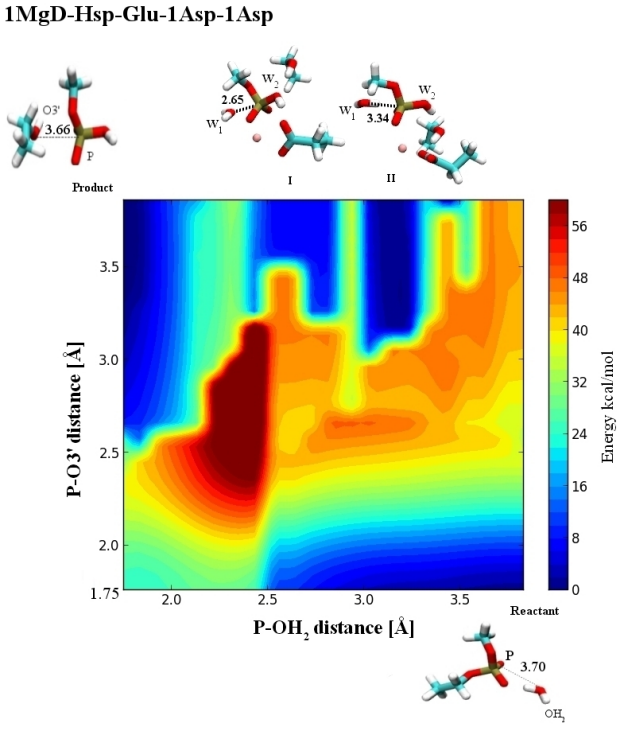

Figure 26: Two-dimensional scan of the potential energy for phosphate hydrolysis in model 1MgD-Hsp-Glu-1Asp-1Asp computed with DFTB. Coordinates are the P–OH_2 and P–O3' distances. Energies are given in kcal/mol.

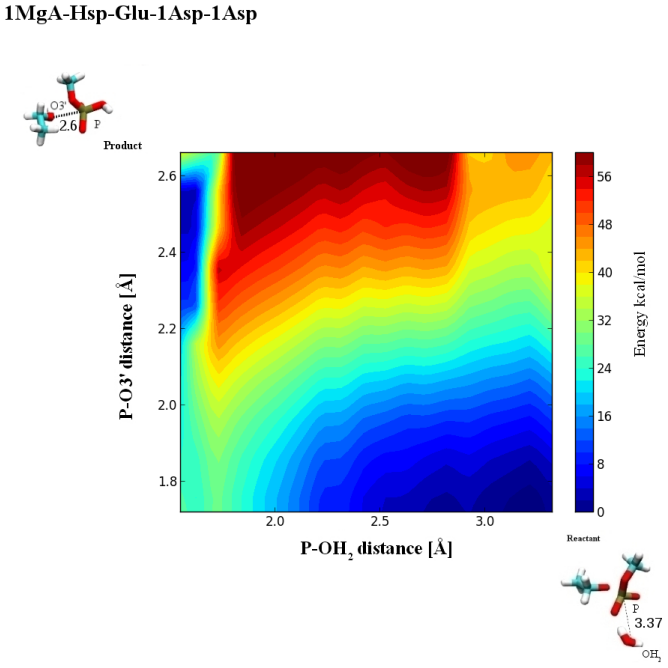

Figure 27: Two-dimensional scan of the potential energy for phosphate hydrolysis in model 1MgA-Hsp-Glu-1Asp-1Asp computed with DFTB. Coordinates are the P–OH_2 and P–O3' distances. Energies are given in kcal/mol.

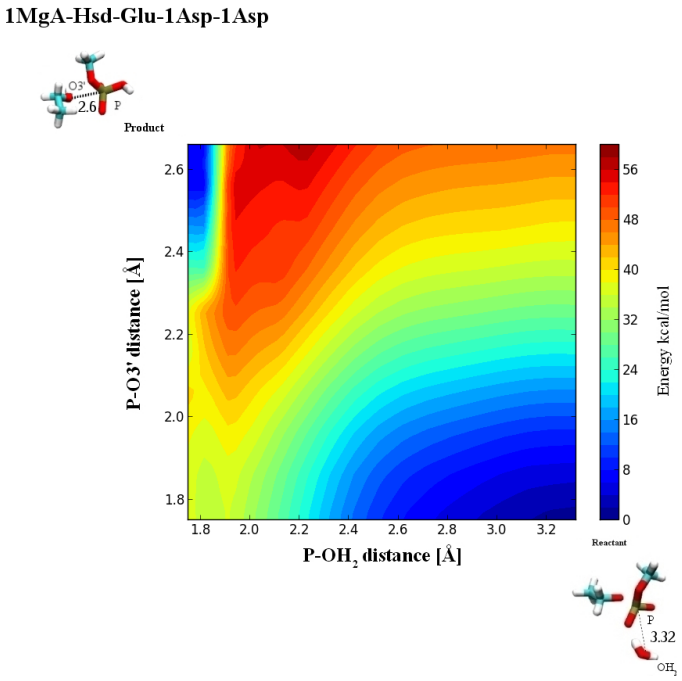

Figure 28: Two-dimensional scan of the potential energy for phosphate hydrolysis in model 1MgA-Hsd-Glu-1Asp-1Asp computed with DFTB. Coordinates are the P–OH_2 and P–O3' distances. Energies are given in kcal/mol.

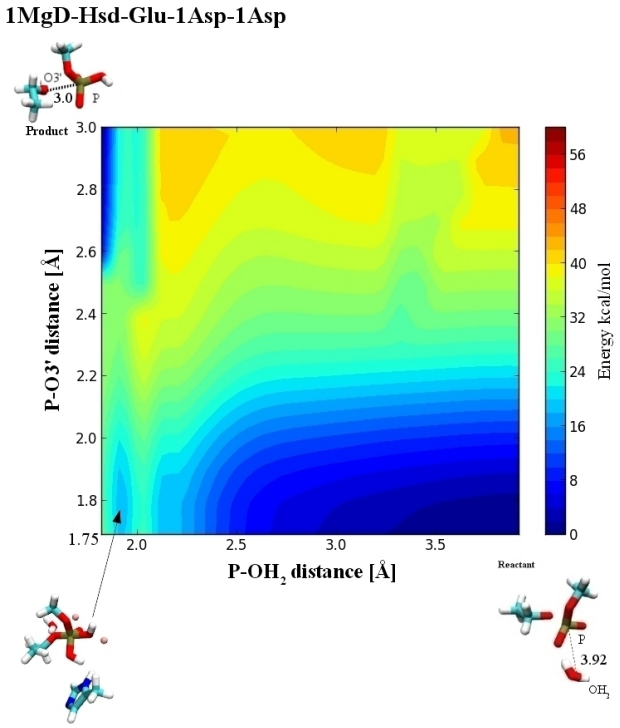

Figure 29: Two-dimensional scan of the potential energy for phosphate hydrolysis in model 1MgD-Hsd-Glu-1Asp-1Asp computed with DFTB. Coordinates are the P–OH_2 and P–O3' distances. Energies are given in kcal/mol.

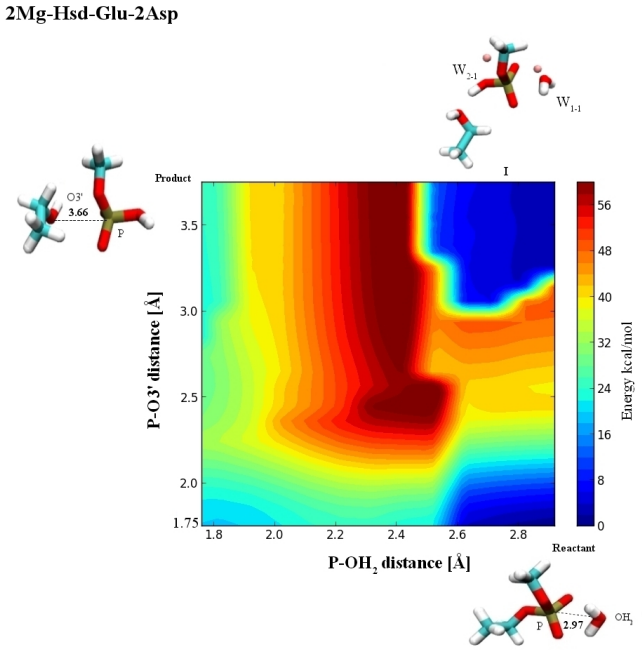

Figure 30: Two-dimensional scan of the potential energy for phosphate hydrolysis in model 2Mg-Hsd-Glu-1Asp-1Asp. Coordinates are the P–OH_2 and P–O3' distances computed with DFTB. Energies are given in kcal/mol.

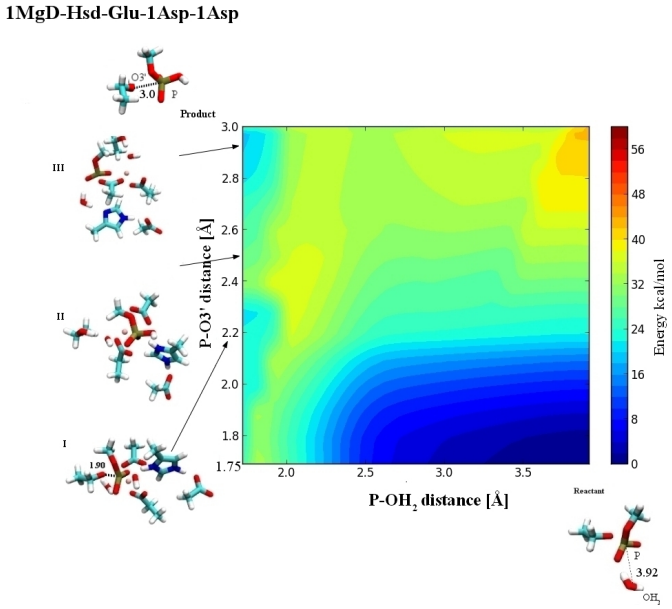

Figure 31: Two-dimensional scan of the potential energy for phosphate hydrolysis in model 1MgD-Hsd-Glu-1Asp-1Asp after proton transfer to the histidine computed with DFTB. Coordinates are the P–OH_2 and P–O3' distances. Energies are given in kcal/mol.

Figure 31 shows the two-dimensional scan of the 1Mg-Hsd-Glu-1Asp-1Asp model. When the distance between the hydroxide ion and the phosphorus atom is about 1.8 Å the hydroxide ion can be considered bound to the phosphorus atom. Incrementing the distance between the O3' atom of leaving group to the phosphorus atom from 1.8 to 3 Å shows two areas with low potential energies. In the first area the P–O3' distance is between 1.87 and 2.36 Å and between 2.77 to 3.0 Å in the

second area. In order to locate optimized minima, structures were picked from these two zones. Minimization of a conformation from the first zone results in a P–O3' distance of 1.90 Å also after optimization with DFTB3/3OB (see pentacovalent structure (I) in Figure 31). This intermediate, however, could not be confirmed by re-optimization on the DFT(B3LYP/6.31G(d,p))-level. Rather, the scissile bond was cleaved during minimization process and no pentacovalent intermediate could not be settled.

In the histidine-assisted mechanism for the 1Mg-Hsd-Glu-1Asp-1Asp model, the nucleophilic water moleculeis activated by transferring a proton to the ND atom of histidine. In the pentacovalent intermediate obtained by DFTB3/3OB, histidine remains in a protonated state and no further proton transfer to the aspartate takes place. The protonated histidine can compensate the developing negative charge on the phosphate group upon nucleophilic attack. In the DFT calculations, by contrast, an aspartate-histidine cluster has formed in which the histidine is not protonated and therefore cannot contribute in the charge stabilization necessary to form a pentacovalent intermediate.

The second structure (II), Figure 31 was chosen with a P–O3' distance of 2.2 Å. Minimization on the DFTB3/3OB level results in an intermediate with cleaved scissile bond similar to the intermediate obtained by DFT calculation. The only difference between the structures obtained by DFT and DFTB3/3OB is the protonation state of the histidine, protonated in the intermediate computed by DFTB3/3OB and un-protonated in the DFT-optimized structure. On both levels of theory, however, a concerted mechanism is possible.

The third intermediate (III), Figure31 has been chosen with a P–O3' distance of 3.0 Å. Minimizing this structure with DFTB3/3OB results in an intermediate with dissociated P–O3' bond before nucleophilic attack, i.e. a dissociative intermediate. The transition state to reach this state, computed with CPR between reactant and Intermediate (III) at the DFTB3/3OB level has an energy of 34.2 kcal/mol. Hence, the barrier is significantly higher than the 25.7 kcal/mol computed for the concerted process.

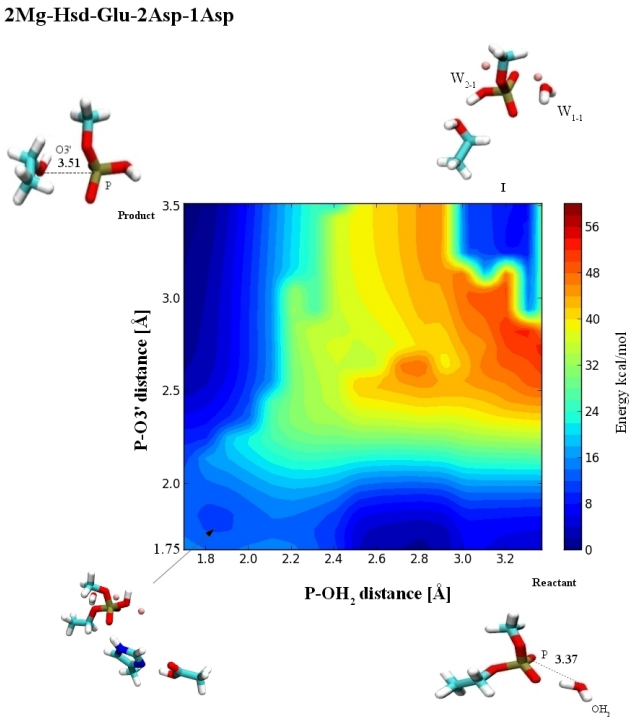

Figure 32: Two-dimensional scan of the potential energy for phosphate hydrolysis in model 2Mg-Hsd-Glu-2Asp-1Asp after proton transfer to the histidine computed with DFTB. Coordinates are the P–OH_2 and P–O3' distances. Energies are given in kcal/mol.

Figure 32 shows the two-dimensional potential energy scan for the 2Mg-Hsd-Glu-2Asp-1Asp after proton transfer from the nucleophile to the histidine. Structure I represents a wrong product in which the metal ligated water has donated a proton to the leaving group but also acted as nucleophile and formed a bond with the phosphate group. The energy barrier to this state in which the nucleophilic water has not attacked (as is the case in the final product) is more than 40 kcal/mol.

4.4 Energies of Stationary points in the different models

4.4.1 One-metal models

Table 7: Energies of stationary points along the phosphate hydrolysis mechanism computed for the 1MgD-Hsp-Glu-2Asp model. All energies are given in kcal/mol, relative to the reactant.

Level	TS1 _1D-p	Int1 _1D-p	TS2 _1D-p	Int2 _1D-p	TS3 _1D-p	Product _1D-p
Vacuum B3LYP/ 6-31G(d,p)	29.1	23.7	35.0	28.0	30 .0	-1.0
corrected by ZPE	27.8	24.5	34.3	28.6	29.3	-0.9
Gibbs free energy at 300 K	31.9	27.4	35.8	31.3	32.3	-1.0
Solvent B3LYP/ 6-31G(d,p)	31.3	26.7	36.2	29.0	31.5	-0.7
Vacuum DFTB	28.9	24.9	29.8	26.5	29.6	-1.7

Table 8: Energies of stationary points along the direct phosphate hydrolysis mechanism computed for the 1MgA-Hsp-Glu-1Asp-1Asp model. All energies are given in kcal/mol, relative to the reactant.

Level	TS1 _1A-p	Int1 _1A-p	TS2 _1A-p	Int2 _1A-p	TS3 _1A-p	Product _1A-p
Vacuum B3LYP/ 6-31G(d,p)	34.9	34.7	34.9	26.6	40.4	7.6
corrected by ZPE	34.6	34.8	34.7	27.5	37.5	7.7
Gibbs free energy at 300 K	35.8	36.1	32.9	28.7	37.4	7.3
Solvent B3LYP/ 6-31G(d,p)	34.8	33.7	34.7	26.7	39.9	7.8
Vacuum DFTB	33.4	33.1	33.1	27.3	38.2	4.5

Table 9: Energies of stationary points along the direct phosphate hydrolysis mechanism computed for the 1MgD-Hsd-Glu-1Asp-1Asp model. All energies are given in kcal/mol, relative to the reactant.

Level	TS1_1D-d	Int1_1D-d
Vacuum B3LYP/ 6-31G(d,p)	35.1	23.1
corrected by ZPE	33.1	23.5
Gibbs free energy at 300 K	36.7	27.2
Solvent B3LYP/ 6-31G(d,p)	39.4	27.7
Vacuum DFTB	36.1	25.9

Table 10: Energies of stationary points along the histidine-assisted phosphate hydrolysis mechanism computed for the 1MgD-Hsd-Glu-1Asp-1Asp model. All energies are given in kcal/mol, relative to the reactant.

Level	TS1_1D-h	Int1_1D-h	TS2_1D-h	Int2_1D-h	TS3_1D-h	Product_1D-h
Vacuum B3LYP/6-31G(d,p)	27.4	-2.61	9.2	-5.2	-2.9	-6.1
corrected by ZPE	24.4	0.0	6.4	-0.1	-1.89	-6.4
Gibbs free energy at 300 K	26.3	-3.1	8.7	-6.8	-3.1	-6.4
Solvent B3LYP/ 6-31G(d,p)	26.2	-1.4	14.2	-4.6	-2.3	-2.6
Vacuum DFTB	25.7	3.5	13.0	-3.7	0.1	-6.6

Table 12: Energies of stationary points along the histidine-assisted phosphate hydrolysis mechanism computed for the 1MgA-Hsd-Glu-1Asp-1Asp model. All energies are given in kcal/mol, relative to the reactant.

								di

Table 11: Energies of stationary points along the direct phosphate phosphate hydrolysis mechanism computed for the 1MgA-Hsd-Glu-1Asp-1Asp model. All energies are given in kcal/mol, relative to the reactant.

Level	TS1 _1A-d	Int1 _1A-d	TS2 _1A-d	Int2 _1A-d	TS3 _1A-d	Product _1A-d
Vacuum B3LYP/6-31G(d,p)	29.6	22.6	33.6	24.7	39.2	-1.2
corrected by ZPE	28.1	22.6	32.4	24.4	37.5	2.4
Gibbs free energy at 300 K	34.3	29.9	40.0	31.8	37.9	5.9
Solvent B3LYP/ 6-31G(d,p)	38.6	28.3	41.1	34.0	39.5	0.2
Vacuum DFTB	33.9	28.6	39.3	36.2	41.2	0.3

Table 13: Energies of stationary points along the phosphate hydrolysis mechanism computed for the 1MgD-Hsd-Glu-1Asp model. All energies are given in kcal/mol, relative to the reactant.

Level	TS1 _1D-d-1	Int1 _1-D-d-1	TS2 _1D-d-1	Product _1D-d-1
Vacuum B3LYP/6-31G(d,p)	36.7	24.2	30.4	-9.3
corrected by ZPE	36.7	24.2	30.0	-9.3
Gibbs free energy at 300 K	34.3	24.6	27.7	-9.4
Solvent B3LYP/ 6-31G(d,p)	37.6	28.5	34.4	-4.3
Vacuum DFTB	36.1	25.9	34.4	-3.3

4.4.2 Two-metal models

Table 14: Energies of stationary points along the phosphate hydrolysis mechanism computed for the 2Mg-Hsd-Glu-3Asp model. All energies are given in kcal/mol, relative to the reactant.

Level	TS1_2-3	Int1_2-3	TS2_2-3	Int2_2-3	TS3_2-3	Product_2-3
Vacuum B3LYP/ 6-31G(d,p)	26.0	25.1	35.8	29.2	37.6	3.2
corrected by ZPE	25.7	25.7	35.7	29.9	36.3	3.3
Gibbs free energy at 300 K	28.1	28.1	38.5	32.3	37.7	1.6
Solvent B3LYP/ 6-31G(d,p)	26.6	24.4	29.1	29.1	34.3	-3.3
Vacuum DFTB	29.4	26.6	34.1	29.7	38.5	0.6

Table 15: Energies of stationary points along the phosphate hydrolysis mechanism computed for the 2Mg-Hsd-Glu-2Asp model. All energies are given in kcal/mol, relative to the reactant.

Level	TS1_2-2	Int1_2-2	TS2_2-2	Int2_2-2	TS3_2-2	Product_2-2
Vacuum B3LYP/ 6-31G(d,p)	22.5	17.6	23.7	20.4	32.0	-6.0
corrected by ZPE	16.6	17.8	23.6	19.1	29.0	-5.9
Gibbs free energy at 300 K	17.9	16.6	24.9	18.4	28.3	-7.6
Solvent B3LYP/ 6-31G(d,p)	22.4	20.3	30.5	23.7	34.9	-2.2
Vacuum DFTB	24.5	19.5	24.2	22.0	27.4	-3.6

Table 16: Energies of stationary points along the phosphate hydrolysis mechanism computed for the 2Mg-Hsd-Glu-1Asp model. All energies are given in kcal/mol, relative to the reactant.

Level	TS1_2-1	Int1_2-1	TS2_2-1	Int2_-1	TS3_2-1	Product_2-1
Vacuum B3LYP/ 6-31G(d,p)	29.2	21.8	23.9	23.8	32.1	-8.6
corrected by ZPE	28.6	22.1	23.2	24.0	30.6	-8.7
Gibbs free energy at 300 K	28.5	22.3	23.7	25.2	30.2	-8.7
Solvent B3LYP/ 6-31G(d,p)	29.1	22.5	25.6	22.2	34.1	-0.1
Vacuum DFTB	30.9	24.9	27.4	23.4	30.7	-6.2

Table 17: Energies of stationary points along the direct phosphate hydrolysis mechanism computed for the 2Mg-Hsd-Glu-2Asp-1Asp model. All energies are given in kcal/mol, relative to the reactant.

Level	TS1_2-21d	Int1_2-21d
Vacuum B3LYP/ 6-31G(d,p)	29.5	18.1
corrected by ZPE	29.2	18.7
Gibbs free energy at 300 K	28.6	19.7
Solvent B3LYP/ 6-31G(d,p)	24.8	19.8
Vacuum DFTB	31.6	17.8

Table 18: Energies of stationary points along the histidine-assisted phosphate hydrolysis mechanism computed for the 2Mg-Hsd-2Asp-1Asp model. All energies are given in kcal/mol, relative to the reactant.

Level	TS1_2-21h	Int1_2-21h	TS2_2-21h	Int2_2-21h	TS3_2-21h	Int3_2-21h	TS4_2-21h	Product_2-21h
Vacuum B3LYP/ 6-31G(d,p)	4.4	0.3	9.7	8.2	12.1	-3.7	25.8	-5.2
corrected by ZPE	0.3	-1	11.1	9.3	11.4	-2	25.7	-3.9
Gibbs free energy at 300 K	2.3	2.2	13.6	10.8	12.7	0.6	30.7	-1.1
Solvent B3LYP/ 6-31G(d,p)	8.1	7.2	16.9	15.6	18.6	1.7	29.2	-3.5
Vacuum DFTB	3.7	3.4	14.2	11.0	19.1	-3.1	21.4	-4.9

4.5 Comparison to other studies of phosphate hydrolysis in cluster models and in enzymes

In all models with two magnesium ions, both MgA and MgD share their positive charge with the phosphate group via the O1P atom. In these models a catalytic effect on the nucleophilic attack, likely through balancing the increasing negative charge, can be observed. Our computed reaction pathways show transferring a proton to the O2P atom of the phosphate group together with attack of the nucleophile to have barriers that are significantly reduced compared to the Gibbs free energy (38 kcal/mol) computed for dimethyl phosphate hydrolysis without any surrounding amino acid residues or metal ions [108]. In that study the energy barrier was dissected into 19.6 kcal/mol for proton transfer from a water molecule to the O2P atom and 18.4 kcal/mol for the attack of the remaining hydroxide ion. The reduced barriers for the direct attack of the nucleophile with concerted proton transfer to the phosphate group in the cluster models is likely a consequence of a facilitated nucleophilic attack due to the close presence of an extra positive charge.

In the histidine-assisted mechanism the attack of the hydroxide ion then has an even lower barrier (only 9.7 kcal/mol in vacuum and 16.7 kcal/mol in solvent). Also in the one-metal models, histidine-assisted nucleophilic attack shows barriers that are significantly lower than those computed for direct attack. In the one-metal model, however, proton transfer for nucleophile generation, nucleophilic attack and leaving group dissociation take place in a concerted step without a pentacovalent intermediate. The barrier for this step ($\sim$27 kcal/mol) is significantly higher than the nucleophilic attack in the two-metal mechanism ($\sim$10 kcal/mol in vacuum and $\sim$17 kcal/mol in solvent, respectively), suggesting a less stabilized transition state with only one metal ion. According to the natural bond (NBO) analysis, the negative charge on the phosphate group increases in the pentacovalent structure to such an extend (see Figure 33) that the scissile bond is cleaved without any further transition state in the one-metal model with the magnesium ion located at the departure site. In the one-metal model with the magnesium located at the attack site, however, the pentacovalent intermediate is stabilized and the leaving group departure appears not to be facilitated and does not occur spontaneously (see Figure 33). In the two-metal model, the negative charge is somewhat compensated by the second metal ion, resulting in a stable pentacovalent intermediate and another transition state for P–O3' bond dissociation with an energy of 12.1 kcal/mol. Calculations of the two-metal histidine-assisted mechanism with a solvent model show an increased barrier for proton transfer to the histidine-aspartate cluster and an intermediate that is about 7 kcal/mol higher in energy than in vacuum. The following transition states and intermediates for nucleophilic attack and leaving group departure are similarly higher in solution than in vacuum, with the leaving group departure showing an energy barrier of 18.6 kcal/mol in solvent. Phosphate hydrolysis by the attack of a hydroxide ion and the difficulty to calculate this step and hydroxide generation accurately has been discussed in the literature where different reaction rates and types of mechanisms, step-wise or concerted fashion, have been proposed [144, 97, 49, 12]. The higher energies calculated for the transition states and intermediates of hydroxide generation by proton transfer to the histidine and subsequent attack with the hydroxide ion as the nucleophile computed in implicit solvent are an indicator for an artifact in the vacuum model: Neutralization of the aspartate residue that is located at the outer part of the cluster model is, in vacuum, preferable than maintaining a neutral water molecule. As suggested by the higher energies in implicit solvent, charge screening effects counteract this preference. The situation may yet be different in explicit solvent or a heterogeneous protein environment with many other charged residues. The present cluster models likely underestimate the barriers for the hydroxide generation but arguably do this in a similar fashion for all mechanisms calculated. The histidine-assisted generation of a hydroxide ion and the subsequent phosphate hydrolysis by attack of this nucleophile is therefore the most probable reaction pathway computed for the cluster models.

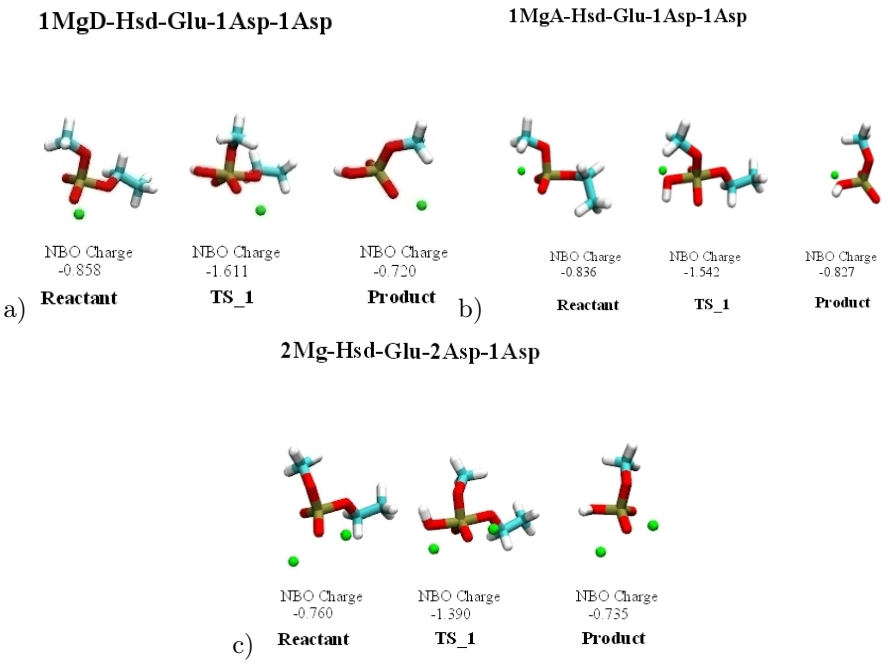

Figure 33: NBO charge (in a.u.) on the phosphate group at different steps of the histidine-assisted reaction pathway in a) one-metal model 1MgD-h, b) one-metal model 1MgA-h, and c) two-metal model 2Mg-Hsd-Glu-2Asp-1Asp .

In the histidine-assisted mechanisms, both for the one-metal and the two-metal model, the leaving group receives a proton from a metal-ligated water molecule due to the lack of a proton on the O2P atom. The barrier for the re-protonation, i.e. proton transfer from the histidine-aspartate cluster via the leaving group to that water molecule, is calculated to be the highest barrier in the histidine-assisted pathway of the two-metal model (∼25 or ∼29 kcal/mol in vacuum and solvent, respectively). It is however, conceivable that such re-protonation steps occur via additional water molecules and with much reduced barriers in a real solvent or in an enzymatic environment. Indeed, many previous mechanistic studies on cleavage of the phosphodiester employing QM/MM methods, and hence taking into account the protein environment, re-protonation steps are computed with lower barriers than the rate-determining nucleophilic attack or leaving group departure steps [30, 94, 108, 100, 34]. Disregarding the reprotonation step, the remaining barrier of ∼19 kcal/mol for the energetically most favorable mechanism is comparable to rates of seconds to minutes as reported for several DNA cleaving enzymes [157, 62, 73, 52, 45]. It is significantly lower than uncatalyzed phosphate hydrolysis[78] and also lower than the phosphate hydrolysis by the direct attack of a hydroxide ion on a phospho-diester [11]. Our models, moreover, have a considerable resemblance with the nuclease domain (Nuc) of 3´hExo complexed with rAMP: two magnesium ions are coordinated by three aspartate residues and one glutamide residue together with one neutral histidine residue which is well located for accepting a proton from a hydrolytic water molecule[174]. This combination of amino acids and metal ligated water molecules offer a catalytic platform for hydrolytic cleavage of bound RNA [174]. Escherichia coli exonuclease III (ExoIII)-like apurinic/apyrimidinic (AP) endonuclease is discussed to start the reaction with the nucleophilic attack of an activated water molecule to phosphate group [94](cf. Figure 13). In a proposed mechanism [30], an aspartate residue positioned close to a histidine aids in stabilizing the development of positive charge on the histidine after proton abstraction from a water molecule. A histidine-assisted mechanism has also been proposed for endonucleolytic DNA cleavage by the homing endonuclease I-PpoI [45]. For that enzyme, however, a single magnesium ion is suggested to activate a water for donation of a proton to the O3' atom of the leaving group. Serratia endonuclease is a magnesium dependent nuclease that is similar to I-PopI in its active site, containing a conserved histidine, likely playing the same role for activating the nucleophilic water [100]. In addition, glutamide and asparagine residues are present for binding the metal ion in both enzymes.

5 The reaction-competent APE-DNA complex

Parts of this chapter are published in "*Batebi, H., Dragelj, J. and Imhof, P. Role of AP-endonuclease (Ape1) active site residues in stabilization of the reactant enzyme-DNA comples. Proteins (2018)*, DOI"

One putative role of the metal cofactor in the phospho-diester hydrolysis reaction of DNA backbone cleavage is positioning a nucleophilic water molecule. Enzymatic phospho-diester hydrolysis by APE has been found to have a bell-shaped pH dependence, suggesting at least two titratable residues involved in the mechanism, one acting as a general base the other as general acid [130]. As a Lewis acid, the metal ion could also activate a coordinated water molecule to transfer a proton to a nearby (general) base, enabling nucleophilic attack as a hydroxide ion. Leaving group departure could likewise be facilitated by direct metal coordination of a negatively charged leaving group or by proton transfer from a metal-bound water molecule or a general acid. Another possible role of the metal ion is the stabilization of the (associative) transition state by balancing the accumulating negative charge. Other amino acid residues can, however, support the metal ions, by proper positioning of the metal ion itself, the water molecules involved, and/or assuming the role of general acid/base [79, 110, 160, 168, 130, 99].

Motivated by the mechanistic ambiguity that arose from the lack of Ape1 active-site structural detail, several structures of Ape1 were published in 2015 [168]. These high-resolution crystal structures include Ape1 bound to the abasic phosphorothioate substrate without metal-ion (PDB ID: 5DFI), an inactive phosphorothioate substrate complex with Mn^{2+} (PDB ID: 5DG0) and a product complex (PDB ID: 5DFF) which diffracted to 1.63 Å, 1.80 Å, and 1.57 Å, respectively.

In this section, the position and number of the metal ion(s) as well as important active-site residues in the substrate complex of Ape1-DNA are analyzed in detail so as to investigate the possible impact of metal ion binding on the stabilization of the active site in the Ape1-DNA complex in a reaction-competent conformation and hence enzymatic activity. To this end, we have performed molecular dynamics simulations of the Ape1-DNA substrate complex with one or two magnesium ions, located at different sites in the active site as motivated by the different available crystal structures. Among all catalytically important amino acid residues in the active site of the Ape1 enzyme, His309 and Asp210 are most likely to play the role of a general base in adopting a proton from a well-positioned water molecule to generate a hydroxyl ion for direct nucleophilic attack to the phosphate group of the abasic (AP) site[79, 168, 138, 31]. Temperature and pH-dependent NMR experiments on ^{15}N-labeled APE, wild type and the D210N mutant, suggest either a high (>8) pKa value of H309, corresponding to this residue being charged [40]. These experiments have, however, been carried out without the abasic DNA substrate bound, leaving the possibility of a different protonation state for His309 in the complex. To our knowledge, the pKa of Asp210 has not been directly assessed by experiments. Hence, complementary pKa calculations have been performed so as to estimate the protonation state of His309 and Asp210 in the active-site of the Ape1-DNA substrate complex.

5.1 pKa Calculations

We evaluated pKa values in the DNA-protein complex by electrostatic energy computations using Karlsberg2+[17, 53, 155] for the crystal structure as well as for some frames (every 1 ns) from the MD simulations [88], as done in previous applications on other protein systems [13, 14]. Our approach of combining MD simulations in a fixed protonation state with subsequent pKa calculations to estimate the most probable protonation pattern has been favored over more sophisticated techniques such as constant pH MD (cpH-MD) simulations [70, 166, 164, 36, 137, 35, 175] or free energy perturbation simulations with mixed quantum mechanical/molecular mechanical potentials [54, 176] because of its computational efficiency. The MD simulations in the different, fixed protonation states allow to explore the conformations favored by the respective protonation states. The subsequent electrostatic energy computations then provide the pKa values for the different conformations. In all pKa computations, only Tyr171, His309, Asp283, Asp70, Asp210, Asp308 and Glu96 were titrated as they are in proximity to the Mg^{2+}-ions. All other residues had a fixed protonation state, as used in the corresponding MD simulations. Protein, DNA, and metal ions were

described by atomic point charges. These were taken from the CHARMM force field [83], employing the same parameters for the pKa calculations as for the MD simulations. For the electrostatic energy computations, bulk water molecules were removed and treated as a dielectric continuum with a dielectric constant of ε=80, as explained in earlier applications on protein-water systems, based on the solution of the linearized Poisson-Boltzmann equation [59, 136]. In the protein-DNA volume the dielectric constant was set to 4 [53]. This choice of dielectric constant is a compromise allowing proper treatment of polarization and small backbone fluctuations, which, if treated explicitly, would lower the dielectric constant even more [104, 167, 136, 139, 105, 145], providing good agreement with experiments [136, 112, 59]. The electrostatic energies were determined by numerically solving the linearized Poisson−Boltzmann equation using the program APBS [112] To obtain reliable results, the finest grid spacing was 0.3 Å. Water molecules that are coordinated to the Mg^{2+}-ions do not exchange with bulk water molecules during the entire simulation time. These conserved water molecules were identified for each MD simulation separately and were kept as explicit molecules, modeled as TIP3P [172] to be consistent with the charge models used in the MD simulations, in all pKa computations and therefore were considered part of the protein-DNA volume. A similar approach regarding functional or coordinated water molecules during the MD simulations has been utilized before in different systems [58, 14, 104, 167, 136, 17, 126, 125, 50, 139, 105, 145]. Whereas it is rare to keep explicit water molecules in pKa computations, the choice of solvent model for explicitly considered water molecules in pKa calculations has been discussed in the literature [33, 84]. However, even though one water model could have an advantage over another in certain aspects (e.g. diffusion constant, solvation) [121, 117], there is no absolute favorite choice. For example, the self-diffusion constant of the TIP3P water model actually gives reasonable values at short cut-offs [81]. This is relevant in this work, as we consider only few water molecules that are coordinated to Mg^{2+} - ions and one pKa computation with a single water molecule in the crystal structure.

In order to properly treat the presence of internal, non-coordinated mobile water molecules in all computations, an advanced internal cavity search was used as an additional feature of the Karlsberg2+ software [156] with a cavity parameter c=0.8. Except for explicitly kept water molecules mentioned earlier, all internal water molecules were removed in the pKa calculations and the resulting cavities were considered as solvent volumes with high dielectric constant (ε=80) as has been proven earlier to give good agreement with experimental results [112].Water penetration is also taken into account in this approach by utilization of the mentioned cavity algorithm, which then allows a lower dielectric for the protein-DNA volume (ε=4). For comparison, we also computed pKa values with other dielectric constants (ε=3 and ε=6).

For the crystal structure, we performed the pKa calculation in two ways regarding the treatment of internal water molecules: The first approach follows the one employed for the snapshots of the MD calculations, that is, all but the metal-coordinated water molecules were removed and represented by cavities with high dielectric constant. In the second approach, one more water molecule, that is resolved in the crystal structure as located between Asp210 and the phosphate groups (the closest O-O distance of the two groups is less than 5 Å) was treated explicitly. Such an approach has been used previously on systems with functional, well-localized water molecules [58].

5.2 Magnesium coordination

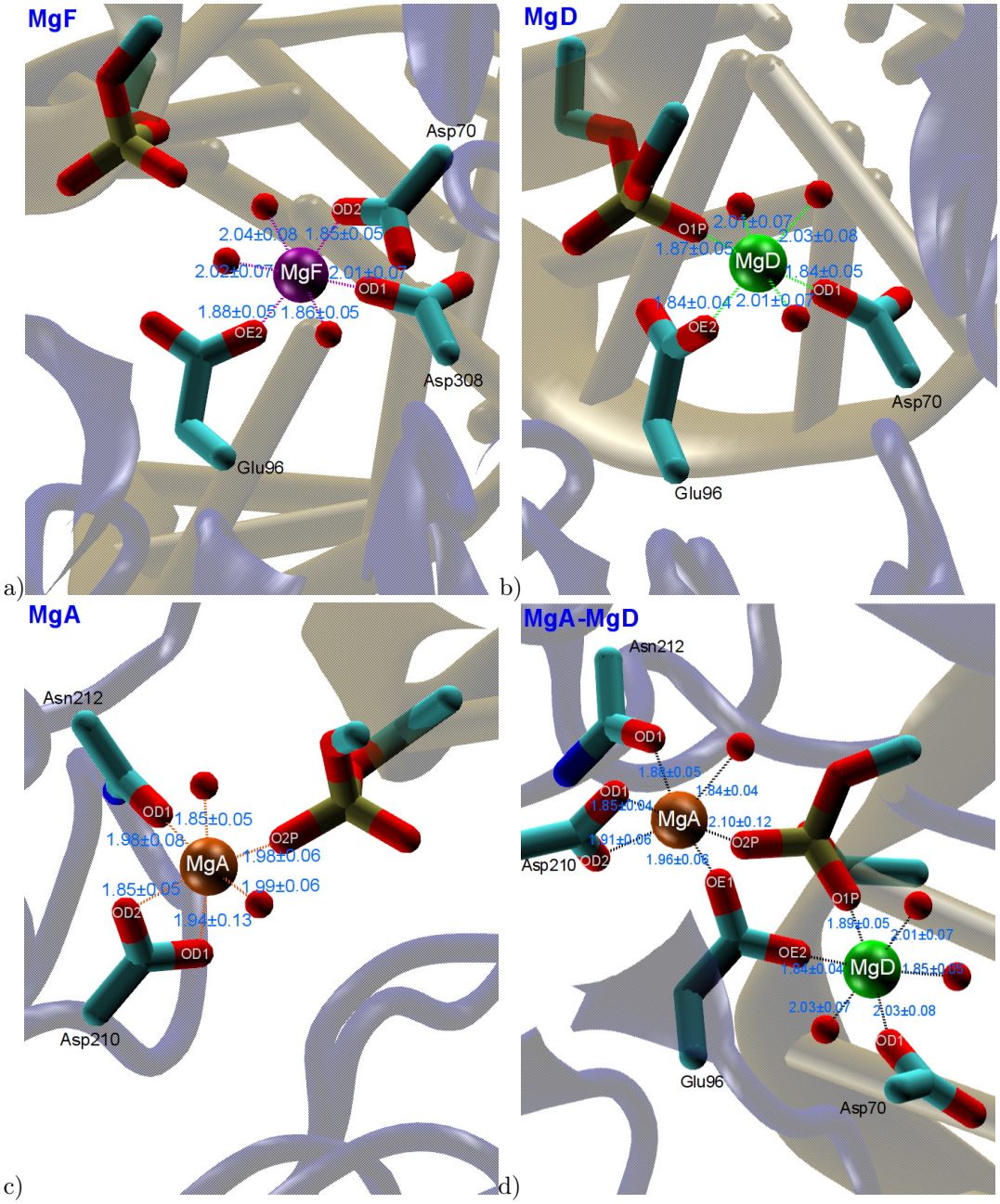

Figure 34: Metal-ion coordination in the active site of the Ape1-DNA substrate complex with a) one magnesium ion at metal binding site F, b) one magnesium ion at metal binding site D, c) one magnesium ion at metal binding site A, d) two magnesium ions at site A and D.

Figure 34 shows the coordination sphere of the Mg^{2+}-ions at the different binding sites F, D, or A, respectively, together with the average distances to their ligands as calculated from the MD simulation trajectories. In all models, the Mg^{2+}-ions remain at their respective positions at different metal binding sites, F, D, or A, respectively, (see Figure 34), and the fluctuations of the active sites, measured as root mean square deviation, are in general only low to moderate (Figures 69 to 72).

Table 19: Average pKa values calculated from snapshots of the different Ape1-DNA complex models.

Model	Asp210	Asp283	His309
Hsp-MgF	3.30 ± 1.05	-6.03 ± 0.99	15.90 ± 1.26
Hsd-MgF	3.94 ± 0.80	-4.16 ± 1.45	10.19 ± 0.80
Hse-MgF	5.08 ± 1.80	-1.30 ± 1.97	6.83 ± 1.86
Hsp-MgD	3.69 ± 1.10	-5.45 ± 1.10	13.34 ± 0.87
Hsd-MgD	3.57 ± 0.98	-2.53 ± 1.90	9.78 ± 0.86
Hse-MgD	5.66 ± 2.48	-2.21 ± 1.18	9.77 ± 1.51
Hsp-MgA	< -10	-3.87 ± 1.14	13.21 ± 0.65
Hsd-MgA	< -10	-3.88 ± 1.11	10.71 ± 0.49
Hsp-MgA-MgD	< -10	-2.86 ± 2.14	11.05 ± 1.13
Hsd-MgA-MgD	< -10	-2.76 ± 2.17	11.73 ± 1.36
Hse-MgA-MgD	< -10	-3.52 ± 2.05	3.80 ± 2.25

5.3 Protonation states

According to the pKa calculations, significant variations in the calculated pKa values, depending on the model are observed only for Asp210 and His309. The computed pKa values are listed in Table 19. Note that pKa values computed for other dielectric constants (ε=3 and ε=6) give qualitatively the same results as those obtained with ε=4.

The pKa calculations on all structures sampled from MD simulations with different positions of Mg^{2+} and protonated His309 suggest protonated His309 with pKa=15.90±1.26, pKa=13.34±0.87, and pKa=13.21±0.65 for Hsp-MgF, Hsp-MgD, Hsp-MgA, respectively, to be the most likely protonation state of this residue (Table 19). This is reasonable since those MD simulations have been carried out with protonated His309 and thus sample conformations that are preferable for Hsp309.

Interestingly, the analysis on structures with two Mg^{2+}-ions, bound to site A and D, respectively, shows His309 pKa values of 11.05±1.13, 11.73±1.36, 3.80±2.25 for Hsp-MgA-MgD, Hsd-MgA-MgD, Hse-MgA-MgD respectively (Table 19). Hence, even with two metal ions, protonated His309 appears to be conceivable. The rather low pKa value computed for Hse-MgA-MgD, however, favors the presence of neutral His309 in this model. All one-metal models simulated with neutral His309 sample conformations which, according to the calculated pKa values, see Table 19, indicate protonated His309 to be favorable or at least probable (pKa=6.83±1.86 for Hse-MgF).

In models that have a metal ion located at site A (single metal models MgA, or the two–metal models MgA-MgD) i.e. in which Asp210 is a bidentate ligand to a Mg^{2+}-ion, Asp210 has a negative pKa value (<-10), i.e. it is highly unlikely to be protonated. The pKa results for the other models exhibit Asp210 close to its standard pKa value in water (4.0 [178]) indicating unprotonated Asp210 to be most probable in these models. In the crystal structure, the pKa value computed for Asp210 with the same implicit water approach is 11.99 representing a protonated Asp210. However, when taking into account the water molecule that is bound between the phosphate group and Asp210 explicitly, in order to allow specific hydrogen bonding with these groups), the pKa calculation performed for the crystal structure points towards an unprotonated Asp210 (pKa=4.63).

Our simulations on protonated Asp210 do not reveal any notable conformational changes compared to the crystal structure of the Ape1-DNA substrate complex (PDB ID: 5DG0). In particular models with the metal ion located at site F and protonated Asp210 are highly similar to the crystal structure of the inactive substrate complex (see Figure 35).

5.3.1 Histidine 309

According to previous studies, His309 is essential for substrate binding and the phosphate hydrolysis step [40, 160, 15]. Hence, it is quite reasonable to assume that a hydrogen-bond interaction between a non-bridging oxygen atom of the AP-site O1P/O2P and His309Nϵ has a significant impact on substrate binding. No such direct hydrogen bonds are observed in the simulations. However, our distance analysis of the Ape1-DNA substrate bound to a single Mg^{2+}-ion at either of the binding sites (F, D, or A) and of the two-metal models show protonated His309 (Hsp) in a distance

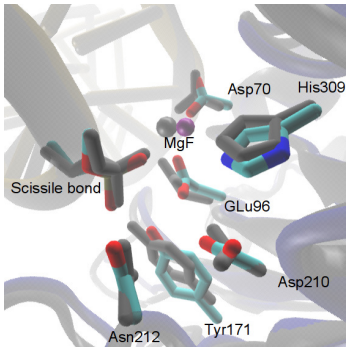

Figure 35: Comparison of the crystal structure of the Ape1-DNA substrate complex (in gray) and a representative snapshot of the MD simulation in a MgF model with protonated Asp210.

to the phosphate group that corresponds to a value of a hydrogen bond to His309's Nε-atom, mediated by a water molecule. In some models (see Table 20), the Nδ-site of His309 is also in a hydrogen-bond distance to Asp283. Furthermore, except for the MgF models, Hsp309 remains close enough to the phosphate group to allow interaction mediated by one water molecule. In contrast, in the simulations of models with Hse309, the histidine residue undergoes a significant conformational change in which it rotates away from the phosphate group and towards Asp283 (see Figures 77 and 80). In the Hse-MgF model, the histidine rotates even further towards Asp308, thereby allowing the formation of a strong hydrogen bond (Table 20)to that residue far on the departure site. This rotation is observed already in the minimization for Hse-MgA and in the equilibration phase for Hse-MgD and Hse-MgF, respectively. Except for the models with Mg^{2+} bound to binding site F, the metal ions are coordinated by a non-bridging oxygen atom of the phosphate group and protein residues, stabilizing the enzyme-DNA binding.

As manifested by the root mean square deviations (RMSD) calculated in the course of the MD simulations, models Hsd-MgF and Hse-MgF experience significant conformational changes in the DNA structure. The RMSD calculated for DNA increases to 5.53 Å and 7.87 Å respectively (see Figures 69 to 72), indicating a loss of stabilization. This suggests a requirement for His309 to be protonated if the Mg^{2+} ion is located at binding site F. Indeed, in the corresponding model, Hsp-MgF, His309 stays close to the scissile phosphate group of the abasic DNA (see also Figure 80).

For models with the Mg^{2+}-ion binding to site D or A, the protonation state of His309 does not have a direct effect on the calculated RMSD of DNA or protein (Figures 69 to 72), in contrast to its impact on the distance of His309 to the phosphate group (see above and Figures 77 and 80). This is likely due to the fact that a Mg^{2+}-ion at binding site D or A is coordinated by the non-bridging oxygen of the AP-site as well as by the protein residues (Asp70 and Glu96 or Asp210 and Asn212, respectively), thus "gluing" together the negatively charged phosphate group and the negatively charged side chains of the aspartate and glutamate residues. According to the pKa calculations, His309 is most likely protonated, except for models Hse-Mg and Hse-MgA-MgD. The large distance of Hse to the AP-site, however, renders all Hse models unlikely. We therefore discussed only Hsp models in the following.

5.4 Hydrogen-Bond interactions

Only model Hsp-MgD exhibits a strong hydrogen bond between a protein residue, Tyr171 and the phosphate group. In the models with a Mg^{2+}-ion at site D, Tyr171 also forms hydrogen bonds with water molecules. Hydrogen bonds between a putative nucleophilic water molecule and the scissile phosphate group are observed almost throughout the entire simulation time in all single-metal models, in the case of MgF to both non-bridging oxygen atoms of the phosphate group. In contrast, in the two-metal models no such strong hydrogen bonds are found. Also for models with a magnesium ion located at the attack site, MgA or MgD-MgA, strong hydrogen bonds between His309 and water molecules are observed. No such bond is found in the MgF models. In all models with only one metal ion, Glu96 forms hydrogen bonds with water. In the two-metal models, this

Table 20: Hydrogen bond occupancies for active-site residues in the Hsp-models. AP is the abasic site. Only values with >60% in at least two of the three individual MD runs are listed. 100% occupancies are reached by several water molecules, one or two at a time, forming a hydrogen bond to the respective residue.

	MgF	MgD	MgA	MgA-MgD
His309–Asp283		95.8±1.5	89.2±4.8	
Tyr171–AP-O2P		86.9±2.5		
Asp70--wat	100.0±0.0	84.9±26.2	100.0±0.0	100.0±0.0
Glu96--wat	100.0±0.0	72.8±20.6	100.0±0.0	
Tyr171--wat		83.5±14.3		86.0±19.1
Asp210--wat	95.9±7.2	100.0±0.0		
Asn212--wat		100.0±0.0		
His309–wat		82.4±8.1	85.9±1.3	80.9±16.5
Asp283–wat				76.0±41.6
AP-O1P--wat	100.0±0.0		100.0±0.0	
AP-O2P--wat	89.7±17.8	100.0±0.0		

is not possible due to the position of Glu96 bridging the two Mg^{2+}-ions. For a similar reason, no hydrogen bonds are observed between Asp210 and water molecules in the models with a Mg^{2+}ion located at site A, in which Asp210 is a bidentate ligand to the metal MgA. The same is observed for Asn212 that is also ligating a metal ion in the models with MgA. In fact, only model Hsp-MgD exhibits hydrogen bonds between Asn212 and water molecules.

5.4.1 Tyrosine 171

In terms of substrate binding, Tyr171 has the possibility to form a hydrogen bond with the non-bridging oxygen atom of the AP-site(O2P)[168, 99]. However, our distance analysis shows that depending on the Mg^{2+}-ion positions, Tyr171 may lose contact of this residue with the scissile phosphate group and instead interact with the side chain of Glu96(OE2) or the O3' atom of the scissile phosphodiester bond (see Figure 78). According to our simulations, hydrogen bonds between Tyr171 and the non-bridging oxygen atom (O2P) of the AP-site or short distances to the O3' atom of the leaving group are frequently observed in model MgD-Hsp whereas model MgF-Hsp exhibits shorter distances between Tyr171 and Glu96 (see Figure78). Models with a single Mg^{2+}-ion bound to site A, show interactions between Tyr171 and Glu96(OE2), in addition to interactions of Tyr171 with the phosphate group. In the Ape1-DNA complex bound to two divalent metal ions, Tyr171 is in short distance to the O3' atom of the leaving group (Figure 78).

5.4.2 Asparagine 212

Our distance analysis shows that Asn212 in models with a single Mg^{2+}-ion bound to site A and in models with two Mg^{2+}-ions at site A and D, respectively, is in a short distance to the AP site that can keep the contact with abasic DNA (see Figure 79). Model Hsp-MgF shows a broad distribution of distances between Asn212 and the phosphate group, rendering a catalytic role, such as DNA binding or keeping a water molecule in a proper position for nucleophilic attack unlikely for Asp210 in this model. In contrast, in model Hsp-MgD there is a very strong hydrogen bond between Asn212 and water molecules, suggesting that in this case, Asn212 can bind and position a nucleophilic water molecule. The distance distribution for the Hsp-MgD model agrees with a water molecule positioned between Asn212 and the phosphate group and indeed such water positions are frequently observed (Figure 37).

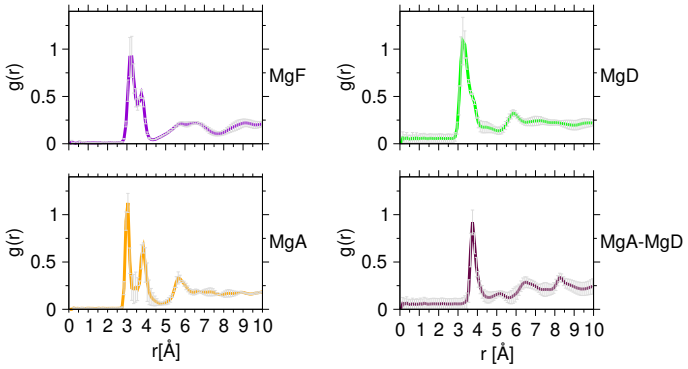

Figure 36: Radial distribution function, g(r), of water oxygen atoms around the CG atom of Asp210 in models Hsp-MgF (purple), Hsp-MgD (green), Hsp-MgA (orange), and Hsp-MgA-MgD (brown).

Simulations with substitution of Asn212 by Alanine show a Mg^{2+}-ion coordination comparable to that of the wild type. In the MgD and MgF models this is of little surprise since these metal ions are coordinated by Asp70, Glu96, the AP-site(O1P), and three water molecules, but not by Asn212. In models with a metal ion at site A, however, the non-coordinating Ala212 is replaced by a metal-ligated water molecule. Another metal-bound water molecule is well-positioned to serve as a nucleophile, albeit held by the metal ion rather than by Asn212. Models with MgD or MgF, in contrast, do not exhibit water molecules at positions suitable for nucleophilic attack (see Figures 37 and 89).

5.4.3 Aspartate 210

All models with a metal ion located at site A, single metal models MgA or two-metal models MgA-MgD, show a very narrow distribution of the distances between Asp210 and the scissile phosphate group (see Figure 80). This is due to the stable bidentate coordination of the MgA metal ion by Asp210. The ~ 6 Å distance observed in models Hsp-MgD and Hsp-MgF model (Figure 80), on the other hand, is compliant with phosphate hydrolysis by a water molecule that is placed in a distance of 3.0 to 5.0 Å between Asp210 and the phosphate group. Indeed, such a putative nucleophilic water molecule is observed as indicated by the radial distribution functions of water molecules around Asp210 (see Figure 36). Models Hsp-MgD and Hsp-MgF show a large probability of a water molecule in 3 to 4 Å from Asp210. In the MgA models, this narrows to two distinct water positions at 3 and $\sim$3.8 Å, the latter of which is the predominant distance in the two-metal models. Similar distributions are observed around Asn212 (see Figure 82), suggesting that both residues play a role in positioning a water molecule. Figure 37 confirms a putative nucleophilic water molecule located between Asp210 and Asn212 for all single-metal models.

Our simulations of the Ape1-DNA mutant complex in which Asp210 has been replaced by Asparagine (D210N) show a coordination sphere of Mg^{2+} that is unaltered after mutation with respect to the position of the metal ion and the ligands surrounding it (see Figure 88). Moreover, throughout the 100ns simulation a water molecule is observed in a position poised for direct attack in models with protonated His309 (Hsp), see Figures 88, and in those models with neutral His309 and a metal ion at position A (single Mg^{2+} ion or two Mg^{2+} ions at sites A and D, respectively).

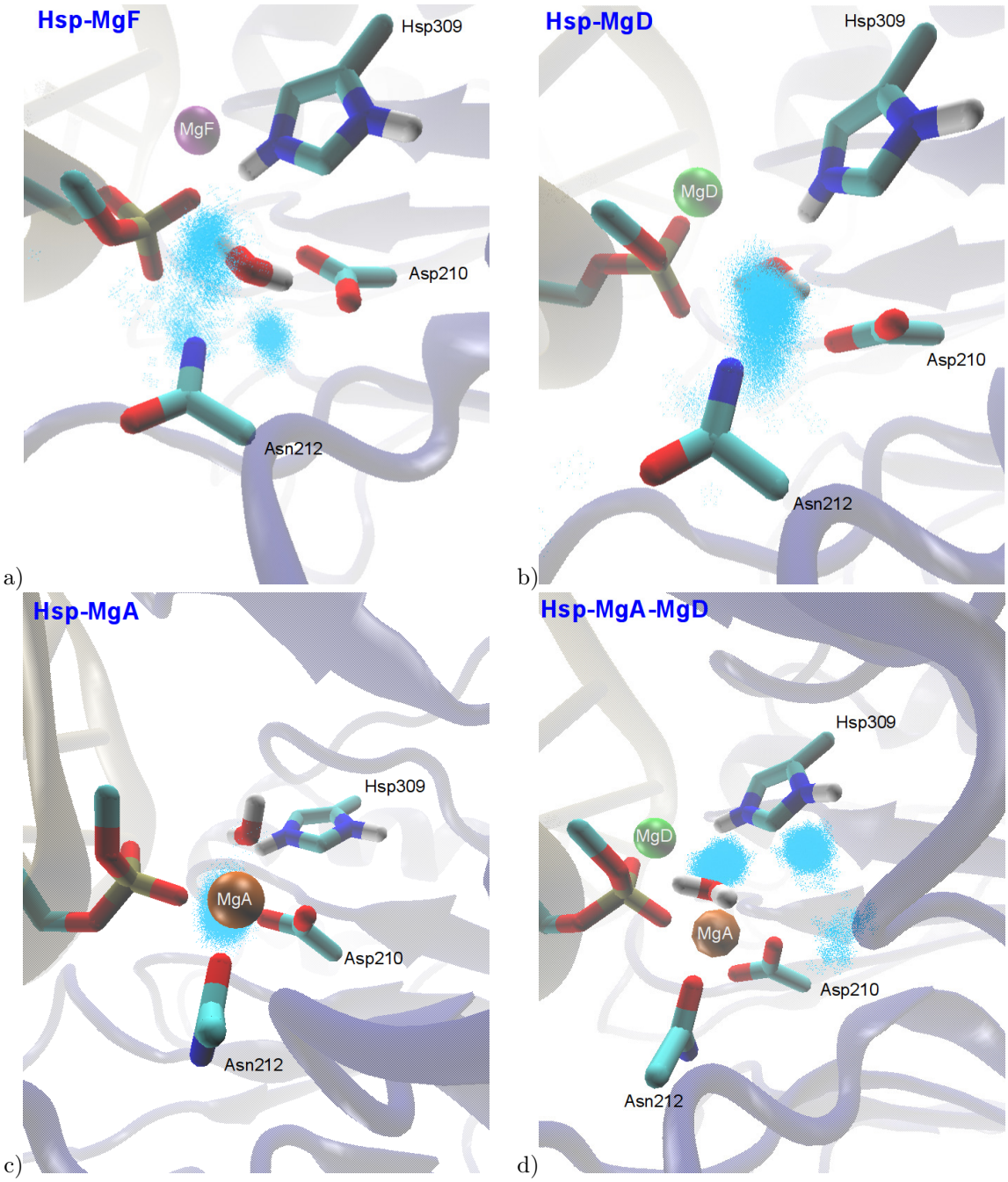

Figure 37: Water oxygen atom positions around Asp210 as shown by cyan dots. For clarity only positions from the last 40 ns of the simulation time are shown.

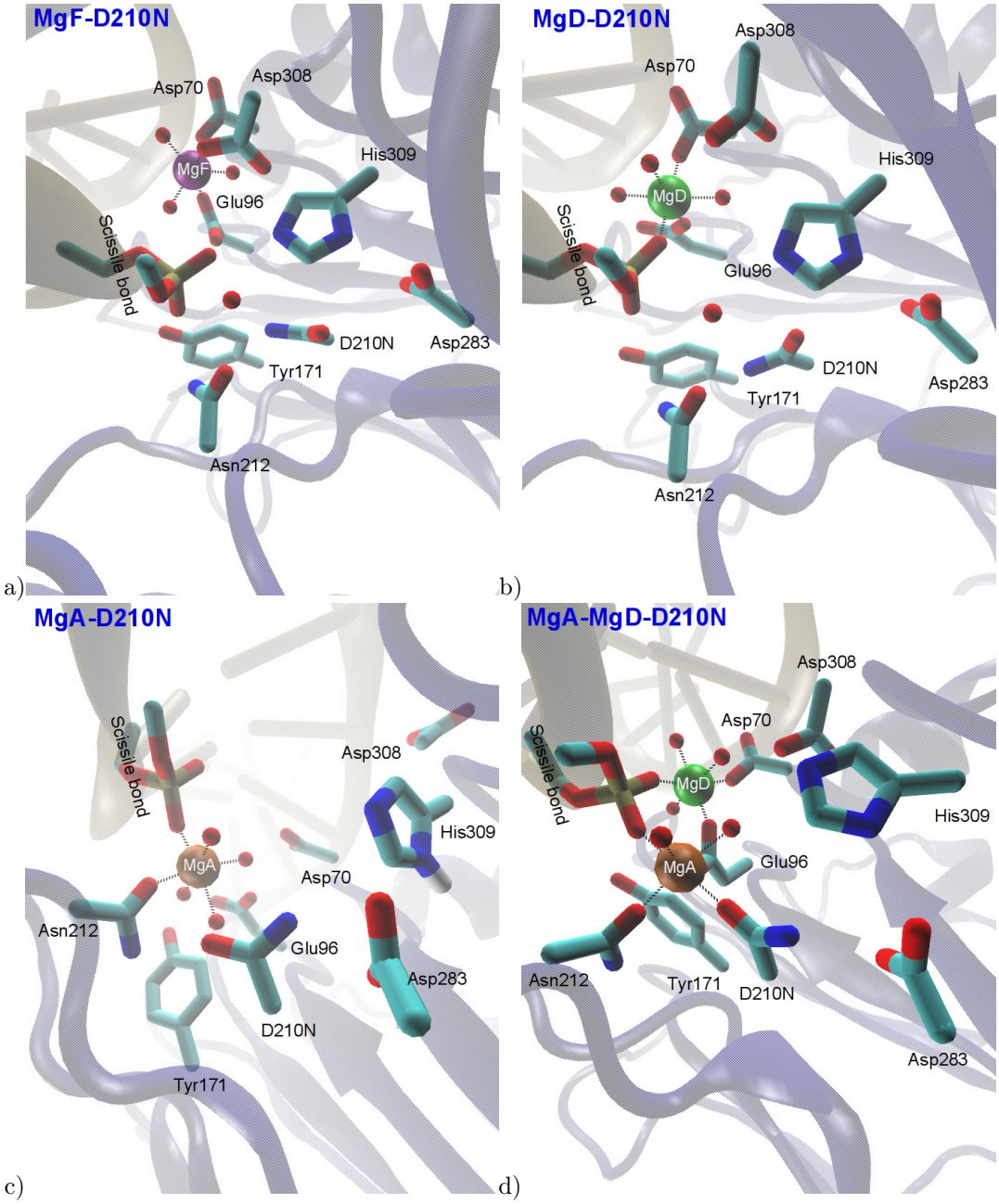

Figure 38: Key amino acid residues in the active site of D210N mutant Ape1-DNA substrate complex with a) one magnesium ion at metal binding site F, b) one magnesium ion at metal binding site D, c) one magnesium ion at metal binding site A, and d) two magnesium ions at site A and D.

Our MD simulations and their analyses allow us a comparison of different models for a APE-DNA reactant-state complex with respect to their active-site conformations. While the simulations provide different, locally equilibrated, conformations, they do not reveal (large) transitions out of or into these conformational states. Thermodynamic stabilization, or binding affinities of the metal ions at the different positions, can thus not be determined from the present data. However, from the structural similarity with the experimental crystal structure, we can infer how well the active-site is stabilized in the different models. That is whether or not local changes in the metal positions, associated ligand exchange, and altered ligand conformations are observed in the simulations. Such local changes can take place on the time scales of our simulations as can be seen in the case of the Hse models where the His309 moves significantly from its initial position. Misplaced metal-

ions, although not leaving the active site, have been observed to alter their position already upon minimization (data not shown) or within a short MD simulations of 500 ps only (see Ref. [111]). On the time scale of the cleavage reaction (k_{cat}=1.8 s^{-1} [31]) which is far beyond the time scale of the present MD simulations, the metal ions may well change their exact positions as is also suggested by the crystal structure of the product complex[168] . Similarly, the time-scales suggested for DNA binding by APE and the formation of a catalytically-competent complex (k =43.1±8.5 s^{-1} [10]) are not captured by the present simulations. However, the metal-ion positions of the reactant complex modeled in this work are inspired by the available X-ray structures of reactant or reactant-like complexes in which specific DNA binding has already occurred. All but one of these models (the Hse-MgA model) "survive" minimization and equilibration phase of our MD simulations, that is the metal-ions stay close to their initially modeled positions, likely due to the strong electrostatic interactions with their coordinating ligands. These coordination spheres remain unaltered in the course of the MD simulations and as such render all models conceivable. Our further analysis then allows us to evaluate whether a particular active-site architecture, in terms of composition and conformation, is probably reaction-competent or not. The putative contribution of the catalytically important residues to the active-site architectures and thereby their role in stabilizing the active reactant complex is discussed in the following.

5.5 Role of amino acid residues in the active-site of APE1-DNA

In addition to the magnesium ion(s), the role of catalytically important amino residues such as Asp70, Glu96, Tyr171, Asp210, Asn212, Asp283, and His309 have been debated in the literature [90, 40, 138, 91]. For instance, mutational studies of key amino acid residues account for the critical impact of them on the enzyme's DNA-binding capacity and/or the catalytic cleavage reaction [149, 93, 148, 91, 160]. Mutation studies on these amino acid residues Y171F, N212A, D212N showedreduction by 1200-fold, 7000 and 10000-fold, respectively, in the processing rate of DNA duplexes containing a substrate-analog [160]. Several mutational studies on Asn212 demonstrated the critical role of this amino acid residue in abasic DNA recognition by Ape1 and in the incision step [91, 138]. N212A does not have a notable impact on the ability of Ape1 to recognize abasic DNA but it reduces the rate of the incision step considerably [91]. In contrast, N212D reduces the ability of Ape1 in abasic DNA recognition although this mutant incises the abasic DNA more efficiently than N212A [91]. Different hypotheses for the role of Asn212 in the incision mechanism have been proposed. A mechanism in which the Mg^{2+}-ion is bound to a site from which the nucleophilic water molecule is likely to attack [111, 130] includes Asn212 as responsible for Mg^{2+}-ion coordination and formation of hydrogen bonds with the abasic DNA.

5.5.1 Role of His309

It has been reported that mutation of His309 to asparagine results in a 2500-fold decrement in enzymatic efficiency relative to wild-type Ape1 [160]. Our simulations indicate a structural role of His309 in stabilizing the conformation of the protein-DNA complex, for which, furthermore, His309 has to be in a protonated form. The significantly different conformation of His309 in the Hse models compared to any of the available crystal structures, renders the Hse-variants of all models studied here unlikely. The large fluctuations of the DNA in the models with a single Mg^{2+}-ion at binding site F and unprotonated His309 further render these models as unstable, and thus disqualify them as candidates for reaction-competent protein-DNA substrate complexes. However, His309's exact role can also go beyond formation of a hydrogen bond with the AP-site and binding of the DNA in the reactant state. For instance, a catalytic triad formed by Asp283, His309, and a water molecule can be envisaged to help catalysis through nucleophile generation by a neutral His309 accepting a proton from the nucleophilic water molecule as found in serine proteases and suggested previously also for APE [99]. This proton acceptance would be facilitated by a hydrogen bond to Asp283, increasing the proton affinity of His309. At the same time a Mg^{2+}-ion coordinated to the phosphate group of the AP-site, either at binding site A or D, could stabilize the abasic DNA for direct attack of the hydroxyl ion [40]. According to previous simulations of cluster models [57], His309 in the neutral Hsd-form can indeed act as a proton acceptor, in particular if supported by a nearby aspartate residue such as Asp283.

Our simulations of the D210N mutant correspond to a catalytically active complex for models with neutral (Hsd) His309. Since experimental studies find the D210N mutant significantly reduced in activity, His309 can therefore be concluded to be predominantly protonated. The very small residual activity of the D210N mutant may be due to a mechanism that involves the unlikely, but not entirely impossible, neutral His309 as general base for nucleophile activation. This is further supported by our pKa calculations for snapshots from the simulations with the metal ion(s) located at different binding sites and in different protonation states of His309. The pKa data strongly suggest His309 to be protonated already in the reactant state and not only after acceptance of a proton from the nucleophile in the course of the reaction. This is in agreement with pH-dependent ^{15}N-NMR experiments of the uncomplexed protein that point to a charged His309, also in the presence of functional (Mg^{2+} or Ca^{2+}) metal ions [40]. This protonated His309 cannot act as a general base.

As an alternative role, His309 in protonated Hsp-form can participate in a His309-ASP283-AP-site hydrogen-bond cluster to catalyze the phosphate cleavage through stabilization of the developing negative charge on a transition-state/intermediate with pentacovalent phosphate [168, 31]. This stabilization can even be envisaged for an initially neutral His309, i.e. after acceptance of a proton from nucleophilic water prior to nucleophilic attack and formation a pentacovalent transition-state/intermediate, in agreement with the essential catalytic role of His309 suggested by mutation of this residue [90, 15, 160]. Whether such a stabilization in the course of the reaction, in addition to the structural role of keeping the phosphate group close to the other catalytically essential residues, is indeed provided by His309 can be answered only in a detailed analysis of the reaction mechanism with electronic structure methods. Such an analysis is, however, beyond the scope of this section.

5.5.2 Role of Tyr171

Based on the initial position of Tyr171 in the Ape1-DNA substrate complex [168] and on the significant reduction in catalytic activity in Y171F (1200-fold) relative to wild-type Ape1 [160, 138], partial neutralization of the developing negative charge on the AP-site upon nucleophilic attack through a hydrogen bond with the phosphate group may be a catalytic role of Tyr171. The only fourfold decrement in catalytic rate for Y171H [160, 138] in which the Histidine can in principle form hydrogen bonds to the phosphate group, supports this hypothesis. According to our simulations, MgA and MgF models show Tyr171 to be close to Glu96 rather than bound to the phosphate group. These models are thus less likely to be catalytically active than models with the metal ion bound at site D. In another mechanism proposed previously [148], Tyr171 in phenolate form generates the nucleophilic hydroxyl ion or attacks the O3' atom of leaving group directly. A similar mechanism has been reported for topisomerases [129]. Neither in the recent crystal structure of the Ape1-DNA substrate complex nor in our simulations is Tyr171 found to be in a proper position to directly attack the leaving group (see Figure 10).

5.5.3 Role of Asn212

Mutation of Asn212 to Alanine shows a significant (7000fold) reduction in the incision step [91]. This has been explained by Asn212 to be responsible for keeping the nucleophilic water molecule in a proper position for direct attack of the phosphate through a hydrogen-bonded network with Asp210 and the nucleophilic water molecule [91]. Our simulations show Asn212 to form a hydrogen bond with which the nucleophilic water molecule in the Hsp-MgD model, thereby keeping it positioned for attack. Hence, according to that model, mutation of Asn212 would severely hamper enzymatic function, in agreement with the mutation experiments. Our simulations of the N212A mutant show an active-site conformation in the MgA models that does have a water molecule positioned to act as nucleophile (see Figures 37), indicating an active complex. This is in contradiction to the experimentally observed reduction in activity upon N212A mutation. Models with Mg^{2+} in position A are therefore unlikely. N212A models with Mg^{2+} located in position D and protonated His309, in contrast, exhibit an active-site geometry that is altered upon mutation in a way that prevents the correct positioning of a water molecule for nucleophilic attack. This leads to the conclusion that the impact of N212A mutation is a distortion of the active site geometry such that no water molecule is properly placed for nucleophilic attack in contrast to the wild-type geometry.

Consequently, the role of Asn212 is this positioning of the nucleophile. Indeed, Asn212 persistently forms a hydrogen bond to such a putative nucleophile in the MgD-Hsp model, keeping it poised for attack on the scissile phosphate group.

5.5.4 Role of Asp210

In addition to supporting Asn212 in conserving the nucleophilic water molecule in a proper position through formation of a hydrogen-bond network with Asn212 and the nucleophilic water molecule, as suggested earlier [91, 168], and supported by our data (see above), several other catalytic functions have been proposed for Asp210 [43, 40, 110, 99]. The essential role of D210 discussed in the literature is that of a general base, accepting a proton from a water molecule and thus activating the nucleophile for attack. Another proposal limits the role of D210 to metal positioning. D210N mutation shows the incision rate reduced by 10000-fold [160, 138, 93]. This argues for a critical function such as activation of the nucleophilic water molecule for attack. Such a role is conceivable for Asp210, albeit rather unlikely with a metal ion placed at site A. In those MgA models both oxygen atoms of Asp210's side chain are involved in the metal-ion coordination and a side-chain rotation should take place before proton acceptance by Asp210 can occur. Although such a rotation with subsequent proton acceptance has been reported in the simulation of the phosphate hydrolysis in the restriction endonuclease EcoRV [119], our present MD simulations do not show any indication of Asp210 leaving the coordination sphere. The bidentate coordination of the metal ion by Asp210 renders proton acceptance by this residue therefore unlikely in the models with Mg^{2+} at site A. In some previous studies, the considerable reduction in incision rate induced by D210N mutation was related to the presence of a second divalent metal ion at site A (in addition to a first one at site D) [110, 111, 130]. In this hypothesis, Asp210 coordinates the Mg^{2+}-ion at binding site A [110, 111, 130] and the Mg^{2+}-ion in turn keeps a water molecule in a proper position for direct nucleophilic attack to the abasic DNA. That way, Asp210 is close to the metal-ligated water from which it may accept a proton (see Figure 10). The proton-accepting oxygen atom of Asp210, however, will likely leave the metal coordination. Our simulations of the D210N mutant with protonated His309 (Hsp) show the coordination sphere of Mg^{2+} to be intact after mutation. The Mg^{2+}-bound water molecule is still in proper position for direct attack throughout the 100ns simulation time (Figure 38). The same observation is made for the D210N mutant with neutral His309 and a single Mg^{2+}-ion at position A, as well as the D212N mutant with two Mg^{2+}-ions bound to sites A and D, respectively. Consequently, Mg^{2+} in A position would still be able to serve in positioning the nucleophile after D210N mutation. Hence, metal-ion coordination cannot be the mere contribution of D210 since this is achieved also in the D210N mutant.

On the assumption that a general base is required for nucleophile activation further, indirect, support for the idea of Asp210 acting as general base (and not His309) can be gathered from the D210N simulations. Protonated H309 cannot accept the proton from the nucleophile leaving this role to Asp210. But in the D210N this role cannot be assumed by the Asn residue, explaining the experimentally found loss of activity upon D210N mutation. Our simulations of the D210N mutant with unprotonated H309, however, show this residue to be in a conformation that allows H309 to act as a general base in models with a single Mg^{2+} bound to site A or D, or with two Mg^{2+}-ions bound to sites A and D, respectively (see Figure 38). Thus, with these combinations, His309 could act as a general base and a D212N mutant would still be catalytically active. Since this is in stark contrast to the considerable rate reduction observed experimentally [15, 79], unprotonated H309 is unlikely - in agreement with the pKa calculations and previous NMR experiments [40]. With protonated His309 obviously not acting as a proton acceptor, an alternative candidate for a general base is Asp210. This important role of Asp210 is in good agreement with the experimental mutation results. According to the pKa calculations, Asp210 is more likely to accept a proton with the metal ion bound to site D than bound to site F. Furthermore, throughout the entire simulation of MgD-Hsp, Asp210 stays hydrogen-bonded to a water molecule that is well positioned for nucleophilic attack. In models with a Mg^{2+}-ion at binding site A, the pKa results and the strong coordination to MgA render proton acceptance by Asp210 unlikely, leaving Hsp-MgF and Hsp-MgD as the most probable models of the Ape1-DNA substrate complex. According to our distance analysis between the carboxyl side chain of Asp210 and the abasic phosphate, Asp210 is in better position for accepting a proton from the nucleophilic water molecule in Hsp-MgD than it

is in Hsp-MgF (see Figure 80).The pKa calculations based on MD snapshots favor an unprotonated Asp210 for all models considered. However, simulations with protonated Asp210 show a strong similarity of the Hsp-MgF model with the crystal structure in the active-site conformation. The pKa calculation for the crystal structure also indicates a protonated Asp210 unless a structured water molecule is taken into account explicitly. In this particular conformation, hydrogen-bound to a water molecule, Asp210 is more likely to be unprotonated. These two scenarios suggest that Asp210 may indeed accept a proton from a well-poised water nucleophile. Keeping in mind that the crystal structure is that of an inactive complex due to the thio-substitution, the observed conformation (in simulation of protonated Asp210 or the crystal structure) may be regarded as a trapped intermediate rather than a reactive complex. An increment by 15-fold has been reported for a D210H mutant. This led to the assumption of a protonated Asp210 that helps the incision process as the proton donor to the O3' atom of the leaving group. Our simulation on protonated Asp210 does not reveal any notable conformational changes and proton transfer to the O3' atom is unlikely to occur due to large distance between the O3' atom and Asp210. On the other hand, the increased activity of the D210H mutant as opposed to the decreased activity in D210N and D210A can be attributed to a neutral histidine that serves as the general base with a proton affinity that is higher than that of an aspartate residue, a mechanism that has been reported for other endonucleases. [100, 45, 74, 57].

5.5.5 Metal ions

Site A as position for a Mg^{2+}-ion can be ruled out on the basis of the mutant simulations contradicting the experimentally observed reductions in activity. With the same arguments, models with two metal ions, one of which is situated at site A, are also unlikely. NMR-based studies also suggest metal binding to Glu96, i.e. the D or F site [40]. Models with MgF are only stable in their active-site as well as in their global conformations with protonated His309. However, also the Hsp-MgF model exhibits significant fluctuations in the DNA conformation, corresponding to unstable interactions between active-site residues of the protein and the abasic site of the DNA. Whereas Asp210 and His309 remain in close enough contact to form a direct or water-mediated hydrogen bond to the phosphate group, interactions with other active-site residues such as Tyr171 and Asn212 are weaker in the Hsp-MgF than in the Hsp-MgD model as manifested by the larger distances observed. In ref. [168], a conformation with a Mg^{2+}-ion coordinated to the phosphate group, corresponding to site D, has been suggested as reaction intermediate/transition state. It is therefore well conceivable that the reaction-competent conformation of an Ape1-DNA complex has indeed the Mg^{2+}-ion situated much closer to the phosphate group (site D) than the Mn^{2+}-ion observed in the thio-substituted Ape1-DNA substrate complex which is bound farther from the phosphate group (site F) with a water molecule in between. On the departure site, the role of the metal-ion cannot be nucleophile positioning nor activation. Instead, the metal-ion may activate one of its coordinating water molecules to donate a proton to the leaving group upon or after departure. The compensation of accumulating negative charge in the transition state/intermediate is also a conceivable role that, however, cannot be determined with the present simulations of the reactant state complex only. These simulations and the analysis of the different models in comparison with experimental findings favor D as the metal-binding site. This position allows the metal to bind the scissile phosphate group and hence place it close to the other catalytic residues is strongly suggested by our simulations. The MD simulations reported here show the most likely model for an Ape1-DNA substrate complex has one Mg^{2+}-ion located at binding site D and His309 in protonated (Hsp) form. At the D site, the metal ion may play a catalytic role in leaving group departure. This scenario allows Tyr171 and His309 to form hydrogen-bonds to the phosphate group that may help DNA binding as well as stabilizing a pentacovalent transition state/intermediate. Asn212 acts in properly positioning the nucleophile which can then transfer a proton to Asp210.

6 Mechanistic insight into enzymatic cleavage of DNA by APE1 endonuclease

Although all the available structural and biochemical information inspired the postulation of a multitude of reaction mechanisms, an understanding at the atomic or even electronic level of detail is still missing for the phosphodiester backbone cleavage by APE1 enzyme. In this section, we investigate the enzymatic hydrolysis reaction pathways by APE1 using a combined QM/MM approach, treating the active site quantum mechanically and the remaining enzyme classically. This combined approach enabled us for an accurate description of the groups directly involved in the reaction while treating the whole enzyme in atomic details.

6.1 Potential energy scans and minimum energy pathways

6.1.1 Finding the best starting point

The abasic DNA cleavage catalyzed by APE1 endonuclease as a phosphate hydrolysis reaction develops through either stepwise or concerted fashion. In phosphate hydrolysis, the stepwise mechanism has two types including dissociative and associative. In the dissociative type, the leaving group departure occurs earlier than the nulcleophilic attack of a water molecule whereas priority of nucleophilic attack to the leaving group departure defines an associative mechanism. In the concerted type, nucleophilic attack and leaving group departure emerge simultaneously. To investigate a viable type of mechanisms in cleavage of abasic DNA catalyzed by APE1, we computed a two-dimensional scan of the potential DFTB/MM energy for phosphate hydrolysis starting from the catalytic active reactant state proposed by X-ray structure and Molecular Dynamic analysis [56, 168]. The two dimensions are the P–OH_2 and P–O3' distances, which correspond to nucleophilic attack and leaving group departure, respectively.

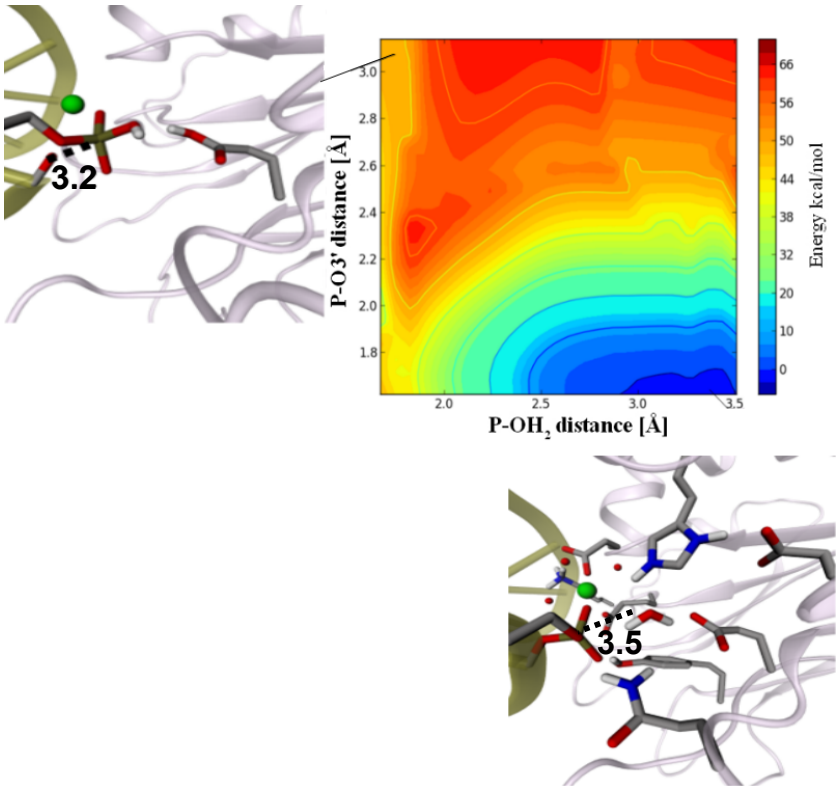

Figure 39: Two-dimensional scan of the potential energy for phosphate hydrolysis starting from WT model computed with DFTB/MM. Coordinates are the P–OH_2 and P–O3' distances. Energies are given in kcal/mol.

In the reactant state, the P-OH2 and P-O3' distances are 3.5 Å and 1.6 Å, respectively. As represented in the two-dimensional scan of thepotential energy surface, shown in Figure 39, the potential energy relative to the reactant state elevates rapidly by driving the P-O3' distance from 1.62 Å to more than 3 Å, suggesting that the formation of a dissociative intermediate is highly unlikely.

According to the two-dimensional potential energy (Figure 39), driving the nucleophilic water molecule towards the phosphorous atom generates an associative-like intermediate, albeit of very high energy with an even higher transition state (over 40 kcal/mol) to pass. Moreover, the relative potential energy for strucutres with a dissociated P-O3' bond are over 50 kcal/mol, rendering this route infeasible. This two-dimensional potential energy surface scan corroborates that the dissociation of the sugar backbone is the most pivotal issue in the phosphate hydrolysis by APE. This is interesting in as much as in the initial active-site architecture, the Mg^{2+} ion located at the departure site is assumed to facilitate leaving group departure by providing positive charge, balancing the negatively charged leaving group, or activating a metal-ligated water molecule so as to donate a proton to the leaving group. However, no such proton transfer has been observed upon driving the P-O3' dissociation.

The accessible candidates with a positive charge that could help protonation of the leaving group on the departure site are K98 or K78. The distance between K98, K78 side chains and the leaving group is too wide (7.8 and 10.3 Å, respectively) to transfer a proton directly to the leaving group. However, proton transfer from K98 or K78 to D70, which is located in the coordination sphere of the Mg^{2+} and therefore near to the metal ligated water molecules, can take place with energy barriers of 6.4 kcal/mol and 11.3 kcal/mol respectively (calculated using CPR). Transferring a proton to D70 through both accessible mentioned pathways (i.e. from K98 or from K78) delivers two intermediate states that resemble each other because of the protonated D70. The intermediate states have an energy of 3.5 kcal/mol and 3.3 kcal/mol, respectively, relative to the reactant state (see Figures, 44 and 41). In addition to a direct hydrogen bond between D70 and K98, the energy barrier for proton transfer from K98 to D70 is lower with an energy of 6.4 kcal/mol relative to the reactant state in comparison to the 11.3 kcal/mol for proton transfer from K78 to D70. Hence, K98 is more entitled to donate a proton to D70 than K78 is.

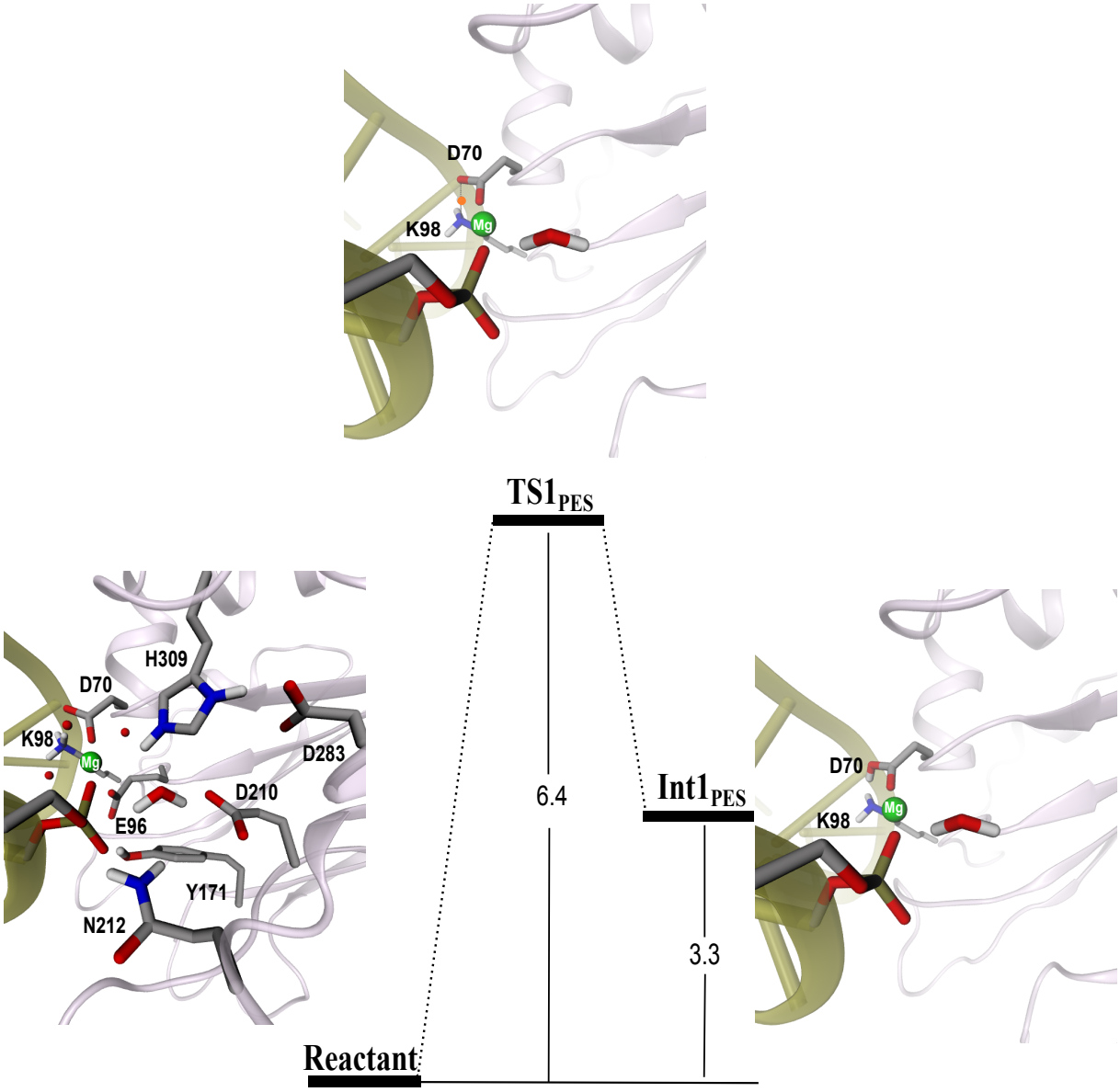

Figure 40: Mechanism for proton transfer from K98 to D70 calculated by DFTB/MM. All calculated potential energies are in kcal/mol relative to the reactant state (the transferred proton is highlighted by an orange circle).

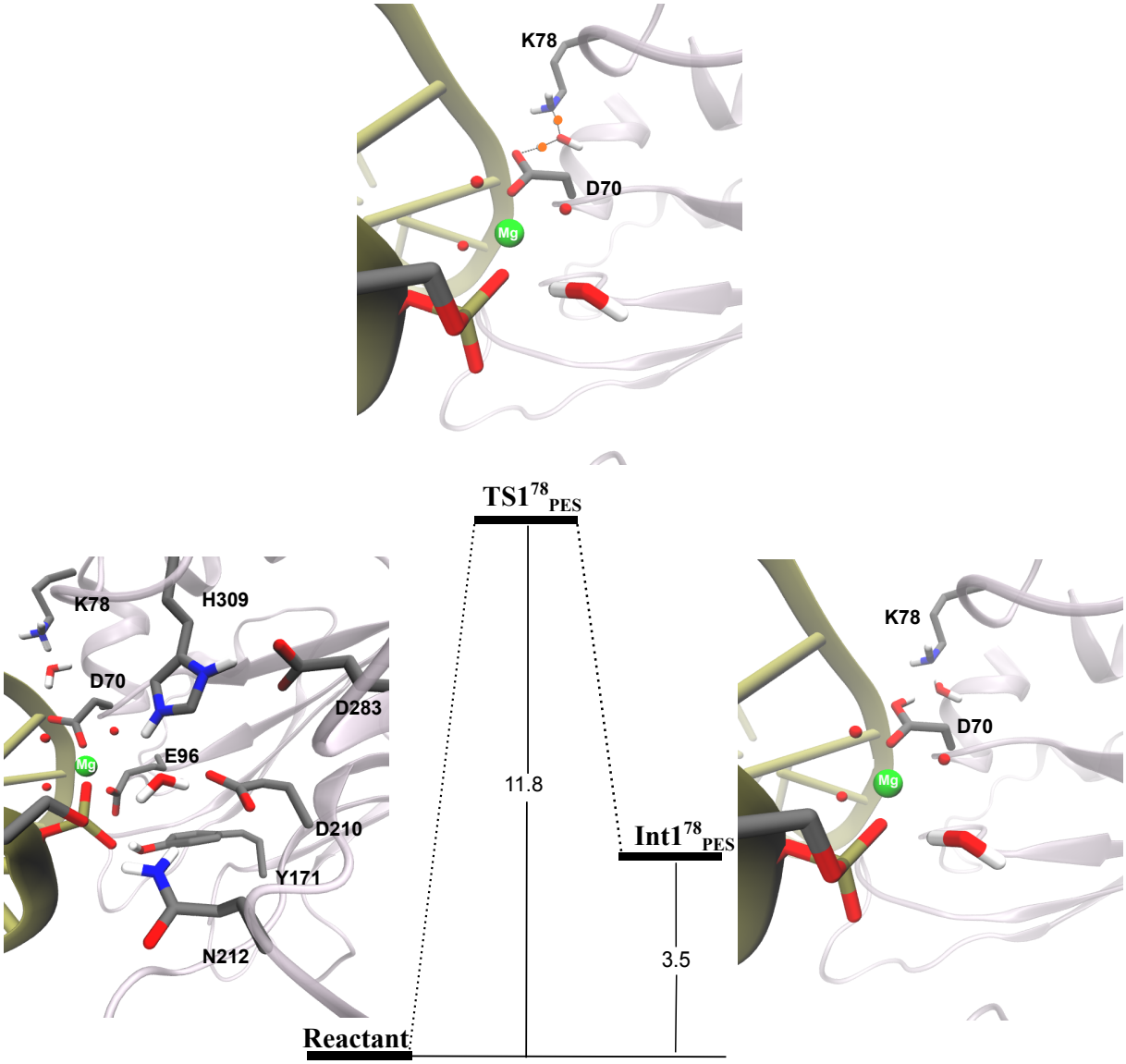

Figure 41: Mechanism for proton transfer from K78 to D70 calculated by DFTB/MM. All calculated potential energies are in kcal/mol relative to the reactant state (the transferred proton is highlighted by an orange circle).

6.1.2 Proposed dissociative mechanism for phosphdiester hydrolysis catalyzed by APE1

Considering now Int1$_{PES}$ as an activated reactant state, another two-dimensional DFTB/MM potential energy scan has been carried out along again the P-O3' and the P-OH2 distances (Figure 42).

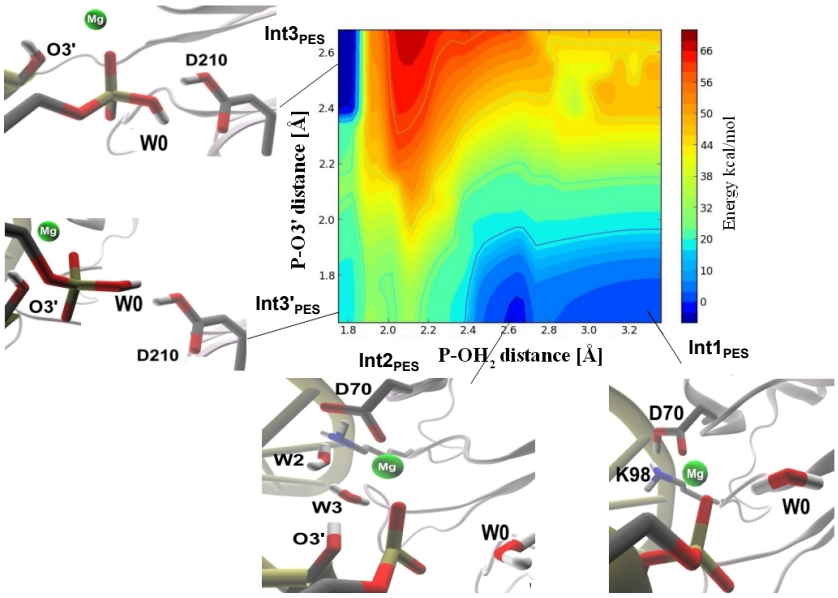

Figure 42: Two-dimensional scan of the DFTB/MM potential energy for phosphate hydrolysis starting from a model with protonated D70 and unprotonated K98. Coordinates are the P–OH_2 and P–O3' distances. Energies are given in kcal/mol relative to Int1

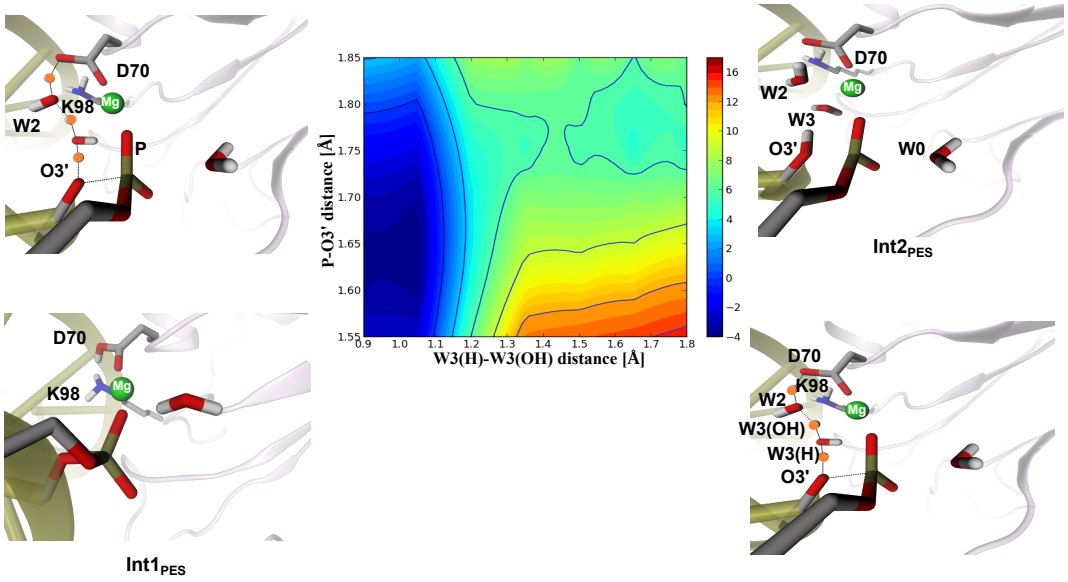

Figure 43: Two-dimensional scan of the DFTB/MM potential energy for phosphate hydrolysis starting from a model with protonated D70 and unprotonated K98. Coordinates are the P–OH_2 and P–O3' distances. Energies are given in kcal/mol relative to Int1

Interestingly, one new minimum state can be observed with a P-O3' distance increased from 1.6 Å to around 1.9 Å and a P-OH2 distance that drops from 3.5 Å to around 2.8 Å, (Figure 42). In this state the O3' atom adopts a proton from a metal ligated water molecule and this water molecule adopts the proton from the protonated D70. The phosphate backbone is cleaved in this minimum state ($Int2_{PES}$ see Figure 42) whereas no nucleophilic attack has happened, yet, qualifying $Int2_{PES}$ as a dissociative intermediate of a step-wise phosphate hydrolysis mechansim. The transition state, $TS2_{PES}$ of this dissociation step is caluclated as 16.9 kcal/mol relative to reactant state (APE1_WT) according to the minimum energy pathway computed with CPR.

After leaving group departure, the nucleophilic water molecule can be activated via donating a proton to D210. According to our potential energy scan this happens simultaneously with the attack of the thus formed hydroxide ion to the phosphorous atom. In the two dimensional potential energy surface scan this step is achieved by passing a transition state of 25.1 kcal/mol ($TS3'_{PES}$). In the resulting intermediate, $Int3'_{PES}$, the P-O3' and P-OH distances are 1.9 Å and 1.7 Å, respectively, and the relative potential energy for this state is 18.6 kcal/mol. Subsequent minimisation of this structure with removing the constraints on the P–OH_2 and P–O3' distances leads to an intermediate, $Int3_{PES}$, with P-O3' and P-OH distances of 3.2 Å and 1.7 Å, respectively and a significantly reduced energy of 9.6 kcal/mol. According to a CPR calculation, connecting $Int2_{PES}$ and $Int3_{PES}$, the transition state has a relative potential energy of 23.6 kcal/mol. The energies of the transition states computed by PES scan and CPR are rather similar. In contrast the spontaneous relaxation of $Int3'_{PES}$ to $Int3_{PES}$ and the differences in energies and geometry show that the former is not a true minimum but a point on or close to the minimum energy path to the actual product of the nucleophilic substitution reaction (see Figure 44 and Table 21).

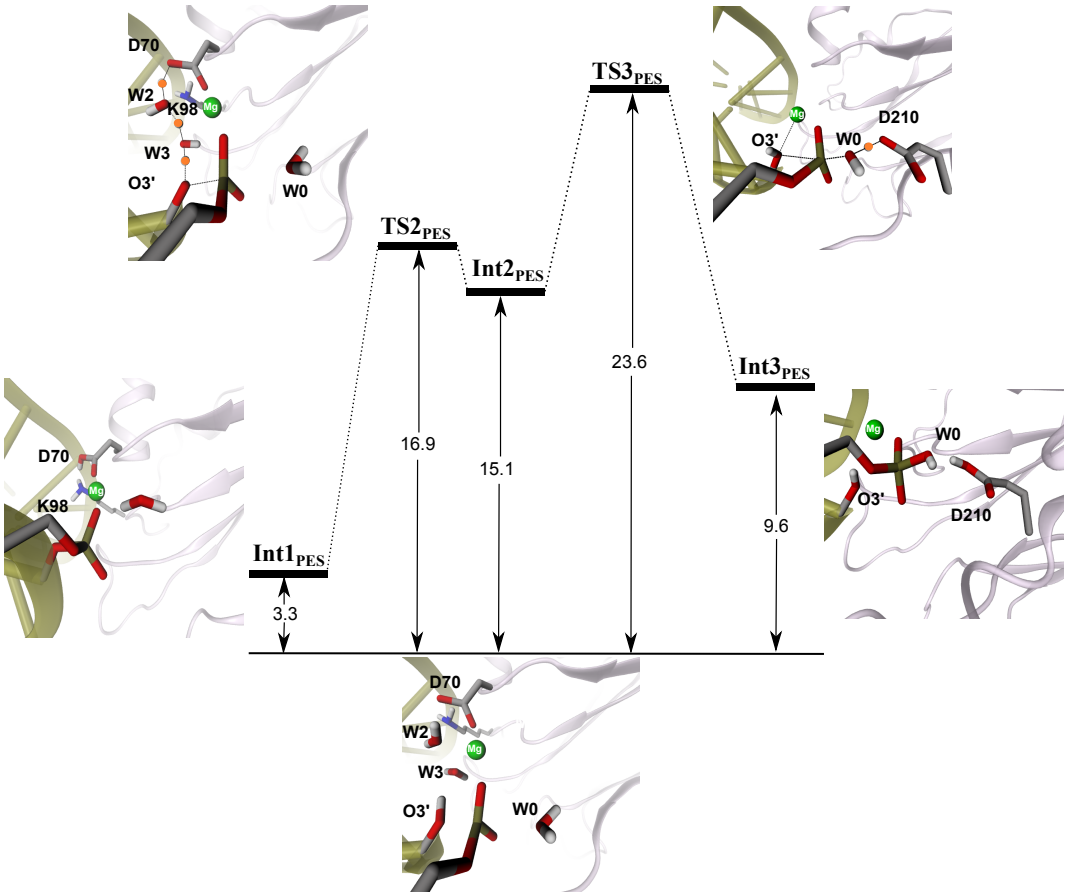

Figure 44: Mechanism for cleaving the abasic DNA backbone and nucleophilic attack of water catalyzed by APE1 enzyme calculated by DFTB/MM. All calculated potential energies are in kcal/mol relative to the reactant state (the transferred proton is highlighted by an orange circle).

State	Rel. Energy
$TS1_{PES}$	6.4
$Int1_{PES}$	3.3
$TS2_{PES}$	16.9
$Int2_{PES}$	15.1
$TS3_{PES}$	23.6
$Int3_{PES}$	9.6

Table 21: Potential energy relative to the reactant state calculated by DFTB/MM using CPR method for placing transition state from reactant to $Int3_{PES}$ in kcal/mol

We assume that, as E96 coordinates the Mg^{2+} ion and it is also near to a charged amino acid side chain such as K98, hence, E96 can also play a base role to adopt a proton from K98. The potential energy profile starting with transferring a proton from K98 to E96 can be seen in Figure 45. In this pathway, the first proton transfer from E96 to D70 relative to the reactant state has a barrier of 7.4 kcal/mol and the nucleophilic attack of activated water called $TS3_{PES}{}^{\mathrm{E96}}$ is the rate determining step with 23.7 kcal/mol transition state energy relative to the reactant state (see 45).

We also computed the potential energy barrier for transferring a proton from K78 to E96 which leads over a transition state of 11.9 kcal/mol energy relative to the reactant state and results in an intermediate with 3.7 kcal/mol as the potential energy barrier to the reactant state. Therefore both D70 and E96 can play a base role to adopt a proton from K98 or K78. However, D70 and E96 can be regarded asin a competition to adopt a proton from K98/K78.

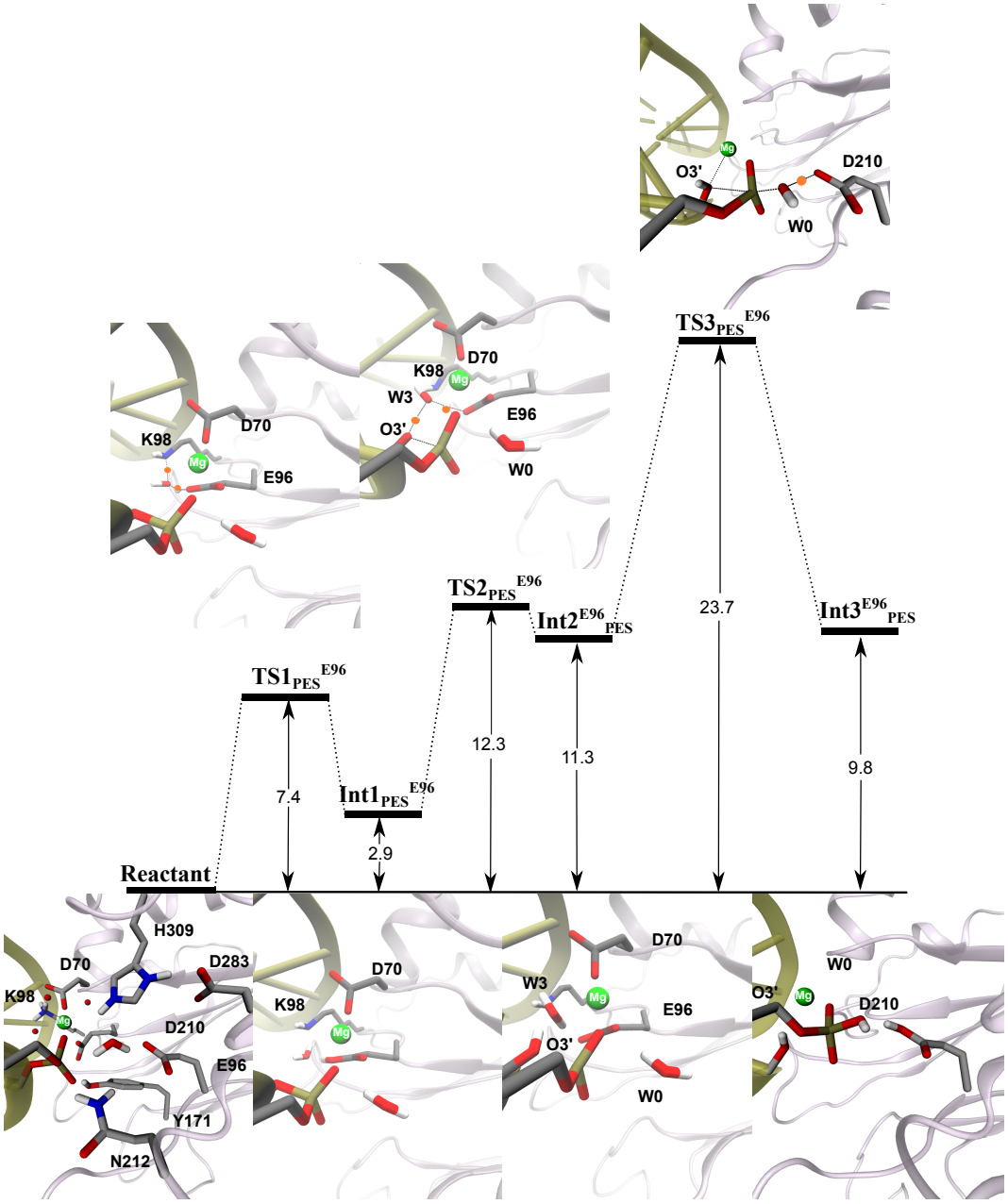

Figure 45: Mechanism for cleaving the abasic DNA backbone and nucleophilic attack of water catalyzed by APE1 enzyme starting from donating a proton from K98 to E96 calculated by DFTB/MM. All calculated potential energies are in kcal/mol relative to the reactant state (the transferred proton is highlighted by an orange circle)..

6.1.3 Reaction path calculations in APE1-mutants

H309N and Y171F

We have done H309N, Y171F mutations here to investigate their catalytic role through calculating the potential energy profile for phosphodiester hydrolysis catalyzed by APE1.

The mechanistic details for each step of phosphodiester hydrolysis in both H309N and Y309F mutant (see Figure 46) are similar to what has been explained for wild type APE1 (see Figures 40,44) According to the calculated relative potential energy pathway in the phosphodiester cleavage of abasic DNA catalysed by APE1 enzyme, the step of nucleophilic attack of water is the rate determining step in WT (see Table 21) as well as in the H309N, Y171F mutant (see Table 22). As

noted before, the rate determining step relative to the reactant state for APE1 complexed with abasic DNA in phosphodiester cleavage reaction is 23.6 kcal/mol for wild type (see Table21) while the rate determining step for H309N , Y171F mutant are 35.1 and 34.9 kcal/mol respectively (see Table 22).

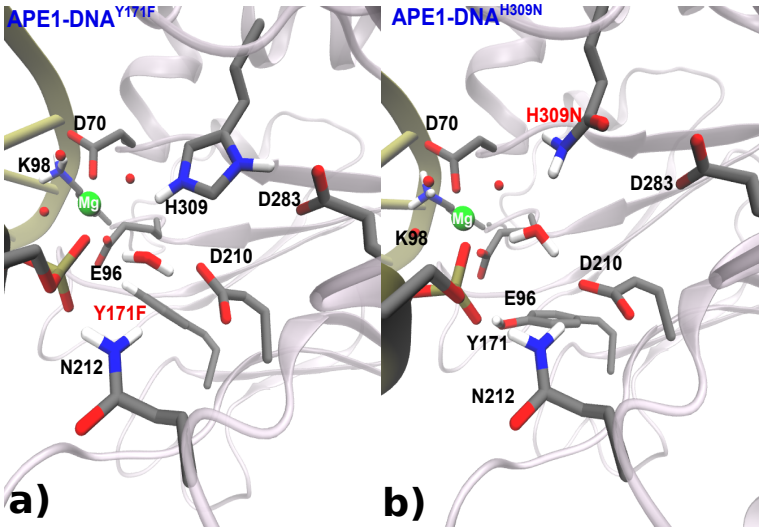

Figure 46: Key amino acid residues in the active site of a) H309N and b) Y171F mutant Ape1-DNA substrate complex

N212A

To investigate the role of Asn212 , we performed potential energy calculation for N212A mutant APE1-DNA substrate. It stands to mention that mechanistic details in calculated pathway for phophodiester hydrolysis of N212A mutant is similar to what we have described in wild type APE1(see Figures 40,44). Mutating Asn212 to Alanine shows a significant increment in the calculated potential energy barrier of the rate determining step (38.9 kcal/mol, see Table 47) in compare with WT APE1 (23.6 kcal/mol, see Table22)

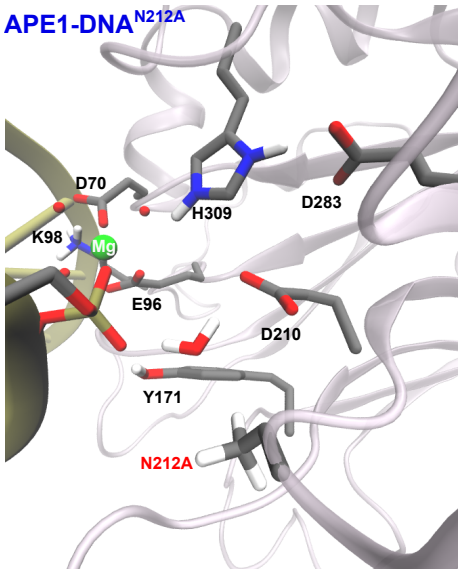

Figure 47: Key amino acid residues in the active site of N212A mutant Ape1-DNA substrate complex

State	WT	H309N	Y171F	N212A
$TS1_{PES}$	6.4	5.4	1.2	10.6
$Int1_{PES}$	3.3	5.2	-2.9	6.5
$TS2_{PES}$	16.9	21.1	20.0	13.6
$Int2_{PES}$	15.1	20.6	19.0	14.1
$TS3_{PES}$	23.6	35.1	34.9	38.9
$Int3_{PES}$	9.6	28.3	30.9	13.6

Table 22: Potential energies (in kcal/mol) relative to the reactant state for N212 mutant Ape1-DNA substrate complex in phosphodiester cleavage reaction in the enzyme APE1 .

D210A

To investigate the catalytic role of D210, we have mutated it to Alanine (D210N) and computed the phosphodiester cleavage reaction (see Figure 49). The calculated energy barrier relative to the reactant state for the rate determining step in the DNA backbone cleavage by the D210A mutant APE1 enzyme, is 50.8 kcal/mol 49. This step is the attack of the nucleophilic water molecule as in the wild type enzyme-

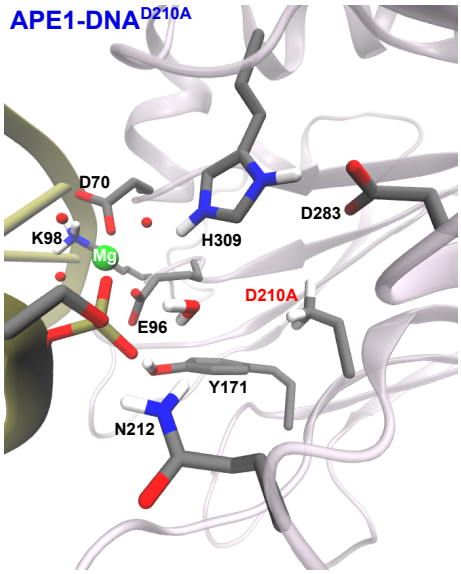

Figure 48: Key amino acid residues in the active site of D210A mutant Ape1-DNA substrate complex

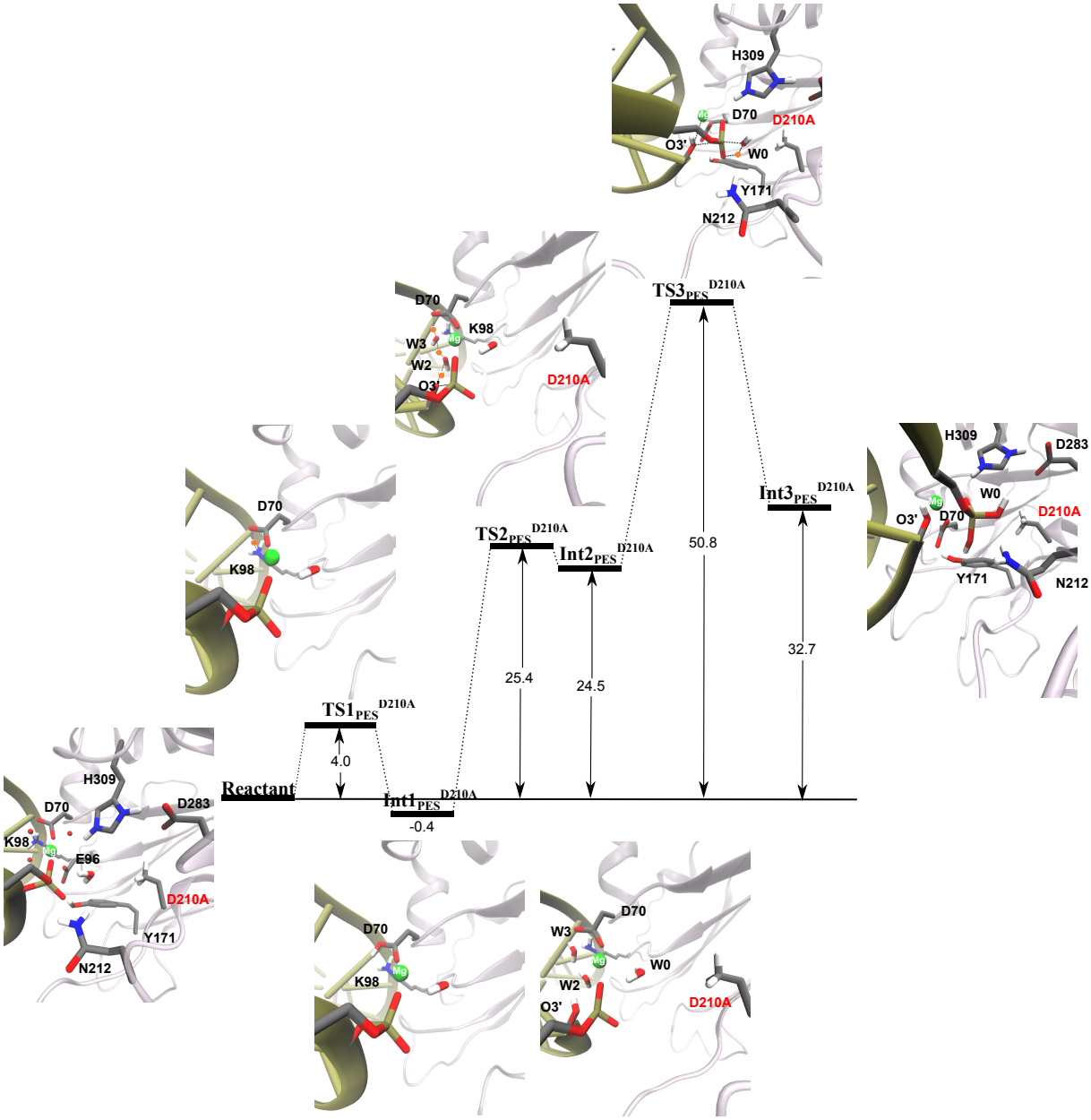

Figure 49: Mechanism for cleaving the abasic DNA backbone and nucleophilic attack of water in D210A mutant APE1 enzyme calculated by DFTB/MM. All calculated potential energies are in kcal/mol relative to the reactant state (the transferred proton is highlighted by an orange circle).

6.2 Potential of Mean Force (PMF) Simulations

6.2.1 From reactant to cleaved intermediate

Our calculated potential energy profile revealed a step-wise mechanism with leaving group departure (P-O3') followed by nucleophilic attack of an activated water molecule (P-OH). (see Figure 42). For each step of this reaction, we have carried out one-dimensional (1D) potential of mean force (PMF) simulations using umbrella sampling [69, 68]. The QM/MM-optimized minima and intermediates along the minimum energy pathways served as starting a points for the subsequent free energy calculations. Between all possibilities of proton transfer from K98 or K78, respectively, to E96 or D70, respectively, a hydrogen bond between the carboxyl group of D70 with the NH3(+) group of K98 in the APE1-DNA substrate complex votes for priority of proton transfer from K98

to D70. However, we also computed the free energy for transferring a proton from K78 to D70 to confirm the eligibility of K78 as proton donor with free energy mehtods sampled by 21 windows showing 13.9 ±0.7 kcal/mol as free energy barrier (see Figure 50).

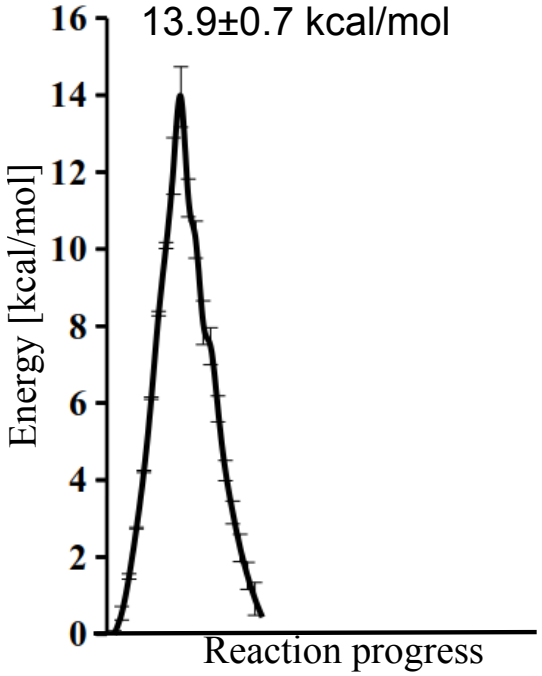

Figure 50: Free energy profile for transferring a proton from K78 to D70 in the APE enzyme calculated by DFTB/MM

According to the free Energy calculations, the nucleophilic attack of the water molecule is indeed the rate determining step (i.e. the step with the highest energy transition state along the mechanism. However, adding dynamical effects reduces the energy barrier from 23.6 kcal/mol on the potential energy surface to a free energy barrier of 18.2 ± 0.7 and 17.3 ± 0.1 for forward and backward respectively (see Figure 51 and Table 23).[1][2]

[1]Note that the free energies relative to the reactant state for intermediate and transition states, respectively) should be unique for one state, regardless of the way that led to this state, i.e. forward or backward. The existing differences can be explained by limited sampling leading to hysteresis effects, and the intrinsic error made by adding free energies that have been obtained from simulations sampling different volumes in phase space.

[2]The free energies as given "relative to the reactant state" assume the calculated relative free energy of e.g. Int1 at the beginning of the umbrella sampling simulation of the step Int1 to Int2 to be the same as the calculated free energy of Int1 at the end of the umbrella sampling simulation of the transition from reactant to Int2.

This is, however, only the case if individual free energies were obtained by an integration over the same phase space volume for all states considered, that is sampled by the same free energy calculation and by following the same reaction coordinates in the simulations with biasing potentials.

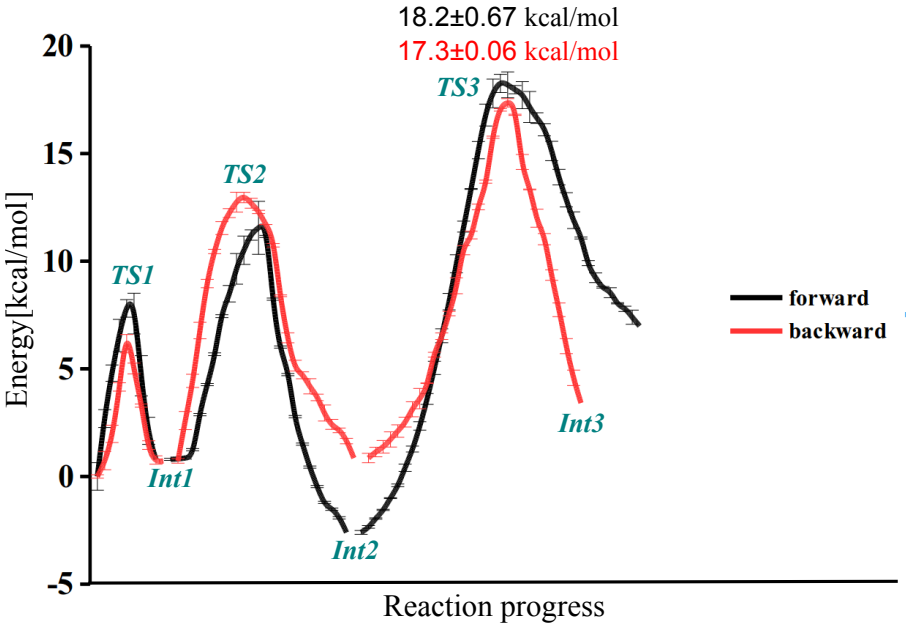

Figure 51: Free energy profile of the phosphodiester hydrolysis in the APE enzyme calculated by DFTB/MM from reactant to Int3 state in kcal/mol relative to the reactant state

During the umbrella sampling calculations from Int2 to Int3 (along a combination of P-O3' and P-OH2 distances), the cleaved sugar backbone migrates toward the Mg^{2+} ion while D70 leaves the Mg^{2+} coordination sphere. The active site architecture including the Mg^{2+} ion coordination sphere in Int3 represented in Figure 54 is perfectly matched with the available experimental crystal structure of the product state from abasic DNA backbone cleavage in the APE1 enzyme [168, 133].

All free energies calculated along the phosphodiester hydrolysis mechanism, from reactant to Int3, are shown in fig 51 and Table 23

The dissociation step, after proton transfer from K98 to D70, has a free energy barrier (11.5 ± 0.2 kcal/mol or 12.9 ± 0.5 kcal/mol, computed forward and backward, respectively) that is also lower than the potential energy barrier (16.9 kcal/mol) by approximately the same amount as the barriers computed for the nucleophilic attack, indicating that both steps are similarly affected by the conformational sampling, aka entropic effects. Furthermore, the significantly lower barrier (free or potential energy) of the leaving group departure step compared to the nucleophilic attack, clearly support a dissociative step-wise mechanism. The entropic contribution to the initial proton transfer step appears to be very small, judged from the very similar potential and free energy barriers (6.4 vs. 7.8 ± 0.4 and 6.1 ± 0.2 kcal/mol, respectively). Comparing the free energy profile of the phosphodiester hydrolysis in the APE enzyme (Figure 51) reveals a minor difference in energies for intermediate step (Int2) computed in forward and backward direction. The reason can be explained by a limited sampling of those residues that are not constrained by the reaction coordinates. Since those residues are free to move they may behave differently in the backward and forward pathways, and as such generate different probability distributions of their conformations. Indeed, as can be seen in Figure 53, conformations of some amino acid residues like H309, ASN212 and etc in the active site of APE1-DNA at the intermediate step (Int2) are different in forward and backward pathways. This can result a slight difference in the calculated free energies for states like Int2. To evaluate our selected reaction coordinate for free energy calculations, we have projected the potential energy extracted from CPR analysis onto the reaction coordinate, i.e on the combination based of distances which define the reaction coordinate, (see Table 3) for each step of the mechanism, Figure 52 shows the potential energy as function of the reaction coordinate combination to be a smooth curve with a reasonable pattern along the CPR pathways for reactant to Int3.

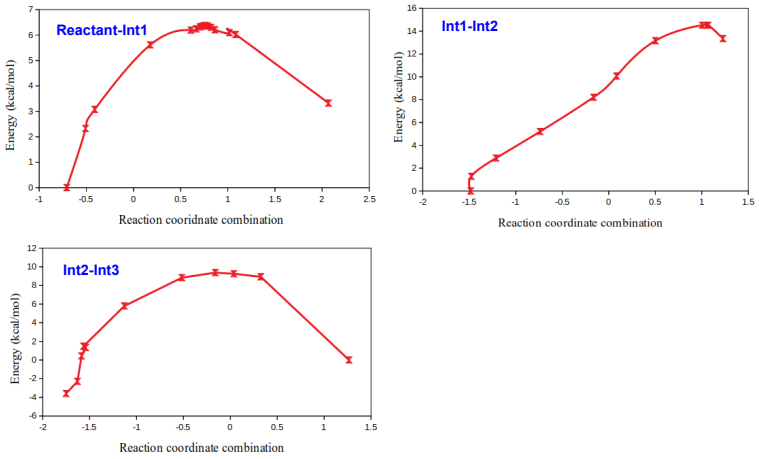

Figure 52: Calculated potential energy relative to the reactant state (in kcal/mol) calculated by DFTB/MM , from reactant to Int3 state to reaction coordinate combination for each reactant step

State	Forward	Backward
TS1	7.8 ± 0.4	6.1 ± 0.2
Int1	0.8 ± 0.6	0.7 ± 0.1
TS2	11.5 ± 0.2	12.9 ± 0.5
Int2	-2.6 ± 0.1	0.9 ± 0.2
TS3	18.2 ± 0.7	17.3 ± 0.1
Int3	7.0 ± 0.3	3.4 ± 0.4

Table 23: Free energy relative to the reactant state (in kcal/mol) calculated by DFTB/MM , from reactant to Int3 state

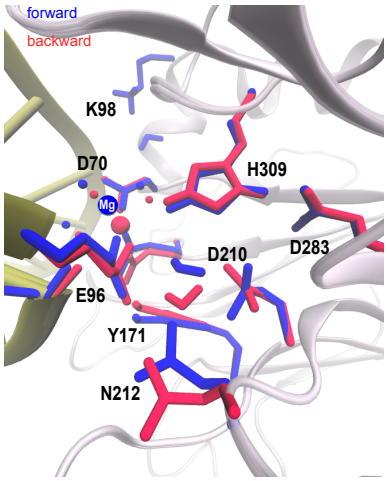

Figure 53: Key amino acid residues in the active site of Ape1-DNA substrate complex in the intermediate step (Int2) for forward(blue) and backward(red) pathways.

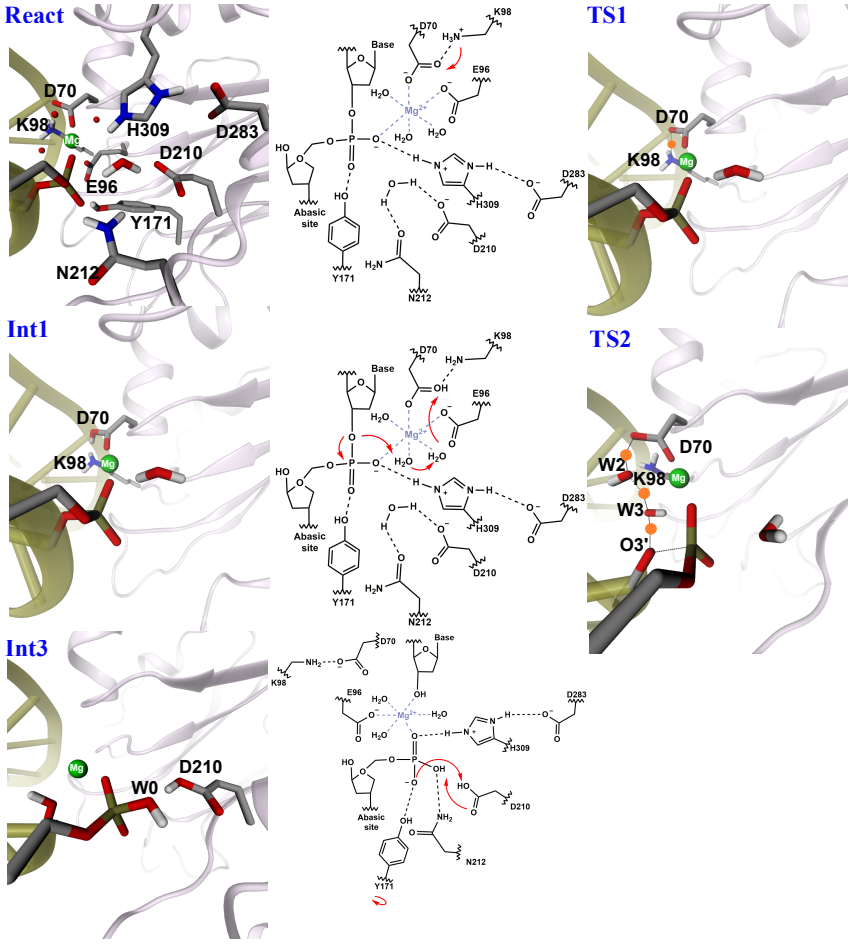

Figure 54: Different steps in Phosphdiester cleavage from reactant state to Int3, only residues involved in respective step are shown for clarity (the transferred proton is highlighted by an orange circle)

6.2.2 Returning to the initial protonation state

The phosphate hydrolysis process is achieved by cleaving the DNA backbone at the abasic site and subsequent nucleophilic attack, that is by reaching Int3 (see Figure 54). The reaction is, however, completed by restoring the protonation state of the amino acid residues in the active site of the APE1 enzyme including D210 and K98. Several probable proton transfer scenarios can be depicted by looking into the active-site geometry of Int3. Therefore for reducing the computational costs of sampling multiple pathways for proton transfer from D210 toward K98, we have first computed minimum energy pathways for a number of possibilities so as to subject only the most tenable ones to subsequent umbrella sampling simulations. Among all calculated pathways only one of them exhibits a transition state with a relative potential energy lower than the 23.6 kcal/mol of the rate determining transition state (TS3).

One of the pathways tested invloves protonated H309 to give a proton to a non-bridging oxygen atom of the phosphate group (O1P) and then accept a proton from D210. This pathway has a barrier of 46 kcal/mol and is thus rather unlikely. Another possibility is transferring the proton located at one oxygen atom of D210 directly to the non-bridging phosphate atom(O2P). No minimum could be trapped along this route. Another possibility is transferring a proton from D210 to the non-bridging oxygen atom of the phosphate group(O2P) while D210 accepts another proton from the former nucleophile OH now part of the phosphate group, (see Figure 55). Our calculations show that these proton transfers take place concertedly with a transition state of 22.4 kcal/mol ($TS4_{PES}$) relative energy. The resulting minimum state ($Int4_{PES}$) has 1.5 kcal/mol potential energy, (see Figure 55 and Table 24). The net effect of this transition is the phosphate group being

protonated at another oxygen atom, O2P. This change in proton position on the phosphate group has an effect on the neighboring Y171. This amino acid donates a hydrogen bond to O2P during the initial steps.

Upon O2P accepting a proton from D210 the hydrogen bond between side chain of Y171 and the O2P atom breaks and the hydroxyl group of Y171 rotates towards the non-coordinated oxygen atom of E96 as new hydrogen bond acceptor. Simultaneously, Y171 accepts a hydrogen atom from the newly protonated O2P (see Figure 55). Furthermore, the phosphate group finally accepts the proton from D210 at the nucleophilic, now deprotonated O2P, (see Figure 55). All these three proton transfer events occur in a concerted fashion, passing a transition state of 14.5 kcal/mol (TS5_P) relative potential energy.

During the next steps, the proton now located on the carboxyl oxygen atom of E96 transfers to the carboxyl oxygen atom of D70, after which the proton is finally transferred from D70 to the K98. These steps take place in a step-wise mechanism with energy barriers of 1.7 and -1.4 kcal/mol relative to the reactant state. The product state is -7.8 kcal/mol exothermic, (see Table 24) and all amino acid residues in the active-site, including K98 and D210, are in the same protonation states as the reactant state but the O3' sugar backbone is cleaved and the nucleophilic water molecule has attacked the phosphate group.

State	Rel.Energy
$Int3_{PES}$	9.6
$TS4_{PES}$	22.4
$Int4_{PES}$	1.5
$TS5_{PES}$	14.5
$Int5_{PES}$	-2.3
$TS6_{PES}$	1.7
$Int6_{PES}$	-3.9
$TS7_{PES}$	-1.4
$Product_{PES}$	-7.8

Table 24: Potential energies (in kcal/mol) relative to the reactant state of intermediates and transitions states from Int3_PES to the product state of the phosphodiester cleavage reaction in the enzyme APE

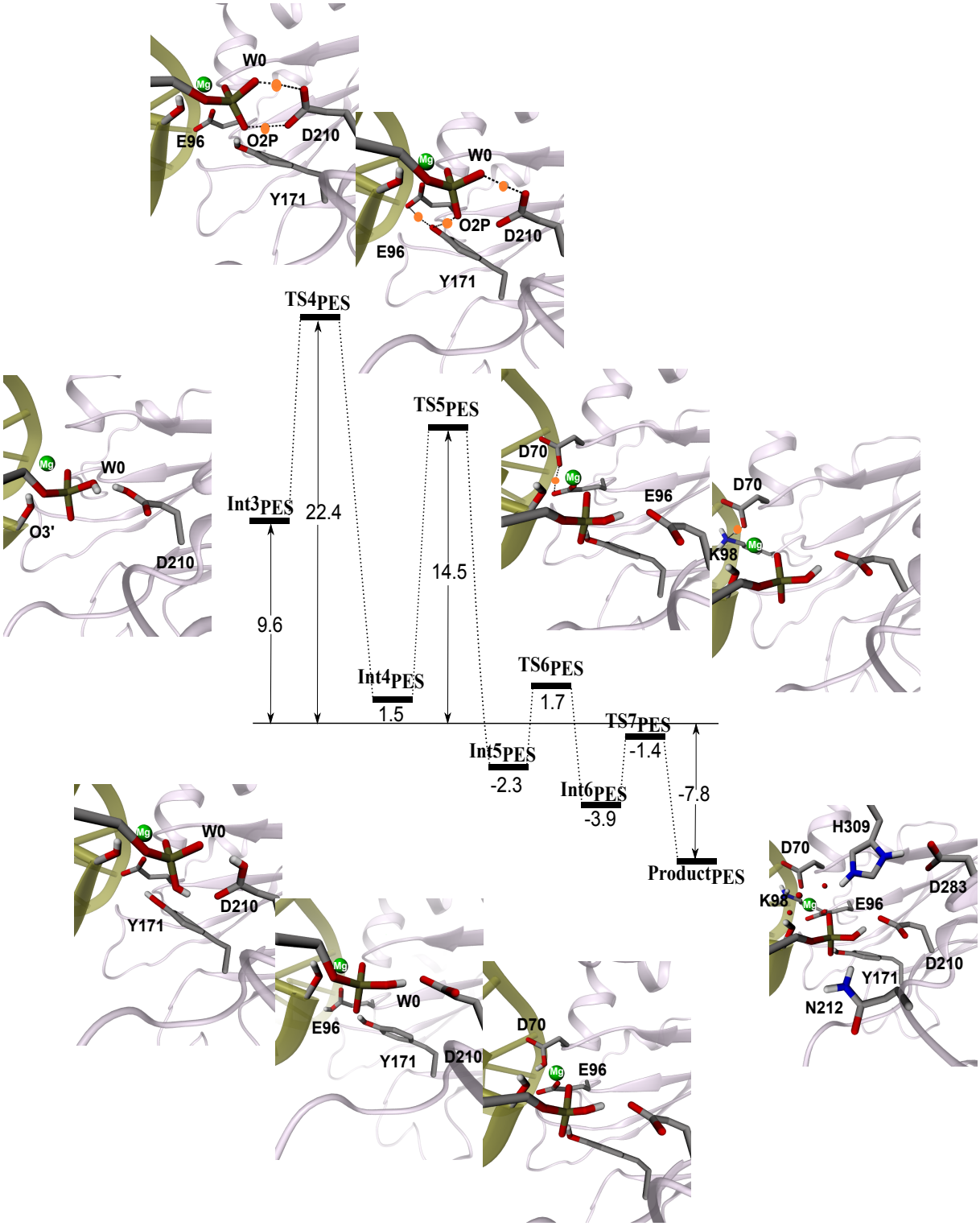

Figure 55: Proton transfer steps in the phosphdiester cleavage mechanism after the actual hydrolysis, i.e. from Int3 to product state (the transferred proton is highlighted by an orange circle)

The same setup and simulation conditions as for the steps of the actual hydrolysis reaction (initial proton transfer, leaving group departure, and nucleophilic attack, i.e. from reactant to Int3, see 6.2) have been used to perform potential of mean force (PMF) calculations for the remaining re-protonation steps (Int3 to product). The reaction coordinates defining the reaction pathway

from Int3 to product state are presented in detail in Table 4 and graphically in Figure 12. The mechanistic details of the proton transfer steps are (by construction) the same as in the potential energy calculations. The free energy barriers for all proton transfer steps (Int3 to Product, i.e. TS4, TS5, TS6, and TS7) are lower than that of the nucleophilic attack, TS3 (see table 25 and Figures 57,56). The free energy of the product state, calculated relative to the reactant state (see Table 25,56) corresponds to an exothermic process with -2.6 ± 0.1 kcal/mol and -4.5 ± 0.0 kcal/mol reaction energy in forward and backward pathways, respectively.

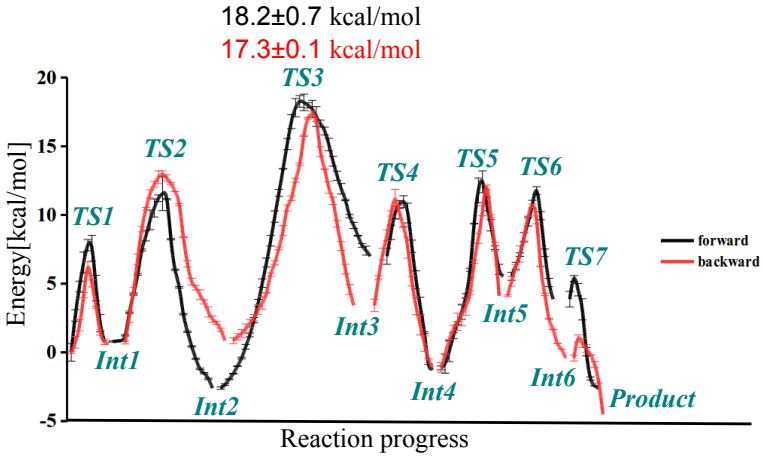

Figure 56: Free energy profile of the phosphodiester cleavage in the APE enzyme calculated by DFTB/MM from reactant to product state in kcal/mol relative to the reactant state

State	Forward	Backward
TS1	7.8 ± 0.4	6.1 ± 0.2
Int1	0.8 ± 0.6	0.7 ± 0.1
TS2	11.5 ± 0.2	12.9 ± 0.5
Int2	-2.6 ± 0.1	0.9 ± 0.2
TS3	18.2 ± 0.7	17.3 ± 0.1
Int3	7.0 ± 0.3	3.4 ± 0.4
TS4	10.9 ± 0.3	11.2 ± 0.2
Int4	-1.13 ± 0.1	-1.1 ± 0.4
TS5	12.4 ± 0.9	10.0 ± 0.4
Int5	5.5 ± 0.4	4.1 ± 0.0
TS6	6.2 ± 0.3	10.7 ± 0.0
Int6	3.9 ± 0.3	-0.4 ± 0.1
TS7	0.1 ± 0.08	1.0 ± 0.2
Product	-2.6 ± 0.1	-4.5 ± 0.0

Table 25: Free energy relative to the reactant state (in kcal/mol) calculated by DFTB/MM, from reactant to product state.

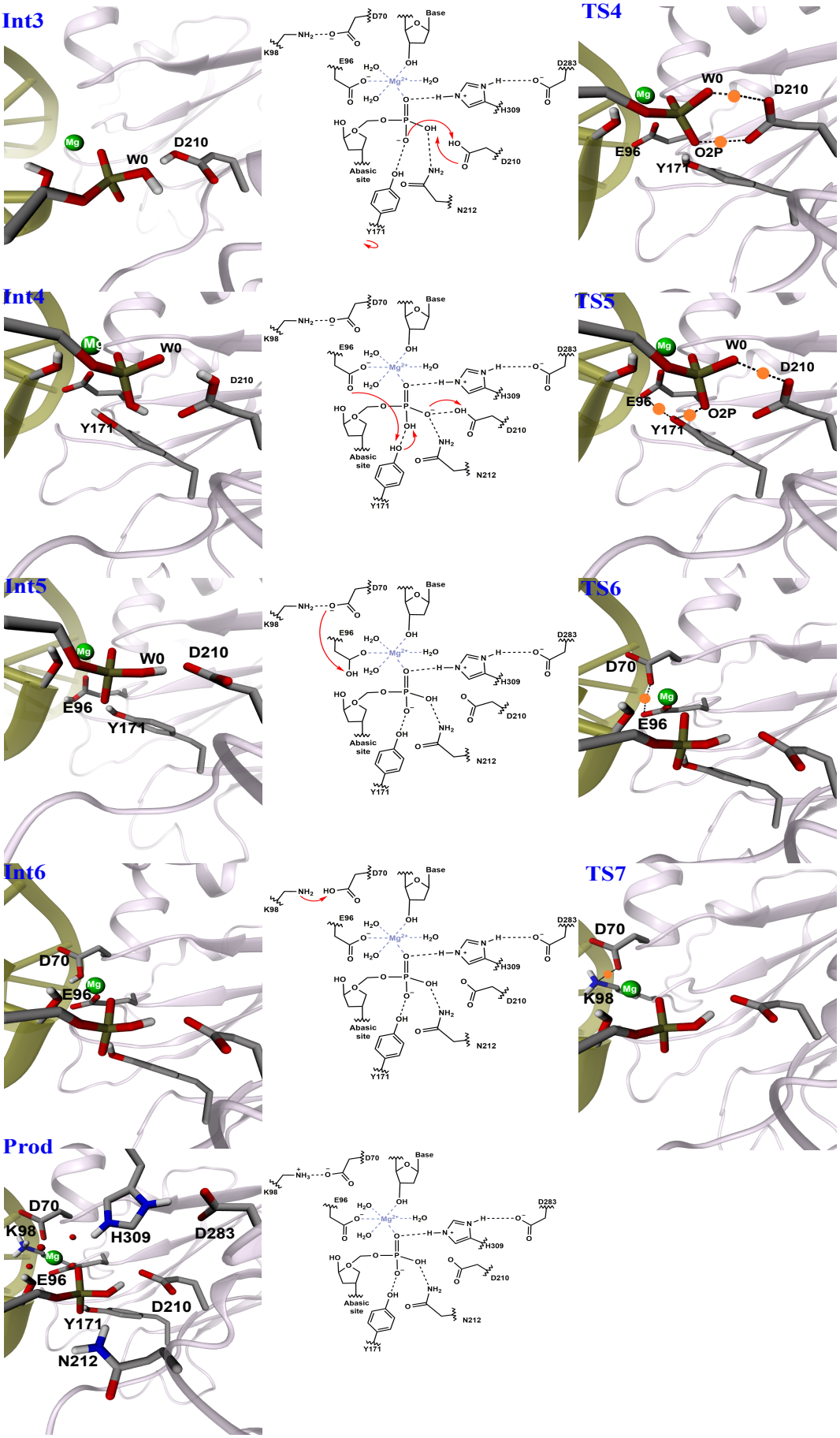

Figure 57: Reaction steps in Phosphdiester cleavage from Int3 to product state (the transferred proton is highlighted by an orange circle)

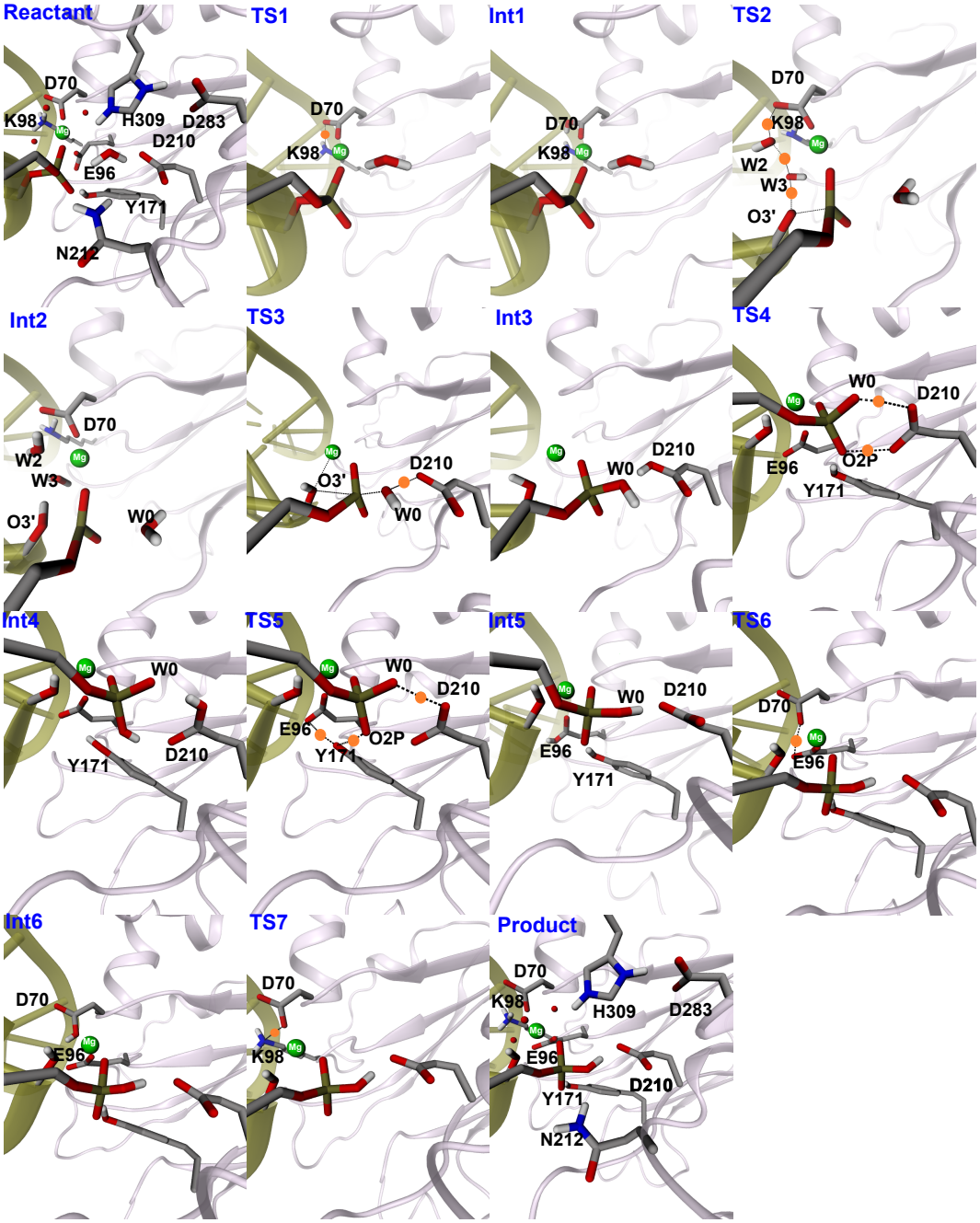

Figure 58: Different reaction steps in Phosphdiester cleavage from reactant to product state calcualted by DFTB/MM.

6.2.3 Charge Analysis

As outlined above, the mechanism for phosphodiester hydrolysis is step-wise and dissociative and contains a number of proton transfer steps. Also the two main steps, P-O3' bond cleavage and P-OH bond formation are accompanied by proton transfers, rendering the phosphate cleavage reaction acid-base catalyzed. The calculated Mulliken charges (shown in Figures 60, 98, 99, 100, and 101) illustrate the charge transfer events along the reaction pathway. The first proton transfer from K98 to D70 could help to accumulate positive charge around the scissile bond, facilitating leaving group departure. There is an accumulation of positive charge around the Mg^{2+} ion, H309, the O3' atom, and the phosphorous atom as well as the whole phosphodiester moiety between Int1 and Int2. This prompts a metal-ligated water molecule to donate a proton to the leaving group.

The elevation of positive charge around D210 (see Figure 98) from React to Int4 corroborates the role of D210 as a general base in accepting a proton from the nucleophilic water molecule.

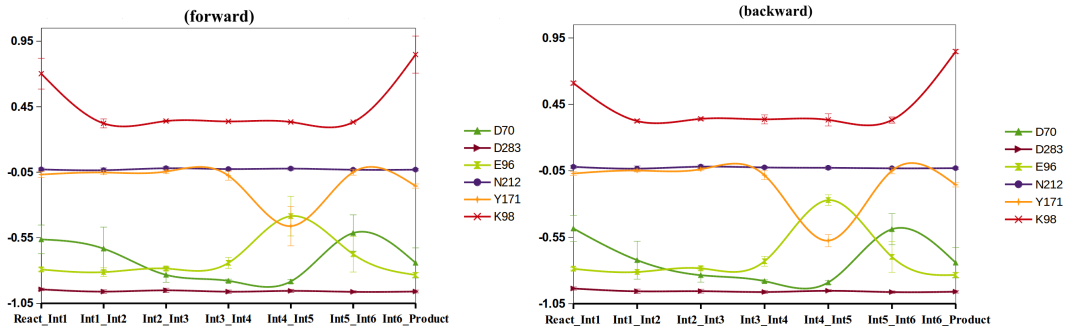

Figure 59: Calculated Mulliken charge for several amino acid residues in the active-site of APE1-DNA complex during the phosphodiester hydrolysis from reactant to product state.

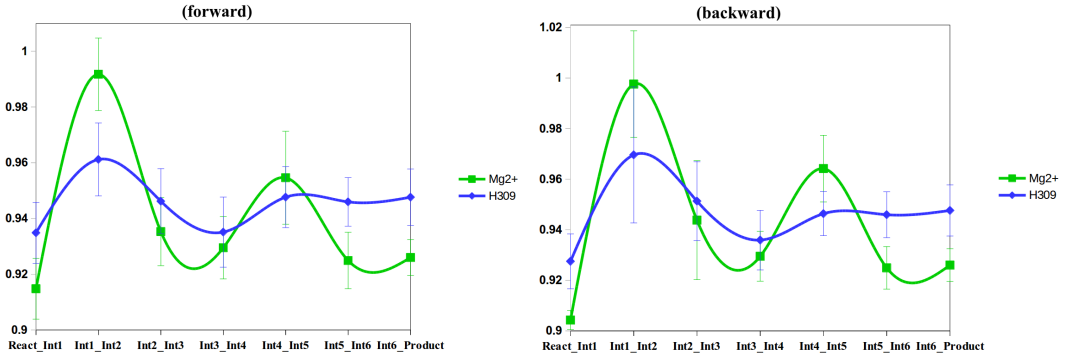

Figure 60: Calculated Mulliken charge for Mg2+ and H309 in the active-site of APE1-DNA complex during the phosphodiaster hydrolysis from reactant to product state.

6.3 Discussion of mechanism and role of active site residues

Phosphodiester hydrolysis catalyzed by the APE1 enzyme follows a step-wise dissociative mechanism in which the nucleophilic attack is the rate determining step. Comparison of the free energy calculated for this rate determining step (18.2 ± 0.67 kcal/mol (forward) or 17.3 ± 0.06 kcal/mol (backward)) to the experimental free energy barrier (18.3 kcal/mol, converted from kinetic rate of DNA cleavege bv APE1)[31, 18, 168] shows a very good agreement supporting our mechanistic proposal. As manifested by the lower energy barriers compared to the uncatalyzed dimethylphosphate hydrolysis (44 kcal/mol as the calculated rate determining step by (B3LYP + 6.311 G**)[120] ([78] and 35.5 kcal/mol as measured experimental activation barrier [75]), the APE1 enzyme achieves a considerable catalytic effect.. In the following, we discuss the individual catalytic contribution of important active site residues.

6.3.1 Role of Lysine98, Lysine78

The initial proton transfer from K98 to D70 and the resulting change in charge distribution has a crucial effect on the reaction by enabling leaving group departure. Substitution of K98A has been shown to result in 12-fold reduction of the 5´-phosphodiester hydrolysis rate in APE [113, 1]. This effect is much less pronounced than those observed for D210, Y171, or H309 mutation, rendering K98 still important but not necessarily essential. In a steady-state kinetic characterization of nucleotide incision repair (NIR)-deficient APE1 mutants [1], K98A, R185A, and double K98A/R185A APE1 mutants exposed a theatrical reduction of both NIR activity on αdA•T by 52-, 220-fold, and an entire loss of activity, respectively. Moreover, the K98A/R185A double mutant resulted in

complete loss of exonuclease activity on Exo20•G [1]. The APE1's K98A/R185A mutant showed 44 fold decrement on THF•T endonuclease activity in BER. As shown by our simulations K78 can provide a proton through transfer to D70 and further via a nearby metal ligated water molecule to the leaving group. This suggests that a dysfunctional K98A mutation can assumably be recompensed with proton transfer from K78 to D70. As our proposed mechanism 40 suggests, K98 or R185 are not indispensable and consequently an K98A/R185A double mutation reduces the endonuclease activity on 5´-phosphodiester hydrolysis not too dramatically. In the reactant state of the APE1-DNA substrate complex, aside from K98 and K78 there are several other positively charged amino acid side-chains such as (H309, K103, R185, R181, R177, R156) within 10 Å of the magnesium ion. These residues have a viable potential to play a catalytic role by altering the charge distribution around the metal ligated water. It should be pointed out that the first proton transfer is not the rate determining step in our proposed mechanism, and hence, any energy barrier in an alternative pathways for providing the positive charge near metal ligated water molecule that is higher than that for the proton transfer from K98 or K78 to D70 but lower than the barrier for nucleophilic attack are still compliant with our proposed mechanism.

6.3.2 Role of Aspartate70, Glutamate96

D70 and E96 have been repeatedly accounted as two amino acid residues which are participating in the Mg^{2+} ion coordination sphere in APE1-DNA substrate complex [99, 91, 168, 133]. In our proposed mechanism for phosphodiester hydrolysis catalyzed by APE1-enzyme showin in Figure 58, D70 and E96 are indeed coordinating the Mg^{2+} ion in the substrate complex, E96 coordinates the Mg^{2+} ion throughout from reactant to product state. However, D70 swaps with the O3' atom of the cleaved sugar backbone in an intermediate state (called Int3) 54 and D70 does not coordinate the Mg^{2+}-ion in the product state. Engagingly, crystal structures showed the same comportment in the activity of D70/E96 in two known structures of the substrate and product state of the APE1-DNA complex [99, 91, 168, 133]. Experimental evidences from D70A, E96A mutations and D70A/E96A double mutation exposed (8.7-,8.4-, 287 fold decrement, respectively) in APE1 endonuclease activity compared with the WT-APE1 [133]. In our proposed mechanism, apart from D70/E96 coordinating the Mg^{2+} ion, we assume that D70 or E96 adopt a proton from K98 or K78 to support a metal ligated water for, in turn, donating a proton to the leaving group (O3') (Figures, 40,41,45). According to our results both D70 and E96 can adopt a proton from K98/K78 smoothly. A hydrogen bond between D70 and K98 can be observed in APE1-DNA substrate complex, rendering D70 more likely to adopt a proton than E96.

As outlined before, in the APE1-DNA substrate complex, the Mg^{2+}-ion coordinates three water molecules, one oxygen atom from the carboxyl group of D70, one of E96 and one non-bridging oxygen atom (O1P) of the AP-site [56, 168], whereas in the APE1-DNA product complex one oxygen atom from the carboxyl group of D70 is replaced as a metal ligand by the cleaved O3' sugar backbone. At the first glance, one of the three water molecules in Mg^{2+}-ion coordination sphere appear to be a more reasonable candidate for swapping place with the O3' atom of the leaving group than a non-bridging oxygen atom of the AP-site or the carboxyl group in both E96/D70 since the latter are negatively charged and as such are supposed to have a stronger interaction with the Mg^{2+}-ion. According to our proposed mechanism the swapping of D70 with the O3' sugar backbone in Mg^{2+}-ion cooridnation sphere, while three metal ligated water molecules are still there, can be explained by the capability of D70 to adopt a proton from K98 to support leaving group departure through donating the adopted proton to a metal ligated water molecule.

6.3.3 Role of Histidine309, Tyrosine171

Both H309 and Y171 are influential in DNA recognition by Ape1 and in the incision step [91, 138, 99, 168, 56]. A catalytic effect of Tyr171 and H309 was suggested mainly based on their initial position in APE, i.e. bound to the abasic DNA. Considerable reduction in catalytic activity in Y171F and H309N (1200-,2500-fold, respectively) relative to wild-type Ape1 [160, 138] depict their notable catalytic functions. According to the initial position of Tyr171 and His309 in the Ape1-DNA substrate complex [168] partial neutralization of the amplifying negative charge through a hydrogen bond with non-bridging oxygen atoms (O1P,O2P) of the AP site is a well-established hypothesis about the catalytic role of H309 and Y171 [168, 56, 91, 99]. The fourfold diminution in

catalytic rate observed for Y171H [160, 138] supports this hypothesis. The (second) Histidine in the Y171H mutant contributes to the mechanisms similarly to Tyr171 by forming hydrogen bonds to the scissile phosphate group .

The reaction path calculations for the H309N and the Y171F mutant show a higher energy barrier for nucelophilic attack of water than in the wild type. As demonstrated in Figure 46, in these mutants non-bridging oxygens (O1P ,O2P) of the AP site can not form hydrogen bonds with H309N, Y171F, that would stabilize the transition state by reducing the partial charge on the AP site.

6.3.4 Role of Asn212

Keeping the nucleophilic water molecule in a right position for direct attack of the phosphate group through a hydrogen-bonded network with D210 and the nucleophilic water molecule has been proposed as the main role of N212 in cleavage of abasic DNA by APE1 enzyme [91, 56], The N212A mutant reveals a notable (7000-fold) deduction in the incision step [91]. This has been discussed by Asn212 to be responsible for keeping water in a proper position for direct nucleophilic attck in our classical molecular dynamic simulations. These data showed that N212 forms a hydrogen bond with the nucleophilic water molecule to preserve it poised for a direct attack [56]. Accordingly, a N212A mutant would rigorously block enzymatic function. Observing the active site of WT APE1 complexed with abasic DNA indicates that N212 can form a hydrogen bond with the non-bridging oxygen atom (O1P/O2P) of abasic DNA.

Our molceular dynamics simulation of wild-type and N212A mutant APE-DNA substrate complex suggest the primary role of N212 to be keeping a water molecule in a right position through hydrogen bonds with the abasic DNA. As our QM/MM calculation show this positioning can also happen in the N212A mutant. However, the significant elevation in the calculated potential energy barrier for the direct nucleophilic attack of a water molecule as the rate determining step in N212A mutant APE1 points towards an additional contribution by Asn212. This corroborates the importance of N212 forming a hydrogen bond with the non-bridging oxygen atom of the abasic DNA (O1P/O2P) and thereby diminishing the elevating charge on abasic DNA upon attack of water.

6.3.5 Role of Asp210

As already mentioned in subsection 5.4.2, D210 acts in positioning the nucleophilic water molecule in a place suitable for direct nucleophilic attack through formation of a hydrogen-bond network with N212 and the nucleophilic water molecule. However, between all roles of D210 discussed in the literature, activating the nucleophile for attack via accepting a proton from a water molecule is the most established one [160, 138, 93, 80]. The big difference between the calculated relative potential energy barrier of D210A (50.8 kcal/mol) and wild type (23.6, see Table 21) emphasizes the important catalytic role of D210 in the active site of APE1 for abasic DNA cleavage. This is in excellent agreement with the experimentally observed 25000-fold reduction in catalytic rate of abasic DNA cleavage in D210A APE1. As can be seen in Figure 49, after mutating D210 to alanine, other accessible amino acid residues around the nucleophilic water are not eligible to play the role of a general base, that is adopting a proton and thus activate the nucelophile. The only attainable choice is an activation of the attacking water molecule via direct transfer of a proton to a non-bridging oxygen atom of the abasic DNA (O1P/O2P). Hence, the presence of D210 is decisive for the nucleophilic attack and thus the entire phosphodiester cleavage mechanism in APE.

Discussion

We have performed quantum mechanical calculations of phosphate hydrolysis reaction pathways based on different cluster models mimicking such active sites of endonuclease enzymes. The similar composition of our cluster models, which are, of course, inspired by the enzymatic active sites, the resulting similarity in the computed most favorable mechanism and the corresponding reaction barrier in the model with experimental, enzymatic values, suggests the active site architecture to be the dominant factor in enzymatic phosphate hydrolysis. Fixing of anchor atoms in our models ensures that the residues of which the models are composed remain positioned as intended, but thereby the flexibility that can be found in an enzymatic active site cannot be mimicked properly. Furthermore, comparison of the reaction barriers for the chemical steps, however, ignores substrate binding (and product release) into the correct position which is an important factor in enzymatic reactions. In addition to investigate the impact of different amino acid residues contributions in the active site of enzyme and generating multiple reaction pathways for phosphodiaster hydrolysis, we have assessed the semi empirical method DFTB3/3OB [132] through comparing calculated reaction pathways in different cluster models with the same reaction pathways calculations on the DFT level of theory (B3LYP/6-31G(d,p)).

Our visions regarding critical impacts of active-site architecture on catalytic effect of endonuclease enzymes for phosphodiester hydrolysis has driven us to start investigations on the exact architecture of Ape1-DNA substrate. By employing Molecular Dynamic simulations, we have simulated and analyzed the APE-DNA reactant complex in different models, varying the position and number of metal ions as well as the protonation state of the essential residues D210 and H309. Augmented by further simulations of the D210N and N212A mutant, which, according to experiments, have a considerably reduced enzymatic activity, our calculations afforded us to rule out some of the simulated models as unlikely due to contradiction with the experimental findings. These are both models with the metal ion located at the attack site, MgA and MgA-MgD, for which the D210N and N212A simulations falsely suggest an active enzyme. From the remaining one-metal models with the Mg^{2+}-ion located at the departure site, MgD appears more favorable than MgF, because of the smaller fluctuations in the DNA conformation, associated with a scissile phosphate group that is closer to the important active-site residues. Our pKa calculations suggest the catalytically essential residue His309 to be predominantly protonated, in agreement with earlier suggestions [40]. This conclusion is further supported by the simulations of the D210N mutant which, in cases of neutral His309, wrongly indicate an active complex. The most probable model is therefore MgD-Hsp. In case of protonated H309, the mutant model does not contain an active-site residue that could activate the nucleophile by accepting a proton, offering an explanation for its inactivity: Asp210 (and not H309) is the general base. On the whole, the most probable model for an Ape1-DNA substrate complex can be seen in Figure 61 wherein His309 is in protonated form and one Mg^{2+}-ion located at binding site D.

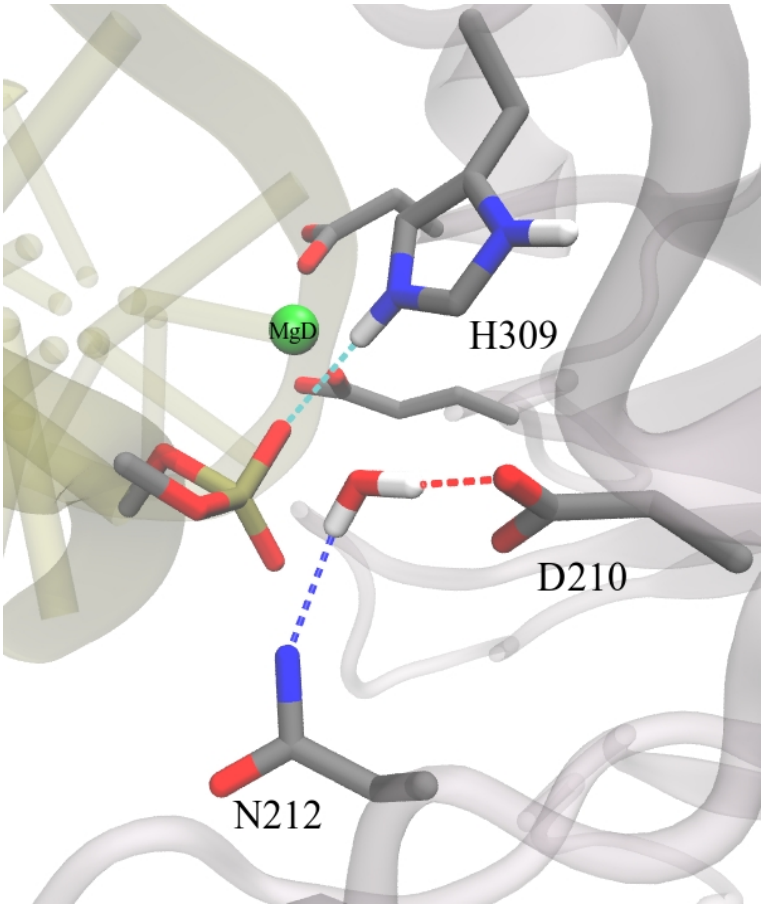

Figure 61: Proposed Active site structure for APE1-DNA complexed with Mg2+ ion [56]

Our widened horizon about the active site architecture in the APE1-DNA complex promoted us to a mechanistic investigation of DNA cleavage catalyzed by APE1 endonuclease. Quantum-mechanical (QM) methods are desired for explaining electronic processes such as charge transfer and chemical reactions. The big hinder for studying bio-molecular systems including more than 1,000 atoms like APE1-DNA complex is the computational cost. Fortunately, force-field-based molecular mechanics (MM) methods provide the opportunity to treat systems including up to 1,000,000 number of atoms for simulating over time scale of hundreds of nano seconds. The combined Quantum Mechanic/Molecular method enabled us to observe the electron transfer during the phosphodiester hydrolysis and treat the surroundings (e.g., protein and solvent) via MM force-field. As we could learn from our cluster model [57] DFTB3/3OB as a semi-empirical method is affordable and accurate enough in comparison with DFT methods to be used in QM/MM calculations. The computed mechanism is as follows: First, in order to provide ample positive charge around the scissile bond, a proton transfers from a charged amino acid residue such as K98 or K78 to the carboxylate group of amino acid residues which are coordinating the magnesium ion such as D70 or E96. Second, the scissile bond is cleaved and a metal ligated water molecule donates a proton to the leaving group whilst it adopts a transferred proton from D70 or E96 (see Figure 54 and Table 23 for more details). Afterwards, the nucleophilic water molecule is activated through donating a proton to D210 with the hydroxyl group attacking the abasic DNA simultaneously. As the calculated free energy relative to the reactant state for this transition is 18.2 $\pm$ 0.7 kcal.mol (forward) or 17.3 $\pm$ 0.1 kcal/mol (backward), this nucleophilic attack can be regarded as the rate determining step of the phophodiester hydrolysis catalyzed by APE1. As suggested by mutational studies and our calculated potential energy pathways for different kind of mutations, Y171, H309, and N212 have a significant catalytic role through balancing the partial charge on the abasic DNA upon nucleophilic attack by forming hydrogen bonds with the non-bridging atoms of the abasic DNA. In addition to keeping the nucleophilic water molecule in a proper position for direct at-

tack together with N212, D210 has the most critical catalytic role in the mechanism. It activates the nueclophilic water molecule by adopting a proton from it. Like depicted in Figure 57, after nucleophilic attack to the dissociative intermediate, several proton transfers occur to restore the protonation states of amino acid residues (including K98 and D210) in the active-site of APE1-DNA (see Table 25 for calculated free energies). Lastly, the relative free energy calculated for the product state is -2.6 ± 0.1 kcal/mol (forward) or -4.5 ± 0.0 kcal/mol (backward) which reveals the exothermic type of reaction.

In section 4, a histidine-assisted associative pathway of the two-metal model with the rate determining step of (∼25 or ∼29 kcal/mol in vacuum and solvent, respectively) relative to the reactant state has been revealed as the most feasible pathway for phosphate hydrolysis by comparison of calculated potential energy pathways based on different cluster models. This result does not confirm or deny whether the dissociative type of mechanism calculated by QM/MM qualifies as the most favorable enzymatic reaction pathway of phosphodiester backbone cleavage by the APE1 enzyme considering that in the active-site of APE1 enzyme substrate multi-architecture can not be observed and it includes H309 in protonated form and only one Mg^{2+}-ion located at binding site D. However, based on quantum calculations on cluster model including His309 in protonated form and one Mg^{2+}-ion located at binding site D proceed through associative type of mechanism with (∼35 or ∼34.3 kcal/mol in vacuum and solvent, respectively) relative to the reactant state (see Figure 15). It implies that the limited number of amino acids mimicking the enzyme environment in the cluster model including H309 in protonated form and one Mg^{2+}-ion located at binding site D could not capture the catalytic effect of the whole enzyme.

In this thesis, combination of classical molecular dynamic and QM/MM calculations have been employed to investigate phosphodiester hydrolysis pathways catalyzed by APE1 endounclease. This approach can be implemented for investigating catalytic effect, active site architecture and chemical reaction pathways for other type of endonuclease enzymes. The semi empirical method DFTB3/3OB as QM part enabled us to significantly reduce computational costs. Hence broad sampling for free energy calculations along the reaction pathway is more feasible in comparison with other quantum mechanical methods like Harte Fock (HF) or Density Functional Theory (DFT).

We hope that further investigations can be carried out on APE1 by development in experimental instruments and computational resources. For instance, double mutations such as K98A, K78A may show the catalytic role of positively charged amino acid residues and their hydrogen bonds network with negatively charged amino acid residues which are coordinating magnesium ion such as E96, D70. To perceive the kinetic basis for the specificity of enzymes involved in BER such as thymine DNA glycosylase (TDG) and APE1 enzyme, each step needs to be fully considered kinetically and thermodynamically. In Base Excision Repair system (BER) the dynamical connections between TDG and APE1 enzymes is still missing. Precisely, the way which TDG enzyme leaves the DNA and APE1 enzyme finds the abasic DNA has not been fully understood yet. Here molecular dynamic simulations combined with a binding analysis may help in better understanding of the swapping process between TDG and APE1 enzyme

Summary

DNA with an abasic site is a cyto-toxic intermediate in the base excision repair (BER) pathway that is handled by the enzyme Apuridinic/Apyrimidinic endonuclease (APE1) [99, 56, 168, 90]. Several kinetics and thermodynamics aspects of the mechanism by which the APE1 enzyme processes its abasic DNA substrate have been discussed in this thesis. APE1 is an endonuclease that is it cleaves the DNA backbone at a non-terminal site, here at the abasic site. To obtain eminent insight about the catalytic role of amino acid residues and magnesium ions which are representatively recognized in active sites of endonuclease enzymes, quantum mechanical calculations of reaction pathways based on various cluster models mimicking such active sites of endonuclease enzymes have been performed and subsequently discussed in section 4. In this light our results underline the importance of an enzymatic active site architecture in the catalytic reaction, given the substrate is properly positioned. As a side-effect, we were able to evaluate the semi empirical method DFTB3/3OB [132] by comparison of reaction pathways calculating in different cluster models with the same reaction pathways calculations on the DFT level of theory (B3LYP/6-31G(d,p)). Comparison of the obtained mechanisms and barriers obtained by DFT and DFTB for the minimalistic (cluster) models may nominate DFTB with reasonable accuracy and computational cost as a potential candidate for quantum method in hybrid QM/MM (quantum mechanics/molecular mechanics) approaches for phosphodiester hydrolysis in an enzymatic environment. However in the "reductionist approach" employed to evaluate multiple plausible types of reaction mechanism, the protein's flexibility and heterogeneous electrostatic environment of protein residues is not taken into account. To enable reaction pathway caculations of the DNA cleavage mechanism in the full APE1 enzyme a model of a reaction-competent APE1-DNA reactant complex has been built, based on available crystal structure data and mutagenesis experiments from the literature [168, 133, 99, 92]. To remedy the lack of lucid information about structural details of APE1/DNA substrate bounded to Mg^{2+} ion molecular dynamic simulation together with pKa calculations for important amino acid residues in the active-site of the enzyme have been carried out (see section 5). To investigate the potential effect of metal ion binding on the stabilization of the active site in the Ape1-DNA substrate complex, the number and position of the metal ion(s) have been varied and single point mutations of vital active-site residues in the substrate complex of Ape1-DNA have been simulated. Taken together, the most likely model for an Ape1-DNA substrate complex has one Mg^{2+}-ion located at binding site D and His309 in protonated form. At the D site, the metal ion may play a catalytic role in leaving group departure. This scenario allows Tyr171 and His309 to form hydrogen-bonds to the phosphate group that may help DNA binding as well as stabilizing a pentacovalent transition state/intermediate. Asn212 acts in properly positioning the nucleophile which can then transfer a proton to Asp210. A combined QM/MM approach wherein the active site was treated quantum mechanically and the remaining enzyme classically empowered us to investigate the enzymatic reaction pathways of phosphodiester backbone cleavage by APE1 enzyme and enabled us to depict a more realistic picture of the enzyme's functions in abasic DNA cleavage (see section 6). Opposed to the "reductionist approach", the QM/MM approach is accurate enough and computationally efficient to gain electronic level insight into the chemical reaction while taking the protein environment explicitly into account and as such allowed the contribution of individual residues in the enzyme environment to be quantified. To find the energetically most favorable pathway for phosphate hydrolysis catalyzed by APE1 enzyme, several possible reaction mechanisms were initially explored by potential energy calculations. As dynamical effects are important in the reaction progress, free energy calculations have subsequently been performed on selected pathways (based on their potential energy profile). According to our QM/MM calculated enzymatic reaction pathways of phosphodiester backbone cleavage by the APE1 enzyme, the dissociative type of mechanism is the most favorable pathway. Herein an important role is played by a metal ligated water molecule that donates a proton to the leaving group while the cleaved sugar backbone migrates toward the Mg2+ ion. This allows a water molecule, activated by proton transfer to Asp210, to attack the phosphorous atom of abasic DNA as nucleophile. The APE1 enzyme as an indispensable key player in BER has been discussed in atomic resolution in this thesis, Our efforts promote the understanding of catalytic and dynamical features of of the APE1 enzyme in the abasic DNA cleavage mechanism, including effects of pH and single-site mutations. The detailed insight thus gained may be helpful in designing inhibitors for APE1 as a potential drug target in cancer chemotherapy.

Zusammenfassung

Abasische DNA ist ein cytotoxisches Intermediat des Basenexizions Reparaturmechanismus (BER), welches durch das Enzym apuridinische/apyrimidinische Endonuklease (APE1) [99, 56, 168, 90] weiter prozessiert wird. In dieser These wurden verschiedene kinetische und thermodynamische Aspekte des Mechanismus, durch welchen APE1 Enzyme ihr abasisches DNA Substrat umwandeln, diskutiert. APE1 ist eine Endonuklease, welche das DNA Rückgrat an nicht-terminaler Position spaltet, in diesem Fall an der abasischen Position. Um tiefere Einsicht in die katalytische Funktion von Aminosäure-Seitenketten und Magnesiumionen, welche representativ für das aktive Zentrum von Endonukleasen stehen, zu erhalten, wurden quantummechanische Berechnungen von Reaktionswegen basierend auf verschiedenen Cluster-Modellen, die das aktive Zentrum von Endonukleasen imitieren, durchgeführt und nachfolgend im Abschnitt 4 diskutiert. In diesem Licht unterstreichen unsere Ergebnisse die Relevanz der Architektur des aktiven Zentrums für die katalytische Reaktion, vorausgesetzt das Substrat ist korrekt positioniert. Als Nebeneffekt waren wir in der Lage die semi-empirische Methode DFTB3/3OB durch den Vergleich von Reaktionswegen berechnet in verschiedenen Cluster-Modellen mit gleichen Reaktionsweg-Parametern auf DFT -Level zu evaluieren (B3LYP/6-31G(d,p)). Der Vergleich, der durch DFT und DFTB erhaltenen Mechanismen und Barrieren für die minimalistischen (Cluster) Modelle, nominiert DFTB [132] mit angemessener Genauigkeit und Rechenkosten als potentiellen Kandidaten für quantummechanische Methoden in hybrid QM/MM (Quantummechaniken/ Molekulare Mechaniken) Ansätzen für Phosphodiester-Hydrolyse in enzymatischer Umbegung. Jedoch wird im oft verwendeten „reduktionistischen Ansatz“ zur Evaluierung multipler plausibler Formen von Reaktions-Mechanismen die Fexibilität des Proteins und die heterogene elektrostatische Umgebung der Protein-Seitenketten nicht berücksichtigt. Um Reaktionsweg-Berechnungen des DNA-Spaltungs-Mechanismus im gesamten APE1 Enzym zu ermöglichen, wurde ein Modell eines Reaktionskompetenten APE1-DNA Edukt-Komplexes gebaut, basierend auf den in der Literatur [168, 133, 99, 92]. vorhandenen Krystallstrukturdaten und Mutagenesis-Experimenten. Um den Fehlen von Informationen über strukturelle Details von APE1/DNA-Substrat gebunden an Mg2+ Ionen entgegenzuwirken, wurden Moleküldynamik-Simulationen zusammen mit pKa Berechnungen für wichtige Aminosäure-Seitenketten im aktiven Zentrum durchgeführt (siehe Abschnitt 5). Zur Untersuchung des potentiellen Effekts der Bindung von Metallionen auf die Stabilisierung des aktiven Zentrum im APE1-DNA Substratkomplex wurde die Anzahl und Position der Metallionen variiert und Punktmutationen wesentlicher Seitenketten des Substratkomplex von APE1-DNA simuliert. Detailierte Analyse dieser Simulationen versorgte uns mit angemessenen Statistiken und Informationen über strukturelle Details des aktiven Zentrums von APE1. Zusammenfassend beiinhalted das wahrscheinlichste Modell für den APE1-DNA Substratkomplex ein Mg2+-Ion lokalisiert an der Bindungsstelle D sowie ein His309 in protonierter Form. And der D Bindungsstelle könnte das Metallion eine katalytische Rolle bei der Abspaltung der Abgangsgruppe einnehmen. Dieses Szenario erlaubt Tyr171 und His309 Wasserstoffbrücken zu den Phosphatgruppen zu formen, welche bei der DNA-Bindung sowie bei der Stabilisierung eines pentakovalenten Überganszustands/ Intermediates behilflich sein könnten. Asn212 spielt bei der korrekten Positionierung des Substrats eine Rolle, welches dann ein Proton auf Asp210 übertragen kann. Ein kombinierter QM/MM Ansatz, in welchem das aktive Zentrum quantummechanisch und das übrige Enzym klassisch betrachtet wurde, ermöglichte uns die enzymatischen Reaktionswege der Phosphodiesterrückgrat-Spaltung durch APE1 zu untersuchen und befähigte uns zu einer realistischeren Abbildung der enzymatischen Funtkionen der abasischen DNA Spaltung (siehe Abschnitt 6). Im Gegensatz zum „reduktionistischen Ansatz“ ist der QM/MM Ansatz präzise genug und rechentechnisch effizient um auf elektorischem Level Einsicht in chemische Reaktionen bei explizieter Berücksichtigung der Proteinumgebung zu erhalten und erlaubt als dieses die Quantifizierung der Beiträge individueller Seitenketten der Enzym-Umgebung. Um den engergetisch favorisierten Reaktionsweg für Phosphathydrolyse katalysiert durch APE1-Enzyme zu indentifizieren, wurden mehrere Reaktionsmechanismen initial durch Potentialenergie-Berechnungen untersucht. Aufgrund der Wichtigkeit dynamischer Effekte für den Reaktionsprozess wurden Berechnungen mit freier Energie stückweise an ausgewählten Reaktionswegen durchgeführt (basierend auf ihrem potentiellen Energieprofil). Zufolge unserer durch QM/MM berechneten enzymatischen Reaktionswege der Phosphodiester Rückgrat-Spaltung durch APE1 Enzyme ist der dissoziative Mechanismus der favorisierte Reaktionsweg. In diesem wird eine wichtige Rolle durch ein Metalliganden gebundenes Wasser-Molekül eingenommen, welches Protonen an die Abgangsgruppe überträgt während das gespaltene Zucker-Rückgrat in Richtung des Mg2+ Ions wandert. Das erlaubt einem Wasser-Molekül, aktiviert durch durch Protonentransfer zu Asp210, als Nukleophil das Phosphoratom der abasischen DNA anzugreifen. Das APE1-Enzym wurde in dieser Thesis als ein unverzichtbarer Schlüsselspieler des BER mit atomarer Auflösung diskutiert. Unsere Bemühungen fördern das Verständnis der katalytischen und dynamischen Eigenschaften des APE1 Enzyms im abasischen DNA Spaltungs-Mechanismus unter Einbeziehung von Effekten durch pH und Punktmutationen. Der hierdurch erhaltene Einblick könnte für das Design von Inhibitoren für APE1 als potentielles Ziel für Therapeutika in der Chemotherapie hilfreich sein.

Supplementary

Section A: Cluster models

Partial atomic (NBO) charges

One-metal models

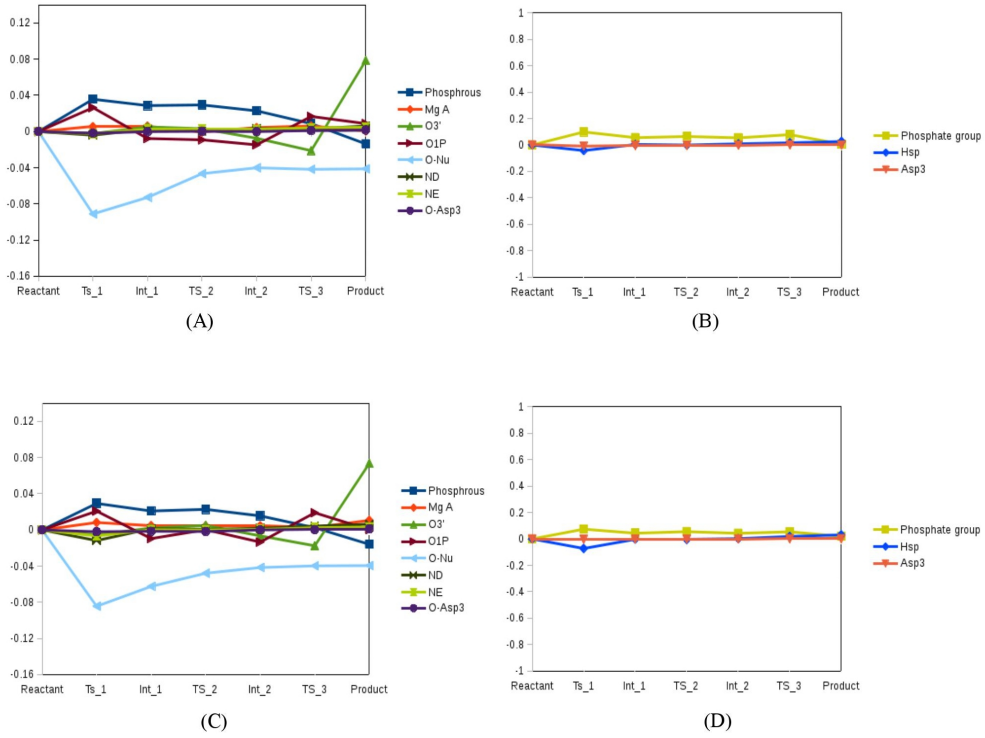

Figure 62: NBO charge (in a.u.) calculated by B3LYP/6-31G(d,p) for the 1MgD-Hsp-Glu-1Asp-1Asp model in vacuum (A-B), and in solvent (C-D). All NBO charges are given relative to the reactant state.

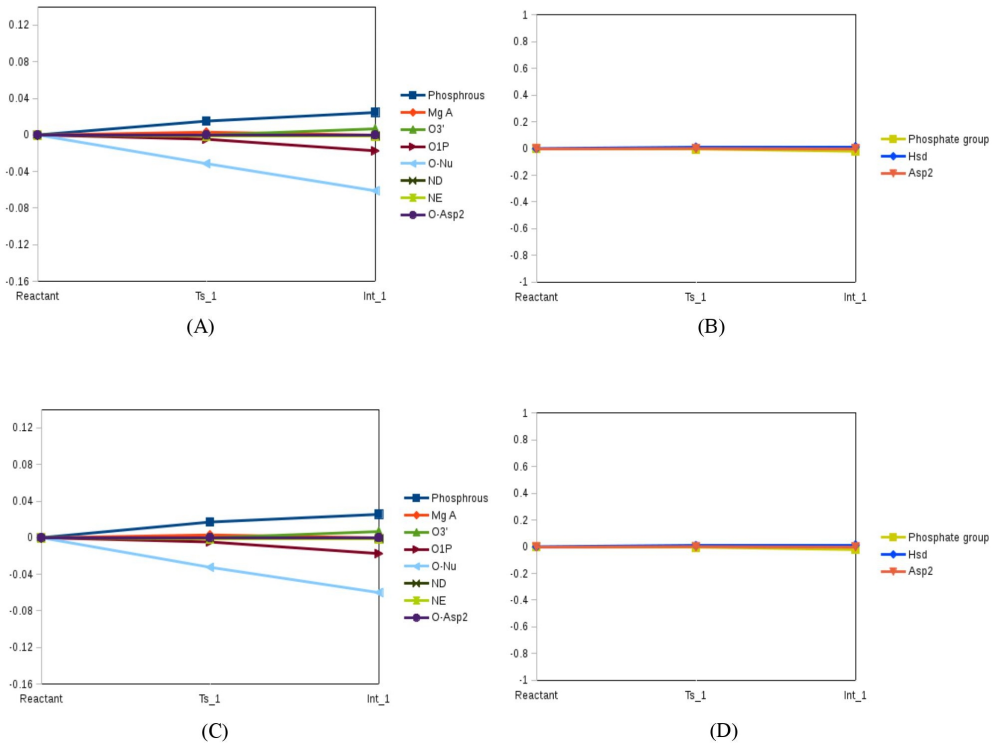

Figure 63: NBO charge (in a.u.) calculated by B3LYP/6-31G** for the 1MgD-Hsd-Glu-1Asp-1Asp model (direct pathway) in vacuum (A-B) and in solvent (C-D). All NBO charges are given relative to the reactant state.

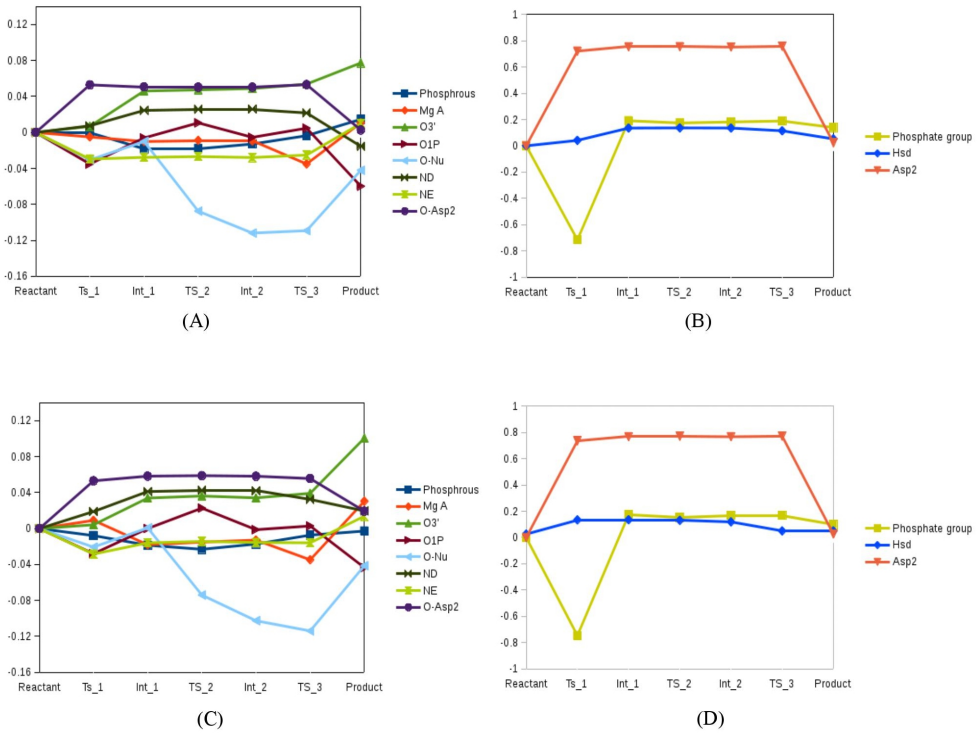

Figure 64: NBO charge (in a.u.) calculated by B3LYP/6-31G(d,p) for the 1MgD-Hsd-Glu-1Asp-1Asp model (histidine-assisted pathway) in vacuum (A-B) and in solvent (C-D). All NBO charges are given relative to the reactant state.

Two-metal models

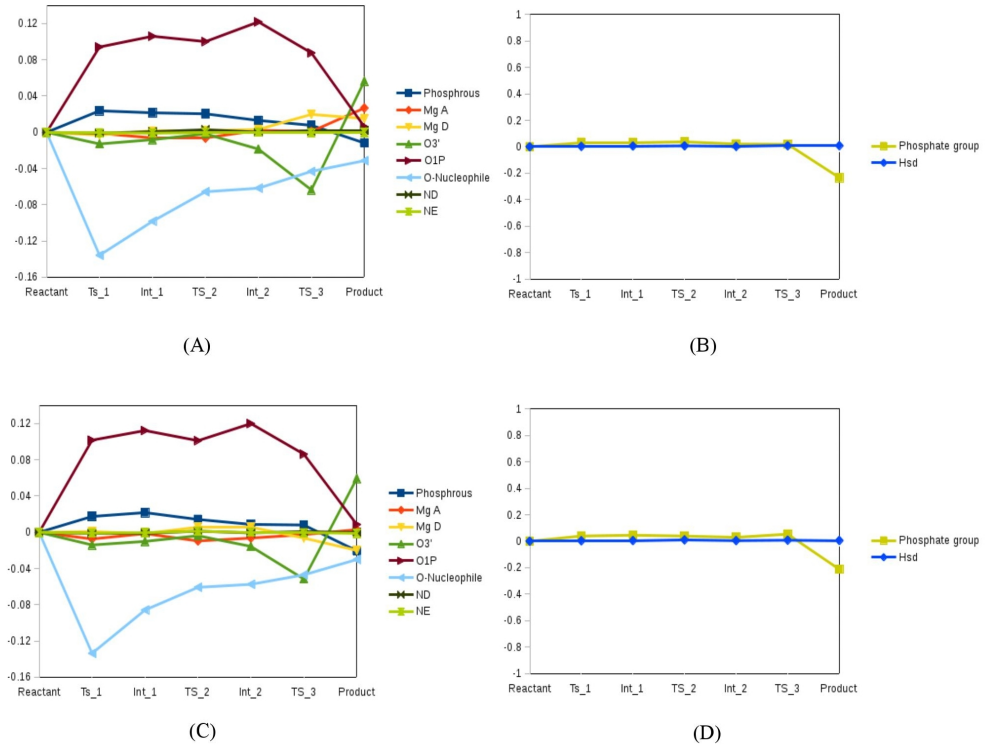

Figure 65: NBO charge (in a.u.) calculated by B3LYP/6-31G(d,p) for the 2Mg-Hsd-Glu-3Asp model in vacuum (A-B) and in solvent (C-D). All NBO charges are given relative to the reactant state.

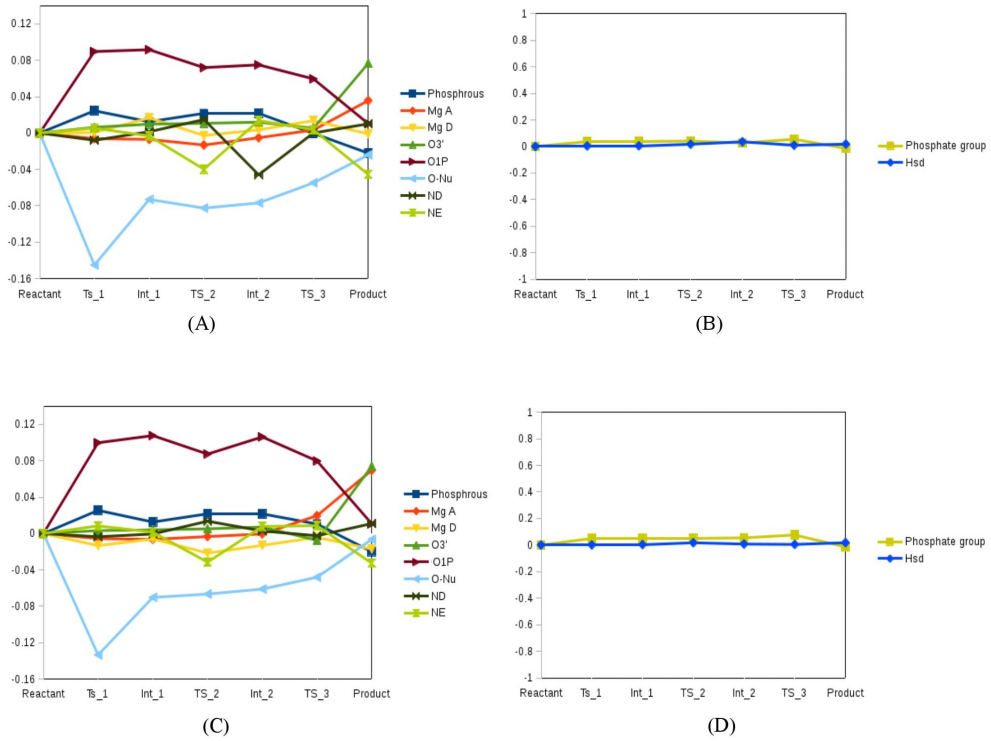

Figure 66: NBO charge (in a.u.) calculated by B3LYP/6-31G(d,p) for the 2Mg-Hsd-Glu-2Asp model in vacuum (A-B) and in solvent (C-D). All NBO charges are given relative to the reactant state.

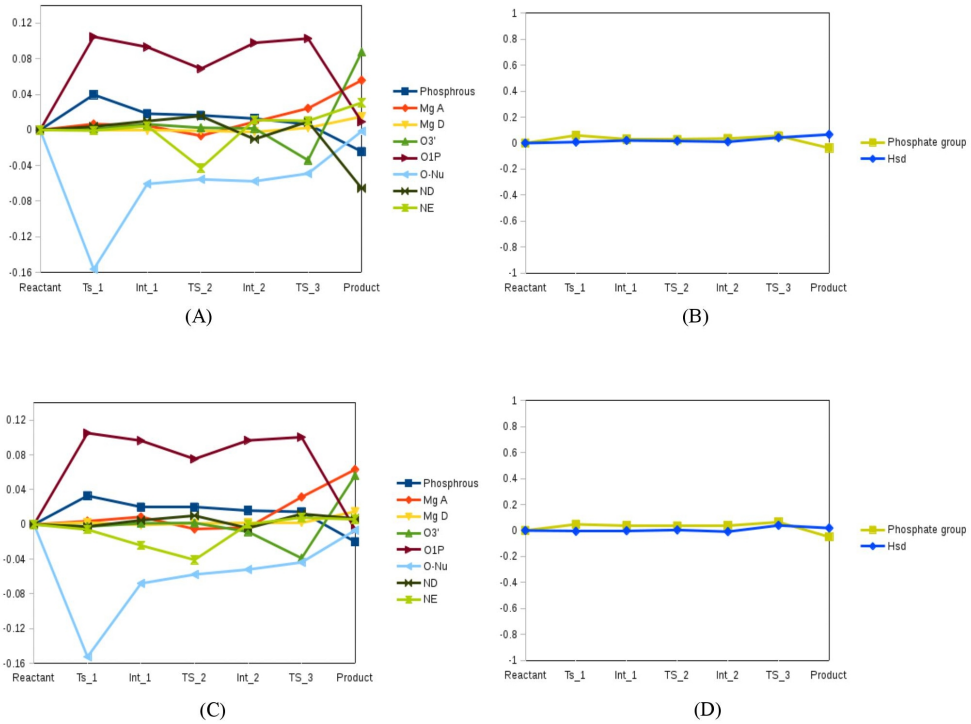

Figure 67: NBO charge (in a.u.) calculated by B3LYP/6-31G(d,p) for the 2Mg-Hsd-Glu-1Asp model in vacuum (A-B) and in solvent (C-D). All NBO charges are given relative to the reactant state.

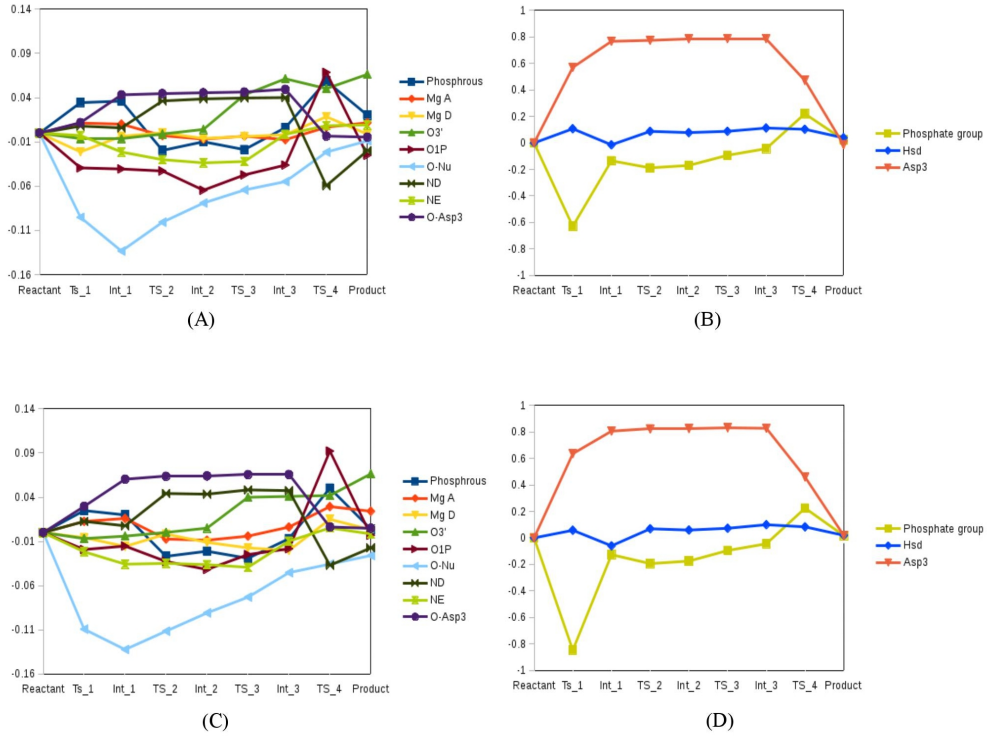

Figure 68: NBO charge (in a.u.) calculated by B3LYP/6-31G(d,p) for the 2Mg-Hsd-Glu-2Asp-1Asp model in vacuum (A-B) and in solvent (C-D). All NBO charges are given relative to the reactant state.

Fluctuations

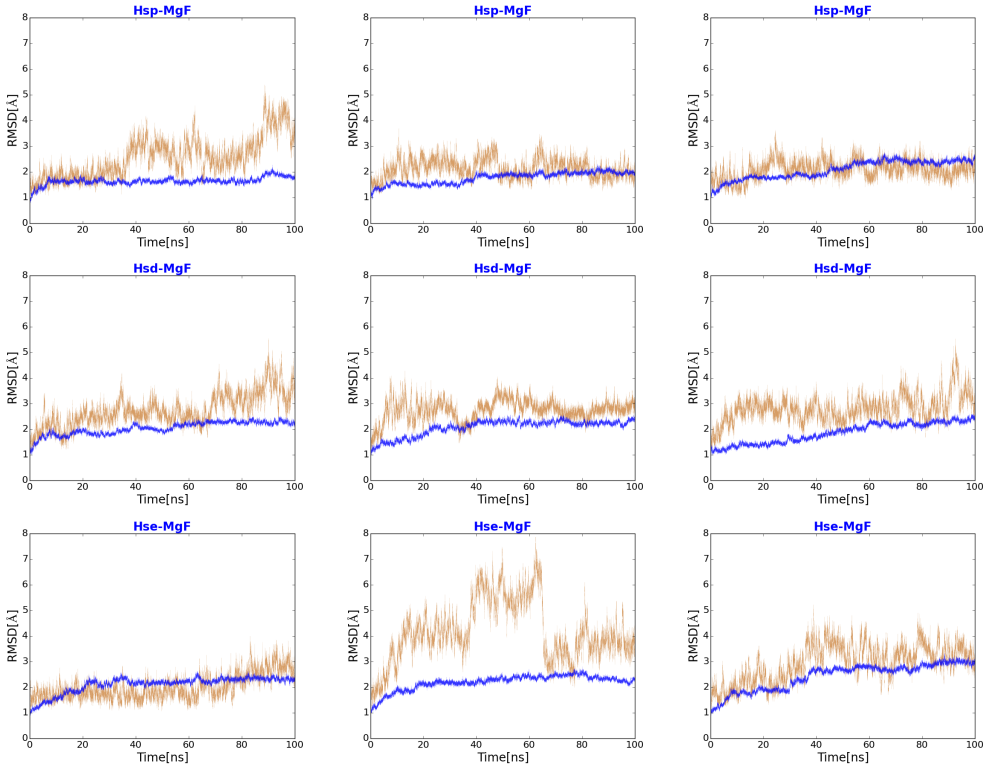

Figure 69: Conformational change from solvated and relaxed crystal structure of the Ape1-DNA substrate complex in the protein (blue) and the DNA (orange) structures along the MD simulation trajectories of the MgF models as quantified by RMSD (in Å). The three plots per model correspond to the three individual simulations performed.

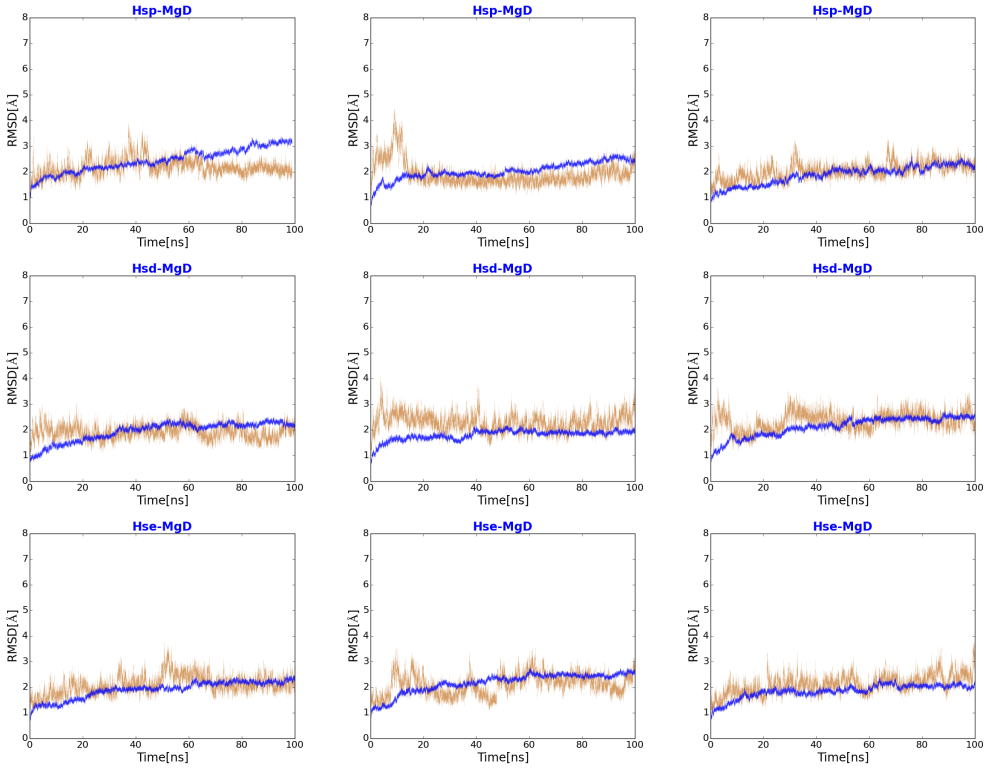

Figure 70: Conformational change from solvated and relaxed crystal structure of the Ape1-DNA substrate complex in the protein (blue) and the DNA (orange) structures along the MD simulation trajectories of the MgD models as quantified by RMSD (in Å). The three plots per model correspond to the three individual simulations performed.

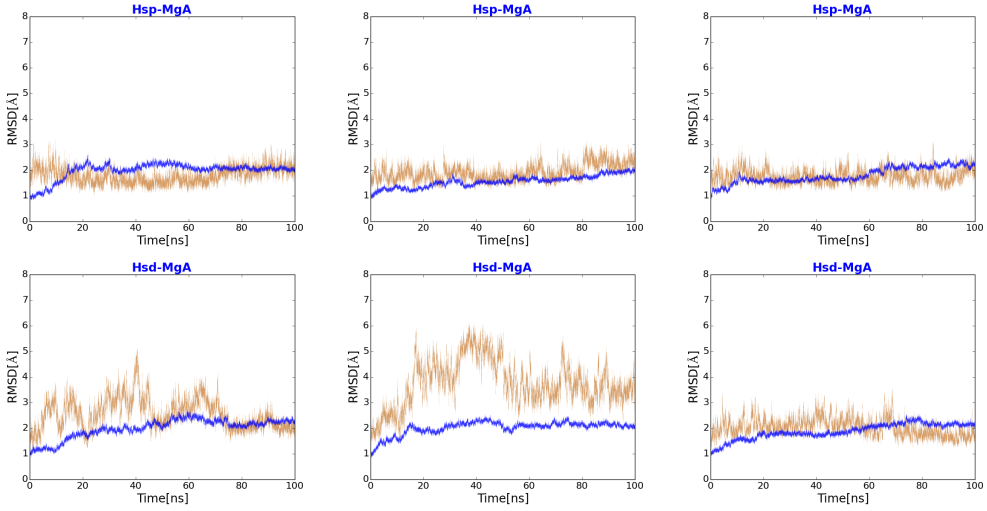

Figure 71: Conformational change from solvated and relaxed crystal structure of the Ape1-DNA substrate complex in the protein (blue) and the DNA (orange) structures along the MD simulation trajectories of the MgA models as quantified by RMSD (in Å). The three plots per model correspond to the three individual simulations performed.

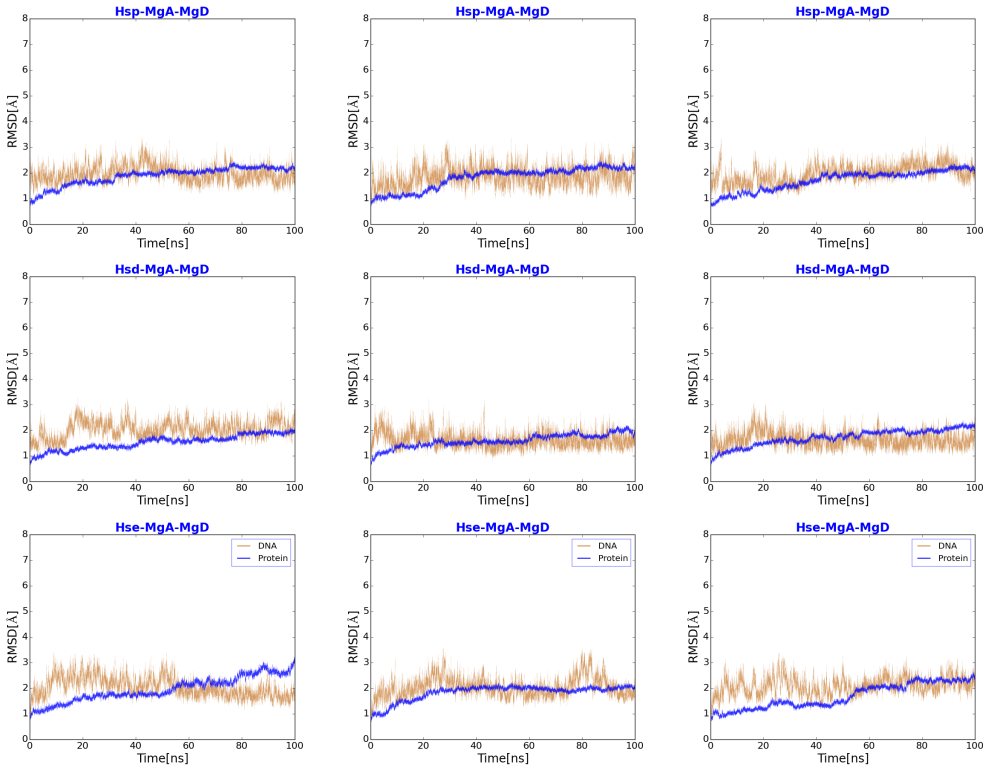

Figure 72: Conformational change from solvated and relaxed crystal structure of the Ape1-DNA substrate complex in the protein (blue) and the DNA (orange) structures along the MD simulation trajectories of the MgA-MgD models as quantified by RMSD (in Å). The three plots per model correspond to the three individual simulations performed.

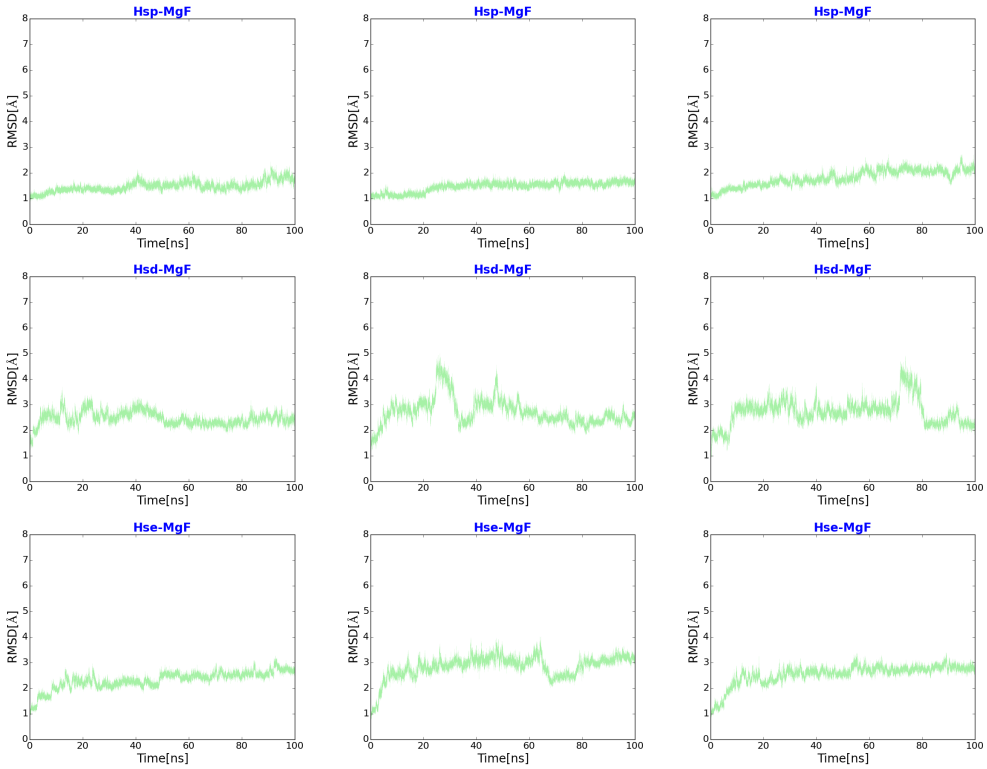

Figure 73: Active-site conformational change from solvated and relaxed crystal structure of the Ape1-DNA substrate complex along the MD simulation trajectories of the MgF models as quantified by RMSD (in Å). The three plots per model correspond to the three individual simulations performed.

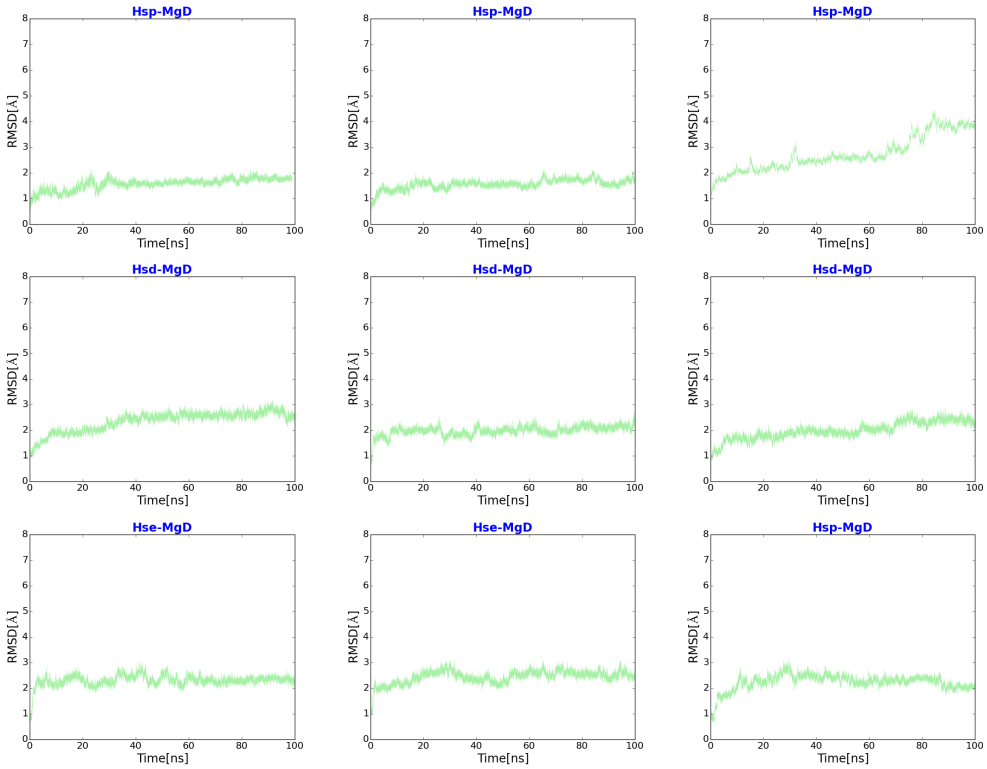

Figure 74: Active-site conformational change from solvated and relaxed crystal structure of the Ape1-DNA substrate complex along the MD simulation trajectories of the MgD models as quantified by RMSD (in Å). The three plots per model correspond to the three individual simulations performed.

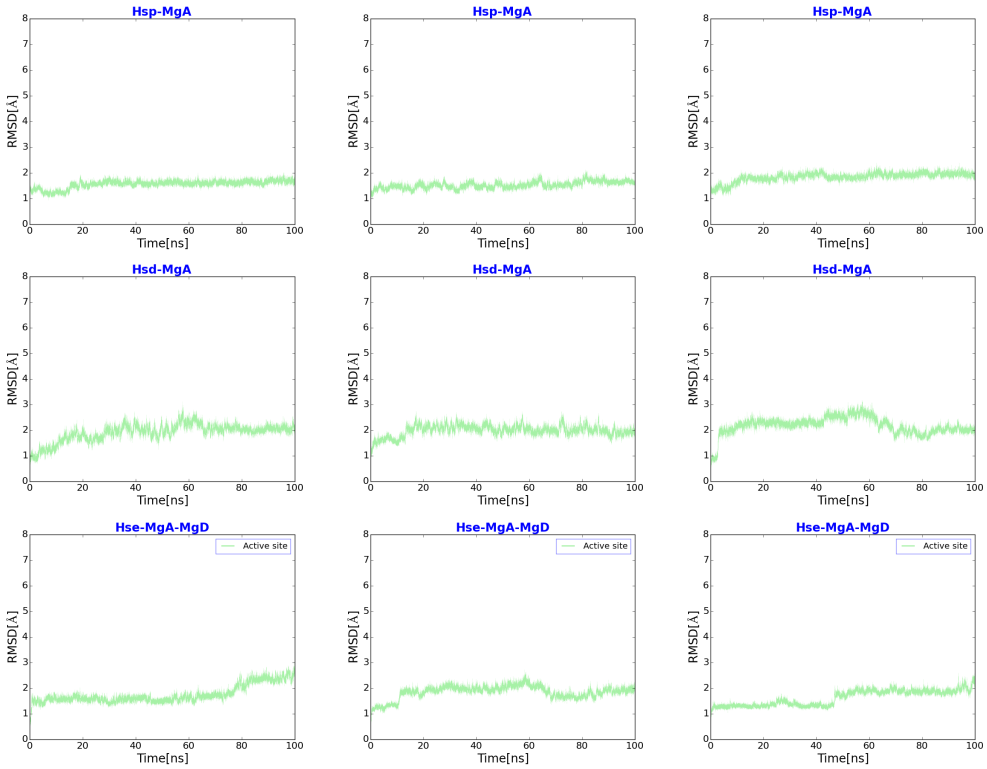

Figure 75: Active-site conformational change from solvated and relaxed crystal structure of the Ape1-DNA substrate complex along the MD simulation trajectories of the MgA models as quantified by RMSD (in Å). The three plots per model correspond to the three individual simulations performed.

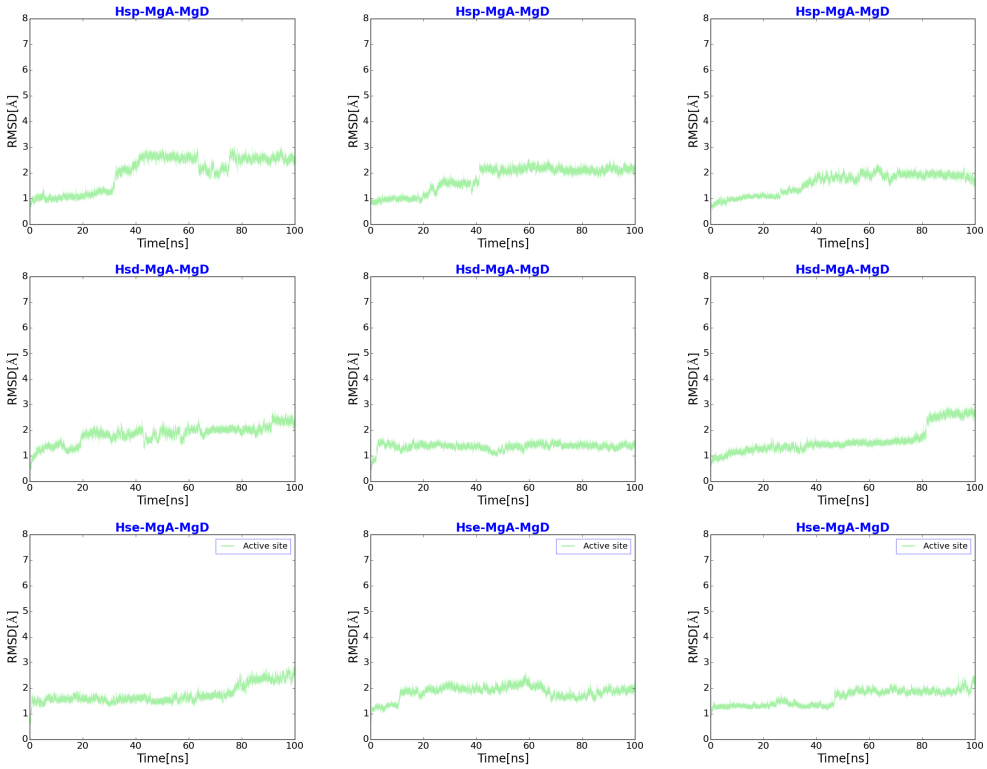

Figure 76: Active-site conformational change from solvated and relaxed crystal structure of the Ape1-DNA substrate complex along the MD simulation trajectories of the MgA-MgD models as quantified by RMSD (in Å). The three plots per model correspond to the three individual simulations performed.

Distances between active-site residues

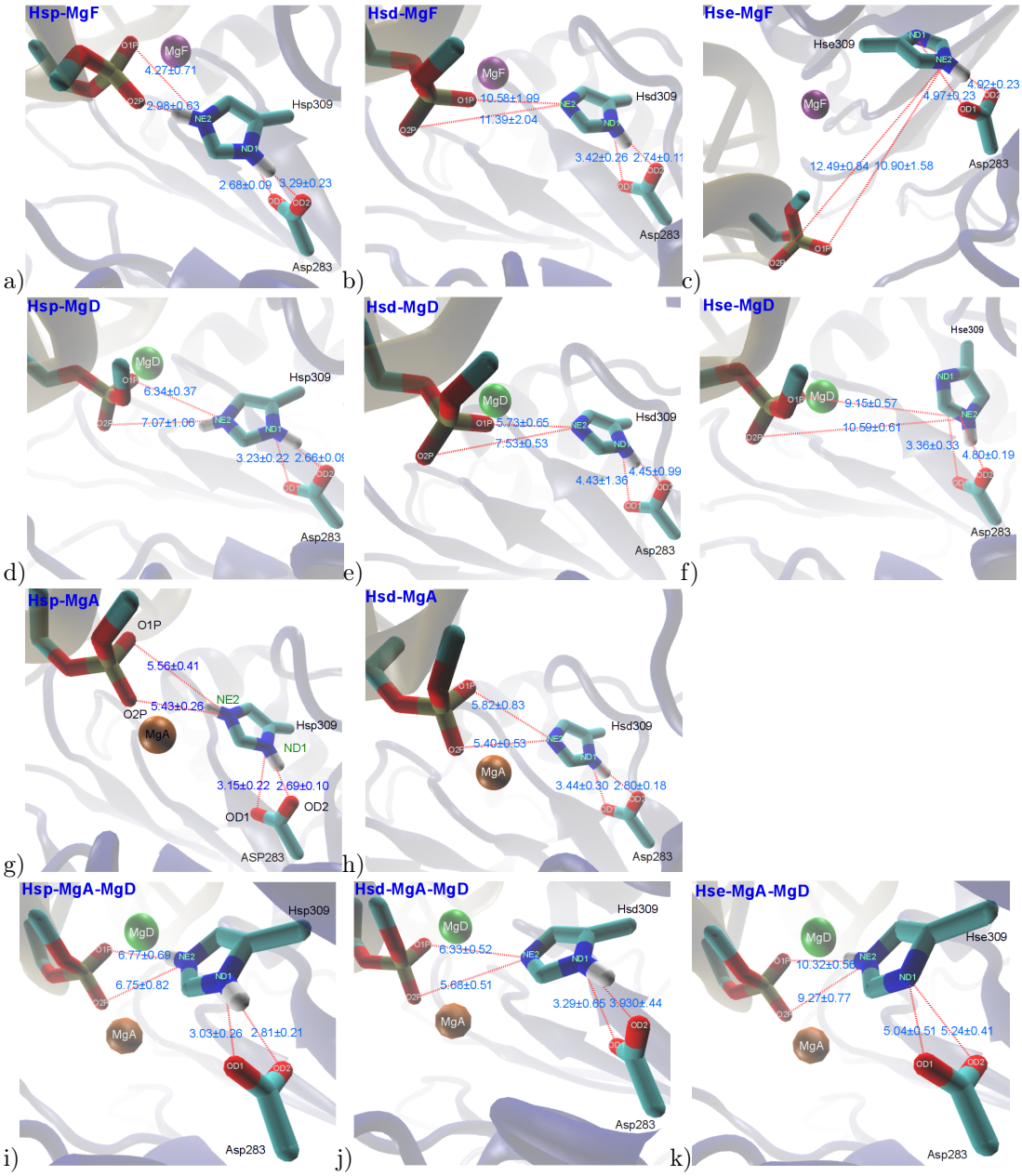

Figure 77: Average distances (in Å) between His309(NE2) and Asp283(OD2,OD1), and between His309(NE2) and non-bridging oxygen atoms of the AP site (O1P,O2P). Other active site residues and water molecules are omitted for clarity

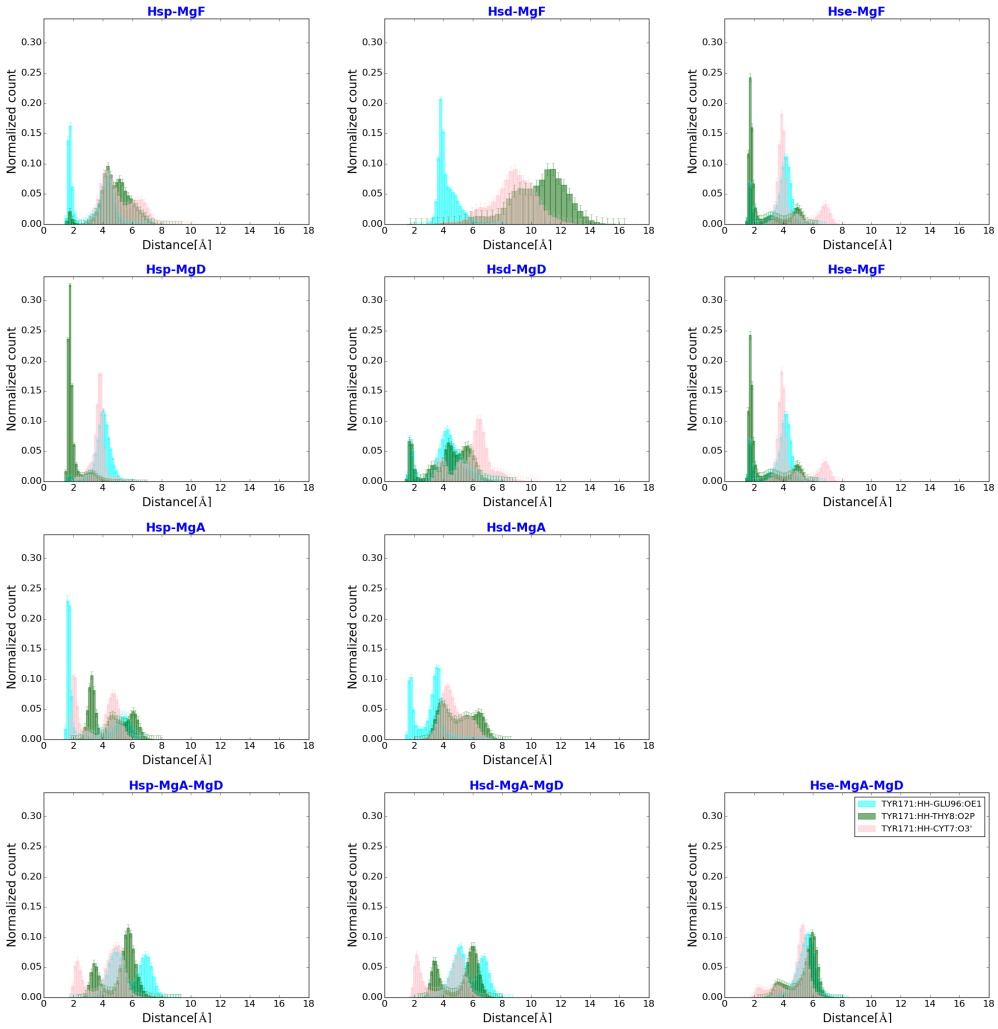

Figure 78: Distribution of the distances (in Å) between hydrogen atom HH of Tyr171 and the O3'-atom of the scissile bond (red) , the nonbridging oxygen atom (O2P) of the AP site (green), and one oxygen atom , OE1 of Glu96 (cyan)..

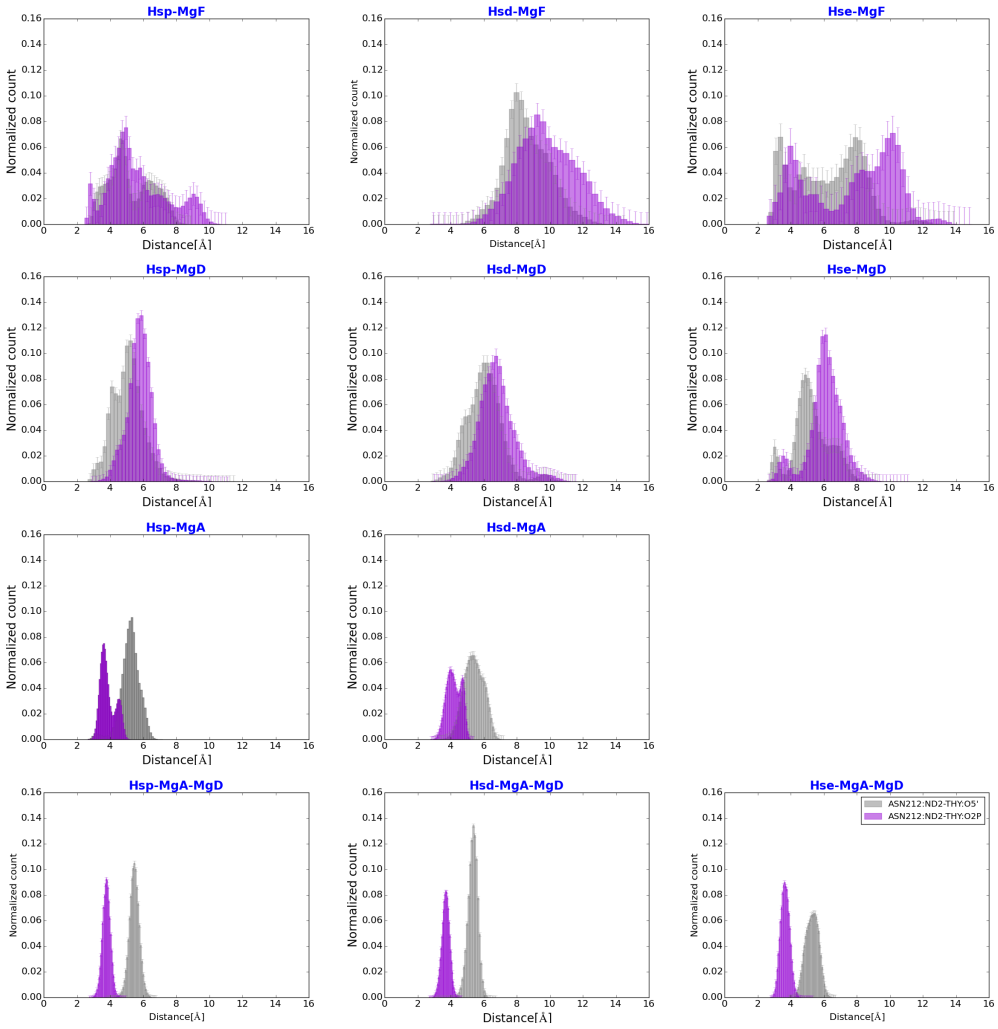

Figure 79: Distribution of the distances (in Å) between the nitrogen(ND2)-atom of Asn212 and the bridging oxygen (O5')-atom of the AP site (grey), and the nonbridging oxygen atom (O2P) of the AP site (purple).

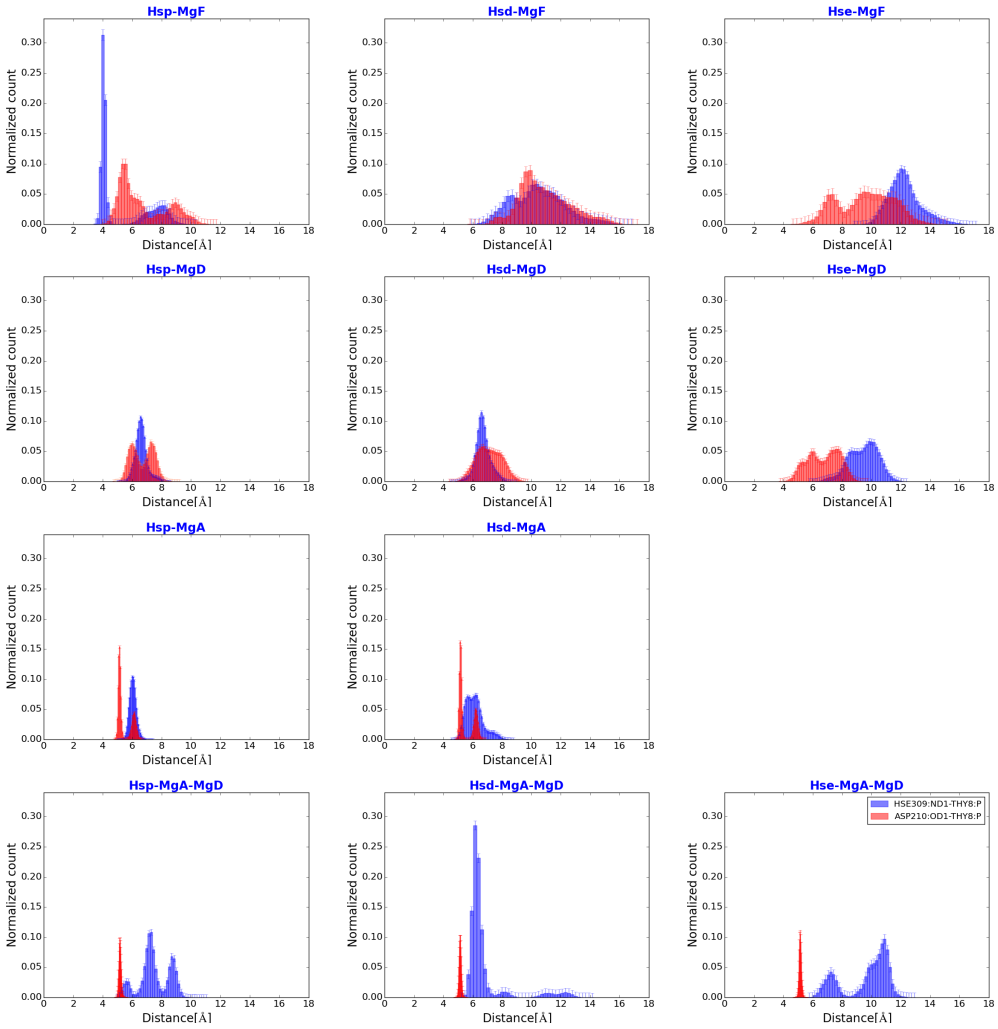

Figure 80: Distribution of the distances (in Å) between the phosphorous atom of the AP site and the oxygen atoms (OD1,OD2) of Asp210 (red), and the nitrogen atomNE2 of His309 (blue).

Hydrogen Bonds

Table 26: Hydrogen bond occupancies for active-site residues. AP is the abasic site. Only values with >60% in at least two of the three individual MD runs are listed. 100% occupancies are reached by several water molecules, one or two at a time, forming a hydrogen bond to the respective residue.

	MgF			MgD		
	Hsp	Hsd	Hse	Hsp	Hsd	Hse
Asn212–Asp210						
His309–Asp308		85.9±1.5	79.5±14.3			
His309–Asp283		95.6±3.0		95.8±1.5		
Tyr171–AP-O2P				86.9±2.5		
Asp70--wat	100.0±0.0			84.9±26.2	96.3±6.5	100.0±0.0
Glu96--wat	100.0±0.0	94.7±9.1		72.8±20.6	100.0±0.0	100.0±0.0
Tyr171--wat				83.5±14.3		74.0±5.0
Asp210--wat	95.9±7.2			100.0±0.0		100.0±0.0
Asn212--wat				100.0±0.0		
His309–wat				82.4±8.1		
Asp283–wat						
AP-O1P–wat	100.0±0.0	100.0±0.0	100.0±0.0			
AP-O2P--wat	89.7±17.8	100.0±0.0	85.6±16.3	100.0±0.0	88.6±17.5	100.0±0.0

	MgA		MgA-MgD		
	HSP	HSD	HSP	HSD	HSE
Asn212–Asp210				68.4±15.7	71.8± 6.0
His309–Asp283	89.2±4.8	95.7±0.6			
Tyr171–AP-O2P					
Asp70--wat	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0	100.0±0.0
Glu96--wat	100.0±0.0	100.0±0.0			
Tyr171--wat			86.0±19.1	76.8±25.7	100.0±0.0
Asp210--wat					
Asn212--wat					
His309–wat	85.9±1.3	100.0±0.0	80.9±16.5	85.1±23.8	
Asp283–wat			76.0±41.6		98.0±3.5
AP-O1P--wat	100.0±0.0	99.2±1.4			
AP-O2P--wat					

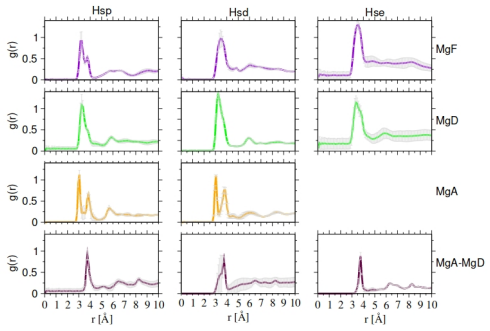

Figure 81: Radial distribution function, g(r) of water oxygen atoms around the CG atom of Asp210 in the different models.

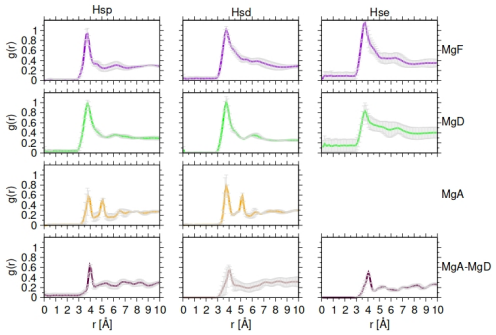

Figure 82: Radial distribution function, g(r) of water oxygen atoms around the CG atom of Asn212 in the different models.

Water distribution

pKa values

$\varepsilon = 4$

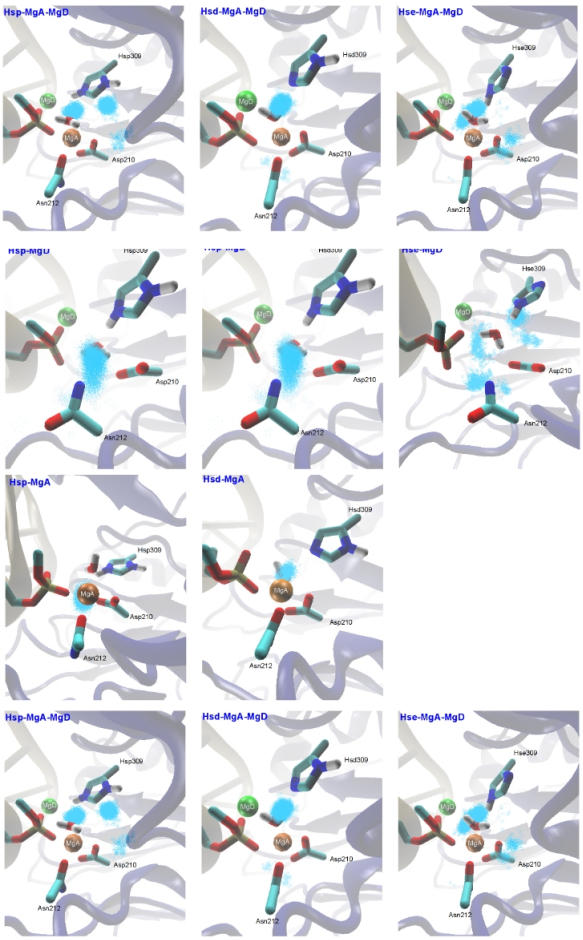

Figure 83: Water oxygen atom positions around Asp210 as shown by cyan dots. For clarity only positions from the last 40 ns of the simulation time are shown.

Table 27: Average pKa values calculated from snapshots of the different Ape1-DNA complex models using a dielectric constant $\varepsilon = 4$.

Model	Asp210	Asp283	His309
Hsp-MgF	3.30 ± 1.05	-6.03 ± 0.99	15.90 ± 1.26
Hsd-MgF	3.94 ± 0.80	-4.16 ± 1.45	10.19 ± 0.80
Hse-MgF	5.08 ± 1.80	-1.30 ± 1.97	6.83 ± 1.86
Hsp-MgD	3.69 ± 1.10	-5.45 ± 1.10	13.34 ± 0.87
Hsd-MgD	3.57 ± 0.98	-2.53 ± 1.90	9.78 ± 0.86
Hse-MgD	5.66 ± 2.48	-2.21 ± 1.18	9.77 ± 1.51
Hsp-MgA	< -10	-3.87 ± 1.14	13.21 ± 0.65
Hsd-MgA	< -10	-3.88 ± 1.11	10.71 ± 0.49
Hsp-MgA-MgD	< -10	-2.86 ± 2.14	11.05 ± 1.13
Hsd-MgA-MgD	< -10	-2.76 ± 2.17	11.73 ± 1.36
Hse-MgA-MgD	< -10	-3.52 ± 2.05	3.80 ± 2.25

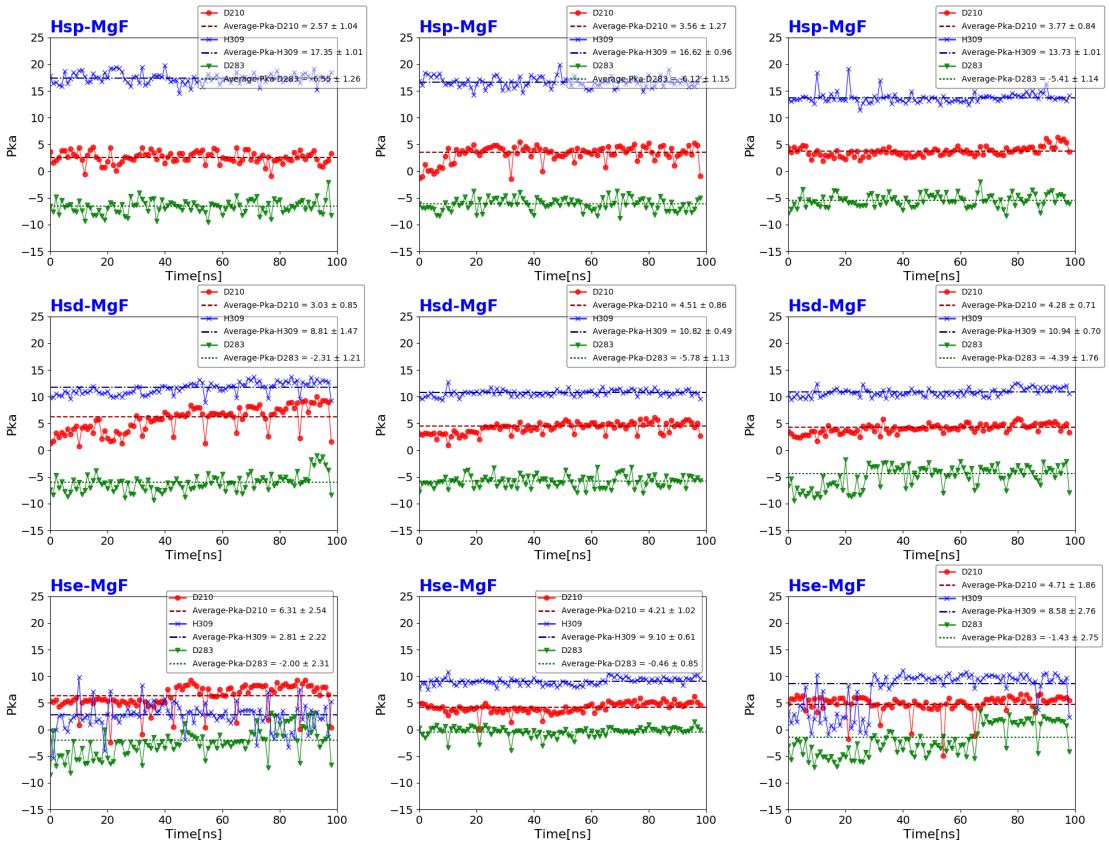

Figure 84: Calculated pKa values for all snapshots of MgF models using a dielectric constant $\varepsilon = 4$. The three plots per model correspond to the three individual simulations performed.

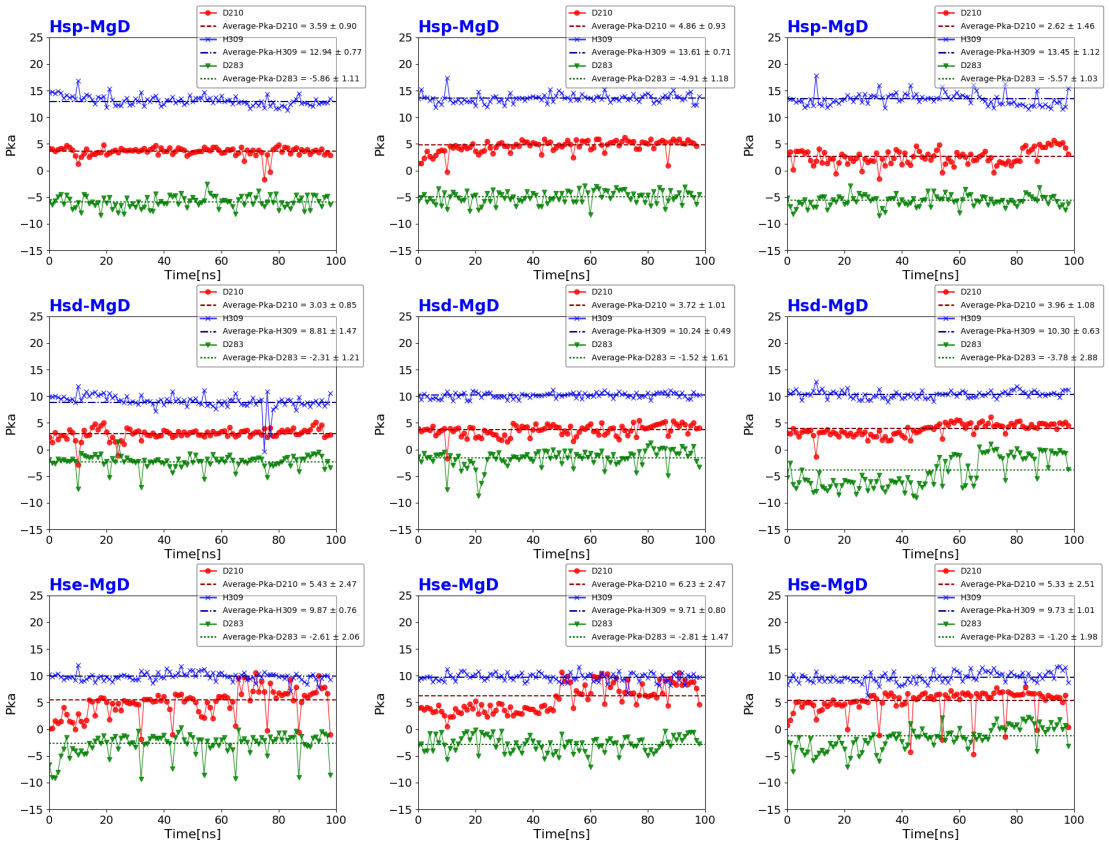

Figure 85: Calculated pKa values for all snapshots of MgD models using a dielectric constant $\varepsilon = 4$. The three plots per model correspond to the three individual simulations performed.

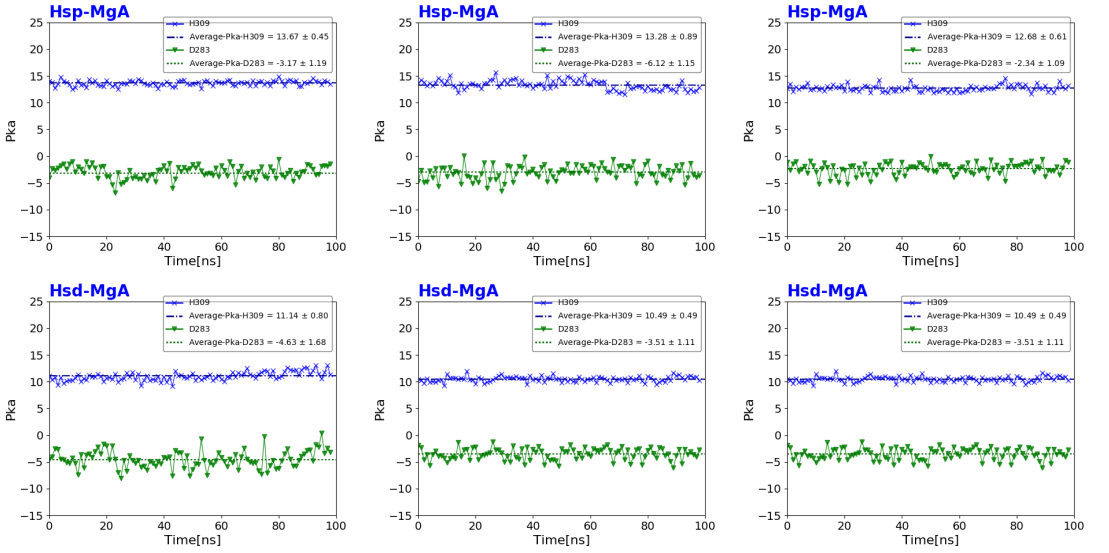

Figure 86: Calculated pKa values for all snapshots of MgA models using a dielectric constant $\varepsilon = 4$. The three plots per model correspond to the three individual simulations performed. For Asp210, the pKa value is outside the titraion range, i.e. <-10.

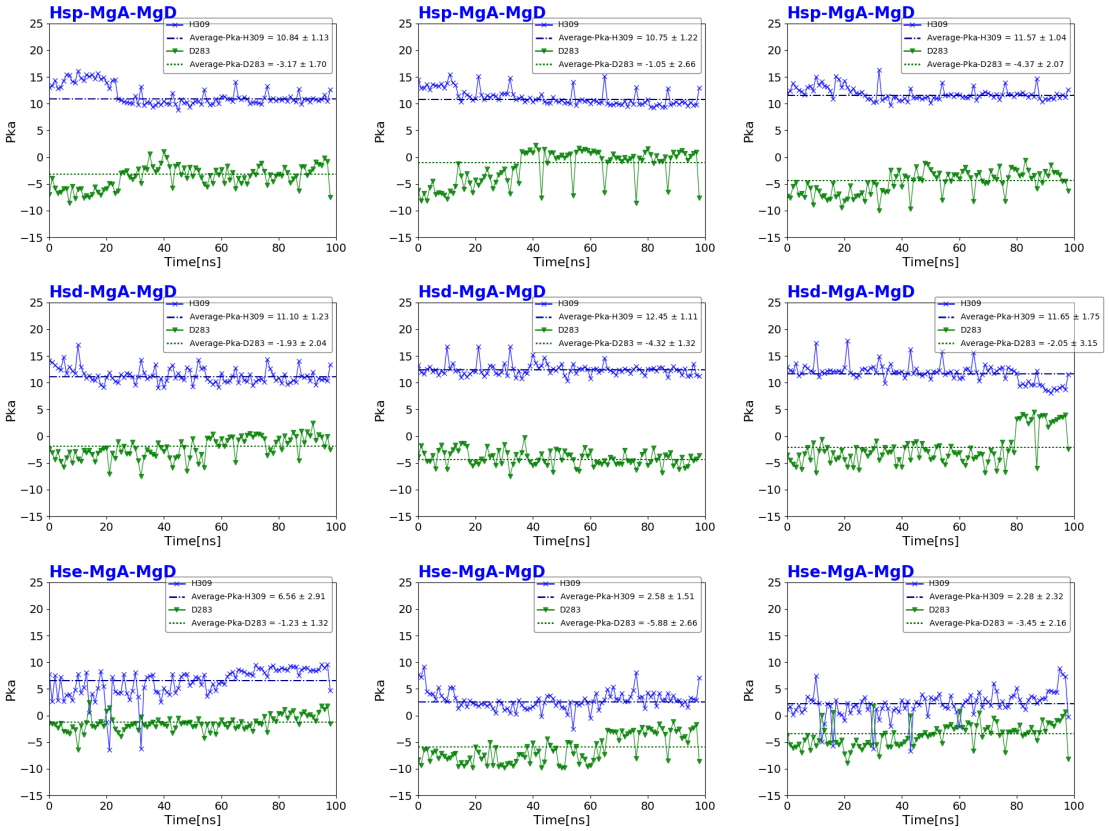

Figure 87: Calculated pKa values for all snapshots of MgA-MgD models using a dielectric constant $\varepsilon = 4$. The three plots per model correspond to the three individual simulations performed. For Asp210, the pKa value is outside the titraion range, i.e. <-10.

$\varepsilon = 3$

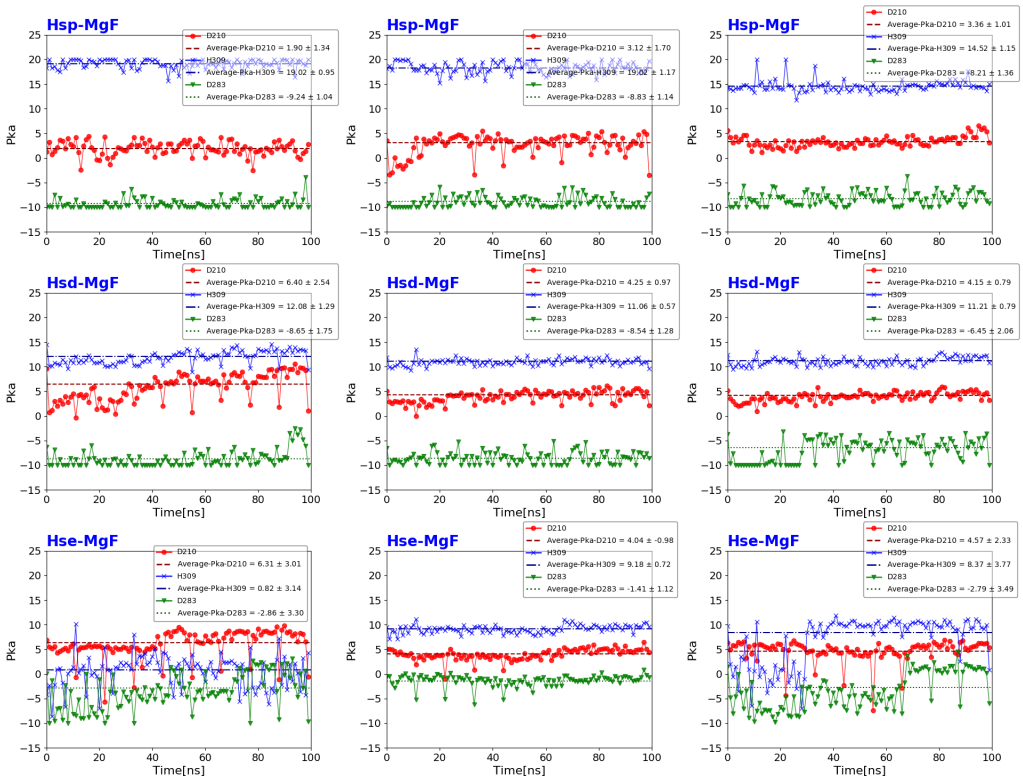

Figure 88: Calculated pKa values for all snapshots of MgF models using a dielectric constant $\varepsilon = 3$. The three plots per model correspond to the three individual simulations performed.

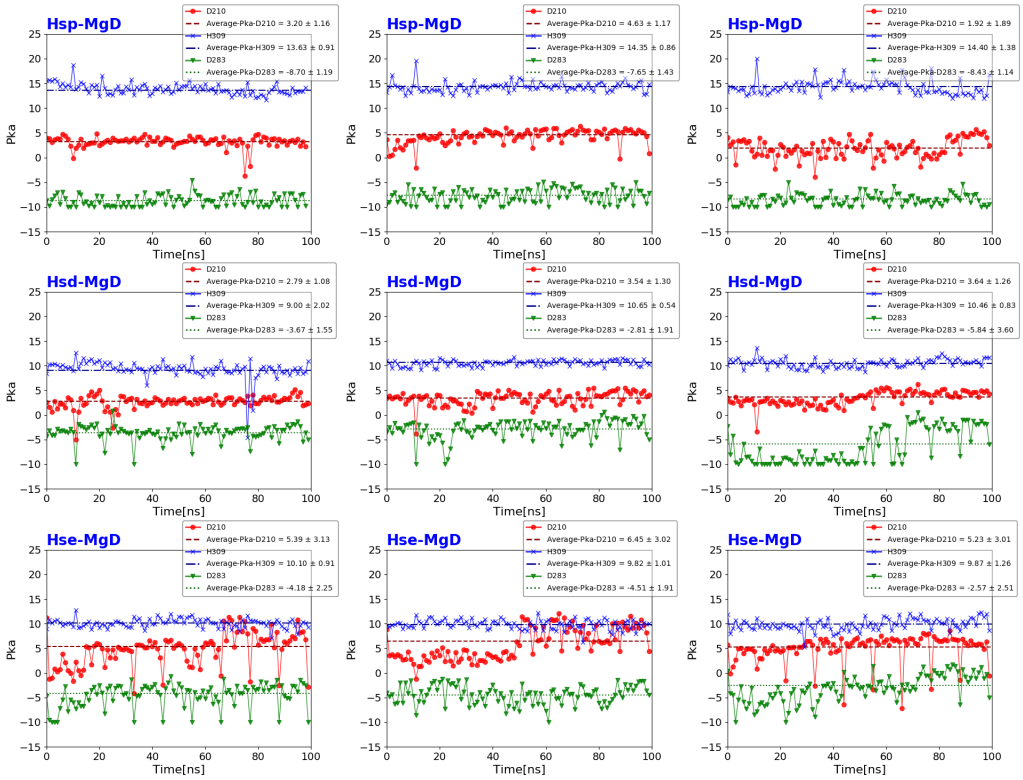

Figure 89: Calculated pKa values for all snapshots of MgD models using a dielectric constant $\varepsilon = 3$. The three plots per model correspond to the three individual simulations performed.

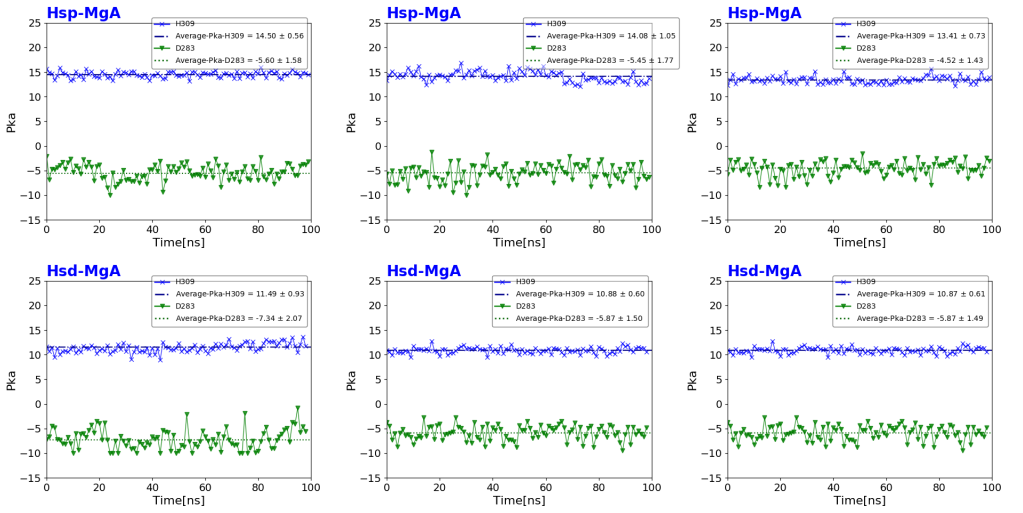

Figure 90: Calculated pKa values for all snapshots of MgA models using a dielectric constant $\varepsilon = 3$. The three plots per model correspond to the three individual simulations performed. For Asp210, the pKa value is outside the titraion range, i.e. <-10.

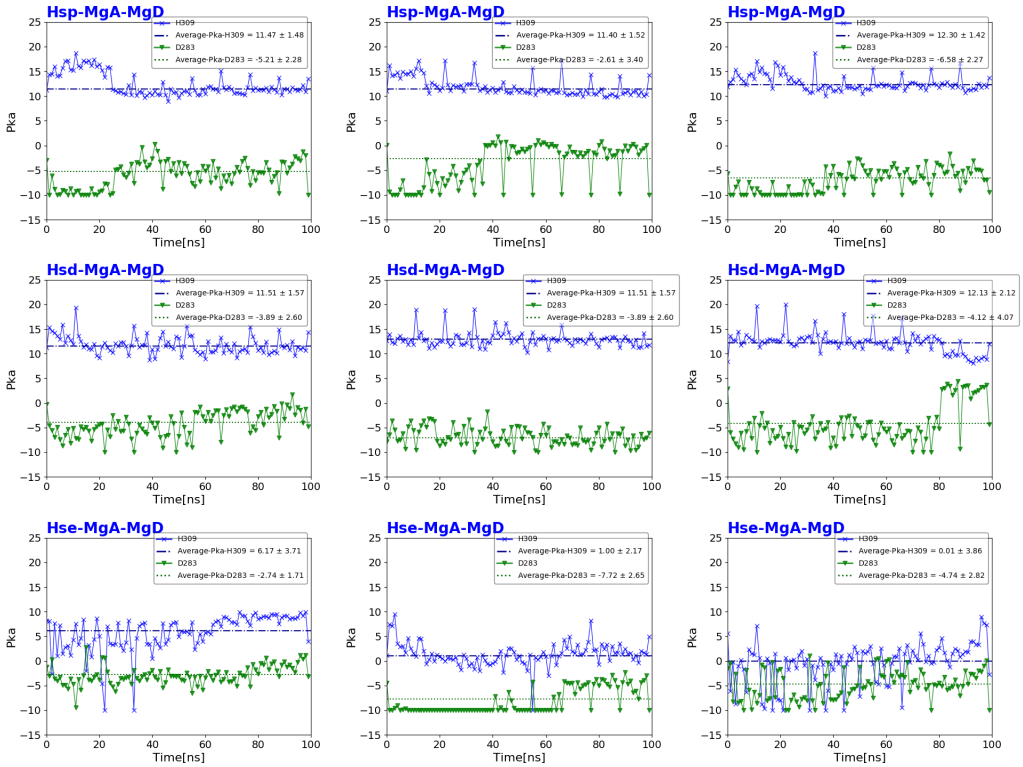

Figure 91: Calculated pKa values for all snapshots of MgA-MgD models using a dielectric constant $\varepsilon = 3$. The three plots per model correspond to the three individual simulations performed. For Asp210, the pKa value is outside the titraion range, i.e. <-10.

$\varepsilon = 6$

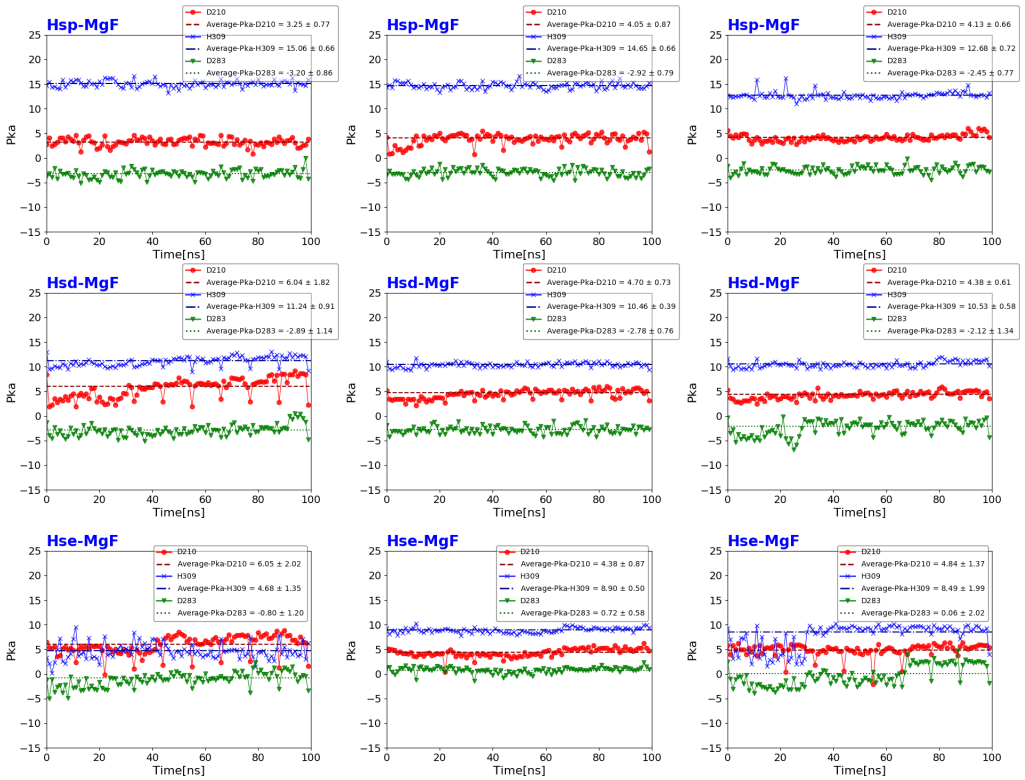

Figure 92: Calculated pKa values for all snapshots of MgF models using a dielectric constant $\varepsilon = 6$. The three plots per model correspond to the three individual simulations performed.

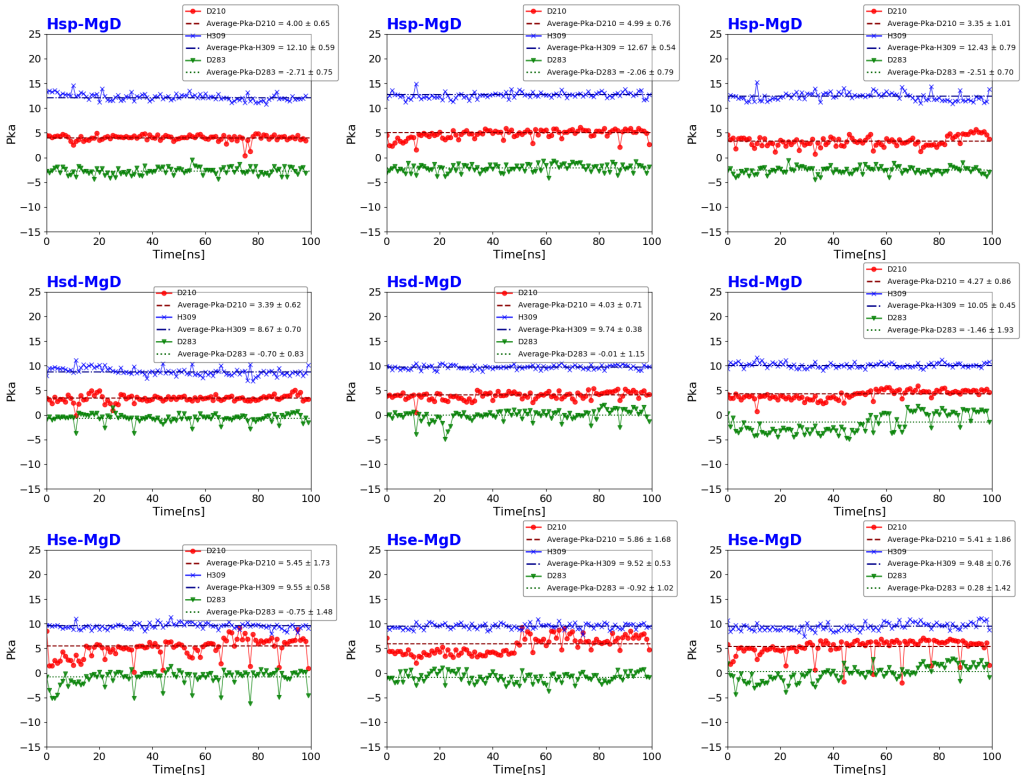

Figure 93: Calculated pKa values for all snapshots of MgD models using a dielectric constant $\varepsilon = 6$. The three plots per model correspond to the three individual simulations performed.

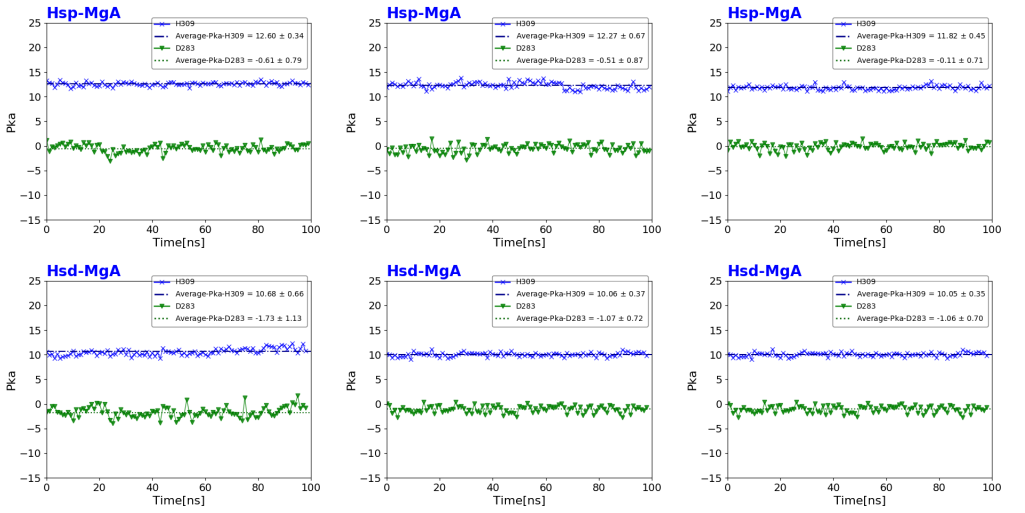

Figure 94: Calculated pKa values for all snapshots of MgA models using a dielectric constant $\varepsilon = 6$. The three plots per model correspond to the three individual simulations performed. For Asp210, the pKa value is outside the titraion range, i.e. <-10.

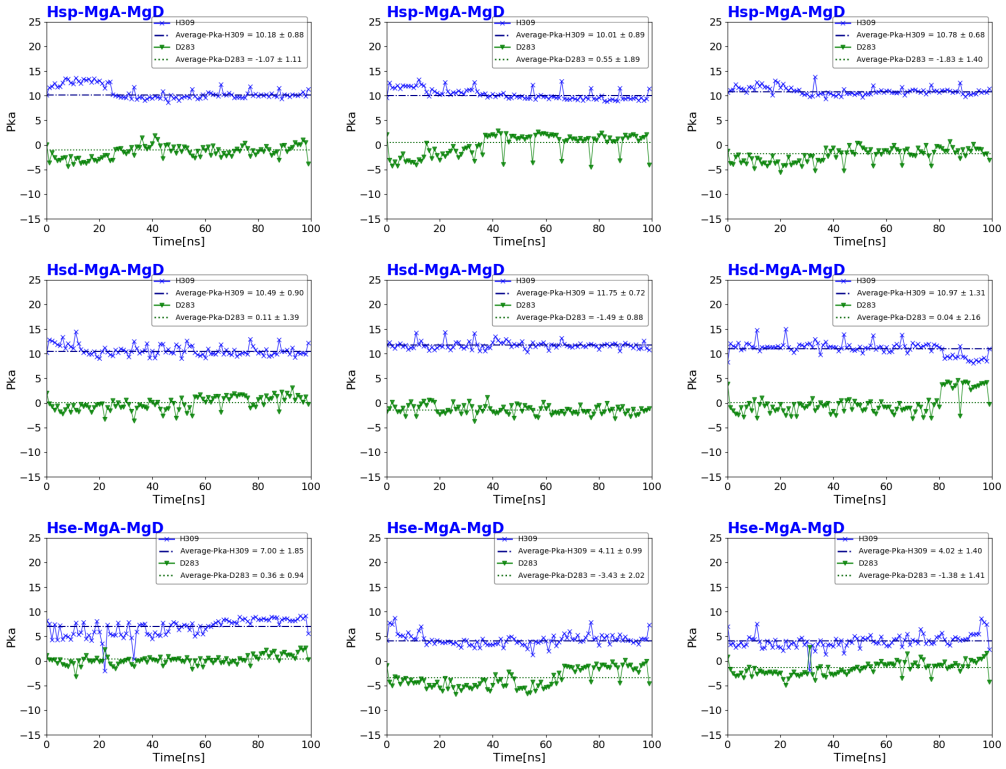

Figure 95: Calculated pKa values for all snapshots of MgA-MgD models using a dielectric constant $\varepsilon = 6$. The three plots per model correspond to the three individual simulations performed. For Asp210, the pKa value is outside the titraion range, i.e. <-10.

Table 28: Average pKa values calculated from snapshots of the different Ape1-DNA complex models using a dielectric constant $\varepsilon = 6$.

Model	Asp210	Asp283	His309
Hsp-MgF	3.81 ± 0.77	-2.86 ± 0.81	14.13 ± 0.68
Hsd-MgF	5.04 ± 1.05	-2.66 ± 1.08	10.44 ± 0.62
Hse-MgF	5.09 ± 1.42	0.00 ± 1.27	7.36 ± 1.28
Hsp-MgD	4.08 ± 0.81	-2.43 ± 0.75	12.40 ± 0.69
Hsd-MgD	3.89 ± 0.73	-0.72 ± 1.30	9.49 ± 0.51
Hse-MgD	5.57 ± 1.76	-0.45 ± 1.31	9.52 ± 0.62
Hsp-MgA	< -10	-0.41 ± 0.79	12.23 ± 0.49
Hsd-MgA	< -10	-1.29 ± 0.92	10.26 ± 0.46
Hsp-MgA-MgD	< -10	-0.78 ± 1.47	10.43 ± 0.82
Hsd-MgA-MgD	< -10	-0.45 ± 1.61	11.07 ± 0.98
Hse-MgA-MgD	< -10	-1.52 ± 1.28	5.04 ± 1.41

D210N mtuant

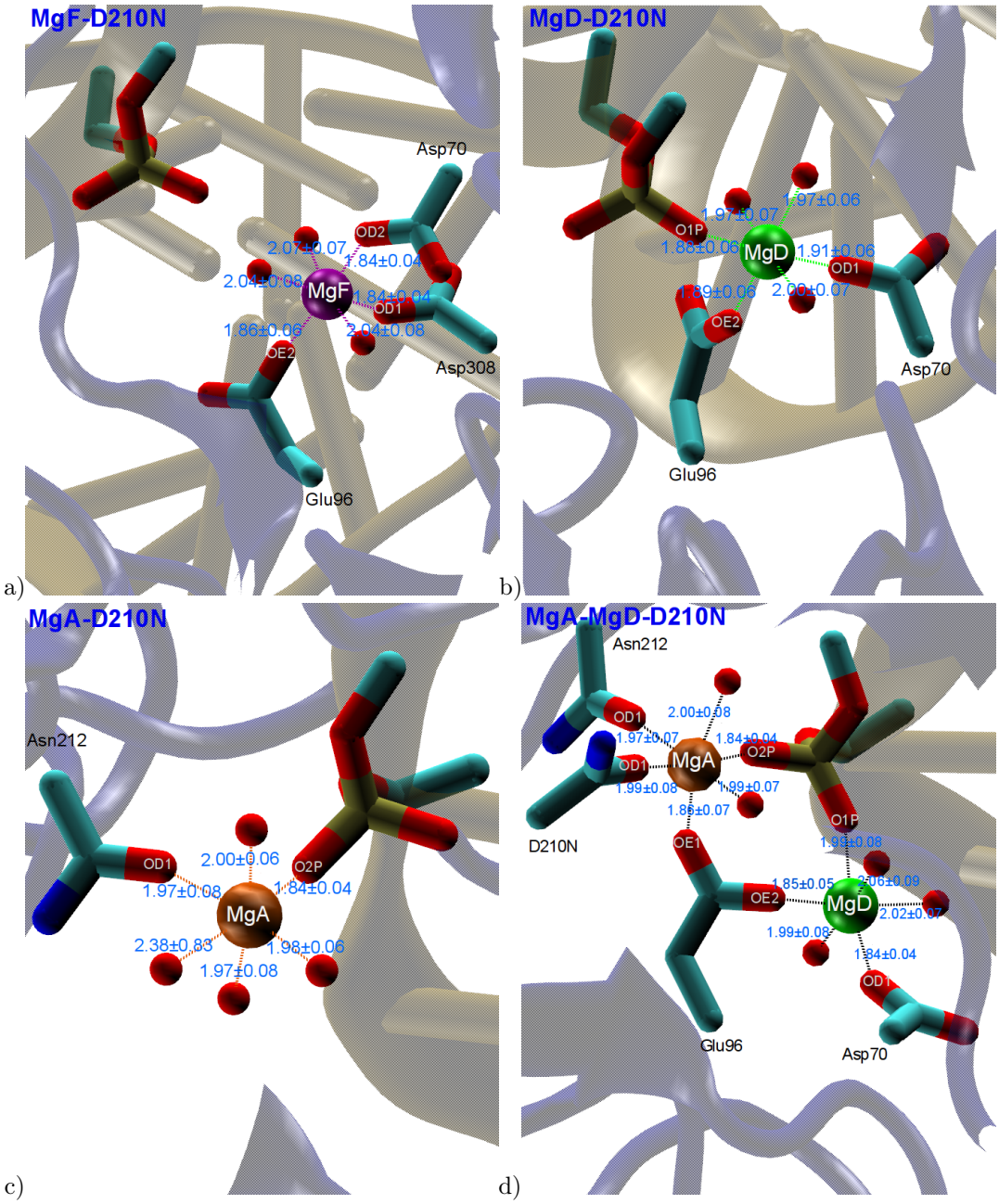

Figure 96: Active site of D210N mutant Ape1-DNA substrate complex with a) one magnesium ion at metal binding site F, b) one magnesium ion at metal binding site D, c) one magnesium ion at metal binding site A, and d) two magnesium ions at site A and D.

N212A mtuant

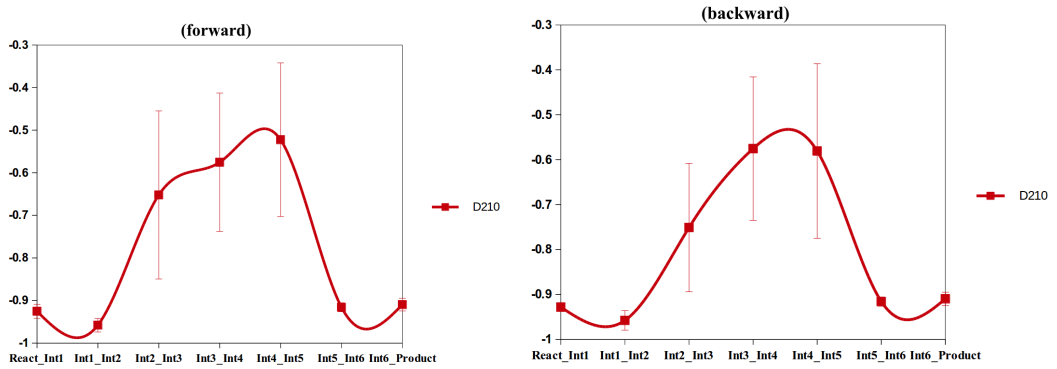

Figure 98: Calculated Mulliken charge for D210 in the active-site of APE1-DNA complex during the phosphodiaster hydrolysis from reactant to product state.

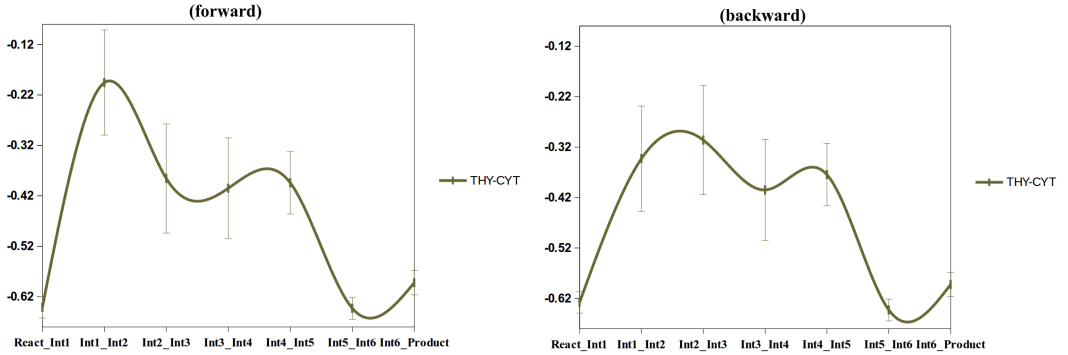

Figure 99: Calculated Mulliken charge for phosphodiester(THY-CYT) during the phosphodiester hydrolysis from reactant to product state.

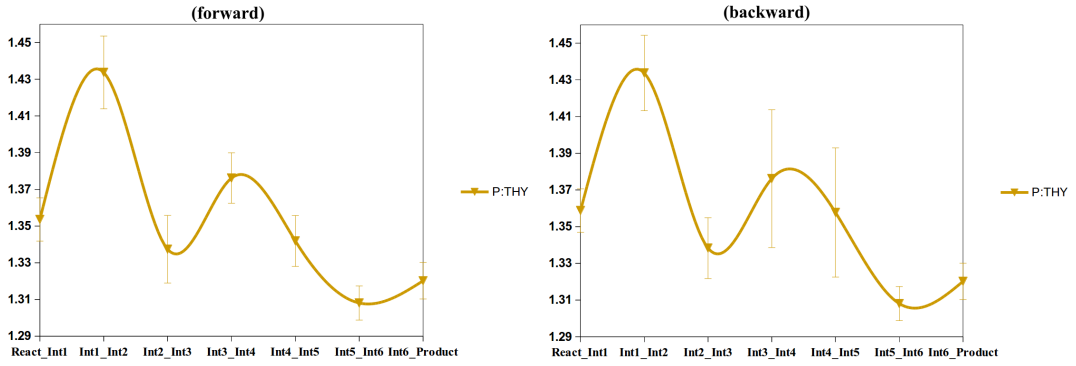

Figure 100: Calculated Mulliken charge for phosphrous atom of phosphodiester(P:THY) during the phosphodiester hydrolysis from reactant to product state.

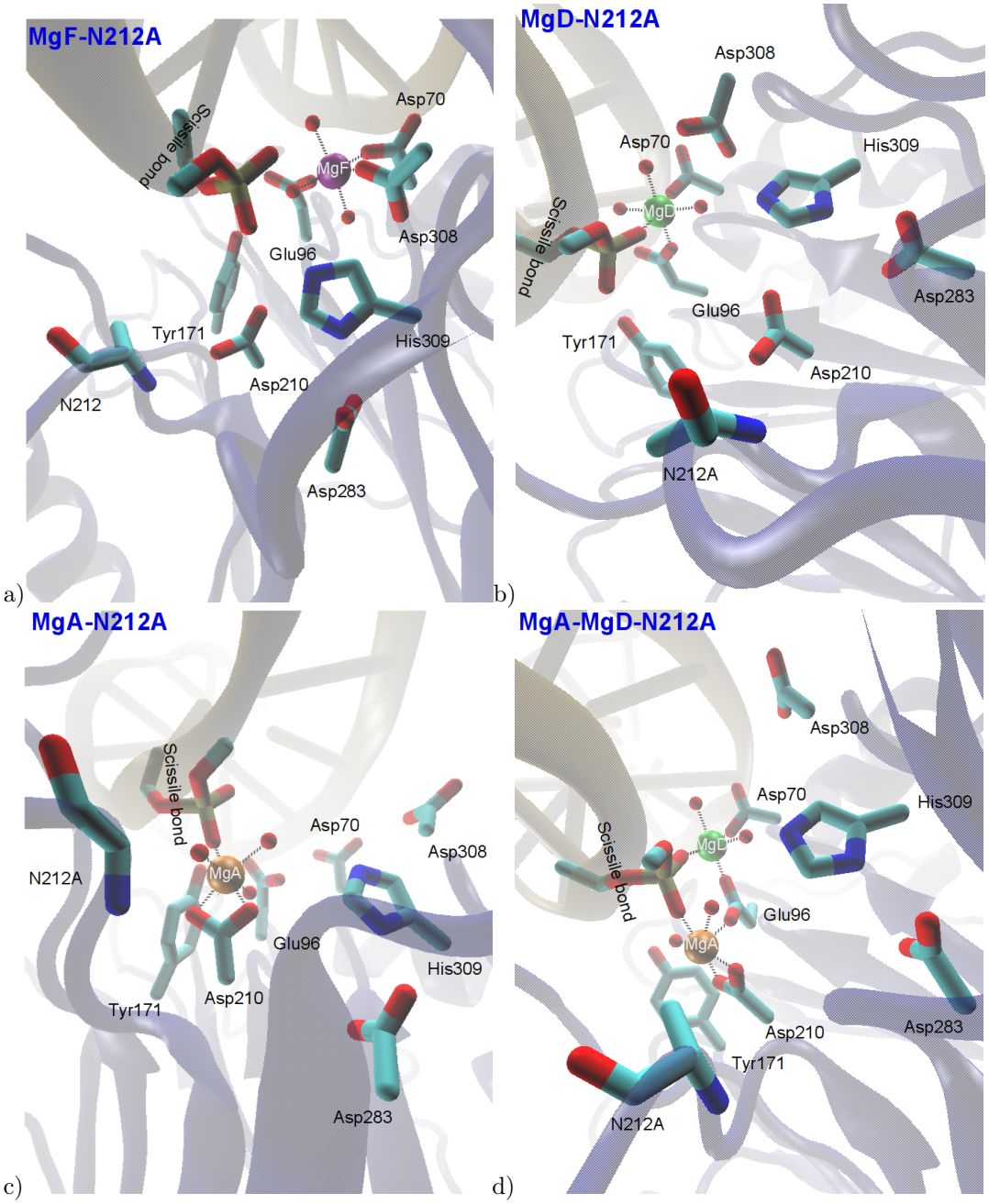

Figure 97: Key amino acid residues in the active site of N212A mutant Ape1-DNA substrate complex with a) one magnesium ion at metal binding site F, b) one magnesium ion at metal binding site D, c) one magnesium ion at metal binding site A, and d) two magnesium ions at site A and D.

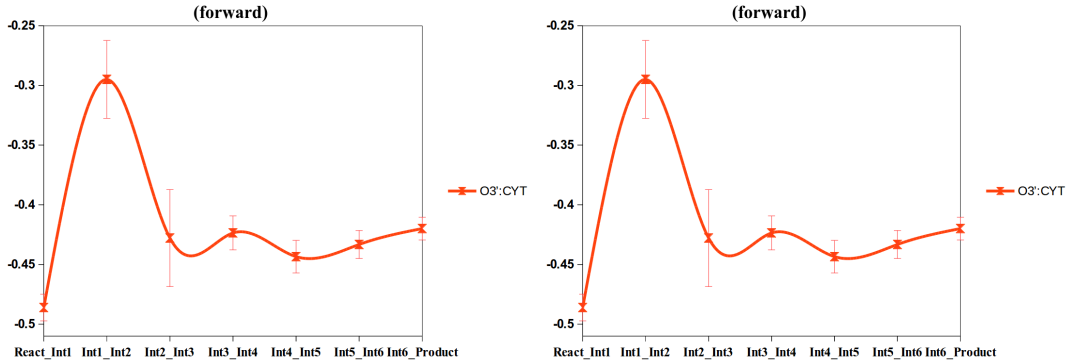

Figure 101: Calculated Mulliken charge for O3' atom of phospodiester (O3':CYT) in the active-site of APE1-DNA complex during the phosphodiester hydrolysis from reactant to product state.

List of Publications

The present thesis is based on the following manuscripts, which have been published or are in preparation for publication in peer-reviewed journals:

[i] Hossein Batebi and Petra Imhof. Phosphodiester hydrolysis computed for cluster models of enzymatic active sites. *Theor Chem Acc* , 135:262, **2016**.

[ii] Hossein Batebi, Jovan Dragelj, and Petra Imhof. Role of ap-endonuclease (ape1) active site residues in stabilization of the reactant enzyme-dna complex. *Proteins: Structure, Function, and Bioinformatics* , **2018**.

[iii] Hossein Batebi and Petra Imhof. A mechanistic simulation study of the enzymatic excision mechanism in AP endonuclease APE1 (*In- preparation*)

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Dissertationsschrift mit dem Titel

A Mechanistic Study Of The Enzymatic Excision Mechanism In AP Endonuclease (APE1)

selbständig angefertigt und hierfür keine anderen als die angegebenen Hilfsmittel verwendet habe. Die Arbeit ist weder in einem früheren Promotionsverfahren angenommen noch als ungenügend beurteilt worden.

Berlin, 4. Oktober 2018

Danksagung

To my love, Parisa: because I owe it all to you. Thank you!

I am deeply grateful to my advisor Petra Imhof. She gave me endless support, I appreciate all her contributions of ideas, time, and funding to make my Ph.D experience productive. I have greatly benefited from her insightful comments and suggestions. I will miss our interesting and long-lasting chats about science and doing research.

I would like to thank my parents who have provided me through emotional, moral and financial support in my life. I am also grateful to my family members and friends, this thesis would not have been possible without their endless support and warm love.

Express my heart-felt gratitude to my lab mates for their useful comments and suggestions. Personal helps,scientific inputs and friendly nature has always made me feel at ease with them.

Advice and comments given by rest of my dissertation committee members has been a great help in making this dissertation sophisticated.

I thank all people who are working in IT-Service at Physik departement, Freie Universität Berlin for their unfailing support.

\

References

[1] Gelin A, Redrejo-RodrÃguez M, Laval J, Fedorova OS, Saparbaev M, and Ishchenko AA. Genetic and biochemical characterization of human ap endonuclease 1 mutants deficient in nucleotide incision repair activity. *PloS one*, 5:e12241, August 2010.

[2] Laio A and Parrinello M. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, 2002.

[3] Leach A. *Molecular Modelling: Principles and Applications (2nd Edition)*. Prentice Hall, 2 edition, April 2001.

[4] Maiti A, Morgan MT, and Drohat AC. Role of two strictly conserved residues in nucleotide flipping and n-glycosylic bond cleavage by human thymine DNA glycosylase. *Journal of Biological Chemistry*, 284(52):36680–36688, 2009.

[5] Maiti A, Morgan MT, Pozharski E, and Drohat AC. Crystal structure of human thymine DNA glycosylase bound to DNA elucidates sequence-specific mismatch recognition. *Proceedings of the National Academy of Sciences*, 105(26):8890–8895, 2008.

[6] Becke AD. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical review A*, 38(6):3098, 1988.

[7] Becke AD. Density-functional thermochemistry. i. the effect of the exchange-only gradient correction. *The Journal of chemical physics*, 96(3):2155–2160, 1992.

[8] Becke AD. Density-functional thermochemistry. iii. the role of exact exchange. *J. Chem. Phys.*, 98:5648, 1993.

[9] Becke AD. A new mixing of hartree–fock and local density-functional theories. *The Journal of chemical physics*, 98(2):1372–1377, 1993.

[10] Miroshnikova AD, Kuznetsova AA, Vorobjev YN, Kuznetsov NA, and Fedorova OS. Effects of mono and divalent metal ions on DNAbinding and catalysis of human apurinic/apyrimidinic endonuclease 1. *Mol Biosyst*, 12(5):1527–1539, 2016.

[11] Cassano AG, Anderson VE, and Harris ME. Evidence for direct attack by hydroxide in phosphodiester hydrolysis. *J. Am. Chem. Soc.*, 124(37):10964–10965, 2002.

[12] Kirby AJ, Medeiros M, Mora JR, Oliveira PS, Amer A, Williams NH, and Nome F. Intramolecular general base catalysis in the hydrolysis of a phosphate diester. calculational guidance to a choice of mechanism. *J. Org. Chem*, 78(4):1343–1353, 2013.

[13] Woelke AL, Wagner A, Galstyan G, Meyer T, and Knapp EW. Proton transfer in the k-channel analog of b-type cytochrome c oxidase from thermus thermophilus. *Biophys. J.*, 107:2177–2184, 2014.

[14] Woelke AL, Galstyan G, Galstyan A, Meyer T, Heberle J, and Knapp EW. Exploring the possible role of glu286 in cco by electrostatic energy computations combined with molecular dynamics. *J. Phys. Chem. B.*, 117:12432–12441, 2013.

[15] Lipton AS, Heck RW, Primak S, McNeill DR, Wilson DM 3rd, and Ellis PD. Characterization of mg2+ binding to the DNA repair protein apurinic/apyrimidic endonuclease 1 via solid-state 25mg nmr spectroscopy. *Journal of the American Chemical Society*, 130(29):9332–9341, 2008.

[16] Demple B and Harrison L. Repair of oxidative damage to DNA: enzymology and biology. *Annual review of biochemistry*, 63(1):915–948, 1994.

[17] Rabenstein B and Knapp EW. ph-dependent population and protonation of carbon-monoxymyoglobin conformers. *Biophys. J.*, 80:1141–1150, 2001.

[18] Freudenthal BD, Beard WA, Cuneo MJ, Dyrkheeva NS, and Wilson SH. Capturing snapshots of APE1 processing DNA damage. *Nat. Struct. Mol. Biol.*, 22:924–931, 2015.

[19] Golden BL. Two distinct catalytic strategies in the hepatitis delta virus ribozyme cleavage reaction. *Biochemistry*, 50(44):9424–9433, 2011.

[20] Ames BN. Dietary carcinogens and anticarcinogens: oxygen radicals and degenerative diseases. In *Risk analysis in the private sector*, pages 297–321. Springer, 1985.

[21] Ames BN. Endogenous oxidative DNA damage, aging, and cancer. *Free radical research communications*, 7(3-6):121–128, 1989.

[22] Ames BN and Gold LS. Endogenous mutagens and the causes of aging and cancer. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 250(1):3–16, 1991.

[23] Ames BN, Shigenaga MK, and Hagen TM. Oxidants, antioxidants, and the degenerative diseases of aging. *Proceedings of the National Academy of Sciences*, 90(17):7915–7922, 1993.

[24] Brooks BR, Brooks CL, MacKerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, and Boresch S et al. Charmm: the biomolecular simulation program. *Journal of computational chemistry*, 30(10):1545–1614, 2009.

[25] Brooks BR, Brooks III CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, T. Lazaridis, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable R. M, Woodcock HL, Wu X, Yang W, York D. M, and Karplus M. The biomolecular simulation program. *J. Comput. Chem.*, 30:1545–1615, 2009.

[26] Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan Sa, and Karplus M. Charmm: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of computational chemistry*, 4(2):187–217, 1983.

[27] Braga C and Travis KP. A configurational temperature nosé-hoover thermostat. *The Journal of chemical physics*, 123(13):134101, 2005.

[28] Jesse C and Scott S. Catalytic strategies of self-cleaving ribozymes. *Accounts of chemical research*, 41(8):1027–1035, 2008.

[29] Lee C, Yang W, and Parr RG. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B.*, 37:785, 1988.

[30] Mol CD, Kuo CF, Thayer MM, Cunningham RP, and Tainer JA. Structure and function of the multifunctional DNA-repair enzyme exonuclease iii. *Nature*, 374:381–386, 1995.

[31] Mol CD, Izumi T, Mitra S, and Tainer JA. DNA-bound structures and mutants reveal abasic DNA binding by APE1 DNA repair and coordination. *Nature*, 403(6768):451, 2000.

[32] Hu CH and Brinck T. Theoretical studies of the hydrolysis of the methyl phosphate anion. *J. Phys. Chem. A.*, 103:5379–5386, 1999.

[33] Gibas CJ and Subramaniam S. Explicit solvent models in protein pKa calculations. *Biophysical J.*, 71:138–147, 1996.

[34] Li CL, Hor LI, Chang ZF, Tsai LC, Yang WZ, and Yuan HS. DNA binding and cleavage by the periplasmic nuclease vvn: a novel structure with a known active site. *EMBO J.*, 22(15):4014–4025, 2003.

[35] Ellis CR, Tsai CC, Lin FY, and Shen J. Conformational dynamics of cathepsin D and binding to a small-molecule BACE1 inhibitor. *J. Comput. Chem.*, 38(15):1260–1269, 2017.

[36] Ellis CR and Shen J. pH-dependent population shift regulates BACE1 activity and inhibition. *J. Am. Chem. Soc.*, 137(30):9543–9546, 2015.

[37] Lu D, Silhan J, MacDonald JT, Carpenter EP, Jensen K, Tang CM, Baldwin GS, and Freemont PS. Structural basis for the recognition and cleavage of abasic DNA in neisseria meningitidis. *Proceedings of the National Academy of Sciences of the United States of America*, 109(42):16852–16857, 2012.

[38] Sheila S D, Valerie L O, and Sucharita K. Base-excision repair of oxidative DNA damage. *Nature*, 447(7147):941, 2007.

[39] Tom D, Darrin Y, and Lee P. Particle mesh ewald: an n.cntdot.log(n) method for ewald sums in large systems. *Journal of Chemical Physics*, 98(12):10089–10092, 1993.

[40] Lowry DF, Hoyt DW, Khazi FA, Bagu J, Lindsey AG, and Wilson DM 3rd. Investigation of the role of the histidine-aspartate pair in the human exonuclease iii-like abasic endonuclease, APE1. *Journal of Molecular Biology*, 329(2):311–322, 2003.

[41] Hosfield DJ, Guan Y, Haas BJ, Cunningham RP, and Tainer JA. Structure of the DNA repair enzyme endonuclease iv and its DNA complex: Double-nucleotide flipping at abasic sites and three-metal-ion catalysis. *Cell.*, 98(3):397–408, August 1999.

[42] Ceperley DM and Alder BJ. Ground state of the electron gas by a stochastic method. *Physical Review Letters*, 45(7):566, 1980.

[43] Wilson 3rd DM, Takeshita M, Grollman AP, and Demple B. Incision activity of human apurinic endonuclease (ape) at abasic site analogs in DNA. *J Biol Chem*, 270:16002–16007, 1995.

[44] Mosbaugh DW and Bennett SE. Uracil-excision DNA repair. *Progress in nucleic acid research and molecular biology*, 48:315–370, 1994.

[45] Galburt EA, Chevalier B, Tang W, Jurica MS, Flick KE, Monnat RJ Jr, and Stoddard BL. A novel endonuclease mechanism directly visualized for i-ppoi. *Nat Struct Biol*, 6(12):1096–1999, 1999.

[46] Friedberg EC, Walker GC, Siede W, and Wood RD. *DNA repair and mutagenesis*. American Society for Microbiology Press, 2005.

[47] Garcin ED, Hosfield DJ, Desai SA, Haas BJ, Bjoras M, Cunningham RP, and Tainer JA. DNA apurinic-apyrimidinic site binding and excision by endonuclease iv. *Nat. Struct. Mol. Biol.*, 15:515–522, April 2008.

[48] Tirel EY and Williams NH. Enhancing phosphate diester cleavage by a zinc complex through controlling nucleophile coordination. *Chemistry - A European Journal*, 21(19):7053–7056, 2015.

[49] Duarte F, Aqvist J, Williams NH, and Kamerlin SC. Resolving apparent conflicts between theoretical and experimental models of phosphate monoester hydrolysis. *J. Am. Chem. Soc*, 137(3):1081–1093, 2015.

[50] Velazquez Escobar F, Lang C, Takiden A, Schneider C, Balke J, Hughes J, Alexiev U, Hildebrandt P, and Mroginski MA. Protonation-dependent structural heterogeneity in the chromophore binding site of cyanobacterial phytochrome cph1. *J. Phys. Chem. B.*, 121:47–57, 2017.

[51] Barzilay G and Hickson ID. Structure and function of apurinic/apyrimidinic endonucleases. *BioEssays.*, 17:713–719, 1995.

[52] Hou G and Cui Q. Qm/mm analysis suggests that alkaline phosphatase (ap) and nucleotide pyrophosphatase/phosphodiesterase slightly tighten the transition state for phosphate diester hydrolysis relative to solution: Implication for catalytic promiscuity in the ap superfamily. *J. Am. Chem. Soc*, 134(1):229–246, 2012.

[53] Kieseritzky G and Knapp EW. Optimizing pka computation in proteins with ph adapted conformations, proteins struct. funct. genet. *Optimizing pKA computation in proteins with pH adapted conformations*, 71:1335–1348, 2007.

[54] Li G and Cui Q. pKa calculations with QM/MM free energy perturbations. *J. Phys Chem. B*, 107(51):14521–14528, 2003.

[55] E.D. Glendening, A.E. Read, J.E. Carpenter, and F. Weinhold. Nbo (version 3.1), 2009.

[56] Batebi H, Dragelj J, and Imhof P. Role of ap-endonuclease (APE1) active site residues in stabilization of the reactant enzyme-dna complex. *Proteins: Structure, Function, and Bioinformatics*, 2018.

[57] Batebi H and Imhof P. Phosphodiester hydrolysis computed for cluster models of enzymatic active sites. *Theor Chem Acc*, 135:262, 2016.

[58] Ishikita H and Knapp EW. Redox potential of quinones in both electron transfer branches of photosystem i. *J. Biol. Chem.*, 278:52002–52011, 2003.

[59] Ishikita H, Morra G, and Knapp EW. Redox potential of quinones in photosynthetic reaction centers from rhodobacter sphaeroides: Dependence on protonation of glu-l212 and asp-l213. *Biochemistry*, 42:3882–3892, 2003.

[60] Korhonen H and Mikkola Sand Williams NH. The mechanism of cleavage and isomerisation of rna promoted by an efficient dinuclear zn2+ complex. *Chemistry-A European Journal*, 18(2):659–670, 2012.

[61] Lodish H, Berk A, Kaiser CA, Krieger M, Scott MP, Bretscher A, Ploegh H, and Matsudaira P. *Molecular Cell Biology (6th Edition)*. W. H. Freeman, 2008.

[62] Ivanov I, Tainer JA, and McCammon JA. Unraveling the three-metal-ion catalytic mechanism of the DNA repair enzyme endonuclease iv. *PNAS.*, 104(5):1465–1470, 2007.

[63] Borden J, Crans DC, and Florian J. Transition state analogues for nucleotidyl transfer reactions: Structure and stability of pentavalent vanadate and phosphate ester dianions. *J. Phys. Chem. B*, 110(30):14988–14999, 2005.

[64] Borden J, Crans DC, and Florian J. Transition state analogues for nucleotidyl transfer reactions: Structure and stability of pentavalent vanadate and phosphate ester dianions. *J. Phys Chem. B*, 30(110):14988–99, 2006.

[65] Florián J and Warshel A. Phosphate ester hydrolysis in aqueous solution: associative versus dissociative mechanisms. *J. Phys. Chem. B.*, 102:719–734, 1998.

[66] Florián J, Åqvist J, and Warshel A. On the reactivity of phosphate monoester dianions in aqueous solution: Bronsted linear free-energy relationships do not have an unique mechanistic interpretation. *J. Am. Chem. Soc.*, 120:11524–11525, 1998.

[67] Florián J, Strajbl M, and Warshel A. Conformational flexibility of phosphate, phosphonate, and phosphorothioate methyl esters in aqueous solution. *J. Am. Chem. Soc*, 120:7959–7966, 1998.

[68] Kastner J. Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(6):932–942, 2011.

[69] Kastner J and Thiel W. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: "umbrella integration". *The Journal of chemical physics*, 123(14):144104, 2005.

[70] Mongan J and Case DA. Biomolecular simulations at constant pH. *Current Opinion in Structural Biology*, 15(2):157 – 163, 2005.

[71] Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, and Schulten K. Scalable molecular dynamics with namd. *J Comput Chem*, 26(16):1781–1802, 2005.

[72] Sinclair JE and Fletcher R. A new method of saddle-point location for the calculation of defect migration energies. *Journal of Physics C: Solid State Physics*, 7(5):864, 1974.

[73] Zalatan JG and Herschlag D. Alkaline phosphatase mono- and diesterase reactions: Comparative transition state analysis. *J. Am. Chem. Soc*, 128(4):1293, 2006.

[74] Eastberg JH, Eklund J, Monnat R Jr, and Stoddard BL. Mutability of an hnh nuclease imidazole general base and exchange of a deprotonation mechanism. *Biochemistry*, 46(24):7215–7225, 2007.

[75] Kim JH and Chin Jik. Dimethyl phosphate hydrolysis at neutral ph. *Journal of the American Chemical Society*, 114(25):9792–9795, 1992.

[76] Knight JL and Brooks CL 3rd. λ-dynamics free energy simulation methods. *Journal of computational chemistry*, 30(11):1692–1700, 2009.

[77] Hadden JM, Declais AC, Phillips SE, and Lilley DM. Metal ions bound at the active site of the junction-resolving enzyme t7 endonuclease i. *EMBO J.*, 21:3505–3515, 2002.

[78] Mercero JM, Barrett P, Lam CW, Fowler JE, Ugalde JM, and Pedersen LG. Quantum mechanical calculations on phosphate hydrolysis reactions. *J. Comput. Chem.*, 21:43–51, 2000.

[79] Erzberger JP and Wilson 3rd DM. The role of mg2+ and specific amino acid residuesin the catalytic reaction of the major human abasic endonuclease: New insights from edta-resistantincision of acyclic abasic site analogs and sitedirected mutagenesis. *J. Mol. Biol.*, 290:447–457, 1999.

[80] Erzberger JP and Wilson, 3rd DM. The role of mg2+ and specific amino acid residuesin the catalytic reaction of the major human abasic endonuclease: New insights from edta-resistantincision of acyclic abasic site analogs and sitedirected mutagenesis. *J. Mol. Biol.*, 290:447–457, 1999.

[81] Blumhagen K, Muegge I, and Knapp EW. Diffusion of two different water models and thermal conductivity in a protein-water system. *Int. J. Quantum Chem.*, 59:271–279, 1996.

[82] Katayanagi K, Okumura M, and Morikawa K. Crystal structure of escherichia coli rnase hi in complex with mg2+ at 2.8 a resolution: proof for a single mg(2+)-binding site. *Proteins.*, 17(4):337–46, December 1993.

[83] Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, and Mackerell AD Jr. Charmm general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields. *J Comput Chem*, 31(4):671–690, 2007.

[84] G. Kieseritzky and EW. Knapp. Improved pKa prediction: Combining empirical and semimicroscopic methods. *J. Comput. Chem.*, 29:2575–2581, 2008.

[85] Chou KM and Cheng YC. An exonucleolytic activity of human apurinic/apyrimidinic endonuclease on 3' mispaired DNA. *Nature*, 415(6872):655–659, 2002.

[86] Walter Kohn and Lu Jeu Sham. Self-consistent equations including exchange and correlation effects. *Physical review*, 140(4A):A1133, 1965.

[87] Nilsen L, Forstrom RJ, Bjoras M, and Alseth I. Ap endonuclease independent repair of abasic sites in schizosaccharomyces pombe. *Nucleic Acids Res.*, 40(5):2000–2009, March 2012.

[88] Nilsson L and Karshikoff A. Multiple ph regime molecular dynamics simulation for pk calculations. *PLoS ONE*, 6(5):e20116, 2011.

[89] Wiesmuller L, Ford JM, and Schiestl RH. DNA damage, repair, and diseases. *Journal of Biomedicine and Biotechnology*, 2(2):45, 2002.

[90] Nguyen LH, Barsky D, Erzberger JP, and DM 3rd Wilson. Mapping the protein-DNA interface and the metal-binding site of the major human apurinic/apyrimidinic endonuclease. *Journal of Molecular Biology*, 298(3):447–459, 2000.

[91] Kanazhevskaya LY, Koval VV, Lomzov AA, and Fedorova OS. The role of asn-212 in the catalytic mechanism of human endonuclease APE1: Stopped-flow kinetic study of incision activity on a natural ap site and a tetrahydrofuran analogue. *DNA Repair*, 21:43–54, 2014.

[92] Boero M, Tateno M, Terakura K, and Oshiyama A. Double-metal-ion/single-metal-ion mechanisms of the cleavage reaction of ribozymes: First-principles molecular dynamics simulations of a fully hydrated model system. *Journal of Chemical Theory and Computation*, 1(5):925–934, 2005.

[93] Chohan M, Mackedenski S, Li WM, and Lee CH. Human apurinic/apyrimidinic endonuclease 1 (APE1) has 3' rna phosphatase and 3' exoribonuclease activities. *J. Mol. Biol.*, 427:298–311, 2015.

[94] Dlakic M. Functionally unrelated signalling proteins contain a fold similar to mg2+dependent endonucleases. *Trends Biochem. Sci.*, 25(272-273):272–273, 2000.

[95] Garcia-Viloca M, Gao J, Karplus M, and Truhlar DG. How enzymes work: analysis by modern rate theory and computer simulations. *Science*, 303(5655):186–195, 2004.

[96] Gaus M, Lu X, Elstner M, and Cui Q. Parameterization of dftb3/3ob for sulfur and phosphorus for chemical and biological applications. *J. Chem. Theory Comput.*, 10(4):1518–1537, 2014.

[97] KlÃ€hn M, Rosta E, and Warshel A. On the mechanism of hydrolysis of phosphate monoester dianions in solution and proteins. *J. Am. Chem. Soc*, 128:15310–15323, 2006.

[98] Zahran M, Berezniak T, Imhof P, and Smith JC. Role of magnesium ions in DNA recognition by the ecorv restriction endonuclease. *FEBS Lett.*, 585(17):2739–2743, September 2011.

[99] Gorman MA, Morera S, Rothwell DG, De La Fortelle E, Mol CD, Tainer JA, Hickson ID, and Freemont PS. The crystal structure of the human DNA repair endonuclease hap1 suggests the recognition of extra-helical deoxyribose at dna abasic sites. *EMBO Journal*, 16(21):6548–6558, 1997.

[100] Miller MD, Tanner J, Benedik M, Alpaugh MJ, and Krause KL. 2.1 å structure of serratia endonuclease suggests a mechanism for binding to double-stranded DNA. *Nat. Struct. Mol. Biol.*, 1:461–468, 1994.

[101] Field MJ, Bash PA, and Karplus M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *Journal of Computational Chemistry*, 11(6):700–733, 1990.

[102] Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, and Scalmani et al G. Gaussian09 Revision E.01, 2009. Gaussian Inc. Wallingford CT 2009.

[103] Giraud-Panis MJ and Lilley DM. T4 endonuclease vii importance of a histidine-aspartate cluster within the zinc-binding domain. *J. Biol. Chem.*, 271:33148–33155, December 1996.

[104] Gilson MK and Honig B. The dielectric constant of a folded protein. *Biopolymers*, 25:2097–2119, 1986.

[105] Gilson MK and Honig BH. Calculation of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies, and conformational analysis. *Proteins: Structure, Function, and Bioinformatics*, 4, 1988.

[106] Cooke MS, Evans MD, Dizdaroglu M, and Lunec J. Oxidative DNA damage: mechanisms, mutation, and disease. *The FASEB Journal*, 17(10):1195–1214, 2003.

[107] Bennett MT, Rodgers MT, Hebert AS, Ruslander LE, Eisele L, and Drohat AC. Specificity of human thymine DNA glycosylase depends on n-glycosidic bond stability. *Journal of the American Chemical Society*, 128(38):12510–12519, 2006.

[108] Iche-Tarrat N, Barthelat JC, Rinaldi D, and Vigroux A. Theoretical studies of the hydroxide-catalyzed p-o cleavage reactions of neutral phosphate triesters and diesters in aqueous solution: Examination of the changes induced by h/me substitution. *J. Phys. Chem. B.*, 109:22570–22580, 2005.

[109] Kanaan N, Crehuet R, and Imhof P. Mechanism of the glycosidic bond cleavage of mismatched thymine in human thymine DNA glycosylase revealed by classical molecular dynamics and quantum mechanical/molecular mechanical calculations. *The Journal of Physical Chemistry B*, 119(38):12365–12380, 2015.

[110] Oezguen N, Mantha AK, Izumi T, Schein CH, Mitra S, and Braun W. Md simulation and experimental evidence for mg2- binding at the b site in human ap endonuclease 1. *Bioinformation*, 7:184–198, 2011.

[111] Oezguen N, Schein CH, Peddi SR, Power TD, Izumi T, and Braun W. A moving metal mechanism for substrate cleavage by the DNA repair endonuclease ape-1. *Proteins: Structure, Function and Genetics*, 68(1):313–323, 2007.

[112] Baker NA, Sept D, Joseph S, Holst MJ, and McCammon JA. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.*, 98:10037–10041, 2001.

[113] Timofeyeva NA, Koval VV, Ishchenko AA, Saparbaev MK, and Fedorova OS. Lys98 substitution in human ap endonuclease 1 affects the kinetic mechanism of enzyme action in base excision and nucleotide incision repair pathways. *PloS one*, 6:e24063, 2011.

[114] Tatsuya Nishino and Kosuke Morikawa. Structure and function of nucleases in DNA repair: shape, grip and blade of the DNA scissors. *Oncogene*, 21(58):9022, 2002.

[115] Patrick J O'Brien and Daniel Herschlag. Alkaline phosphatase revisited: Hydrolysis of alkyl phosphates. *Biochem.*, 41:3207–3225, 2001.

[116] Altoè P, Stenta M, Bottoni A, and Garavelli M. A tunable qm/mm approach to chemical reactivity, structure and physico-chemical properties prediction. *Theoretical Chemistry Accounts*, 118(1):219–240, 2007.

[117] Florová P, Sklenovský P, Banáš P, and Otyepka M. Explicit water models affect the specific solvation and dynamics of unfolded peptides while the conformational behavior and flexibility of folded peptides remain intact. *J. Chem. Theory Comput.*, 6:3569–3579, 2010.

[118] Hohenberg P and Kohn W. Inhomogeneous electron gas. *Physical review*, 136(3B):B864, 1964.

[119] Imhof P, Fischer S, and Smith JC. Catalytic mechanism of DNA backbone cleavage by the restriction enzyme ecorv: A quantum mechanical/molecular mechanical analysis. *Biochem.*, 48(38):9061–9075, September 2009.

[120] Imhof P, Fischer S, Krämer R, and Smith JC. Density functional theory analysis of dimethylphosphate hydrolysis: effect of solvation and nucleophile variation. *Journal of Molecular Structure: THEOCHEM*, 713(1):1–5, 2005.

[121] Mark P and Nilsson L. Structure and dynamics of the tip3p, spc, and spc/e water models at 298 k. *J. Phys. Chem. A.*, 105:9954–9960, 2001.

[122] Hariharan PC and Pople JA. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chem. Acc.*, 28(3):213–222, September 1973.

[123] A. Peck, F. Sunden, L. D. Andrews, V. S. Pande, and D. Herschlag. Tungstate as a transition state analog for catalysis by alkaline phosphatase. *J Mol Biol*, 428:2758–2768, 2016.

[124] Grzyska PK, Czyryca PG, Purcell J, and Hengge AC. Transition state differences in the hydrolysis reactions of alkyl versus aryl phosphate monoester monoanions. *J. Am. Chem. Soc*, 125:13106–13111, 2003.

[125] Dragan M. Popović, Snežana D. Zarić, Björn Rabenstein, and EW. Knapp. Artificial cytochrome b: Computer modeling and evaluation of redox potentials. *J. Am Chem. Soc.*, 123(25):6040–6053, 2001.

[126] Dragan M. Popović, Aleksandra Zmirić, Snežana D. Zarić, and EW. Knapp. Energetics of radical transfer in DNA photolyase. *J. Am Chem. Soc.*, 124(14):3775–3782, 2002.

[127] Henrik E Poulsen, Steffen Loft, Helene Prieme, Kirsten Vistisen, Jens Lykkesfeldt, Kristiina Nyyssonen, and Jukka T Salonen. Oxidative DNA damage in vivo: relationship to age, plasma antioxidants, drug metabolism, glutathione-s-transferase activity and urinary creatinine excretion. *Free radical research*, 29(6):565–571, 1998.

[128] Domenico Praticò. Peripheral biomarkers of oxidative damage in alzheimer's disease: the road ahead. *Neurobiology of Aging*, 26(5):581–583, 2005.

[129] Kingma PS, Corbett AH, Burcham PC, Marnett LJ, and Osheroff N. Abasic sites stimulate double-stranded DNA cleavage mediated by topoisomerase ii. DNA lesions as endogenous topoisomerase ii poisons. *Journal of Biological Chemistry*, 270(37):21441–21444, 1995.

[130] Beernink PT, Segelke BW, Hadi MZ, Erzberger JP, Wilson DM 3rd, and Rupp B. Two divalent metal ions in the active site of a new crystal form of human apurinic/apyrimidinic endonuclease, APE1: implications for the catalytic mechanism. *Journal of molecular biology*, 307(4):1023–1034, 2001.

[131] Doetsch PW and Cunningham RP. The enzymology of apurinic/apyrimidinic endonucleases. *Mutation Research/DNA Repair*, 236(2-3):173–201, 1990.

[132] Cui Q, Elstner M, Kaxiras E, Frauenheim T, and Karplus M. A qm/mm implementation of the self-consistent charge density functional tight binding (scc-dftb) method. *J. Phys. Chem. B.*, 105(2):569–585, 2001.

[133] He H& Chen Q& and Georgiadis MM. High-resolution crystal structures reveal plasticity in the metal binding site of apurinic/apyrimidinic endonuclease i. *Biochemistry*, 53(41):6520–6529, 2014.

[134] Abbotts R and Madhusudan S. Human ap endonuclease 1 (APE1): From mechanistic insights to druggable target in cancer. *Cancer Treatment Reviews*, 36(5):425–435, 2010.

[135] Kovall RA and Matthews BW. Type ii restriction endonucleases: structural, functional and evolutionary relationships. *Curr. Opin. Chem. Biol.*, 3(5):578–583, October 1999.

[136] B. Rabenstein, G. Matthias Ullmann, and EW. Knapp. Energetics of electron-transfer and protonation reactions of the quinones in the photosynthetic reaction center of rhodopseudomonas viridis. *Biochemistry*, 37:2488–2495, 1998.

[137] Harris RC, Tsai CC, Ellis CR, and Shen J. Proton-coupled conformational allostery modulates the inhibitor selectivity for β-secretase. *J. Phys. Chem. Lett.*, 8(19):4832–4837, 2017.

[138] Maher RL and Bloom LB. Pre-steady-state kinetic characterization of the ap endonuclease activity of human ap endonuclease 1. *Journal of Biological Chemistry*, 282(42):30577–30585, 2007.

[139] D. Rosen. Dielectric properties of protein powders with adsorbed water. *Trans. Faraday Soc.*, 59:2178–2191, 1963.

[140] Hooft RWW, van Eijck BP, and Kroon J. An adaptive umbrella sampling procedure in conformational analysis using molecular dynamics and its application to glycol. *The Journal of chemical physics*, 97(9):6690–6694, 1992.

[141] Fischer S and Karplus M. Conjugate peak refinement: an algorithm for finding reaction paths and accurate transition states in systems with many d egrees of freedom. *Chem. Phys. Lett.*, 194:252–261, 1992.

[142] Kawanishi S, Hiraku Y, and Oikawa S. Mechanism of guanine-specific DNA damage by oxidative stress and its role in carcinogenesis and aging. *Mutation Research/Reviews in Mutation Research*, 488(1):65–76, 2001.

[143] Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, and Kollman PA. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of computational chemistry*, 13(8):1011–1021, 1992.

[144] Kamerlin SC, Williams NH, and Warshel A. Dineopentyl phosphate hydrolysis: Evidence for stepwise water attack. *J. Org. Chem*, 73(18):6960–6969, 2008.

[145] Claudia N. Schutz and Arieh Warshel. What are the dielectric "constants" of proteins and how to validate electrostatic models. *Proteins: Structure, Function, and Bioinformatics*, 44(4):400–417, 2001.

[146] Roland K. O. Sige and Anna Marie Pyle. Alternative roles for metal ions in enzyme catalysis and the implications for ribozyme chemistry. *Chem. Rev*, 107(1):97–113, 2007.

[147] Marc Souaille and Benoit Roux. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Computer physics communications*, 135(1):40–57, 2001.

[148] Mundle ST, Fattal MH, Melo LF, Coriolan JD, O'Regan NE, and Strauss PR. Novel role of tyrosine in catalysis by human ap endonuclease 1. *DNA Repair*, 3(11):1447–1455, 2004.

[149] Barnes T, Kim WC, Mantha AK, Kim SE, Izumi T, Mitra S, and Lee CH. Identification of apurinic/apyrimidinic endonuclease 1 (APE1) as the endoribonuclease that cleaves c-myc mrna. *Nucleic Acids Research*, 37(12):3946–3958, 2007.

[150] Humphry T, Forconi M, Williams NH, and Hengge AC. An altered mechanism of hydrolysis for a metal-complexed phosphate diester. *J. Am. Chem. Soc*, 124:14860–14861, 2002.

[151] Izumi T, Brown DB, Naidu CV, Bhakat KK, Macinnes MA, Saito H, Chen DJ, and Mitra S. Two essential but distinct functions of the mammalian abasic endonuclease. *Proceedings of the National Academy of Sciences of the United States of America*, 102(16):5739–5743, 2005.

[152] Lindahl T. DNA repair enzymes. *Annual review of biochemistry*, 51(1):61–87, 1982.

[153] Lindahl T. Instability and decay of the primary structure of DNA. *nature*, 362(6422):709–715, 1993.

[154] Lindahl T and Ljungquist S. Apurinic and apyrimidinic sites in DNA. *Molecular mechanisms for repair of DNA*, pages 31–38, 1975.

[155] Meyer T and Knapp EW. pka values in proteins determined by electrostatics applied to molecular dynamics trajectories. *J. Chem. Theory Comput.*, 11, 2015.

[156] EW. Knapp T. Meyer, G. Kieseritzky. Electrostatic pk a computations in proteins: Role of internal cavities. *Proteins Struct. Funct. Bioinforma.*, 79:332–03332, 2011.

[157] M. Takeuchi, R. Lillis, B. Demple, and M. Takeshita. Mammalian abasic site base excision repair. *J. Biol. Chem.*, 269(34):21907–21914, August 1998.

[158] Bruice TC and Holmquist B. Question of the importance of electrostatic catalysis. i. comparison of the reactivity of o-nitrophenyl hydrogen oxalate and ethyl o-nitrophenyl oxalate toward nucleophiles. *Journal of the American Chemical Society*, 89(16):4028–4034, 1967.

[159] S.E. Tsutakawa, H. Jingami, and K. Morikawa. Recognition of a tg mismatch: the crystal structure of very short patch repair endonuclease in complex with a DNA duplex. *Cell.*, 99(6):615–623, December 1999.

[160] S.E. Tsutakawa, D.S. Shin, C.D. Mol, T. Izumi, A.S. Arvai, A.K. Mantha, B. Szczesny, I.N. Ivanov, D.J. Hosfield, B. Maiti, M.E. Pique, K.A. Frankel, K. Hitomi, R.P. Cunningham, S. Mitra, and J.A Tainer. Conserved structural chemistry for incision activity in structurally non-homologous apurinic/apyrimidinic endonuclease APE1 and endonuclease iv DNA repair enzymes. *Journal of Biological Chemistry*, 288(12):8445–8455, 2013.

[161] Barone V, Bencini A, Cossi M, Matteo AD, Mattesini M, and Totti F. Assessment of a combined qm/mm approach for the study of large nitroxide systems in vacuo and in condensed phases. *Journal of the American Chemical Society*, 120(28):7069–7078, 1998.

[162] Bohr VA. Repair of oxidative DNA damage in nuclear and mitochondrial DNA, and some changes with aging in mammalian cells1, 2. *Free Radical Biology and Medicine*, 32(9):804–812, 2002.

[163] Loup Verlet. Computer" experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Physical review*, 159(1):98, 1967.

[164] Chen W, Morrow BH, Shi C, and Shen JK. Recent development and application of constant ph molecular dynamics. *Molecular Simulation*, 40(10-11):830–838, 2014.

[165] Martinet W, Knaapen MW, De Meyer GR, Herman AG, and Kockx MM. Elevated levels of oxidative DNA damage and DNA repair enzymes in human atherosclerotic plaques. *Circulation*, 106:927–932, 2002.

[166] Jason A. Wallace and Jana K. Shen. Continuous constant pH molecular dynamics in explicit solvent with pH-based replica exchange. *J. Chem. Theory Comput.*, 7(8):2617–2629, 2011.

[167] Arieh Warshel and Arno Papazyan. Electrostatic effects in macromolecules: fundamental concepts and practical modeling. *Current Opinion in Structural Biology*, 8(2):211 – 217, 1998.

[168] Bret D Freudenthal& William A Beard& Matthew J Cuneo& Nadezhda S Dyrkheeva& Samuel H Wilson. Capturing snapshots of APE1 processing DNA damage. *Nature Structural & Molecular Biology*, 22:924–931, 2015.

[169] David M Wilson. Properties of and substrate determinants for the exonuclease activity of human apurinic endonuclease APE1. *Journal of molecular biology*, 330(5):1027–1037, 2003.

[170] David M Wilson and Daniel Barsky. The major human abasic endonuclease: formation, consequences and repair of abasic lesions in DNA. *Mutation Research/DNA Repair*, 485(4):283–307, 2001.

[171] Lutz WK. Endogenous genotoxic agents and processes as a basis of spontaneous carcinogenesis. *Mutation Research/Reviews in Genetic Toxicology*, 238(3):287–295, 1990.

[172] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, and Klein ML. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys*, 79:926–935, 1983.

[173] Jencks WP. *Catalysis in Chemistry and Enzymology.* McGraw Hill., New York, 1969.

[174] Cheng Y and Patel DJ. Crystallographic structure of the nuclease domain of 3'hexo, a deddh family member, bound to ramp. *J. Mol. Biol*, 343(2):305–312, 2004.

[175] Huang Y, Chen W, Wallace JA, and Shen J. All-atom continuous constant ph molecular dynamics with particle mesh ewald and titratable water. *Journal of Chemical Theory and Computation*, 12(11):5411–5421, 2016.

[176] Liu Y, Fan X, Jin Y, Hu X, and Hu H. Computing pKa values with a mixing hamiltonian quantum mechanical/molecular mechanical approach. *J. Chem. Theory Comput.*, 9(9):4257–4265, 2013.

[177] Masuda Y, Bennett RA, and Demple B. Dynamics of the interaction of human apurinic endonuclease (APE1) with its substrate and product. *Journal of Biological Chemistry*, 273(46):30352–30359, 1998.

[178] Nozaki Y and Tanford C. Examination of titration behavior. *Methods Enzymol*, 11:715–734, 1967.

[179] Zhang Y. Pseudobond ab initio qm/mm approach and its applications to enzyme reactions. *Theoretical Chemistry Accounts*, 116(1-3):43–50, 2006.

[180] Zhang Y, Liu H, and Yang W. Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio qm/mm potential energy surface. *The Journal of Chemical Physics*, 112(8):3483–3492, 2000.