



Institute for Physical Chemistry

Computational Chemistry and Biochemistry Group

Bachelor Thesis

Proton mediation in human 20S proteasome catalysis

Name: David Prekel
21674993
david.prekel@stud.uni-goettingen.de
from Freren

First Examiner: Prof. Dr. Ricardo Mata
Second Examiner: Prof. Dr. Kai Tittmann

Contents

1	Introduction	3
2	Theory	6
2.1	Potential Energy Surface (PES)	6
2.2	Electronic Structure Methods	7
2.2.1	Density Functional Theory	7
2.2.2	DLPNO-CCSD(T)	8
2.3	Hybrid QM/MM Calculations	9
2.4	Nudged Elastic Band (NEB) method	11
2.4.1	Improved Tangent NEB	12
2.4.2	Climbing Image NEB (CI-NEB)	12
3	Computational Details	14
3.1	System Setup	14
3.2	Generation of minimum energy states	17
3.2.1	Structure generation	17
3.2.2	L-BFGS energy minimizations	20
3.3	Setup of NEB calculations	20
3.4	Wave function based calculations	20
4	Results and Discussion	22
4.1	General consideration of the inhibition mechanism	22
4.2	Reaction pathways for the active site with neutral charge	25
4.2.1	First step: Activation of the nucleophile	25
4.2.2	Second step: Protonation of the carbonyl group	28
4.3	Reaction pathways for the active site with charge -1	31
4.4	Reaction pathways for the active site with charge $+1$	33
4.5	NEB convergence behaviour	35
4.6	Single point energy calculations	37
4.7	Conclusions for the reaction mechanism	39
5	Summary and Outlook	42
6	Appendix	44

Chapter 1

Introduction

A wide variety of proteins can be found in each human cell, which are part of a complex yet sophisticated network, that guarantees all vital functions. Fine tuning of these intracellular processes is necessary for maintaining homeostasis and the cell cycle. Under the influence of various environmental conditions, the interaction of the cell components can be irreversibly disrupted which can lead to the development of many different diseases. In this context, the so-called ubiquitin-proteasome system (UPS) plays a crucial role being responsible for the constant hydrolysis of proteins that are either misfolded or are no longer needed.^[1] The identification of these proteins to be degraded is pursued through a prior labelling with ubiquitins, small proteins consisting of only 76 amino acids with a molecular mass of 8.5 kDa. Ubiquitinated proteins are then transported to the 26S proteasome, a large molecular machine of 2500 kDa, which is located in the cell nucleus and cytoplasm, and are degraded there.^[2, 1] Interestingly, it could be observed that if the cell is unable to remove specific proteins due to a breakdown of the UPS, apoptosis is usually initiated. This observation is of significant importance for diseases like cancer. Cancer cells are known to divide at high speed and are therefore more sensitive to a dysfunction of the UPS.^[3, 4, 5]

The 26S proteasome consists of one central 20S particle and two 19S particles, which in turn are composed of several proteins. In total, the complex consists of about 31 subunits. The size of the proteasome is 150 Å × 115 Å.^[6] The 20S subunit has the shape of a hollow cylinder and acts as a multicatalytic protease. It consists of four rings, which in turn are composed of seven subunits each (α_1 to α_7 and β_1 to β_7). The two inner rings consist of β subunits, the two outer ones of α subunits. While the outer rings are responsible for substrate recognition and substrate access, the actual proteolytic activity is located at the inner side of the β subunits 1, 2 and 5, which show caspase-,

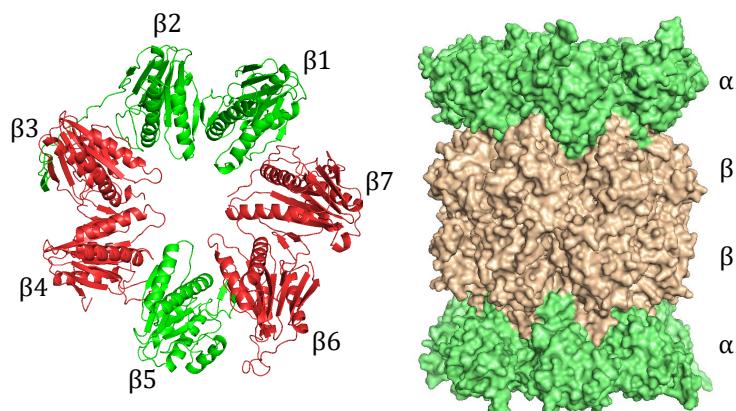


Fig. 1.1: Structure of the 20S Proteasome. Right: The α and β rings highlighted in green and brown. Left: Structure of one β ring with the catalytically active subunits highlighted in green.

tryptic- and chymotryptic-like activity, respectively. The different catalytic subunits have different substrate specificities and are believed to cleave for acidic ($\beta 1$), basic ($\beta 2$) and hydrophobic ($\beta 3$) amino acids, respectively.^[1] Nevertheless, their active sites are highly conserved and always contain an aspartate (Asp17), a lysine (Lys33) and an N-terminal threonine (Thr1). From now on, the 1-letter code will be used for the amino acid residues (e.g. D17, K33, T1).

It is generally accepted that the first step in the reaction is to activate T1, which then attacks the peptide bond and forms a covalently bound, tetrahedral intermediate. This is followed by an autocatalytic cleavage of the tetrahedral intermediate, which releases the N-terminal part of the cleaved peptide. The acyl intermediate that is formed is then attacked nucleophilically by water, creating a new tetrahedral intermediate. This also breaks down and releases the C-terminal part of the cleaved peptide, as a result of which the original active center is restored.^[7] A proposed reaction mechanism for the protonation state T1-NH₂;K33-NH₃⁺ is shown in Fig. 1.2.

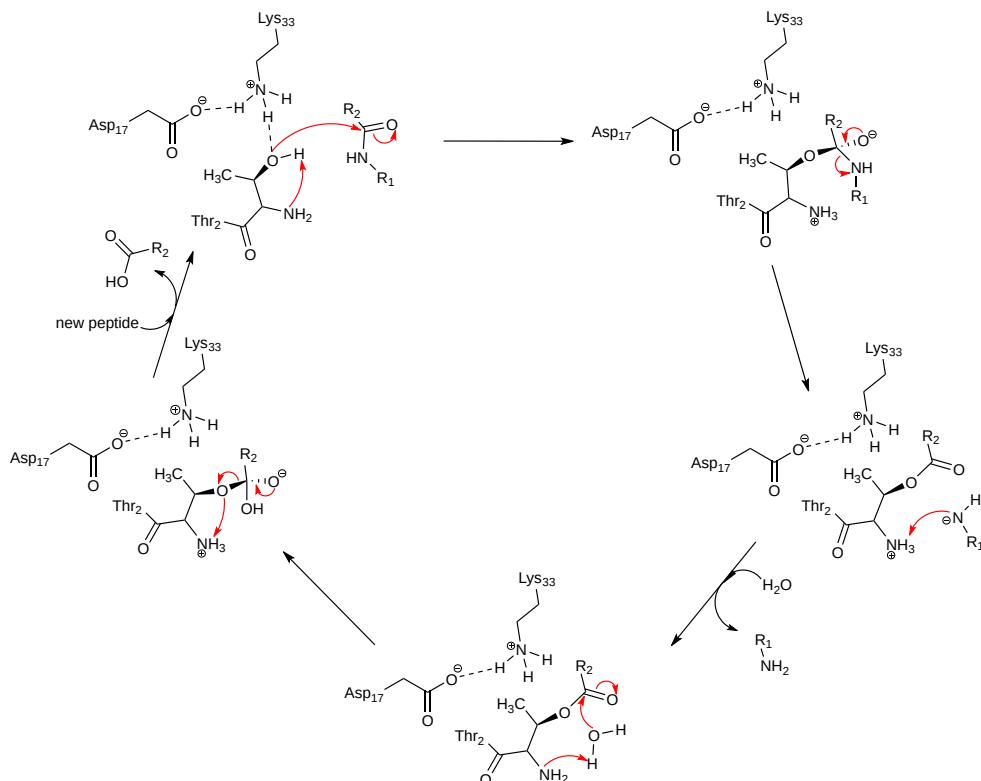


Fig. 1.2: One possible reaction mechanism for substrate processing in the $\beta 1$ active site of the 20S proteasome. A protonation state T1-NH₂;K33-NH₃⁺ is assumed here.

The covalent catalysis that takes place in the 20S proteasome makes it an attractive target for the design of inhibitor molecules that bind irreversible to the active site. Covalently binding inhibitors have the advantage of a high binding affinity and can therefore be administered as medicament in lower doses.^[8] This fact, as well as the recognition of the central importance of the UPS in the cancer process, have led to a strong interest in potential inhibitors for the human 20S proteasome as cancer drugs. Many different proteasome inhibitors have been developed over the years, especially those with epoxyketone warheads such as Epoxomicin, Dihydroeponemycin, and Oprozomib. Oprozomib, for example, the subject of this work, was developed by Onyx

Pharmaceuticals Inc. and is currently being tested in clinical studies.^[9] Epoxyketone-containing inhibitors bind covalently to the N-terminal T1 in the form of a 1,4-oxazepane-7 ring.^[10]

In most cases it is assumed that the first step of the inhibition reaction, as in substrate processing, is a nucleophilic attack of the T1 hydroxyl group on the carbonyl group of the epoxyketone region of the inhibitor,^[11] although it must be mentioned that other mechanisms for epoxyketone inhibitors were postulated.^[12] Knowledge of the exact mechanism is necessary in order to develop even more specific, high-affinity inhibitor molecules with fewer side effects in the future. However, the exact mechanism is still unknown. In this context, the protonation states of the catalytically active amino acids play a decisive role. Both the N-terminal amino group and the K33 amino group can appear either as NH₂ or NH₃⁺ during catalysis. As a result, very different electrostatic environments can be generated in the active site, which could cause both a ground-state destabilization and a transition-state stabilization. In this context, the question arises to what extent the surrounding amino groups influence the pK_a value of the T1 hydroxyl group. The optimal pH value for the activity of the 20S proteasome is 7-8,^[13] so that a significant reduction in the pK_a value for the T1 alcohol appears to be necessary.

In this work different possible protonation states within the active region of the caspase site ($\beta 1$) are considered. The nudged elastic band (NEB) method will be applied to different potential educt, intermediate and product states in order to better understand how a changed protonation state affects the energy barriers of the respective partial reactions. The focus is on the investigation of the first important step in catalysis, namely the nucleophilic attack of the T1 alcohol on the carbonyl group of the Oprozomib molecule and the subsequent formation of the tetrahedral intermediate.

Chapter 2

Theory

2.1 Potential Energy Surface (PES)

In the non-relativistic limit, the dynamics of a molecular system is completely described by the time-dependent Schrödinger equation

$$\mathrm{i}\hbar\frac{\partial}{\partial t}|\Xi(t)\rangle = \hat{H}|\Xi(t)\rangle \quad (2.1)$$

with the molecular structure Hamiltonian

$$\hat{H} = \hat{T}_n + \underbrace{\hat{T}_e + \hat{V}_{nn} + \hat{V}_{ee} + \hat{V}_{ne}}_{\hat{H}_{\mathrm{el}}}. \quad (2.2)$$

Here \hat{T}_n and \hat{T}_e are the kinetic energy operators of nuclei and electrons, \hat{V}_{nn} is the Coulomb repulsion between the nuclei, \hat{V}_{ee} is the Coulomb repulsion between the electrons and \hat{V}_{ne} the Coulomb attraction between electrons and nuclei. Since the Schrödinger equation (2.1) is impossible to solve for actual molecules, a separation into nuclei- and electron-dependent parts is carried out first. For this, the total wavefunction $|\Xi(t)\rangle$ is expanded in terms of eigenfunctions of \hat{H}_{el} , followed by an integration over all electronic coordinates. Finally, this procedure results in a time-independent Schrödinger equation

$$\hat{H}_{\mathrm{el}}|\Psi_j\rangle = E_j|\Psi_j\rangle \quad (2.3)$$

for the electrons, where the nuclear coordinates only appear as parameters, and another time-dependent set of coupled differential equations

$$(\hat{T}_n(\vec{R}) + E_j(\vec{R}))|\chi_j(t)\rangle + \sum_k \hat{C}_{jk}|\chi_k(t)\rangle = \mathrm{j}\hbar\frac{\partial}{\partial t}|\chi_j(t)\rangle, \quad (2.4)$$

which depend on nuclear coordinates as well as electronic coordinates.^[14] The expression $E_j(\vec{R})$ can in principle be obtained by solving equation (2.3) for all possible values of the $3N$ -dimensional vector \vec{R} . It is called the potential energy surface (PES) for eigenstate j . The operator \hat{C}_{jk} is responsible for the non-adiabatic coupling between the PESs of different eigenstates E_j . Since

the electronic ground state and the first excited states are usually clearly separated from each other, the coupling is often neglected. This is the heavily used Born-Oppenheimer approximation, the foundation of quantum chemistry. The resulting simplified nuclear Schrödinger equation describes the dynamics of the nuclear wave packets on a single potential energy surface. Due to the relatively strong localization of the wave packets, the nuclear dynamics can also be described by purely classical molecular dynamics simulations in most cases.^[14]

Another possibility of investigating chemical systems is to ignore the nuclear dynamics completely and instead to try to find minima and saddle points on the PES with the help of certain algorithms and thus to be able to estimate energy barriers. Rate constants can then be determined using Eyrings transition state theory.^[15]

2.2 Electronic Structure Methods

2.2.1 Density Functional Theory

Density functional theory allows theoretically to circumvent the disadvantageous scaling of wave function-based methods with the system size, since its aim is not to determine the wave function ($3N$ dimensional) but the 3-dimensional electron density. The basis is the Hohenberg-Kohn theorem, which states that the ground state energy of a system is directly determined by its electron density. Therefore the ground state energy can be defined as a functional of the electron density $\rho(\vec{r})$:^[16]

$$E_{\text{tot}}[\rho(\vec{r})] = T_{\text{S}}[\rho] + J[\rho] + E_{\text{XC}}[\rho] + E_{\text{Ne}}[\rho]. \quad (2.5)$$

Now the ground state energy of a many-electron system can be approximately determined by solving the Kohn-Sham equations. The Kohn-Sham equations are based on the introduction of a fictitious reference system in which the electrons do not interact directly with each other, but only indirectly via an effective potential $V_{\text{eff}}(\vec{r})$. Thus an effective one-particle operator can be introduced, the Kohn-Sham operator:^[16]

$$\hat{f}^{\text{KS}} = -\frac{1}{2}\nabla^2 + V_{\text{eff}}(\vec{r}). \quad (2.6)$$

It is a formal analog of the Fock operator. Similar to the Hartree-Fock method, a one-electron Schrödinger equation (or Schrödinger-like equation)

$$\hat{f}^{\text{KS}}|\phi_i\rangle = \varepsilon_i|\phi_i\rangle \quad (2.7)$$

can be defined with it. The Kohn-Sham orbitals are similar to the molecular orbitals from HF calculations, but they do not have the same physically descriptive meaning. To establish a connection between the energy functional (2.5) and the Kohn-Sham operator (2.6), the electron density for the system is taken as the sum of all occupied squared Kohn-Sham orbitals:^[16]

$$\rho(\vec{r}) = \sum_{i=1}^N |\phi_i(\vec{r})|^2. \quad (2.8)$$

Now a large part of the density functional (2.5) can be expressed as a function of (2.8) and can be minimized using the method of undetermined Lagrange multipliers. This results in N uncoupled one-electron eigenvalue equations of the form

$$\underbrace{\left(-\frac{1}{2}\nabla^2 + \int \frac{1}{r_{12}}\rho(\vec{r}_2)d^3r_2 + \frac{\delta E_{XC}}{\delta\rho(\vec{r}_1)} - \sum_A^M \frac{Z_A}{r_{1A}} \right)}_{=V_{\text{eff}}(\vec{r})} |\phi_i\rangle = \varepsilon_i |\phi_i\rangle. \quad (2.9)$$

The effective potential depends explicitly on the Kohn-Sham orbitals, so that the Kohn-Sham equations have to be solved iteratively to self-consistency.^[16] By using a suitable basis set, $|\phi_i\rangle = \sum_{\nu=1}^K |\xi_{\nu}\rangle \langle \xi_{\nu}| \phi_i\rangle$, the Kohn-Sham equations can be solved with standard methods of linear algebra.

2.2.2 DLPNO-CCSD(T)

In contrast to density functional theory, wavefunction-based post-Hartree-Fock methods are computationally more demanding. But they have the advantage that their accuracy can be systematically improved by using a larger basis set and including more parts of the correlation energy, although it is not straightforward to do so in a computationally efficient way. A variety of advanced wave function methods have been developed, such as Configuration Interaction (CI),^[17] Coupled Cluster (CC), and Møller-Plesset perturbation theory.^[18] CCSD(T) is regarded as the gold standard of electronic structure methods, which allows for a relatively cost-efficient solution of the time-independent Schrödinger equation for the electronic ground state.^[19]

The Pauli principle requires that the total wave function must be totally antisymmetric with respect to the interchange of two electrons. Since Slater determinants have the desired symmetry properties, the wave function can be set up as a linear combination of Slater determinants.^[20] In Coupled Cluster theory, the wave function is formally constructed as the exponential of the so called excitation operator applied to the reference determinant:

$$|\Psi\rangle = e^{\hat{T}}|\Psi_0\rangle. \quad (2.10)$$

For the CCSD method, the excitation operator \hat{T} is composed of two parts

$$\hat{T} = \hat{T}_1 + \hat{T}_2, \quad (2.11)$$

which generate linear combinations of all possible 1- and 2-fold excited determinants. The insertion of (2.10) into the electronic Schrödinger equation and projection onto the determinant states yields the CCSD equations

$$\langle\Psi_0|e^{-(\hat{T}_1+\hat{T}_2)}\hat{H}_{\text{el}}e^{(\hat{T}_1+\hat{T}_2)}|\Psi_0\rangle = E, \quad (2.12)$$

$$\langle\Psi_i^a|e^{-(\hat{T}_1+\hat{T}_2)}\hat{H}_{\text{el}}e^{(\hat{T}_1+\hat{T}_2)}|\Psi_0\rangle = R_i^a = 0, \quad (2.13)$$

$$\langle\Psi_{ij}^{ab}|e^{-(\hat{T}_1+\hat{T}_2)}\hat{H}_{\text{el}}e^{(\hat{T}_1+\hat{T}_2)}|\Psi_0\rangle = R_{ij}^{ab} = 0. \quad (2.14)$$

This is a matrix eigenvalue problem.^[21] Before the CCSD equations are solved, a Hartree-Fock

calculation is always carried out first; in the case of singlet ground states, these are restricted Hartree-Fock calculations. The obtained molecular orbitals are then used as components of the Slater determinants in the CCSD calculations.

The exact solution of the CCSD equations, however, scales with $O(N^6)$ and therefore very unfavorably with system size. The DLPNO ("Domain based local pair natural orbital") method has been developed to circumvent this problem. It enables terms that result in zero to be excluded from the calculation from the outset. Almost linear scaling can thus be achieved. The Hartree-Fock molecular orbitals are not used directly in the DLPNO method. Instead, localized PAOs ("projected atomic orbitals") are used. In order to construct the PAOs, the molecular orbitals are first localized and then the localized MOs $\{|i\rangle\}$ are projected onto atomic orbitals $|\mu\rangle$ ^[21] to yield

$$|\tilde{\mu}\rangle = \left(1 - \sum_i |i\rangle\langle i|\right) |\mu\rangle. \quad (2.15)$$

In a second step, a rather complicated so called "electron pair prescreening" is carried out. For this, a second order Møller Plesset (MP2) calculation is done in order to obtain pair densities $d_{\tilde{\mu}}^{ij}$ from which the "pair natural orbitals" (PNOs) are calculated as

$$|\tilde{a}_{ij}\rangle = \sum_{\tilde{\mu} \in ij} d_{\tilde{\mu}}^{ij} |\tilde{\mu}\rangle. \quad (2.16)$$

Once the PNOs are constructed, they can be used to solve the CCSD equations (2.13) to (2.14) using a direct-inversion in iterative subspace algorithm.^[21] The triples correction is calculated on top of the solution of the CCSD equations to allow for the inclusion of the important triple excitations. This is done in a perturbative fashion.^[22]

2.3 Hybrid QM/MM Calculations

In order to correctly describe chemical reactions in enzymes, it is necessary to describe the active site using methods from electronic structure theory. In this way, around 100 atoms can often be treated quantum mechanically within a manageable period of time.^[23] But the protein environment and the solvent also have an influence on the reactivity in the active site and must therefore be taken into account. Molecular mechanics force fields are used for this. A classical force field is based upon a rather simple model of interatomic interactions and consists of two parts: On the one hand, of terms that describe the covalent (bonded) interactions between the atoms and, on the other hand, of terms that describe the non-covalent (nonbonded) interactions:

$$E_{\text{MM}}(\{\vec{R}\}) = E_{\text{MM}}^{\text{cov}}(\{\vec{R}\}) + E_{\text{MM}}^{\text{noncov}}(\{\vec{R}\}). \quad (2.17)$$

The covalent terms usually have a simple functional form. They depend on all atoms $\{\vec{R}\}$ of the system:

$$E_{\text{MM}}^{\text{cov}}(\{\vec{R}\}) = \sum_{\text{bonds}} \frac{1}{2} k_\alpha (d_\alpha - d_{\alpha,0})^2 + \sum_{\text{angles}} \frac{1}{2} \tilde{k}_\alpha (\theta_\alpha - \theta_{\alpha,0})^2 + \sum_{\text{dihedral}} \frac{1}{2} V_\alpha (1 + \cos(n_\alpha \omega_\alpha - \omega_{\alpha,0}))^2. \quad (2.18)$$

The first term is a two-body term describing the interaction of bonded atoms, modelled by harmonic potentials that give the increase in energy as the bond length d_α deviates from the reference value $d_{\alpha,0}$. The second term is a three-body term that describes the opening and closing of all possible angles between three neighboring atoms, again modelled by harmonic potentials. The third term is a four-body term containing torsional potentials that model how the energy changes as bonds rotate.^[24]

In contrast to the covalent interactions, the non-bonded contributions to the energy are calculated between all pairs of atoms (I and J) that are in different molecules or in the same molecule but separated by at least three bonds.^[24] In a simple force field these contributions usually look like

$$E_{\text{MM}}^{\text{noncov}}(\{\vec{R}\}) = \sum_I \sum_{J>I} \frac{q_I q_J}{4\pi\varepsilon_0 R_{IJ}} + \sum_I \sum_{J>I} 4\varepsilon_{IJ} \left[\left(\frac{\sigma_{IJ}}{R_{IJ}} \right)^{12} - \left(\frac{\sigma_{IJ}}{R_{IJ}} \right)^6 \right]. \quad (2.19)$$

Here the first term is the Coulomb interaction that describes the (long-range) electrostatic interactions. A frequently used method for its evaluation is the "Particle Mesh Ewald" (PME) method. The second term is a Lennard-Jones potential which is responsible for the inclusion of the Pauli repulsion ($1/R_{IJ}^{12}$ term) and the van-der-Waals attraction force ($1/R_{IJ}^6$ term).^[24]

There are different methods of calculating the total energy in a combined QM/MM system such as an enzyme. The additive method is used most often. This contains three components that are added together:

$$E_{\text{tot}} = E_{\text{QM}}^{\text{vac}} + E_{\text{MM}} + E_{\text{QM/MM}}. \quad (2.20)$$

The term $E_{\text{QM}}^{\text{vac}}$ corresponds to the total energy of the isolated QM region in the vacuum. It is often the case that the dividing line between the QM and MM regions runs along a covalent bond. In this case, the QM system is saturated at the points where the bond is broken by adding virtual hydrogen atoms (linker atoms) in order to ensure correct spin multiplicity.^[23, 25]

The most complicated part is correctly describing the interaction of the QM and MM regions, which is predominantly electrostatic in nature. Various methods for calculating $E_{\text{QM/MM}}$ have been developed. The most frequently used method is the so-called electrostatic embedding. Here the polarization of the QM system is taken into account through the interaction with the point charges of the MM system using the two additional terms

$$E_{\text{QM/MM}}^{\text{el}} = \sum_I \left(\int \frac{q_I \rho(\vec{r})}{\|\vec{R}_I - \vec{r}\|} d^3r + \sum_{I'} \frac{q_I q_{I'}}{\|\vec{R}_I - \vec{R}_{I'}\|} \right). \quad (2.21)$$

The I index labels all MM nuclei while I' labels all QM nuclei. So the symbols q_I are the MM point charges located at \vec{R}_I , while $q_{I'}$ are the nuclear charges of the QM atoms located at $\vec{R}_{I'}.$ ^[25, 26] Note that \vec{R} is a 3-dimensional vector here and not a $3N$ dimensional one as in the rest of this thesis. The electron density is given by $\rho(\vec{r})$. Taking the expression $E_{\text{QM/MM}}^{\text{el}}$ into account, one obtains the energy of the polarized QM system by adding the polarization correction to the energy of the isolated QM system to yield

$$E_{\text{QM}}^{\text{pol}} = E_{\text{QM}}^{\text{vac}} + E_{\text{QM/MM}}^{\text{el}}. \quad (2.22)$$

The van-der-Waals interaction between QM and MM atoms and the bonded contributions between QM and MM atoms also contribute to the interaction term, so that the overall interaction term has the form

$$E_{\text{QM/MM}} = E_{\text{QM/MM}}^{\text{el}} + E_{\text{vdW}} + E_{\text{bond}} + E_{\text{ang}} + E_{\text{dih}}. \quad (2.23)$$

2.4 Nudged Elastic Band (NEB) method

The Nudged Elastic Band (NEB) method is a method for finding saddle points and Minimum Energy Paths (MEPs) between known educt and product states on the potential energy surface. An elastic band can be described by a set of vectors $[\vec{R}_0, \vec{R}_1, \vec{R}_2, \dots, \vec{R}_n]$ with $3N$ entries each (N : number of atoms). The basic idea is to optimize the components of this elastic band (images) in the coordinate space so that the force \vec{F}_i^{NEB} specifically defined for this converges to zero for all images. This force is broken down into two components

$$\vec{F}_i^{\text{NEB}} = \vec{F}_{i\parallel}^s + \vec{F}_{i\perp} = \vec{F}_{i\parallel}^s - \nabla E(\vec{R}_i)_\perp. \quad (2.24)$$

Here $\vec{F}_{i\perp}$ is the perpendicular force component and $\vec{F}_{i\parallel}^s$ is the parallel force component.^[27, 28, 29] The perpendicular force can be calculated by first taking the energy gradient of image i at point \vec{R}_i , and then subtracting the gradient component parallel to $\hat{\vec{\tau}}_i$

$$\nabla E(\vec{R}_i)_\perp = \nabla E(\vec{R}_i) - \langle \nabla E(\vec{R}_i), \hat{\vec{\tau}}_i \rangle \cdot \hat{\vec{\tau}}_i. \quad (2.25)$$

The vector $\hat{\vec{\tau}}_i$ is the normalized tangent vector to the image \vec{R}_i , that is $\hat{\vec{\tau}}_i = \vec{\tau}_i / \|\vec{\tau}_i\|$.^[27, 28, 29] In the vicinity of the MEP, the directions of the energy gradient $\nabla E(\vec{R}_i)$ and the tangent vector converge more and more, so that their difference vector becomes smaller and smaller and converges to zero.

There are different variants of the NEB method, which differ mainly in how the tangent vector and the parallel force component are calculated. In the original formulation of the NEB method by Mills and Jonsson, the tangent vector at point \vec{R}_i was approximated as the normalized difference vector between the previous point \vec{R}_{i-1} and the next point \vec{R}_{i+1} within the elastic band,

$$\hat{\vec{\tau}}_i = \frac{\vec{R}_{i+1} - \vec{R}_{i-1}}{\|\vec{R}_{i+1} - \vec{R}_{i-1}\|}. \quad (2.26)$$

With the help of the tangent vector the parallel force $\vec{F}_{i\parallel}^s$ can be calculated. This is an artificially added force that has the role of creating distances between the images that are as equidistant as possible. It is defined by Hooke's law with a force constant k and reads^[27, 28, 29]

$$\vec{F}_{i\parallel}^s = k[(\vec{R}_{i+1} - \vec{R}_i) - (\vec{R}_i - \vec{R}_{i-1})] \cdot \hat{\vec{\tau}}_i \hat{\vec{\tau}}_i. \quad (2.27)$$

For each cycle in the NEB algorithm, an energy calculation and an energy gradient calculation must be carried out on each image. In principle, any sufficiently precise electronic structure method is suitable for this. In reality, however, one is dependent on less computationally expensive methods with a good cost-performance ratio. This makes density functional theory the method

of choice.

2.4.1 Improved Tangent NEB

The NEB method by Mills and Jonsson has the disadvantage that the energy path develops kinks in the course of the calculation, which have a negative effect on the convergence of the method. Program packages such as Chemshell therefore use an improved version, the Improved Tangent method by Henkelmann and Jonsson.^[28, 30]

In the Improved Tangent Method, the tangent is calculated using the image i and the image of higher energy (either $i + 1$ or $i - 1$). The tangent vector is calculated via

$$\vec{\tau}_i = \begin{cases} \vec{\tau}_i^+ & \text{if } E_{i+1} > E_i > E_{i-1} \\ \vec{\tau}_i^- & \text{if } E_{i+1} < E_i < E_{i-1} \end{cases} \quad (2.28)$$

with

$$\vec{\tau}_i^+ = \vec{R}_{i+1} - \vec{R}_i \quad \text{and} \quad \vec{\tau}_i^- = \vec{R}_i - \vec{R}_{i-1} \quad (2.29)$$

and E_i being the energy of image i , $E(\vec{R}_i)$.^[28] In the case that image i lies at an energy maximum or minimum compared to the adjacent images $i - 1$ and $i + 1$, the calculation of the tangent is a bit more complicated. The tangent is then taken to be a weighted average of the vectors to the two neighboring images. So if $E_{i+1} > E_i < E_{i-1}$ or $E_{i+1} < E_i > E_{i-1}$ applies, the tangent vector is calculated as^[28]

$$\vec{\tau}_i = \begin{cases} \vec{\tau}_i^+ \Delta E_i^{\max} + \vec{\tau}_i^- \Delta E_i^{\min} & \text{if } E_{i+1} > E_i > E_{i-1} \\ \vec{\tau}_i^+ \Delta E_i^{\min} + \vec{\tau}_i^- \Delta E_i^{\max} & \text{if } E_{i+1} < E_i < E_{i-1} \end{cases} \quad (2.30)$$

where

$$\Delta E_i^{\max} = \max(|E_{i+1} - E_i|, |E_{i-1} - E_i|) \quad (2.31)$$

and

$$\Delta E_i^{\min} = \min(|E_{i+1} - E_i|, |E_{i-1} - E_i|) \quad (2.32)$$

So in this case there are a total of four options for the tangent vector to be calculated.

2.4.2 Climbing Image NEB (CI-NEB)

The main goal of the Improved Tangent NEB method described above is to make sure that the elastic band converges to the MEP, but it is unlikely that any of the images actually lies exactly on the transition state. To achieve this, a small modification of the NEB algorithm can be introduced. This modified NEB method is called Climbing Image NEB (CI-NEB).^[31] In this procedure, the normal Improved Tangent NEB procedure is carried out up to a certain degree of convergence. When a specified multiple of the convergence criterion is reached, the algorithm begins to change the procedure: The image with the highest energy is selected as the climbing image and optimized with a modified force so that its energy is maximized in the direction of the transition state.^[28] This modified force is given by

$$\vec{F}_{i_{\max}} = -\nabla E(\vec{R}_{i_{\max}}) + 2\langle \nabla E(\vec{R}_{i_{\max}}), \hat{\vec{\tau}}_{i_{\max}} \rangle \hat{\vec{\tau}}_{i_{\max}} \quad (2.33)$$

In cases in which it is relatively easy to calculate the MEP and one wants to gain more precise knowledge about the transition state, this method is useful. On the other hand, it makes less sense when the primary aim is to calculate the height of energy barriers for a chemical reaction.

Chapter 3

Computational Details

This thesis centers around the investigation of the influence of different protonation states on the reactivity in the active site of the $\beta 1$ subunit of the human 20S proteasome. In the following, the computational details of the theoretical calculations will be presented. The main focus of this work is on the computation of minimum energy states and the associated energy paths that connect these states with each other. After that, wave-function based methods were employed in order to refine the energy paths that were obtained with the NEB method. Before the calculations could be started, the initial PDB files used had to be modified and the system had to be equilibrated with short MD simulations. These two preparatory steps were carried out by Jon Uranga and are also described here for the sake of completeness.

3.1 System Setup

PDB file modification

The object of investigation is the $\beta 1$ subunit of the human 20S proteasome, based on the PDB structure 5LF1. The 3D structure determined by means of X-ray crystallography has a resolution of 2.0 Å and shows the 20S proteasome in complex with the epoxyketone inhibitor Dihydroeponemycin. Crystal structures of the 20S proteasome in complex with Oprozomib have also been published in the past.^[10] Unfortunately, Oprozomib is not present in the catalytic pocket of the $\beta 1$ subunit, so that the crystal structure with Dihydroeponemycin had to be used. The caspase-like chain (chain N) of the Dihydroeponemycin-bound proteasome was used as a starting structure. The equivalent parts of the ligand were kept, modifying only the minimal parts that lead to Oprozomib using the Chimera software.^[32] The covalent bond between T1 and the ligand was removed. At this point, PROPKA 3.1^[33, 34] was employed on the enzyme without any substrate (PDB 5LE5) in order to find out the most favorable protonation states for the hydrophilic residues. Then the system was neutralized by adding the necessary counterions and solvated in a periodic cuboid box, using the Amber18 program package.^[26] The TIP3P water potential was used for the solvent water molecules.^[35] The ff14SB force field^[36] was employed for the protein and the GAFF force field^[37] for the ligand. This procedure was done for all four systems that result from the different possible protonation states the T1 and K33 amino groups can have: T1-NH₂;K33-NH₂, T1-NH₂;K33-NH₃⁺, T1-NH₃⁺;K33-NH₂ and T1-NH₃⁺;K33-NH₃⁺.

Equilibration

All four systems were then equilibrated with the Amber18 program package according to the following procedure: The respective system was first minimized with all hydrogens restrained using a restraint weight of $10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. For this, the Steepest Descent and Conjugate Gradient algorithms were used for 1000 cycles in each case. The system was then minimized, restraining the backbone atoms of the protein with a force constant of $100 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, using the Steepest Descent and Conjugate Gradient algorithms for 1000 and 2000 cycles, respectively. The following minimization was performed by allowing only the substrate to move. All other atoms were restrained with a force constant of $100 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, using the Steepest Descent algorithm for 4000 cycles and the Conjugate Gradient algorithm for 6000 cycles. Then, the substrate was restrained with a force constant of $100 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, using the Steepest Descent algorithm for 4000 cycles and the Conjugate Gradient algorithm for 6000 cycles. A final minimization was performed allowing all the atoms to relax, using the Steepest Descent algorithm for 2000 cycles and the Conjugate Gradient algorithm for 8000 cycles.

At this point molecular dynamics simulations were performed using the Amber18 GPU code. The system was heated in the canonical ensemble (NVT) for 4 ns, slowly increasing the temperature in the first half of the dynamics from 0 to 300 K. For the next half it remained stable at 300 K. Temperature control was done by the Langevin thermostat^[38] with a collision frequency of 5 ps^{-1} . The system was then equilibrated for 1 ns at 300 K and 1 atm using the isothermal-isobaric ensemble (NPT) in order to obtain a stable density (using the CPU code). Berendsen's barostat^[39] was employed with a collision frequency of 5 ps^{-1} . A production run was then performed for 10 ns, using the same ensemble and parameter settings employed along the equilibration phase, but now using the GPU code. All molecular dynamics simulations were performed using a cutoff of 8 Å for the non-bonded interactions, updating the neighbour list for the "Particle Mesh Ewald" (PME)^[40] calculation every 50 ps. The SHAKE algorithm was enabled to constrain all bonds involving hydrogen atoms.

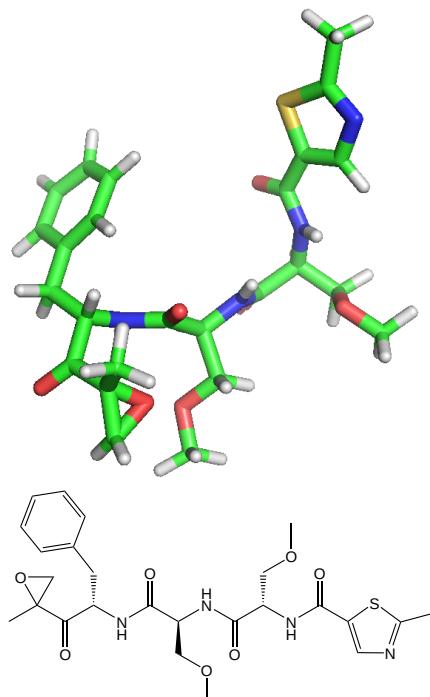


Fig. 3.1: Above: Oprozomib in its β_1 subunit-bound conformation (stick representation). Below: Oprozomib in Lewis representation.

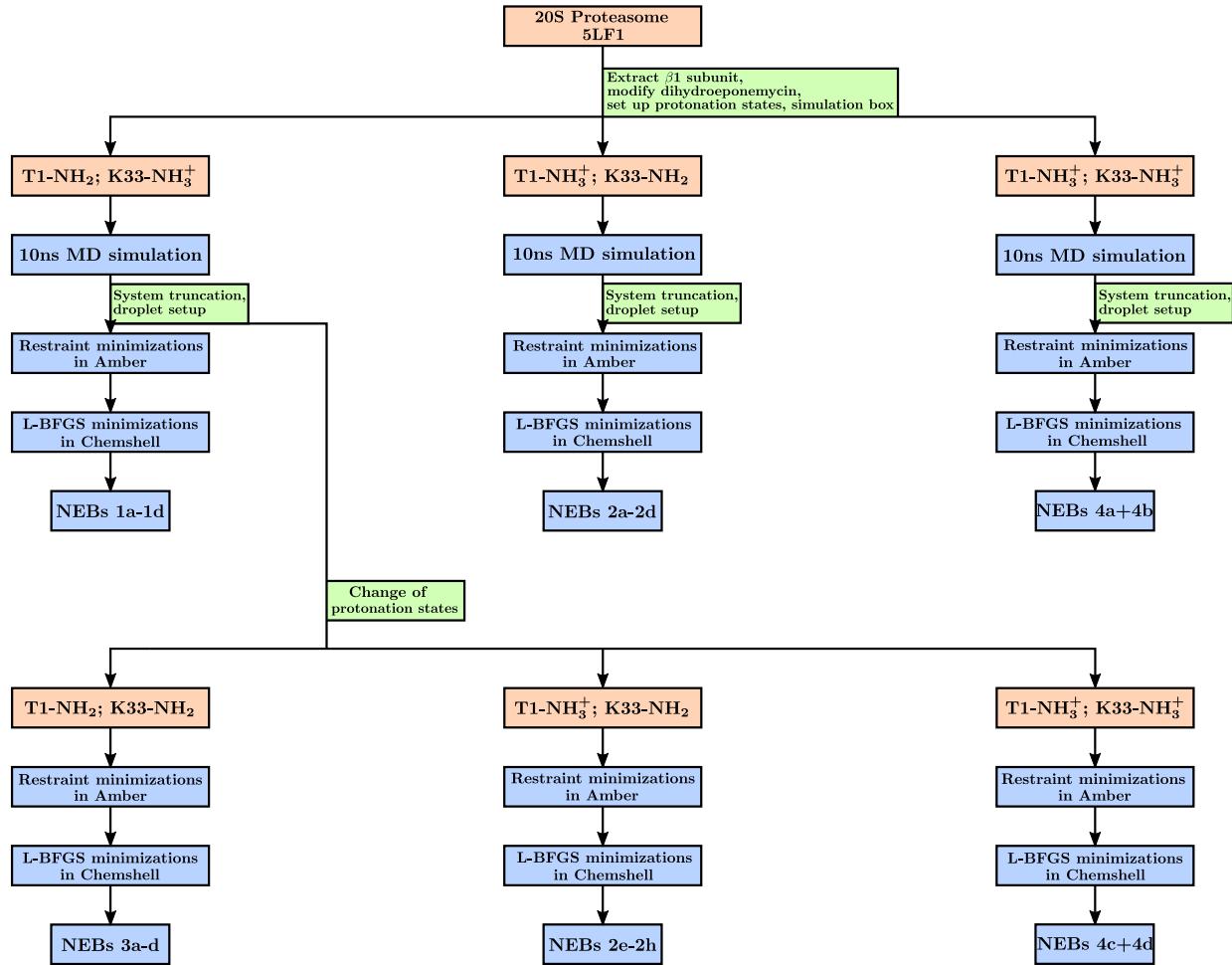


Fig. 3.2: Overview about the system setup and workflow for the different protonation states. System descriptions are colored in red, system preparation steps in green and actual calculations in blue.

Truncation and droplet setup

For each of the four protonation states, the coordinates of the last MD snapshot were used and processed as a PDB file in Chimera. First, all solvent molecules and ions which are further than 8 Å away from the protein backbone were removed. The original cuboid simulation box with periodic boundary conditions was reduced to a droplet with the $\beta 1$ subunit in the middle. For the resulting new PDB structures (`Non_periodic.pdb`), the parameter/topology file for the force fields (`.prmtop` file) and the coordinate file readable for Amber18 (`.inpcrd` file) were generated with tleap. These files formed the basis for all subsequent calculations. The used atomistic potentials were the ff14SB force field (for the protein), the GAFF force field (for Oprozomib) and the TIP3P water model (for the water solvent). In the further course of the work the `Non_periodic.pdb` file for the protonation state $T1-NH_2;K33-NH_3^+$ was used as a template for the generation of new protonation states. For this purpose, the protonation state of the `Non_periodic.pdb` file was changed with tleap. The `.prmtop` file was created in the same way as described above. The motivation for this approach was to be able to more accurately compare the structures with different protonation states but the same atomic arrangement and to create more redundancy in the reaction paths.

Choice of the QM region

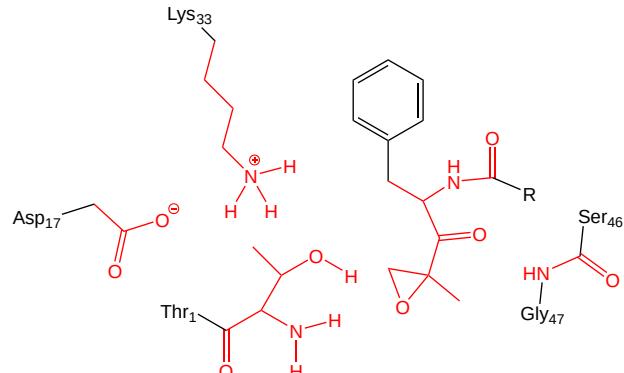


Fig. 3.3: The QM region used for all calculations is shown in red. For many calculations a nearby water molecule was also included.

For the calculations described below (minimizations, NEB calculations), a region within the active site was selected that was treated quantum mechanically (Fig. 3.3). The entire T1 amino acid, the functional group of K33, the functional group of D17 and the part of Oprozomib which is closest to the active site were placed in the QM region. In addition, the peptide bond between Ser46 and Gly47 was included, because there is a stable hydrogen bond between the protein backbone and the Oprozomib carbonyl group nearby the epoxy ketone

warhead. In addition, a nearby water molecule was included in the QM region for some calculations, which could serve as a potential proton shuttle. Overall, the size of the QM region is 64 ± 1 atoms, depending on the protonation state of the amino groups. In addition, there are three more atoms when including a water molecule and seven hydrogen linker atoms for the saturation of bonds connecting the QM region to the MM region.

3.2 Generation of minimum energy states

To carry out an NEB calculation, two end points are required, one for the educt and one for the product. The last available snapshot from the respective 10 ns molecular dynamics simulations was used as the starting structure for this. For the generation of educt and product states, the Amber18 program package was first used to create the atomic structures by carrying out energy minimizations with artificial constraints. In a second step, further energy minimizations were carried out for each structure with the Chemshell 3.5.0 program package.^[41, 42]

3.2.1 Structure generation

The Amber18 minimizations were performed for all required structures (see Fig. 3.2) according to the following procedure: If restraints were applied, which was the case in the majority of cases, five minimization runs were carried out in succession, steadily increasing the force constant from 3 to 5, 10, 50 and 350 kcal mol⁻¹ Å⁻². This should ensure that the structures remain stable and that no artifacts arise from very large gradients. 200 Steepest Descent cycles were performed in each individual minimization, followed by 300 cycles with the Conjugate Gradients algorithm. For structures without restraints, only a single minimization run was carried out, using the Steepest Descent and Conjugate Gradient algorithms for 500 and 2000 cycles, respectively. Only atoms within a radius of 7.0 Å around the hydroxyl oxygen atom of T1 were included in the minimizations. All other atoms were kept frozen by applying harmonic restraint potentials with a force constant of 1000 kcal mol⁻¹ Å⁻². The QM region was treated with the "Density Functional Tight Binding" (DFTB) method,^[43] while the surrounding protein was treated with the molecular mechanics

ff14SB force field and the solvent water molecules were treated with the TIP3P water potential. The PME method was used for the accurate description of long-range electrostatic interactions between QM atoms and between QM and MM atoms.

In the following, the choice of restraints for the products/educts/intermediates of the different protonation states will be shown. In general, all structures, including those of the products, were created directly from the modified MD snapshots. The restraints were chosen with the idea of using as few restraints as possible for the generation of a specific molecular structure.

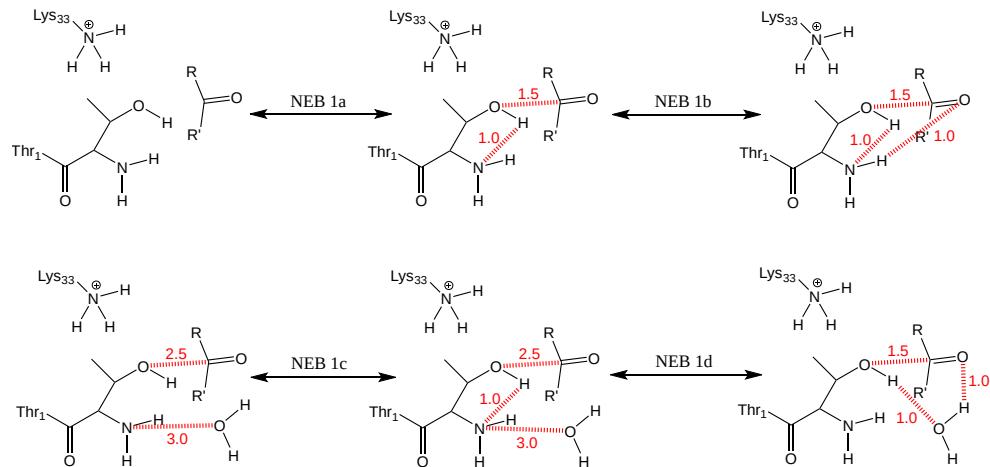


Fig. 3.4: Restraints used for the initial structure generation with Amber18 (NEBs 1a-d)

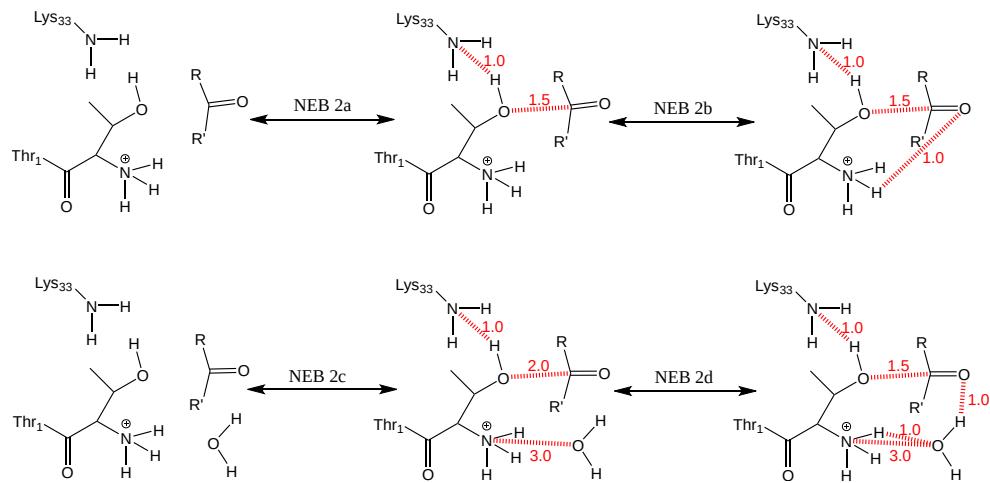


Fig. 3.5: Restraints used for the initial structure generation with Amber18 (NEBs 2a-d)

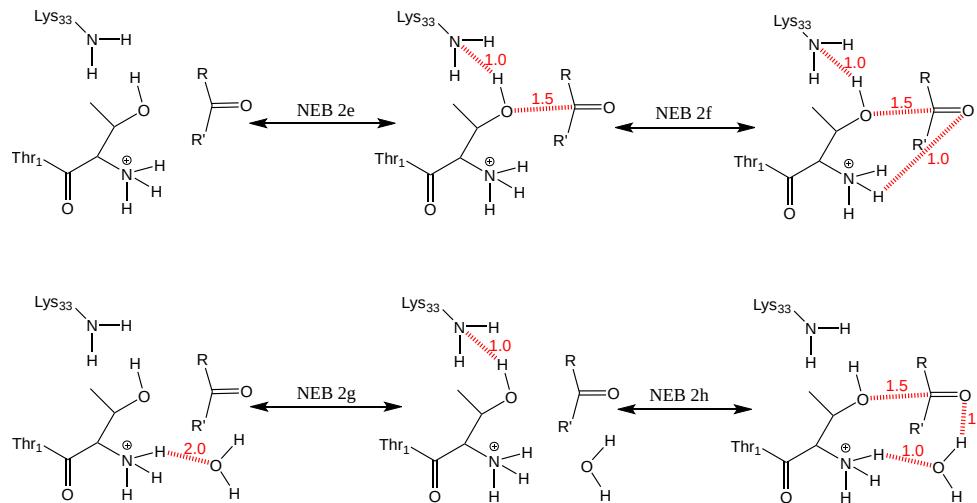


Fig. 3.6: Restraints used for the initial structure generation with Amber18 (NEBs 2e-h)

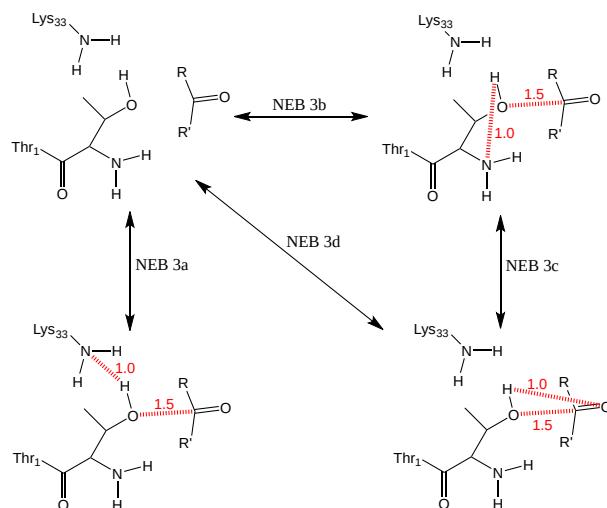


Fig. 3.7: Restraints used for the initial structure generation with Amber18 (NEBs 3a-d)

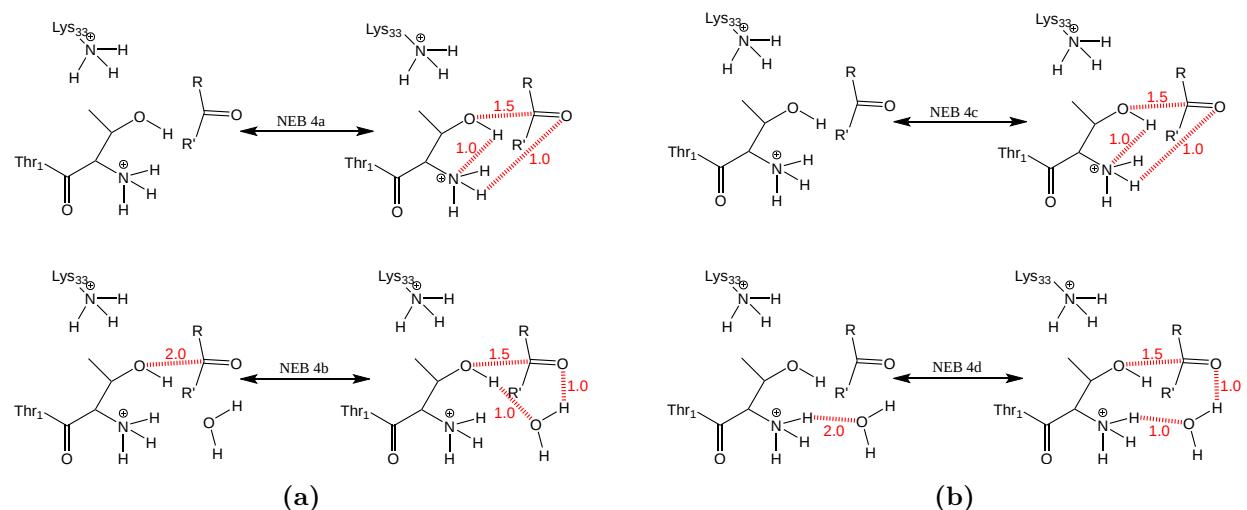


Fig. 3.8: Restraints used for the initial structure generation with Amber18 (NEBs 4a-d)

3.2.2 L-BFGS energy minimizations

After creating the endpoint structures, the energy of the obtained structures was minimized with the Chemshell program package. The energy calculations of the molecular mechanics region were carried out via the Chemshell interface to DL_POLY.^[44] The topology and the force field parameters were adopted from the previous Amber minimizations. The scaling factors for electrostatic and van-der-Waals interactions were set to 1/1.2 and 0.5, respectively. The electrostatic interactions were fully calculated for all atoms by disabling the cutoff with the `use_pairlist=0` setting. The energies of the QM region were calculated via the Chemshell interface to the ORCA 4.0.1 program package^[45, 46] using dispersion-corrected density functional theory (DFT-D3).^[47] The B3LYP hybrid functional^[48] and a def2-SVP basis set^[49] were used. The QM/MM minimization was carried out with the geometry optimization library DL-FIND^[30] which is integrated into Chemshell. For this the L-BFGS algorithm ("Limited memory Broyden-Fletcher-Goldfarb-Shanno")^[50] was used. The maximum number of optimization cycles was set to 500.

3.3 Setup of NEB calculations

After the endpoint structures had been tested for convergence, attempts were made to find the MEPs between the obtained educt and product states using the nudged elastic band method. For the energy evaluations of the QM and MM regions the same settings and methods were used as for the previous Chemshell minimizations. Electrostatic embedding (`coupling=shift`) was used to couple the QM and MM regions. The NEB calculations were carried out with the Improved Tangent implementation of DL-FIND, using the L-BFGS algorithm for the corresponding energy evaluations. Seven images were used for each NEB calculation. For all 20 observed reaction paths, the NEB calculations were first carried out using the climbing image method. The `neb=free` setting was used to minimize the start and end point during the NEB calculation. The maximum number of optimization cycles was set to 500. In addition, a NEB calculation was carried out for all 20 reaction paths, in which the CI method was disabled. For this, the threshold values for the start of the CI calculation and for the freezing of the images (as a multiple of the convergence criterion for the maximum energy gradient `max_grad`) were reduced from 3.0 and 1.0 to 0, respectively. The convergence criterion for the No CI calculations was increased from $4.5 \times 10^{-4} E_h \text{\AA}^{-1}$ to $1.0 \times 10^{-3} E_h \text{\AA}^{-1}$. After completing the NEB calculations, the atomistic structures of the images and the corresponding energy values were analyzed. The NEB energy paths with the lowest transition states were used for further analysis.

3.4 Wave function based calculations

To verify the correctness of the energy paths obtained, single-point energy calculations were carried out for all 20 reaction paths. First, QM/MM single point calculations were carried out at the DFT level for all individual images of the respective NEB energy path in order to obtain the point charges (`pointcharges.xyz` file). The DFT-D3 calculations were performed using the B3LYP functional and the def2-SVP basis set.^[48, 49] The subsequent wave function-based calculations were

carried out without explicitly including the MM region. For all images of the 20 reaction pathways, CCSD(T) calculations were carried out using the DLPNO method in ORCA.^[21] The influence of the MM region on the energies was included through a polarization correction with the MM point charges from the previous DFT calculations. The DLPNO method is largely implemented in ORCA as a black box method. A correlation-consistent, polarized cc-pVTZ basis^[51] was used as the basis set. The numerical precision for the "Self Consistent Field" (SCF) cycle was set to 1.0×10^{-8} au (`TightSCF`). For the calculation of the pair densities, from which the PNOs ("Pair natural orbitals") are constructed, complete "second order Møller-Plesset perturbation theory" (MP2)^[52] calculations were carried out using the `TightPNO` setting. The "Resolution of Identity" (RI) method was used to calculate all occurring Coulomb integrals.^[53] 8 processor cores with 15GB RAM/core were used for each DLPNO-CCSD(T) calculation. Depending on the convergence behavior, one DLPNO-CCSD(T) calculation took between 5 hours and 2 days. After completing the calculations, the DLPNO-CCSD(T) energies were added to the MM energies from the DFT single point energy calculations. For each reaction path, all energy values obtained were subtracted from the energy of the first image in order to obtain energy differences that are directly comparable to the NEB reaction path energies. The MP2 results were used in the same way for the calculation of reaction path energies.

Chapter 4

Results and Discussion

4.1 General consideration of the inhibition mechanism

In recent years, it could be shown by X-ray crystallographic studies that Oprozomib forms a 1,4-oxazepane 7-ring with the N-terminal T1 amino acid residue, whereby the active site is irreversibly inhibited.^[10] The exact mechanism of this reaction with Oprozomib as well as with related epoxyketone inhibitors has still not been clarified and is subject of current research.

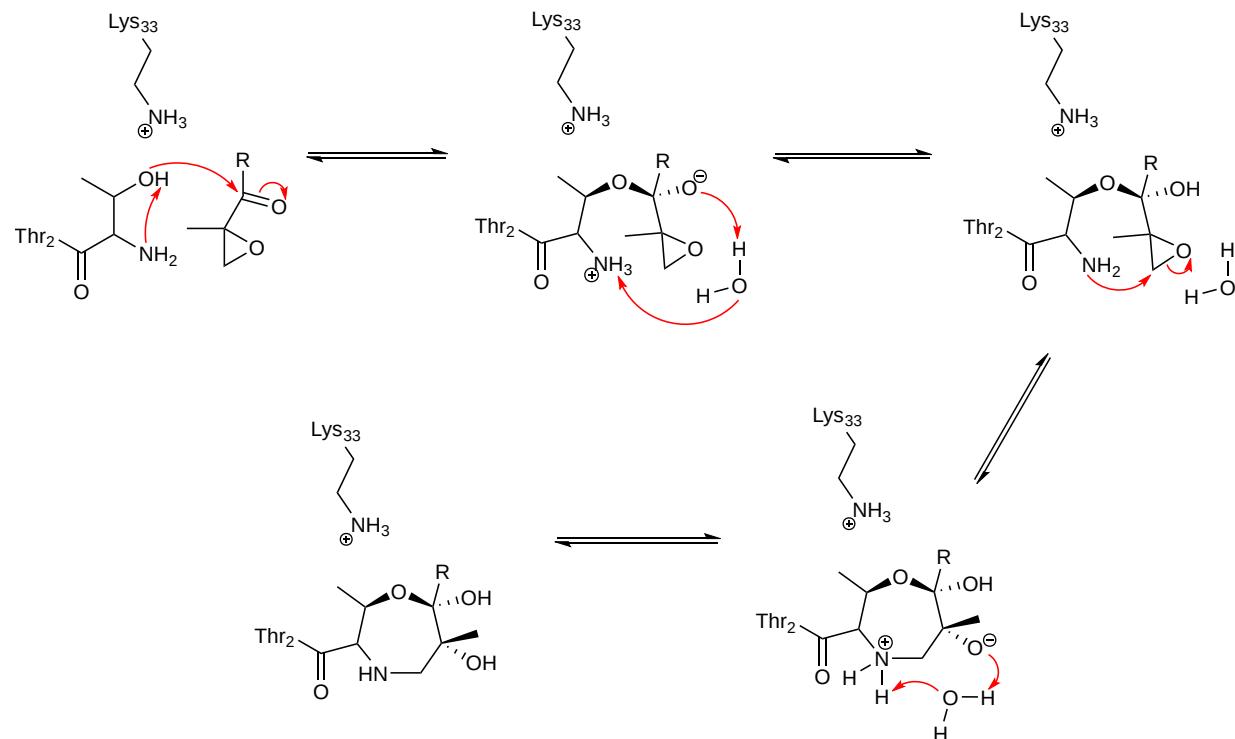


Fig. 4.1: The proposed inhibition mechanism of the $\beta 1$ subunit by Oprozomib. The mechanism is shown for the $\text{T1-NH}_2;\text{K33-NH}_3^+$ protonation state.

In most of the theoretical investigations of the reaction mechanism carried out so far, it is assumed that the first step in this reaction is the nucleophilic attack of the T1 hydroxyl group on the carbonyl group of the epoxyketone inhibitor, which would result in the formation of a tetrahedral

intermediate.^[11] The atomic structure of the proposed tetrahedral intermediate together with the atom names used in the following sections are depicted in Fig. 4.3. This first step is comparable to the general hydrolysis of peptides by proteases, in which a tetrahedral intermediate is also initially formed (see Fig. 1.2).^[7] The second step in the inhibition mechanism then consists in the nucleophilic attack of the N-terminal amino group on the β C atom of the epoxyketone warhead. A complete atomistic picture of this inhibition mechanism is shown in Fig. 4.1. However, it could also be possible that these steps take place in reverse order so that the first step consists in the nucleophilic attack of the N-terminal amino group. Umbrella Sampling-based molecular dynamics simulations at least do not exclude this alternative mechanism for epoxyketone inhibitors.^[12] On the other hand, there exist boron-based inhibitor molecules such as Bortezomib, which can only bind to the T1 alcohol via a single covalent bond to the boron-containing warhead.^[54] This adduct can only arise from an initial nucleophilic attack by the T1 alcohol, which underlines its high reactivity. Consequently, in this thesis the focus will be on the formation of the tetrahedral intermediate as a potential first, rate-limiting step. Based on these investigations, a general picture for the activation mechanism of alcohol proteases may be drawn.

An important factor in studying mechanistic details is the dynamics of proton movement during the reaction. Due to their positive charge and high mobility, protons have a strong effect on the electrostatic environment in enzymes and can thus influence the pK_a values of surrounding amino acid residues. In the active site of the β 1 subunit, the carboxyl group of D17, the amino groups of K33 and T1 as well as the hydroxyl group of T1 are potential proton donors and/or acceptors. Also, water molecules in the vicinity of the active site must be considered, which presumably play a role as proton shuttles in the reaction catalyzed by proteasomal subunits.^[11, 7]

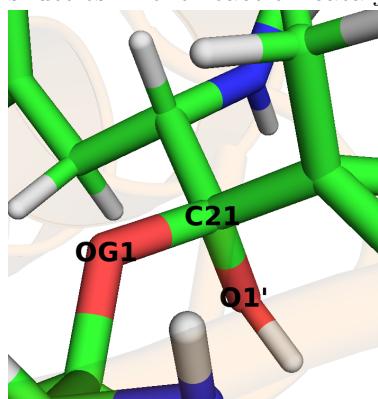


Fig. 4.3: The tetrahedral intermediate formed between Oprozomib and the T1 residue.

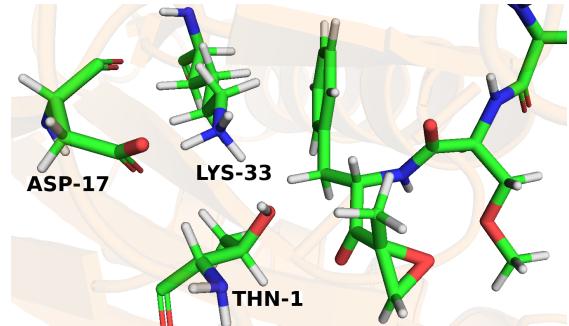


Fig. 4.2: The active site of the β 1 subunit bound to the Oprozomib inhibitor. The most important residues for the catalytic activity of this site are Lys33, Thr1 and Asp17.

The catalytic triad is an important functional unit in all proteases. Proteases with serine-histidine-aspartate triads are very common in nature and are among the best-studied enzymes. They provide a good reference for investigating related protease classes. Nevertheless, even in this case some mechanistic details are still open.^[55] The Ser-His-Asp triad consists of a centrally located histidine-imidazole ring, which is connected to an aspartate via a hydrogen bridge. This leads to an increase in the pK_a of the imidazole nitrogen, making histidine a strong base. This in turn increases the nucleophilicity of the serine residue significantly and thus lowers the barrier for the formation of a tetrahedral intermediate with the substrate.^[56] The lowered pK_a value of the serine

alcohol raises the question of whether serine in the resting state can be present in a stable manner as a deprotonated alkoxide ion (charge relay model) or whether this deprotonation can only occur

through the binding of a substrate to the active site and a subsequent nucleophilic attack (modified charge-relay model).^[55] These adaptations of these two possible mechanisms for the proteasome active site is shown in Fig. 4.4.

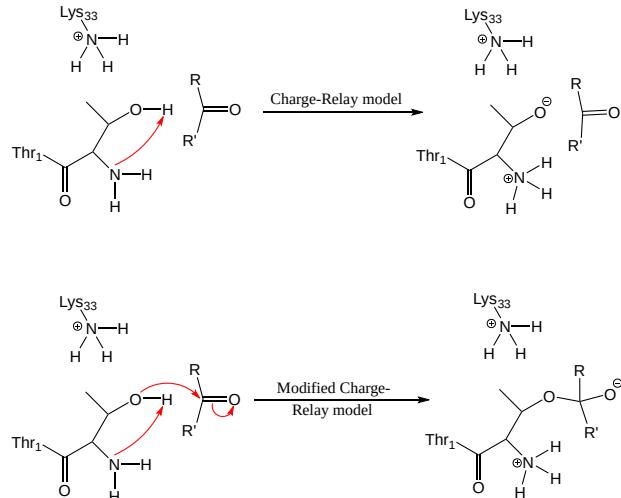


Fig. 4.4: Comparison between the charge-relay model and the modified charge-relay model. In the charge-relay model, it is assumed that a stable alkoxide ion can be formed while the modified charge-relay model assumes a concerted mechanism where deprotonation and nucleophilic attack take place at the same time.

The catalytically active β subunits of the 20S proteasome contain a threonine residue which performs the hydrolysis. They have a special structure of the active site that rarely occurs in other proteases. As with serine proteases, an aspartate is present. The role of histidine is taken by K33. And the T1 hydroxyl group takes on the role of the nucleophile instead of a serine. Theoretical calculations carried out in the Mata group using different methods (PROPKA, Replica Exchange MD, QM cluster calculations) indicate that the K33 amino group in the caspase site ($\beta 1$) has a pK_a value in the range of 10-11, which makes a NH_3^+ state likely. It can therefore be assumed that the positive charge of the protonated K33 amino group has a similar effect on the nucleophile as the histidine imidazole ring in serine proteases. A factor that makes it more complicated is the additional T1 amino group. Theoretical calculations in the Mata group suggest that its pK_a value is in the range of 7-9 if the K33 amino group is protonated. The T1 amino group can therefore be present in both protonated and deprotonated form at neutral pH. Consequently, in numerous studies it has been assumed that the T1 amino group is the primary proton acceptor during the first nucleophilic attack,^[11, 7, 57, 58, 59] whereby a water-mediated proton transfer was partially assumed. The question has also been raised whether alternatively the K33 amino group could serve as the primary proton acceptor. Due to the presumably high pK_a value of the K33 amino group, this possibility has so far been viewed rather critically.^[60] Nonetheless, this is still a controversial topic. A recently published study used the "Protein Dipoles Langevin Dipoles" method (PDLP) to calculate the K33 and T1 pK_a values in order to determine the reaction mechanism for hydrolysis of a model peptide.^[61] They report a pK_a value for the K33 amino group of 5.4 ± 2 and a T1 amino group pK_a of 3.0 ± 2 . This would mean that both amino groups are deprotonated at neutral pH, which is in contradiction to previous findings described above.

In order to find the optimal reaction path for the formation of the tetrahedral intermediate, it is therefore necessary to consider different protonation states. NMR studies on serine proteases have also shown that the protons in the catalytic triad have high mobility and are in constant exchange with water molecules of the solvent.^[55] It therefore makes sense to take into account different protonation states of amino acids that presumably have high pK_a values, such as K33. Another open question is under which circumstances a stable tetrahedral intermediate can be formed at all. The protonation state of the carbonyl function of the substrate after nucleophilic attack is a controversial topic. Since hydrogen atoms cannot be resolved in X-ray crystallographic

images, one has to rely on NMR spectroscopy or theoretical calculations. In common textbooks, the example of serine proteases is often used to describe the formation of an anionic tetrahedral intermediate that is stabilized by hydrogen bonds with the protein backbone (oxyanion hole).^[62] In computer simulations of the reaction mechanism of the 20S proteasome carried out to date, it is mostly assumed that the tetrahedral intermediate is present in anionic form and is stabilized by a conserved hydrogen bond with the Gly47 backbone (compare Fig. 3.3).^[12] In the active sites of the 20S proteasome, however, there are water molecules in the vicinity of the substrate carbonyl function, so that the formation of a stable, neutral intermediate cannot be ruled out.

However, to ensure the formation of a neutral tetrahedral intermediate, a proton transfer from a nearby proton donor to the carbonyl oxygen atom of Oprozomib is necessary. The only residue that is in a position to act as a proton donor is the N-terminal amino group. In principle, this proton transfer can take place in a water-mediated or non-water-mediated manner. The calculation of the MEPs that lead to the formation of the tetrahedral intermediate were therefore carried out with and without the inclusion of water and can provide interesting insights into the energetics of the reaction.

4.2 Reaction pathways for the active site with neutral charge

As explained in the previous section, different protonation states of the two amino groups in the active site can make a big difference to the electrostatics contribution. In the following, the results of the energy path calculations for the system with neutral charge are shown. So either the T1 amino group or the K33 amino group can be described as being protonated (NH_3^+). The choice of the protonation state has a great influence on the course of the reaction, which ultimately leads to the formation of the first enzyme-Oprozomib intermediate.

4.2.1 First step: Activation of the nucleophile

The first step in the reaction mechanism is the activation of the T1 hydroxyl group, i.e. the nucleophile, by a primary proton acceptor. In the case of the protonated K33, the direct proton transfer to the N-terminal amino group was investigated, which is shown in Fig. 4.5a. The obtained MEP shows the formation of a stable intermediate. Interestingly, the product (Image 7) is not in the form of a tetrahedral intermediate. It is a T1 alkoxide zwitterion, the alkoxide oxygen of which is still 2 Å apart from the carbon atom of the Oprozomib carbonyl function. The energy of the transition state is around 6.5 kcal mol⁻¹. The zwitterion, however, has an energy that is about 4 kcal mol⁻¹ above that of the educt. So even though the NEB calculation did not converge (see Table 4.1), a very low energy barrier could be found. This indicates that the reaction path is very near to the actual MEP. Also it must be mentioned that an alcohol usually has a higher pK_a value than an amine. So it's reasonable that the product energy is higher than the educt energy.

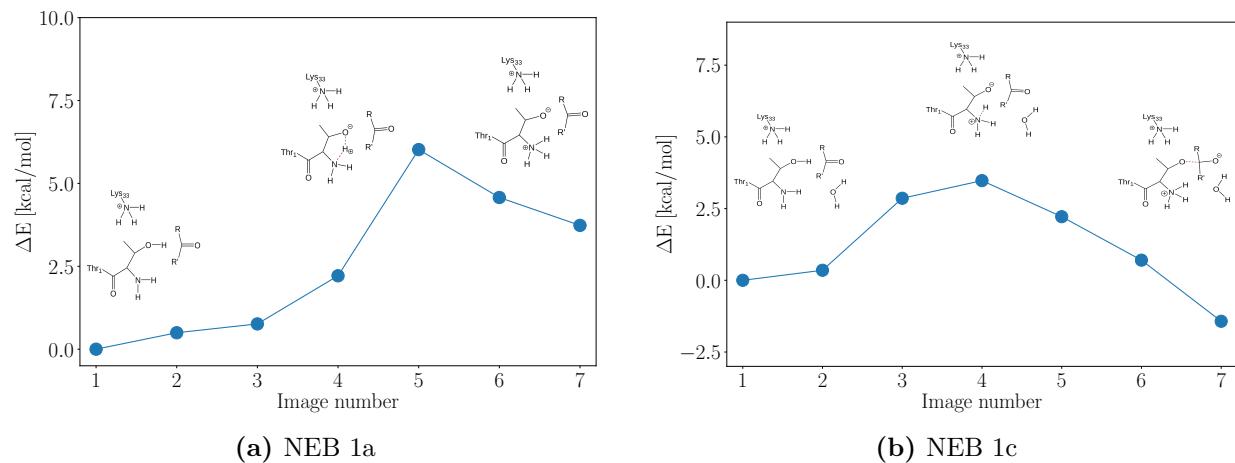


Fig. 4.5: Calculated reaction pathways for the deprotonation of the T1 alcohol by the T1 amino group without (a) and with (c) inclusion of a water molecule into the QM region

In a second step, the energy barrier for the same system was calculated by including a water molecule in the QM region (see Fig. 4.5b). The energy of the transition state is, as one can expect, in the same order of magnitude (approx. $3.5 \text{ kcal mol}^{-1}$), but the product state has a slightly lower energy than the educt state. This is surprising because it would mean that the pK_a value of the T1 alcohol is lower than the pK_a of the T1 amine. On the other hand, wave function based single point energy calculations (discussed in section 4.6) show that the obtained energy for the product state is underestimated in the DFT-based NEB calculation and is in fact about 4 kcal mol^{-1} higher than the educt state energy. Another difference to the reaction path without the inclusion of water in the QM region is the distance between C21 and OG1, which is 1.6 \AA instead of 2.0 \AA in the NEB 1a product state. This indicates that the solvent water molecules have a surprisingly large influence on the formation of a tetrahedral intermediate.

Fig. 4.6a and Fig. 4.6b show the calculated energy paths for the reverse protonation state, where T1 occurs as NH_3^+ and K33 as NH_2 . Again, the calculations were carried out with and without the inclusion of water in the QM region. Only K33 is in a position to act as the primary proton acceptor for the activation of the nucleophile. The minimized educt states already show a hydrogen bond between K33-N and the T1 hydroxyl proton. It has already been discussed above that theoretical calculations performed in the Mata group indicate the K33 pK_a to be 10-11 if the T1 amine is protonated. So a low energy barrier can be expected which is confirmed by the NEB calculations. In both cases (NEB 2a and 2c), the product state is less than 3 kcal mol^{-1} above the educt state. Again it can be observed that the QM water molecule has an influence on the C21-OG1 bond length (see Table 6.1). The reason for this could be a further stabilisation of the attacked carbonyl group upon the inclusion of a water molecule which makes direct contact with it.

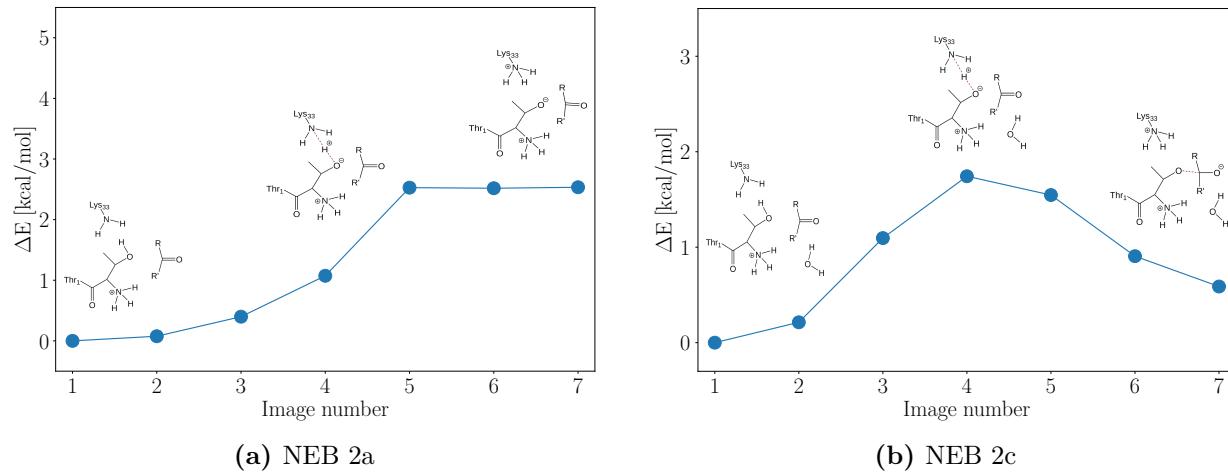


Fig. 4.6: Reaction pathways for the deprotonation of the T1 alcohol by the K33 amino group without (a) and with (c) inclusion of water into the QM region.

Furthermore, the same reactions as in Fig. 4.6a and Fig. 4.6b were repeated with a different starting structure by changing the protonation states of the T1-NH₂;K33-NH₃⁺ structure, as described in Chapter 3. The results of these calculations are shown in Fig. 4.7a and Fig. 4.7b. The proton transfer to the amino group without the inclusion of water in the QM region in Fig. 4.7a has a peculiarity: No energy barrier is recognizable, although the educt state (Image 1) was obtained by a converged energy minimization. This can possibly be explained by the fact that it is a very small local minimum on the PES which could not be resolved by the NEB method.

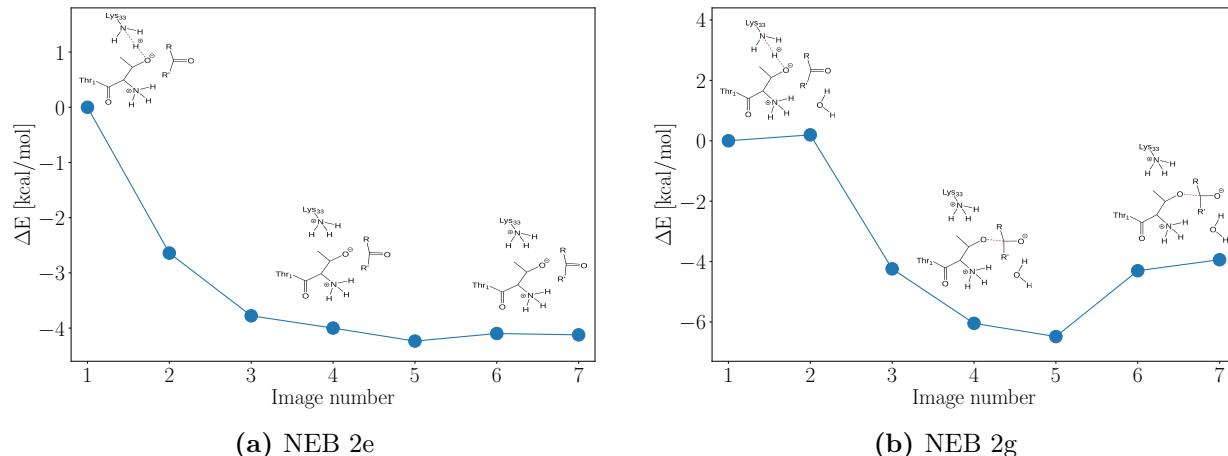


Fig. 4.7: Reaction pathways for the deprotonation of the T1 alcohol by the K33 amino group without (e) and with (g) inclusion of QM water. The starting structures for these NEBs were obtained by modifying the starting structure of the protonation state T1-NH₂:K33-NH⁺.

It is remarkable, however, that the product state has a slightly lower energy than the educt state, although the product state contains a deprotonated alcohol (alkoxide ion) and a protonated K33 amino group. This means that an atomic configuration in the active site was found in which the T1 alcohol has a lower pK_a value than the K33 amino group. However, one has to be aware that the starting structure was generated by changing an originally different protonation state, without performing a renewed equilibration of the system by means of a MD simulation. Thus,

the chemical environment of the T1 and K33 residues could still be adapted to the original atomic configuration with K33-NH_3^+ . The calculation, which was carried out with one QM water molecule (see Fig. 4.7b) provided a similar energy path. Here, too, the product state has an energy approx. 4 kcal mol^{-1} lower than the educt state. There is also another minimum around image 5. However, a particular molecular structure cannot be assigned to this minimum. In fact, the single point energy calculations using the DLPNO method (section 4.6) show that it is more of an energy plateau, similar to NEB 2e.

The NEB reaction pathways shown up to this point demonstrate that the surrounding amino groups in the active site can activate the nucleophile, i.e. the T1 alcohol, by abstraction of the proton. Furthermore, they indicate that the resulting structures represent more or less stable intermediates. This conclusion results from the fact that (local) minima can be found on the PES, at least if a starting structure is used that has previously been equilibrated by short molecular dynamics simulations. It is particularly worth mentioning that these intermediates usually have a C21-OG1 distance between 1.6 and 2.0 Å. No stable C-O bond is formed, for which a bond length of about 1.4 to 1.5 Å would be necessary. Instead, the T1 residue appears to form an intermediate that still has alkoxide ion character. This assumption is supported by the change of the C21-O1' bond length, which is the carbonyl group being attacked. While the educt always shows a bond length of 1.22 Å, which is typical for a carbonyl group, the product states have a bond length between 1.24 and 1.30 Å (see Table 6.1 in the Appendix). So the stretching of the C=O bond is small which leads to the conclusion that this bond still has double bond character after the activation of the alcohol.

4.2.2 Second step: Protonation of the carbonyl group

Since no stable tetrahedral intermediate could be generated in the first step, the second step studied is focused on whether the saturation of the negative charge by a proton transfer to the carbonyl group leads to the formation of a new, possibly more stable intermediate. The height of the energy barriers for proton transfer is of particular interest here. The products obtained from the first reaction step were used as starting structures. The first reaction, described in subsection 4.2.1, leaves the active site in a protonation state $\text{T1-NH}_3^+;\text{K33-NH}_3^+$, regardless of whether the positive charge in the initial structure was located at T1 or K33. It can therefore be expected that the mechanism is similar in all three available structures.

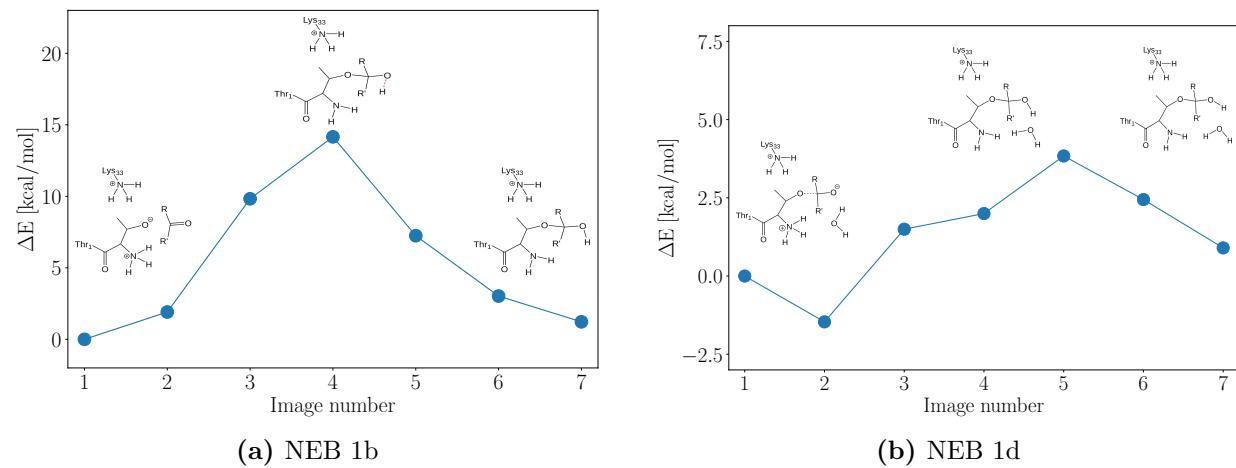


Fig. 4.8: Reaction pathways for the formation of a stable tetrahedral intermediate through protonation of the Oprozomib carbonyl group. On the right, the included QM water molecule acts as a proton shuttle. The starting structures for the calculations were obtained from the product structures of NEB 1a and 1c, respectively.

In Fig. 4.8a and Fig. 4.8b the reaction paths obtained for the proton transfer from the N-terminal amino group to the Oprozomib carbonyl group are shown. A clear difference can be found here if the reaction takes place in a water-mediated or non-water-mediated manner. The proton transfer via a water molecule provides a very small energy barrier of only approx. 4 kcal mol^{-1} , while the reaction without the inclusion of water has an energy barrier of 15 kcal mol^{-1} . The reason for this is that the water molecule is positioned directly between the T1 amino group and the carbonyl group (hydrogen bond to the carbonyl group) and can thus optimally act as a proton shuttle. A proton is transferred from the T1 amino group to the water molecule and another proton, which forms the hydrogen bond to the carbonyl group, is transferred to the carbonyl group almost instantaneously. So it is a concerted process similar to the Grotthuß mechanism.^[63] In NEB 1d this concerted mechanism can be recognized by the fact that the entire proton transfer only provides one single transition state. The analysis of the underlying structures also shows that the two proton transfers take place at the same time. A proton transfer without the participation of water means, however, that the proton has to overcome a distance of several Å, which leads to a higher energy barrier.

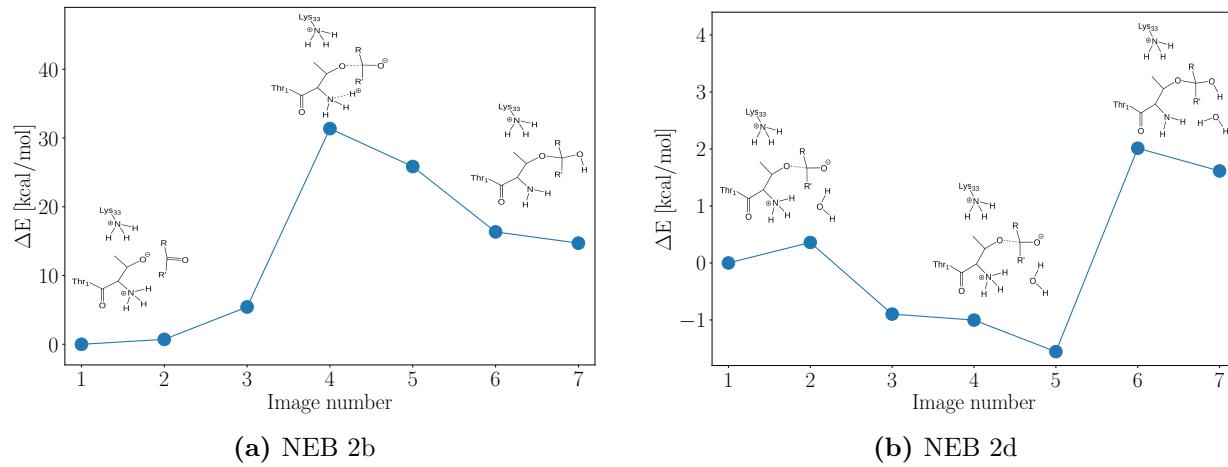


Fig. 4.9: Reaction pathways for the formation of a stable tetrahedral intermediate through protonation of the Oprozomib carbonyl group. On the right, the included QM water molecule acts as a proton shuttle. The starting structures for the calculations were obtained from the product structures of NEB 2a and 2c, respectively.

Fig. 4.9a and Fig. 4.9b show a qualitatively similar behavior. Here the energy barrier without the participation of water is even higher than 30 kcal mol^{-1} . The water-mediated proton transfer takes place with very low energy barriers. The NEB in Fig. 4.9b shows that the minimum from which the reaction path was started is only a local minimum. Viewed from the point of lowest energy in Image 5, the energy barrier is almost 4 kcal mol^{-1} and is thus comparably low as in NEB 1d. When comparing the NEBs 2b and 2d, the large relative energy difference between the endpoints is particularly striking. In NEB 2b the energy difference between educt and product is more than 15 kcal mol^{-1} , while this difference in NEB 2d is under 2 kcal mol^{-1} . The structure comparison of the two product states shows that the orientation of the reactive water molecule is different. In the product of NEB 2d the water molecule forms an additional hydrogen bond with the T1 amino group, which could help stabilize this state.

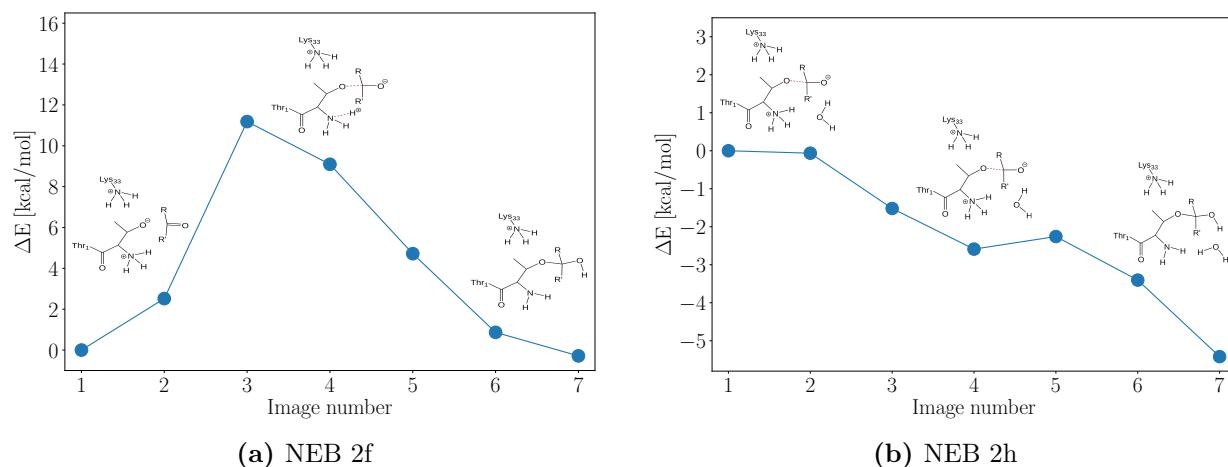


Fig. 4.10: Reaction pathways for the formation of a stable tetrahedral intermediate through protonation of the Oprozomib carbonyl group. On the right, the included QM water molecule acts as a proton shuttle. The starting structures for the calculations were obtained from the product structures of NEB 2e and 2g, respectively.

The NEBs in Fig. 4.10a and Fig. 4.10b also show a qualitatively similar course of the energy profile. Without the water proton shuttle, an energy barrier of about 12 kcal mol⁻¹ was calculated. The water-mediated reaction in this case did not provide any significant energy barrier at all. Instead, there is a more or less continuous decrease in energy up to the product, which has an energy about 5 kcal mol⁻¹ lower than the educt state.

All in all, it can be stated that it has a big influence on the energetics of the reaction whether the proton transfer takes place in a water-mediated manner or not. In addition, protonation of the carbonyl group appears to be necessary to generate a stable tetrahedral intermediate. After the proton transfer, a significant increase in the C21-O1' bond length to 1.37-1.41 Å was found, which clearly shows the loss of the double bond character. The OG1-C21 bond length, however, decreased to 1.41-1.46 Å.

4.3 Reaction pathways for the active site with charge -1

In view of the basic character of the amino groups in the active site, an overall protonation state with T1-NH₂ and K33-NH₂ in the native state is rather unlikely. In order to get a more complete picture of the influence of different protonation states, however, it makes sense to calculate energy barriers for this case as well. Fig. 4.11a shows the calculated energy profile for the proton transfer from the T1 alcohol to the K33 amino group. The NEB calculation did not converge, however, and a relatively high energy barrier of over 30 kcal mol⁻¹ resulted. This is in contrast to the same proton transfer reaction in the previous section in which the T1 amino group was assumed to be NH₃⁺ and not NH₂. A low energy barrier could be obtained there. Even if the meaningfulness of the energy path in Fig. 4.11a is limited, the result is still in line with chemical intuition: A T1 amino group with an additional proton increases the positive charge in the immediate vicinity and thus lowers the pK_a value of the T1 alcohol, which can then be deprotonated more easily. If this positive charge is absent, it becomes very difficult to deprotonate the alcohol.

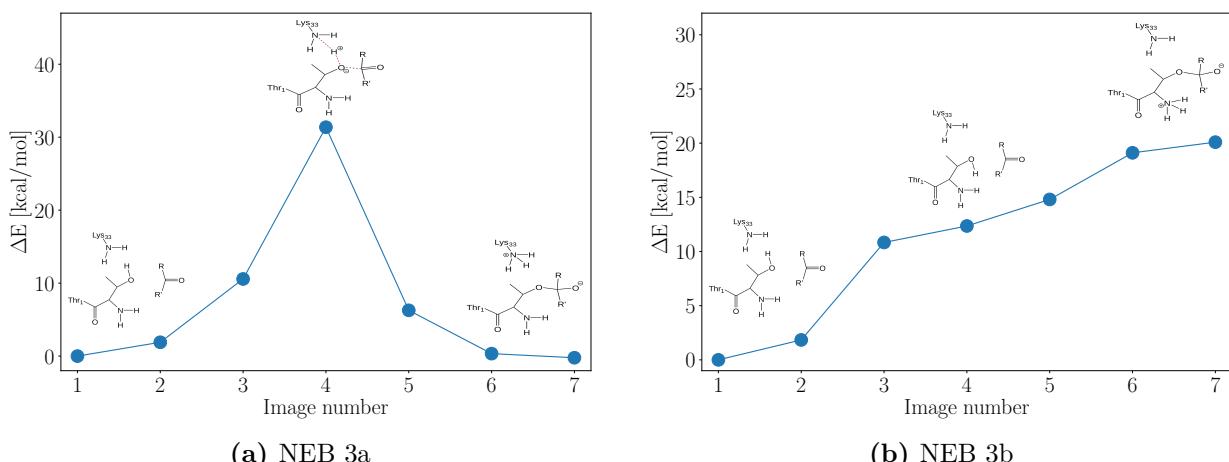


Fig. 4.11: (a): Reaction pathway for the deprotonation of the T1 alcohol by the K33 amino group. (b): Reaction pathway for the deprotonation of the T1 alcohol by the T1 amino group. The starting structures for both NEBs were obtained by modifying the starting structure of the protonation state T1-NH₂;K33-NH₃⁺.

Furthermore, the proton transfer from the T1 alcohol to the T1 amino group was investigated. The converged energy profile in Fig. 4.11b shows that the transfer is energetically very unfavorable. There is no high energy barrier, but the product state is about 20 kcal mol^{-1} above the educt state. Here, too, one can assume that the lack of protons in the active site is responsible for this unfavorable course.

An additional proton transfer to the Oprozomib carbonyl group is only possible if the proton is transferred to the T1 amino group. Therefore, an additional energy profile was calculated for this case, which is shown in Fig. 4.12a. The reason for the occurrence of two maxima is that the proton is not transferred directly to the carbonyl group, but first moves back to the T1 alcohol and then the jump to the C=O oxygen takes place. One can learn from this energy profile that any proton transfer involving the T1 amino group is very unfavorable for this protonation state.

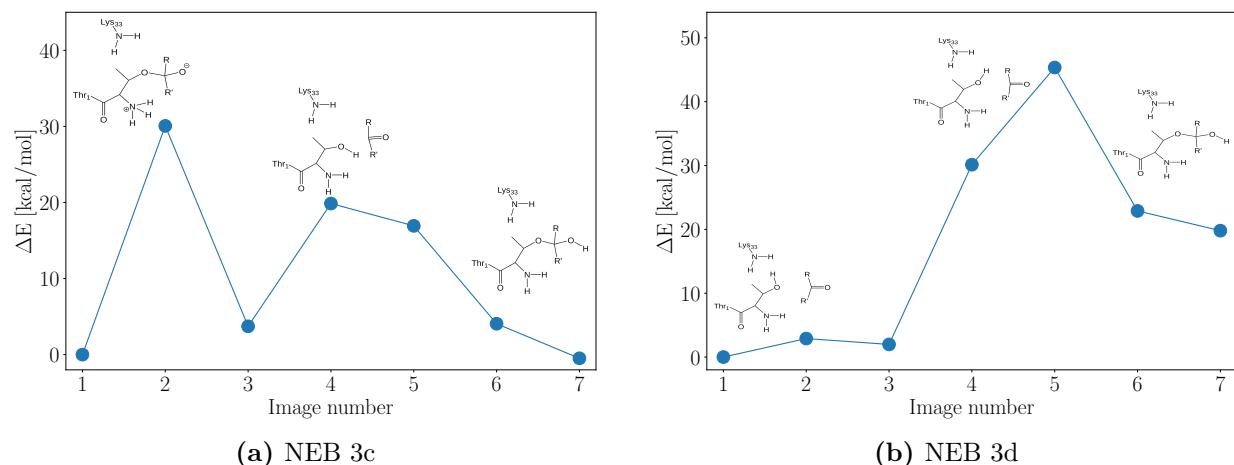


Fig. 4.12: Reaction pathways for the formation of a stable tetrahedral intermediate through protonation of the Oprozomib carbonyl group. The left reaction path is based on NEB 3b to transfer the T1 amino group proton to the carbonyl group. The right reaction path shows the direct proton transfer from the T1 alcohol to the carbonyl group.

Therefore, in a final step, the energy profile for the direct transfer of the T1 alcohol proton to the carbonyl group was calculated. However, no reduction in the activation barrier could be achieved for this reaction. The reaction path, shown in Fig. 4.12b, has an energy barrier of almost 50 kcal mol^{-1} .

In summary, it can be stated that the capability of the active site to act as a catalyst is significantly lower if neither of the two amino groups is in the protonated state. The behavior of the C-O bond lengths is similar to that of the other protonation states discussed so far: The products of NEBs 3a and 3b do not have a protonated carbonyl group and have C21-O1' bond lengths of 1.30 \AA , while the same bonds in NEBs 3c and 3d are protonated and have bond lengths of 1.42 \AA .

4.4 Reaction pathways for the active site with charge +1

Furthermore, attempts were made to determine reaction pathways in the event that both amino groups of T1 and K33 are in the NH_3^+ protonation state. The difficulty here is that the saturation of both amino groups with protons makes the formation of a T1 alkoxide ion difficult. The reason is that there is not a base nearby the alcohol which will abstract its proton. Thus, the only possibility in this case is a concerted proton transfer to the carbonyl that is being attacked. Therefore, a possible pathway was assumed in which one proton is simultaneously transferred from the T1 alcohol to the T1 amino group and another proton from the T1 amino group to the Oprozomib carbonyl group. In this way, the deprotonation of the T1 alcohol would be directly associated with the formation of an alcohol on Oprozomib, which would ensure the formation of a tetrahedral intermediate.

The associated NEB is shown in Fig. 4.14a. It turns out that the proton transfer is not completely concerted. The first observable energy barrier of 30 kcal mol⁻¹ around Image 2 is caused by a proton exchange between the T1 amino group and the T1 alcohol. The second energy barrier around Image 5 reflects the proton transfer from the T1 alcohol to the Oprozomib carbonyl group. So the first maximum does not really contribute to the progress of the reaction and is in fact caused by the choice of the constraints during the structure generation process (compare Fig. 3.8). It appears that the proton transfer from the T1 alcohol to the carbonyl group is preferred over the transfer from the T1 amino group to the carbonyl group. The reason for this is that the T1 alcohol is positioned directly between the T1 amino group and the carbonyl oxygen. So it's a much shorter distance the proton has to overcome. Nevertheless, the energy barrier of 30 kcal mol⁻¹ is still very high.

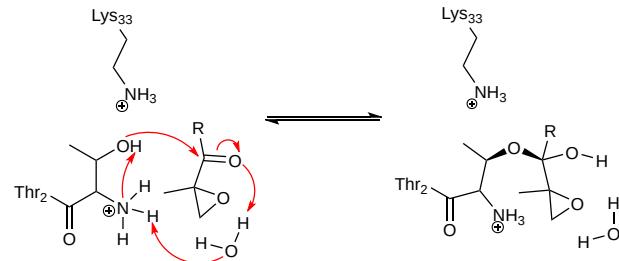


Fig. 4.13: The proposed mechanism for the formation of the tetrahedral intermediate in the case of a $\text{T1-NH}_3^+;\text{K33-NH}_3^+$ protonation state. In order to produce this intermediate, several concerted proton transfers must occur.

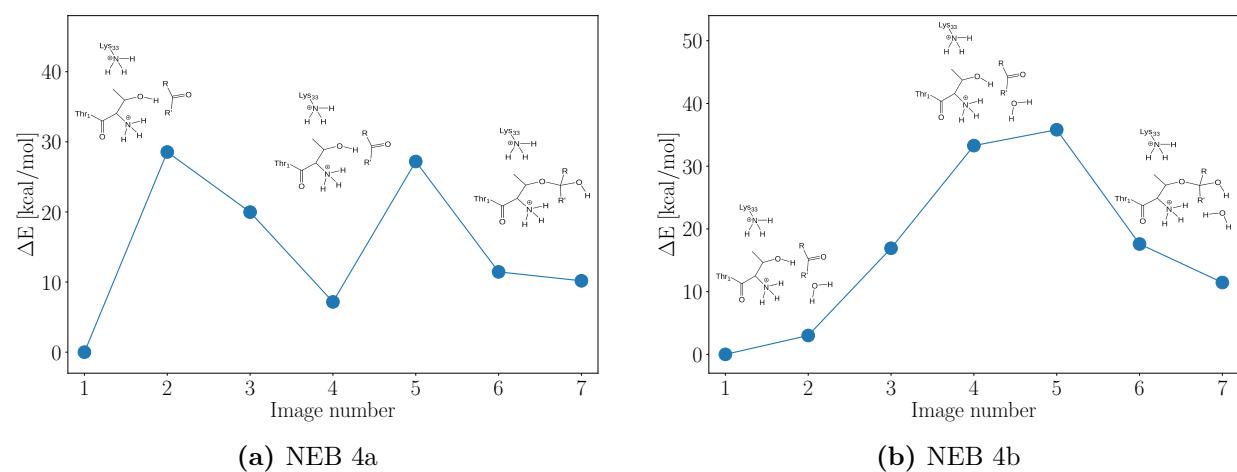


Fig. 4.14: Computed reaction pathways for the concerted proton transfers shown in Fig. 4.13. The right pathway shows the energy barrier of the water-mediated transfer, while the left one shows the barrier without included water.

Therefore, in a second step, an attempt was made to find a MEP for a water-mediated proton transfer between the T1 alcohol and the Oprozomib carbonyl group, omitting the amino group. A look at the initial structure, resulting from a 10 ns equilibration MD simulation, showed that two water molecules in the immediate vicinity of the active site could be used as proton shuttles. The first water molecule was bridged by hydrogen bonds with both the T1 amino group and the T1 alcohol proton. The other water molecule formed a hydrogen bond with the Oprozomib carbonyl group. For the water-mediated proton transfer, the first water molecule was chosen because of its somewhat more central location. The resulting reaction path is shown in Fig. 4.14b. Here, too, the proton transfer proceeds concerted as in the Grotthuß mechanism.^[63] One proton is transferred from the T1 alcohol to the water molecule and at the same time the other proton is transferred from the water molecule to the carbonyl group. The reaction path, which reflects the result of a converged NEB calculation, shows an energy barrier of almost 40 kcal mol⁻¹. Although the proton transfer is water-mediated, the energy barrier is still very high.

It must be taken into account that the active sites of enzymes are highly complex molecular systems in which a large number of atoms have an influence on the overall reactivity. This leads to a high-dimensional PES in which numerous local minima, maxima and saddle points can occur. It can therefore not be said with absolute certainty that the obtained energy paths are near to the actual MEPs. Nevertheless, Fig. 4.14a and Fig. 4.14b suggest that the deprotonation of T1 alcohol is generally very difficult for this protonation state.

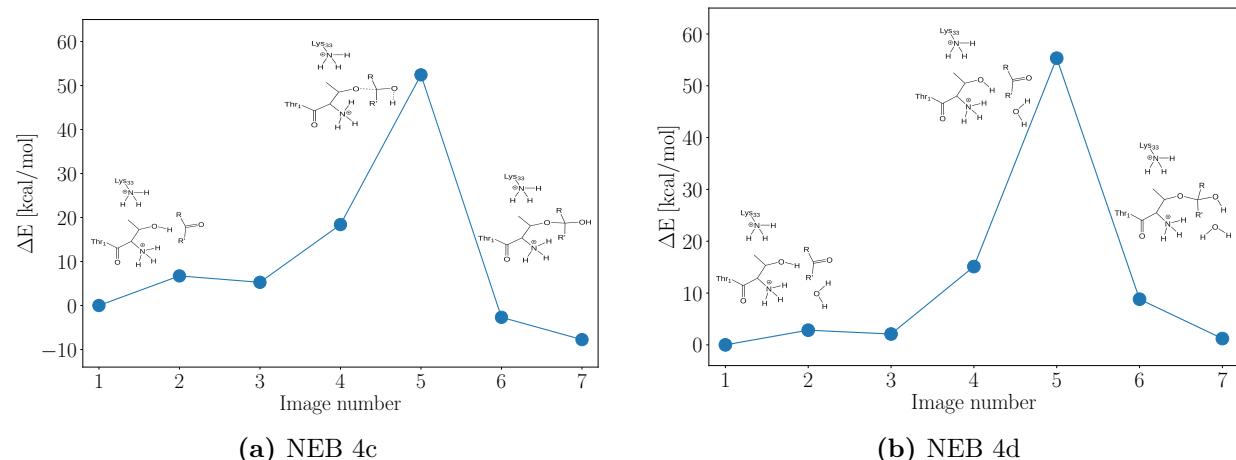


Fig. 4.15: Computed reaction pathways for the concerted proton transfers shown in Fig. 4.13. The right pathway shows the energy barrier of the water-mediated transfer, while the left one shows the barrier without included water. In contrast to NEBs 4a and 4b, the starting structures were obtained by modifying the protonation state T1-NH₂;K33-NH₃⁺.

In order to make a comparison between different MD snapshots, the T1-NH₂;K33-NH₃⁺ protonation state was modified to T1-NH₃⁺;K33-NH₃⁺ and used for the calculation of NEBs. The results are shown in Fig. 4.15a and Fig. 4.15b. The proton transfer without the direct involvement of water was investigated first (Fig. 4.15a). A T1 amino group proton is transferred towards the Oprozomib carbonyl group, although in this case no intermediate step is taken via the T1 alcohol. The distance of about 4 Å between the T1 amino group and the Oprozomib carbonyl group is overcome directly. The energy barrier of NEB 4c is over 50 kcal mol⁻¹.

The inclusion of a water molecule as a proton shuttle (NEB 4d) does not change the high energy barrier at all. In the initial structure, only one water molecule is suitable as a nearby proton shuttle. It is located midway between the Oprozomib carbonyl group and the T1 amino group and should therefore be well suited as a shuttle based on this consideration. Since in NEB 4c the T1 amino group proton is transferred directly (and not via the T1 alcohol) to the carbonyl group, the Grotthuß-like proton transfer from the T1 amino group to the water and from the water to the carbonyl group was investigated here. This is a concerted transfer of three protons at the same time. Although the protons are transferred over very short distances, the NEB 4d still shows an energy barrier of almost 60 kcal mol^{-1} . It must be mentioned that both NEB 4c and 4d were very difficult to converge and could not be brought to convergence even after 500 cycles. One possible reason could be that the starting structures were not equilibrated by previous MD simulations. There could also be other factors causing the high energy barriers: On the one hand, factors that are related to the NEB algorithm itself. It happens that after a certain number of cycles the calculation diverges so that the nudged elastic band penetrates into areas of the PES that are very far away from the MEP.^[64] Another factor could be that the higher number of protons and thus positive charges compared to other protonation states is generally unfavorable for the course of the reaction (e.g. due to repulsive electrostatic effects). This would provide an explanation for the fact that even the inclusion of water as a proton shuttle does not lead to a significant reduction in the energy barrier, although the transfer distance between the individual protons decreases significantly.

All in all, it can be said that, based on the current state of knowledge and the performed NEB calculations, it seems unlikely that a high level of protonation of both amino groups has a beneficial effect on the course of the reaction in nature. Nonetheless, the calculations carried out so far are of course not sufficient to make a final statement. For example, it would make sense to examine the proton shuttle with and without the inclusion of the T1 amino group for both starting structures.

4.5 NEB convergence behaviour

For all potential reaction pathways, NEBs with different settings were calculated. Those that provided the lowest energy barriers were used for further analysis. It is known that the NEB calculations are often difficult to converge, especially with older implementations of the NEB method. The application of the NEB method to the active site of the $\beta 1$ subunit is characterized by the fact that a large number of atoms must be taken into account in the optimization, which increases the complexity of the underlying potential energy surface. A large number of local minima and maxima are therefore to be expected. The Table 4.1 shows the convergence behavior of all NEBs shown in this work.

Slightly different convergence criteria were used for the CI and No CI-NEBs: For the CI-NEBs, the default settings from Chemshell have been used, which corresponds to a maximum gradient of $4.5 \times 10^{-4} E_h \text{\AA}^{-1}$ as convergence criterion (`tolerance=0.00045`). For the No CI-NEBs carried out later, the convergence threshold value for the maximum gradient was set to $1.0 \times 10^{-3} E_h \text{\AA}^{-1}$. The convergence criterion for the root-mean-square gradient is defined as $2/3 \times \text{tolerance}$. The

convergence criterion for the energy (more precisely: the energy difference between the last two NEB cycles) is defined as `tolerance`/450.

Table 4.1: Convergence behaviour of the different reaction paths (RPs) computed by the NEB method. Depicted are the NEB type (usage of the climbing image method (CI) or disabled CI), number of requested (Req.) and accomplished (Acc.) NEB cycles, the energy difference between the last two cycles for a specific image, the highest gradient norm value from the seven images in the last cycle and the root-mean-square value of the gradient norm for all seven images in the last cycle. The convergence based on the respective criterion is colored in red or green.

RP number	NEB type	Req. cycles	Acc. cycles	Energy [E _h]	Max grad [E _h Å ⁻¹]	RMS grad [E _h Å ⁻¹]
1a	CI	100	100	1.11×10^{-4}	1.93×10^{-3}	1.52×10^{-4}
1b	No CI	500	500	6.02×10^{-5}	9.74×10^{-3}	3.59×10^{-4}
1c	No CI	500	100	1.32×10^{-10}	2.86×10^{-4}	2.63×10^{-5}
1d	No CI	500	500	9.36×10^{-3}	5.51×10^{-2}	1.51×10^{-3}
2a	CI	500	66	9.51×10^{-7}	1.64×10^{-4}	2.39×10^{-5}
2b	No CI	500	500	5.80×10^{-3}	1.08×10^{-2}	3.71×10^{-4}
2c	CI	500	143	9.87×10^{-7}	2.61×10^{-4}	1.86×10^{-5}
2d	No CI	500	456	1.81×10^{-6}	7.19×10^{-4}	3.89×10^{-5}
2e	No CI	500	68	0.0	9.84×10^{-3}	2.67×10^{-3}
2f	CI	500	371	5.73×10^{-7}	2.64×10^{-4}	2.73×10^{-5}
2g	No CI	500	246	1.54×10^{-7}	3.06×10^{-4}	1.87×10^{-5}
2h	No CI	500	279	0.0	6.15×10^{-4}	3.83×10^{-5}
3a	No CI	500	500	3.32×10^{-2}	1.69×10^{-1}	3.20×10^{-3}
3b	No CI	500	500	0.0	2.94×10^{-3}	1.32×10^{-4}
3c	No CI	500	500	1.73×10^{-4}	1.29×10^{-1}	2.56×10^{-3}
3d	No CI	300	300	2.35×10^{-2}	3.29×10^{-2}	1.59×10^{-3}
4a	No CI	500	500	1.29×10^{-3}	1.70×10^{-2}	7.10×10^{-4}
4b	No CI	500	413	3.24×10^{-7}	2.96×10^{-4}	1.34×10^{-5}
4c	No CI	100	100	7.76×10^{-2}	1.02×10^{-1}	3.89×10^{-3}
4d	No CI	300	200	3.02×10^{-2}	5.73×10^{-2}	1.64×10^{-3}

The analysis of the convergence behavior of the NEBs, shown in Table 4.1, shows that especially the NEBs of the protonation state T1-NH₃⁺;K33-NH₂ could easily be brought to convergence. For all other protonation states this was rather difficult to achieve. Also, it was mostly easier to get low energy barriers when the climbing image method was disabled.

4.6 Single point energy calculations

When examining different energy paths, it is not only important whether the NEB calculations actually converge and whether the energy path found is the MEP. Also important is the accuracy of the underlying electronic structure methods with which the potential energies of the system under investigation are calculated. In order to save time and costs, mainly semi-empirical methods and density functional theory are used. Density functional theory usually delivers good results for transition state energies, which deviate from the exact energies by only a few kcal mol^{-1} , but chemical accuracy cannot be achieved in general, which would require deviations not larger than 1 kcal mol^{-1} .^[65] Although the energy differences between different atomic configurations show significantly fewer deviations from the exact energies than their corresponding total energies, they are not always qualitatively correct. The height of energy barriers for example is often underestimated by density functional calculations.^[66]

Much more precise energies are provided by wave function-based methods such as the DLPNO-CCSD(T) method. Therefore, additional single point energy calculations were carried out for all reaction paths using the DLPNO-CCSD(T) method and then compared with the DFT-based energies. Because the differences in the absolute energies from DFT and CC calculations were very large, only the energy differences between the individual images were plotted for each reaction path, with the reactant state being set to zero for both DFT and CC. In addition, Møller-Plesset 2 energies were added to Fig. 4.16, Fig. 4.17 and Fig. 4.18. The MP2 energies are more or less a by-product of the DLPNO calculations. For the NEBs 3a-d and 4a-d, these values were not included, since relatively high energy barriers occur here and the energy differences between different electronic structure methods are therefore relatively small.

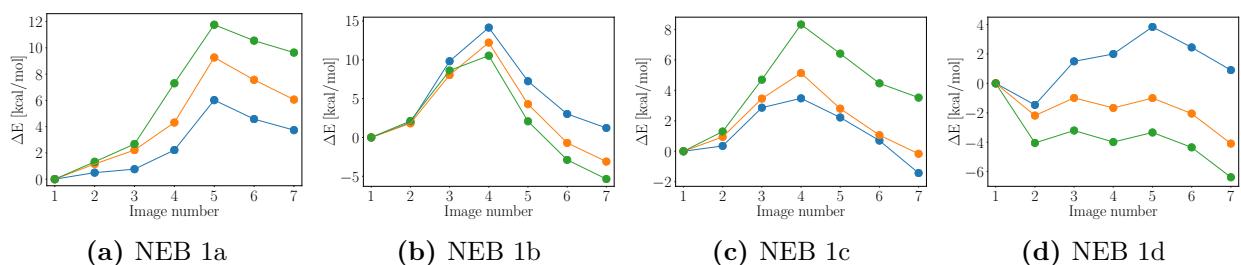


Fig. 4.16: Energy values of the reaction pathways from initial protonation state T1-NH₂;K33-NH₃⁺ calculated at different levels of theory: Original DFT from the NEB calculation (blue), MP2 (orange) and DLPNO-CCSD(T) (green).

The reaction paths 1a-d show that the energy barriers calculated with the DLPNO method are not necessarily higher than the DFT-based energy barriers. Instead, the reaction paths 1a-d show that the energy differences between the educt and product states deviate by a few kcal mol^{-1} . These energy differences are up to 8 kcal mol^{-1} . It is also noticeable that the largest energy differences are between DFT and CC, while the MP2 energies are more in between. Since the DLPNO-CCSD(T) energies usually achieve chemical accuracy, one can conclude that the MP2 method gives more precise results than DFT, but is still not in the range of chemical accuracy.

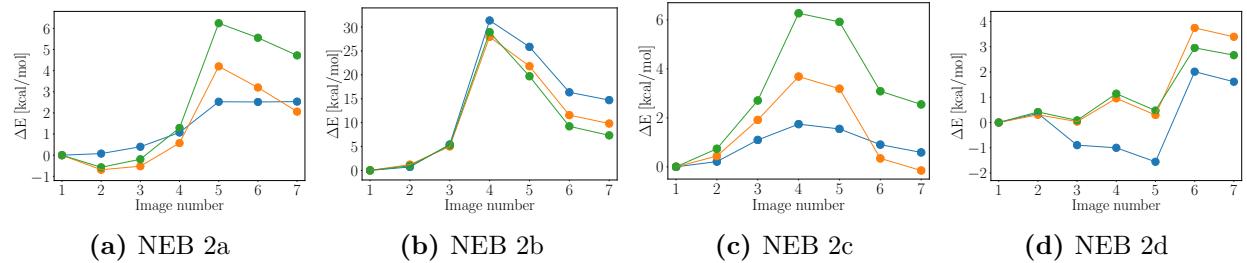


Fig. 4.17: Energy values of the reaction pathways from initial protonation state T1-NH₃⁺;K33-NH₂ calculated at different levels of theory: Original DFT from the NEB calculation (blue), MP2 (orange) and DLPNO-CCSD(T) (green).

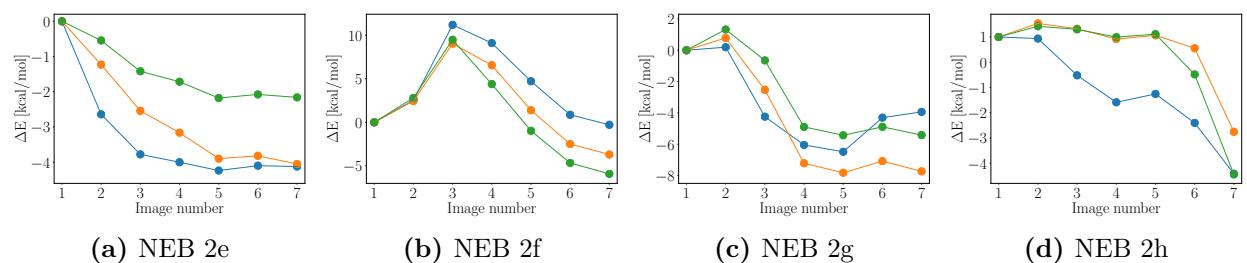


Fig. 4.18: Energy values of the reaction pathways from initial protonation state T1-NH₃⁺;K33-NH₂ calculated at different levels of theory: Original DFT from the NEB calculation (blue), MP2 (orange) and DLPNO-CCSD(T) (green).

The reaction paths 2a-h show a similar behavior. Reaction path 2a, for example, is interesting here: While the DFT-based NEB calculation did not provide an appreciable energy barrier, this is certainly the case for the DLPNO-CCSD(T) and MP2 energies. Both methods provide a small energy barrier between 1 and 2 kcal mol⁻¹. Reaction path 2c is also interesting: Here the height of the energy barrier is clearly underestimated by DFT: While DFT predicts an energy barrier of almost 2 kcal mol⁻¹, the DLPNO method delivers an energy barrier of over 6 kcal mol⁻¹. In many other cases, however, the energy barriers calculated with the DLPNO method are not higher than the DFT-based barriers.

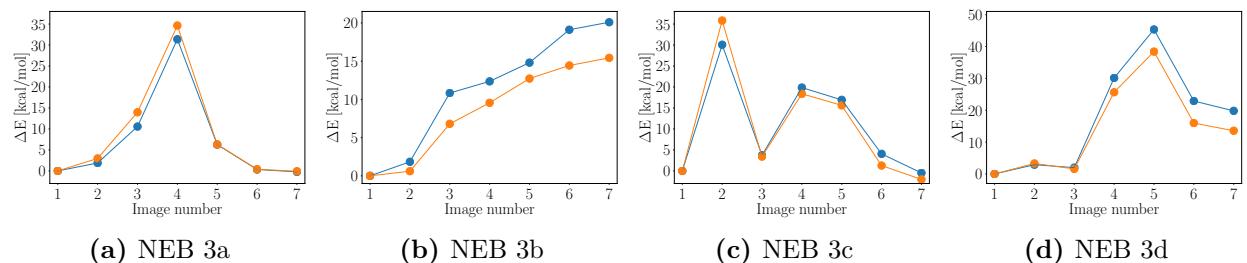


Fig. 4.19: Energy values of the reaction pathways from initial protonation state T1-NH₂;K33-NH₂ calculated at different levels of theory: Original DFT from the NEB calculation (blue) and DLPNO-CCSD(T) (green).

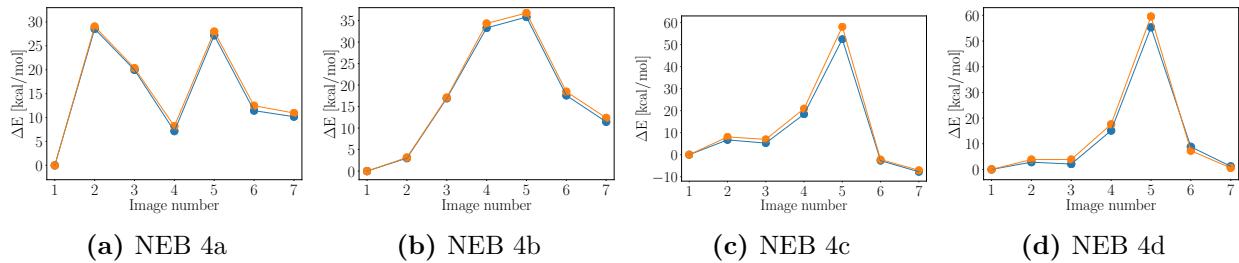


Fig. 4.20: Energy values of the reaction pathways from initial protonation state T1-NH₃⁺;K33-NH₃⁺ calculated at different levels of theory: Original DFT from the NEB calculation (blue) and DLPNO-CCSD(T) (green).

The reaction paths of NEBs 3a-d and 4a-d are characterized by very high energy barriers. Therefore, the differences between DFT and DLPNO-CCSD(T) energies are relatively small. However, for the sake of completeness, they are shown in Fig. 4.19 and Fig. 4.20.

All in all, it can be stated that the DLPNO-CCSD(T) and MP2 calculations can confirm the correctness of the reaction paths. The reaction paths that were found with the NEB method are recorded very well quantitatively with the DFT method. Ultimately, the much greater difficulty lies in finding the correct MEP and not in calculating the energies of that reaction path with chemical accuracy. The variability of the energy paths depends to a much greater extent on which NEB settings are used than on the electronic structure method used. Therefore, the application of more sophisticated electronic structure methods especially makes sense if there is a high probability that a MEP has already been found with the DFT-based NEB method.

4.7 Conclusions for the reaction mechanism

The first-principles NEB calculations carried out in this thesis have shed light on the detailed mechanism of the first chemical steps in the inhibition of the 20S proteasome by Oprozomib. Based on the NEB calculations, it became very clear that the protonation states of the two amino groups in the active site have a major influence on the reactivity. In particular, it can be stated that sufficient reactivity is only guaranteed if either the T1 or the K33 amino group is in the NH₂ state, while the other one is in the NH₃⁺ state. This means that the positive charge of the protonated amino group plays the role of lowering the pK_a value of the T1 alcohol so much that its proton can be more easily transferred to the other amino group, which acts as a base. This mechanism does not work if both amino groups are deprotonated, since the necessary lowering of the pK_a of the T1 alcohol can no longer be achieved. This leads to high energy barriers for the deprotonation of the alcohol. Furthermore, the reaction cannot take place if both amino groups are in the NH₃⁺ state, since no functional group is available in the vicinity of the T1 alcohol that could act as a proton acceptor. These two unfavorable protonation states led to extremely high energy barriers between 20 and 60 kcal mol⁻¹ in all calculations, while the protonation states that were favorable for the reaction always provided extremely low energy barriers.

The question of whether nature uses a protonated T1 amino group or a protonated K33 amino group for the first reaction step is somewhat more difficult to answer. The reaction pathways in Fig. 4.7a and Fig. 4.7b have shown that the pK_a value of the K33 amino group could even be

higher than the pK_a value of the T1 alcohol under certain conditions. This indicates that under standard conditions the K33 amino group is more likely to be in the NH_3^+ state, while the T1 amino group is more involved in the exchange of protons with surrounding water molecules and thus has a variable protonation state.

Furthermore, knowledge could be gained about the circumstances under which a stable tetrahedral intermediate is formed. It could be shown that the deprotonation of the T1 alcohol leads to a nucleophilic attack on the Oprozomib carbonyl function, but that a stable bond is not formed immediately. The C21-OG1 bond lengths after deprotonation of the alcohol are between 1.6 and 2.0 Å, which suggests that the OG1 atom still has an anionic character. However, it could also be observed that the surrounding solvent molecules have an influence on the C21-OG1 bond length, so that a correct description of the surrounding water molecules is necessary. The calculated mechanisms therefore partly coincide with the charge relay model, according to which a semi-stable anionic alkoxide ion is formed after the deprotonation of the alcohol. The calculations also show that a further shortening of the C21-OG1 bond length occurs after the transfer of a proton to the carbonyl function. This proton transfer seems to be important for the stabilization of the tetrahedral intermediate.

It has become clear that the use of different starting structures from previous MD simulations leads to different reaction paths in the NEB calculations. This is not surprising, since in complex systems such as in enzymes, in addition to the reactive degrees of freedom (DOFs) in the active site, other flexible DOFs, the so called "spectator DOFs", must also be taken into account. This leads to a rugged potential energy landscape, in which a whole ensemble of energy paths similar to the MEP may exist that can drive the system from an educt state to a product state.^[64] In order to get a more precise picture of the possible reaction paths, it is therefore useful to carry out NEB calculations for many different starting structures.

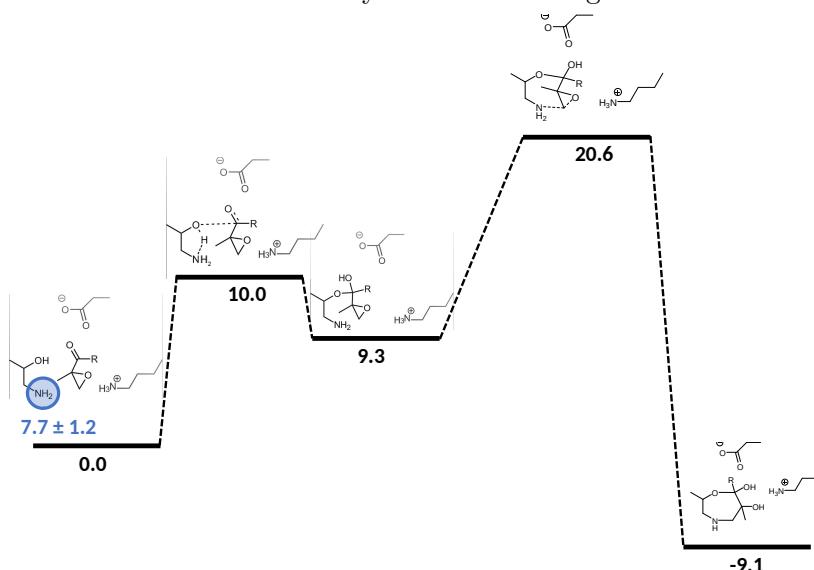


Fig. 4.21: The whole inhibition mechanism of the $\beta 1$ active site by Oprozomib for the $\text{T1-NH}_2;\text{K33-NH}_3^+$ protonation state. The energy values are based on pure QM cluster calculations.

carried out with neglect of the surrounding solvent and the protein environment, deviations from the

The Mata group also carried out further studies on the inhibition mechanism of Oprozomib. Fig. 4.21 shows the results of cluster calculations on the active site of the $\beta 1$ subunit, which were carried out by Jon Uranga. Shown here are the results for the protonation state $\text{T1-NH}_2;\text{K33-NH}_3^+$, which was also identified as the most promising initial state by the calculations carried out in this thesis. Since the cluster calculations are car-

results of this thesis are to be expected. The cluster calculations result in an energy barrier of 10 kcal mol^{-1} and an intermediate that has an energy $9.3 \text{ kcal mol}^{-1}$ higher than the educt. In the present work, significantly lower energy barriers could be found through the combined QM/MM approach. This indicates that the formation of the tetrahedral intermediate is not rate limiting for the overall reaction. For a final statement on this, of course, entropic contributions to the reaction must also be examined. Instead, the second major step after the formation of the tetrahedral intermediate, namely the attack of the T1 amino group on the epoxyketone moiety, seems to be rate-limiting, as shown in Fig. 4.21 with an energy barrier of $20.6 \text{ kcal mol}^{-1}$. However, this was not subject of the present work and needs to be investigated further.

Chapter 5

Summary and Outlook

Enzymes are among the most complicated chemical systems and a deeper understanding of how the precise catalytic abilities come about at the atomic level is absolutely necessary, especially in medicine, in order to be able to develop targeted inhibitors in the future that intervene precisely and highly specific in certain cellular processes. Clinical research in particular is still dependent on time-consuming and expensive studies in the development of new drugs that are based on the "trial and error" principle. A more systematic approach ultimately requires a much more precise, holistic understanding of all cellular processes in the cell. There is still a long way to go.

In the present work the highly dynamic role of protons in the catalysis of the human 20S proteasome was investigated. The energetics of individual proton movements and their effects on the reaction mechanism can be investigated relatively precisely and in a controlled manner with the NEB method within the accuracy of the underlying electronic structure methods. Taking into account all possible protonation states and proton movements, it is possible over time to gain a precise picture of the energy landscape in the active site.

The subject of this work was the first catalytic step in the irreversible binding of the Oprozomib inhibitor to the active site of the $\beta 1$ subunit of the 20S proteasome. Even if the resulting formation of the tetrahedral intermediate is only the first step in this reaction, the investigation of its proton dynamics is very important for the overall reaction. The results of the NEB calculations indicate that one of the two amino groups in the active site must be protonated for the reaction to proceed optimally. Both too many and too few protons in the active site impede the catalytic activity. So the results give a clear tendency as to which protonation states are preferred by Nature. Since Oprozomib is a peptide analogue, statements about the mechanism of proteolytic cleavage of natural substrates are also possible based on the available results. However, it has also been shown that the calculation of MEPs with the NEB method is associated with frequent convergence problems in such large molecular systems. Also, the exact height and shape of the energy barriers will depend on which starting structure is used.

In further mechanistic investigations of the 20S proteasome, as well as of other proteases, other tools are available in addition to methods that aim to analyze the PES and MEPs. Calculations of the potential energy of a system cannot, for example, make a statement about the role entropic effects play in catalysis. Sampling methods based on MD simulations (for example Umbrella Sampling) are necessary here.

Chapter 6

Appendix

Table 6.1: The change of the OG1-C21 bond length and the C21-O1' bond length during the calculated reactions. The bond lengths are shown for Image 1 and Image 7, respectively.

NEB	Image 1		Image 7	
	$d(\text{OG1} - \text{C21})$ [Å]	$d(\text{C21} - \text{O1}')$ [Å]	$d(\text{OG1} - \text{C21})$ [Å]	$d(\text{C21} - \text{O1}')$ [Å]
1a	2.90	1.22	1.98	1.25
1b	1.98	1.25	1.43	1.39
1c	2.76	1.22	1.60	1.30
1d	1.60	1.30	1.43	1.38
2a	2.58	1.22	2.00	1.24
2b	2.00	1.24	1.44	1.40
2c	2.56	1.22	1.61	1.29
2d	1.61	1.29	1.45	1.37
2e	2.46	1.22	1.98	1.25
2f	1.98	1.25	1.41	1.41
2g	2.56	1.22	1.62	1.29
2h	1.62	1.29	1.46	1.38
3a	2.43	1.22	1.54	1.30
3b	1.54	1.30	1.38	1.42
3c	2.43	1.22	1.38	1.42
3d	2.43	1.22	1.58	1.30
4a	3.26	1.21	1.47	1.39
4b	3.21	1.21	1.47	1.38
4c	3.01	1.22	1.43	1.40
4d	2.97	1.22	1.46	1.38

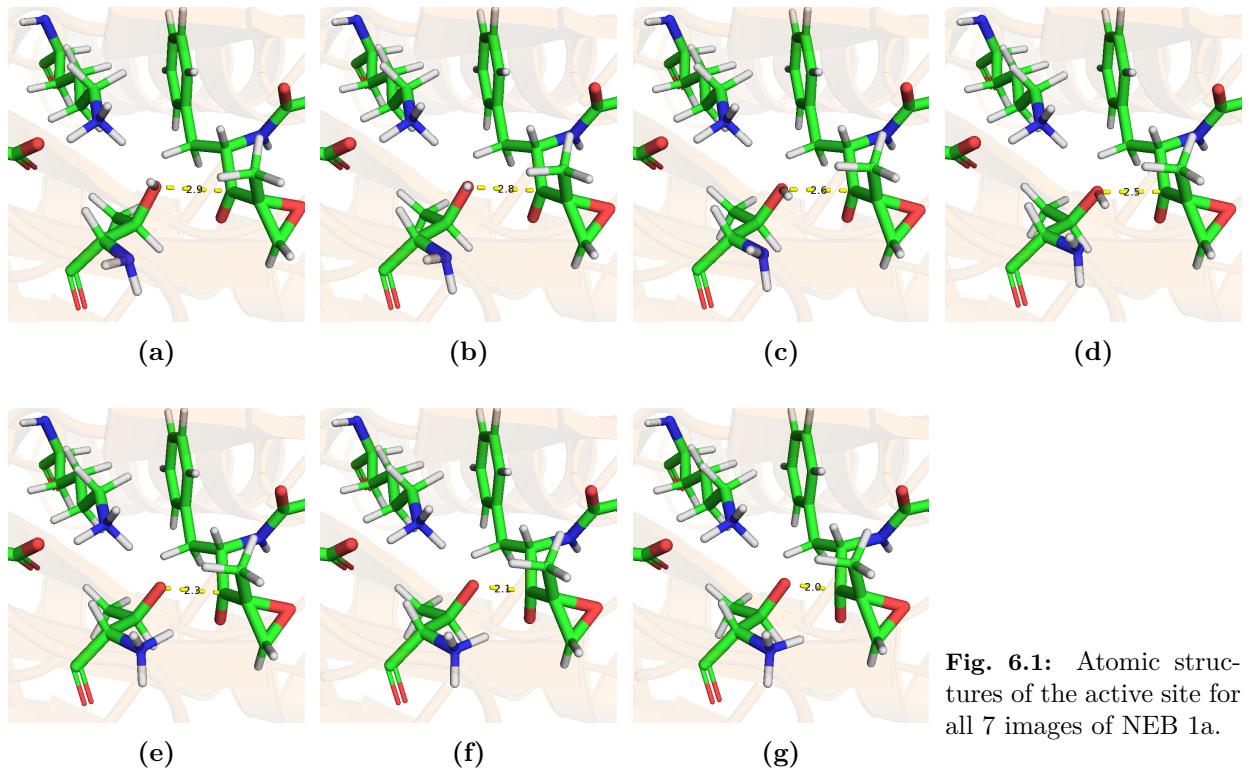


Fig. 6.1: Atomic structures of the active site for all 7 images of NEB 1a.

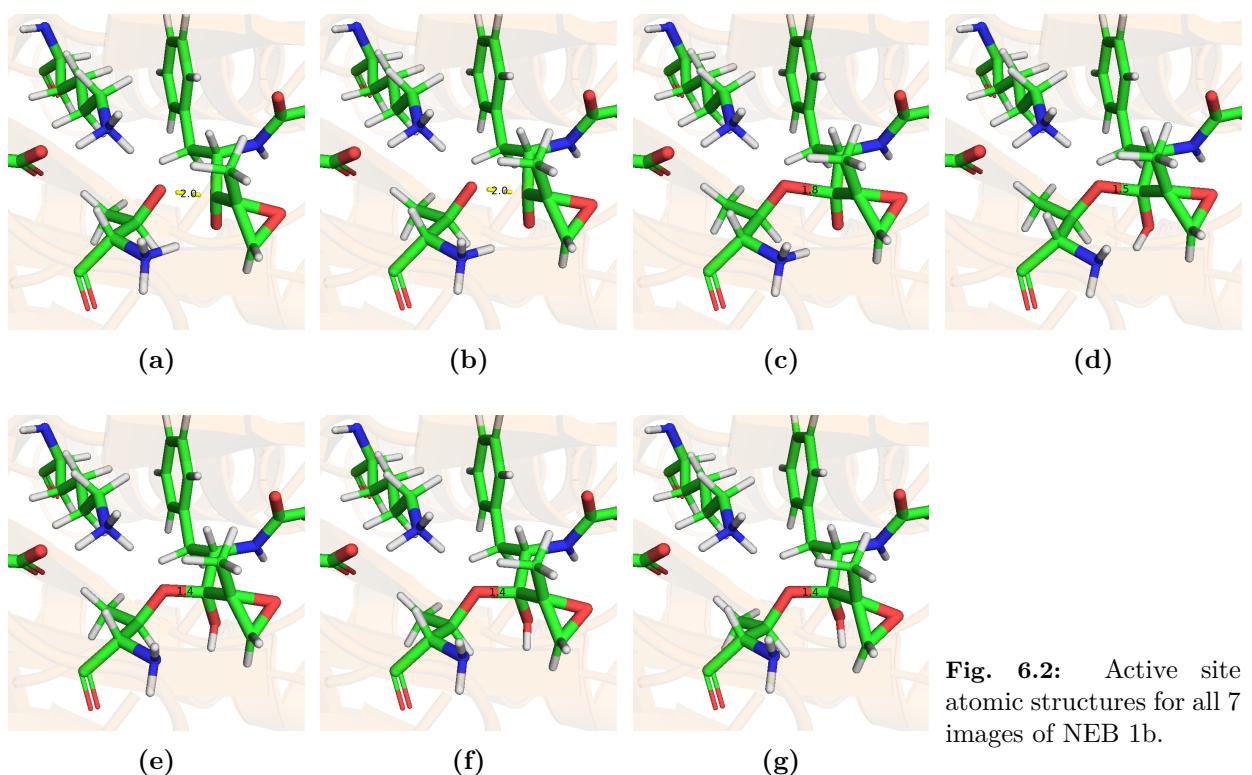


Fig. 6.2: Active site atomic structures for all 7 images of NEB 1b.

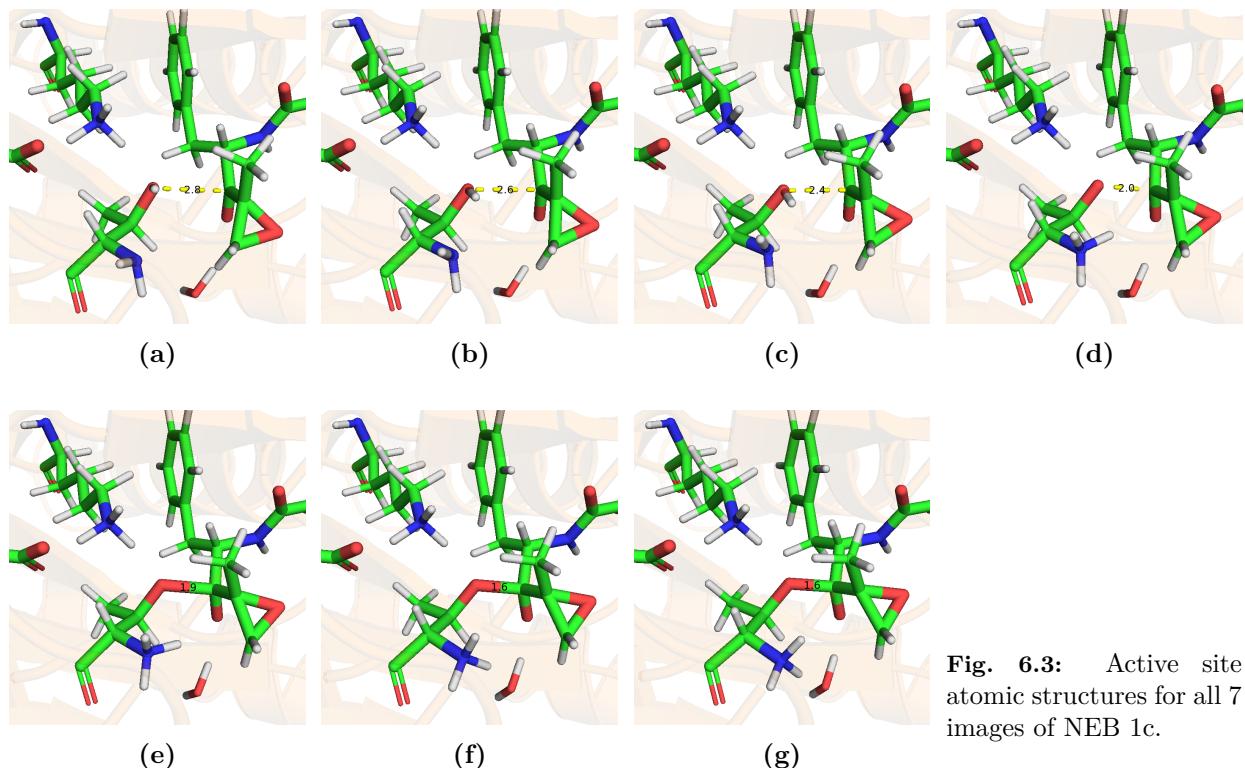


Fig. 6.3: Active site atomic structures for all 7 images of NEB 1c.

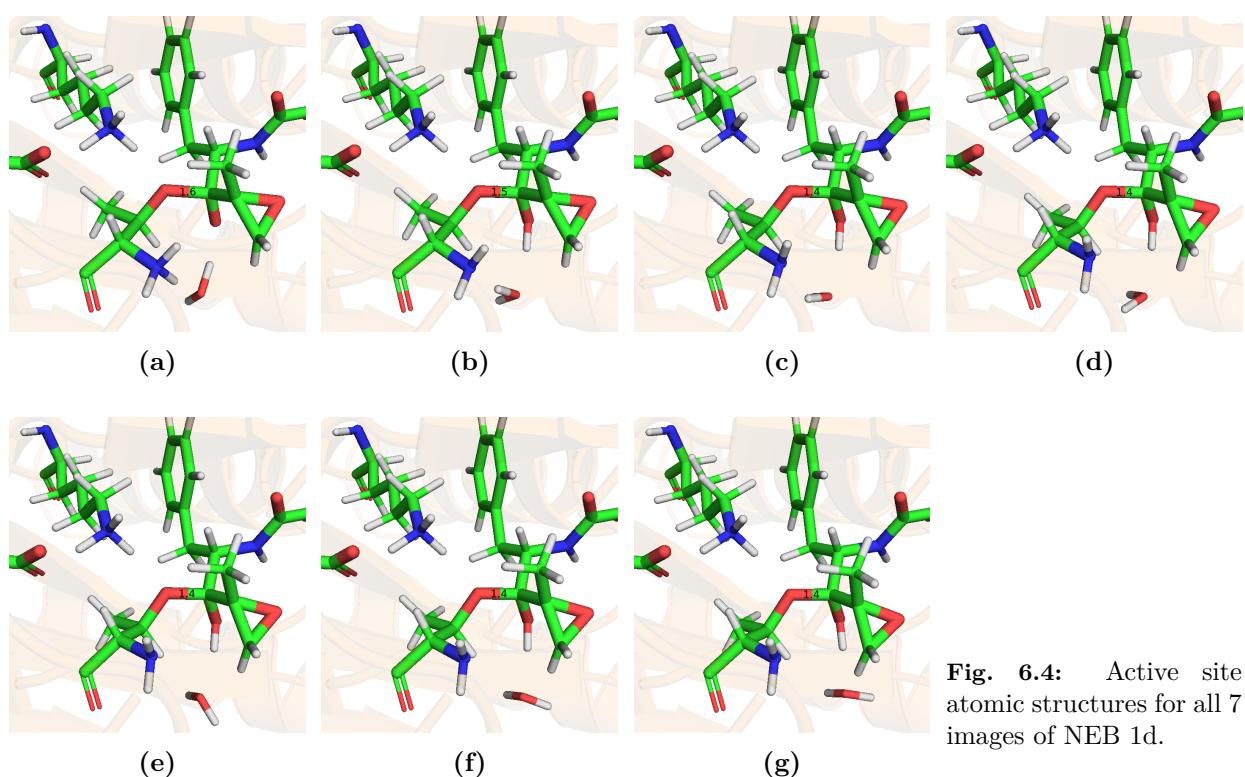


Fig. 6.4: Active site atomic structures for all 7 images of NEB 1d.

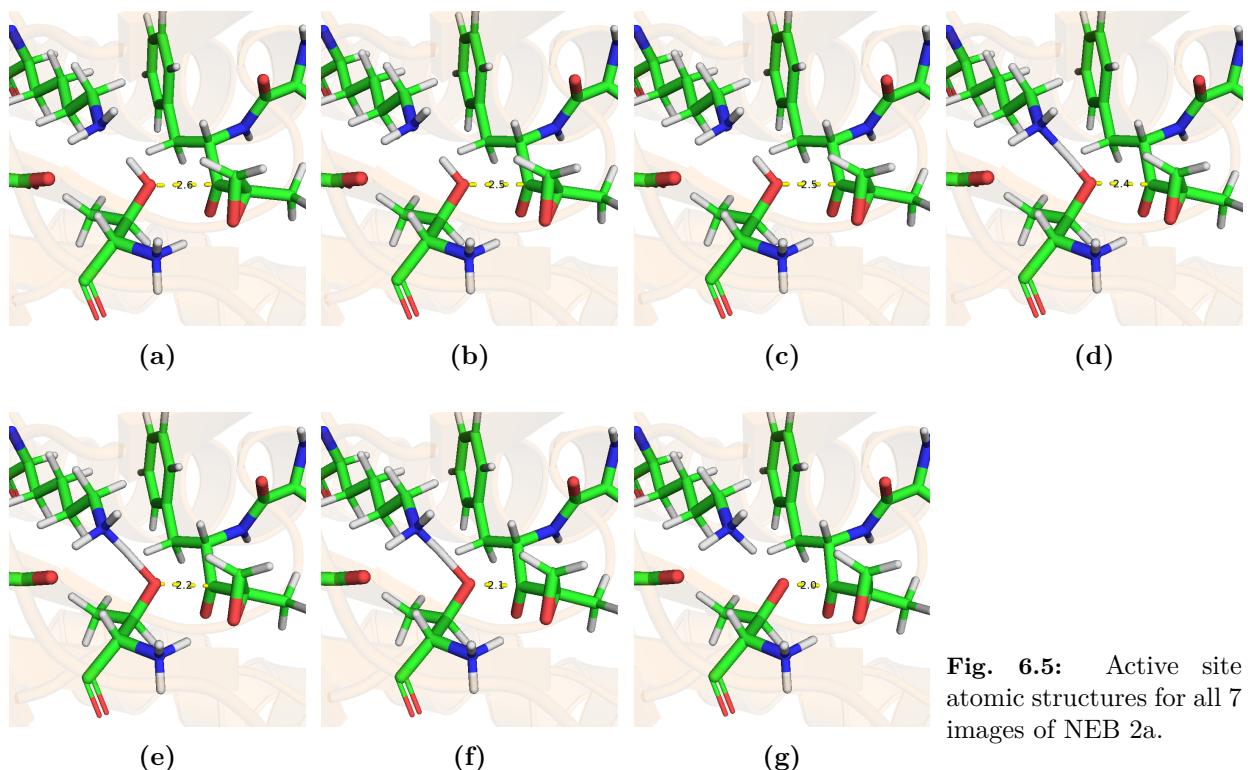


Fig. 6.5: Active site atomic structures for all 7 images of NEB 2a.

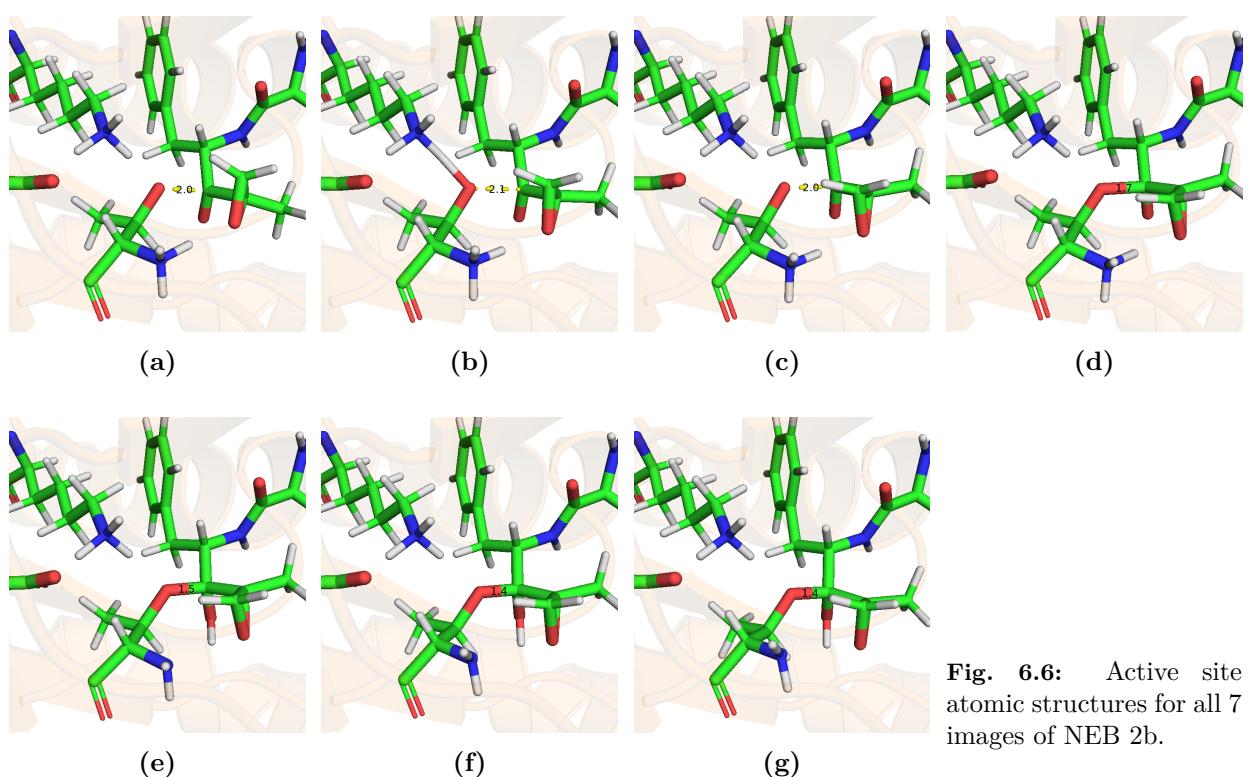


Fig. 6.6: Active site atomic structures for all 7 images of NEB 2b.

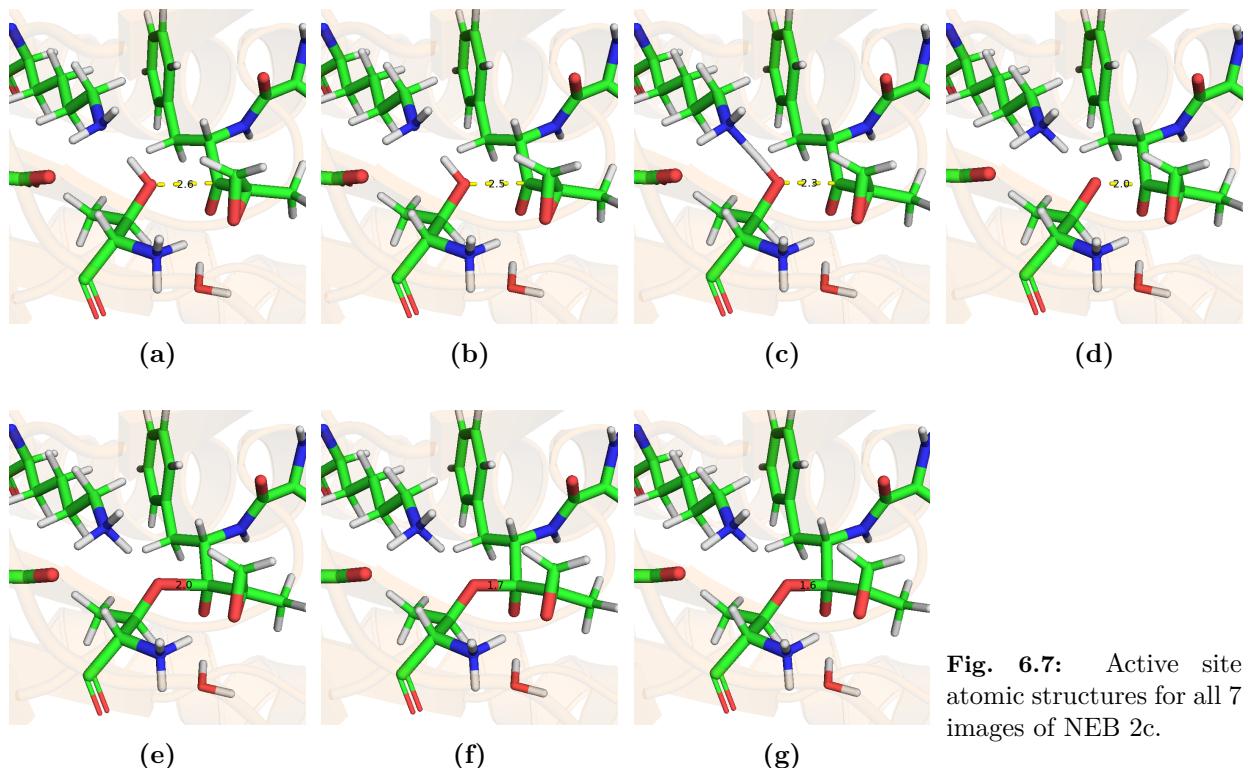


Fig. 6.7: Active site atomic structures for all 7 images of NEB 2c.

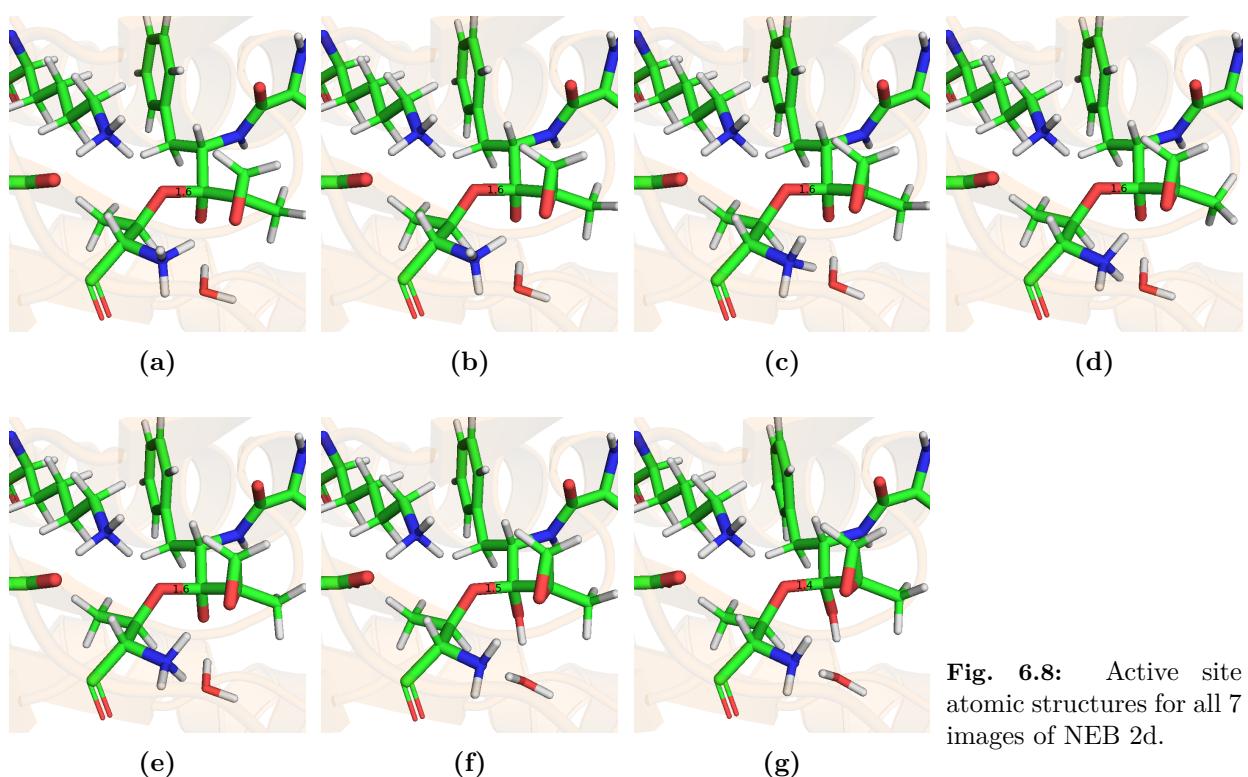


Fig. 6.8: Active site atomic structures for all 7 images of NEB 2d.

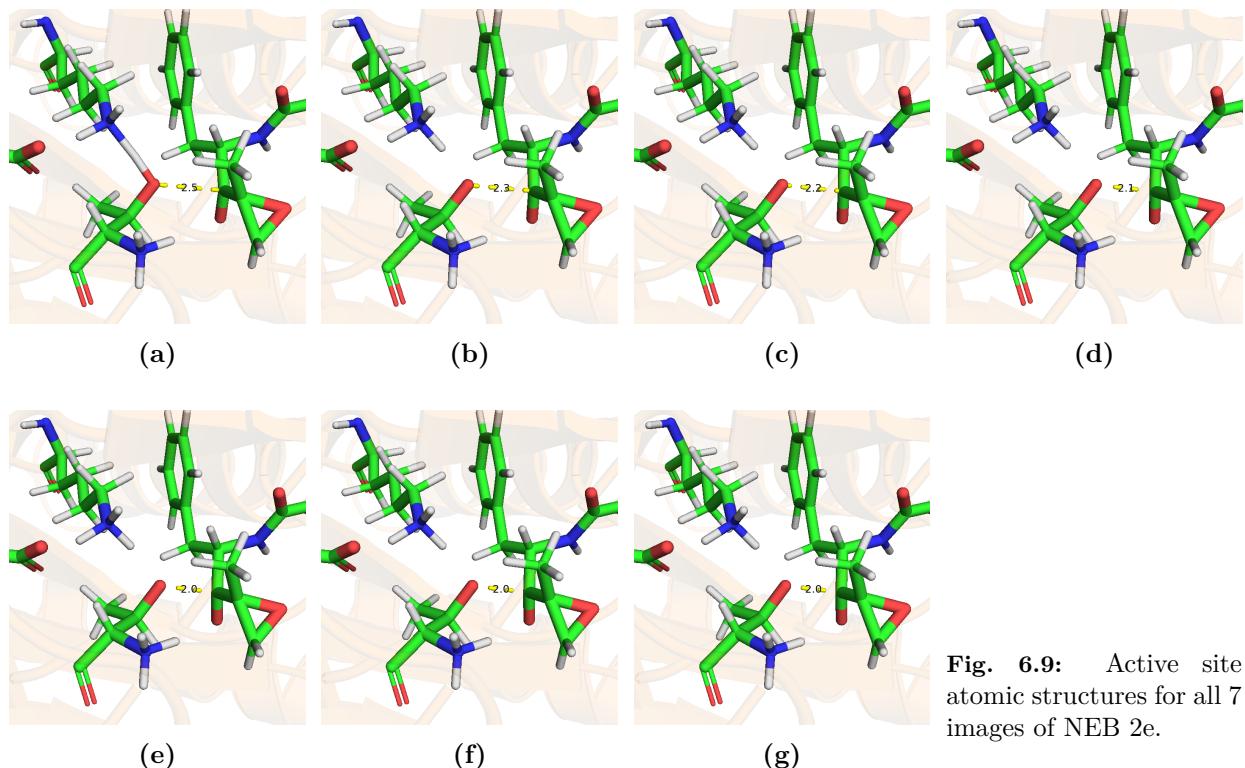


Fig. 6.9: Active site atomic structures for all 7 images of NEB 2e.

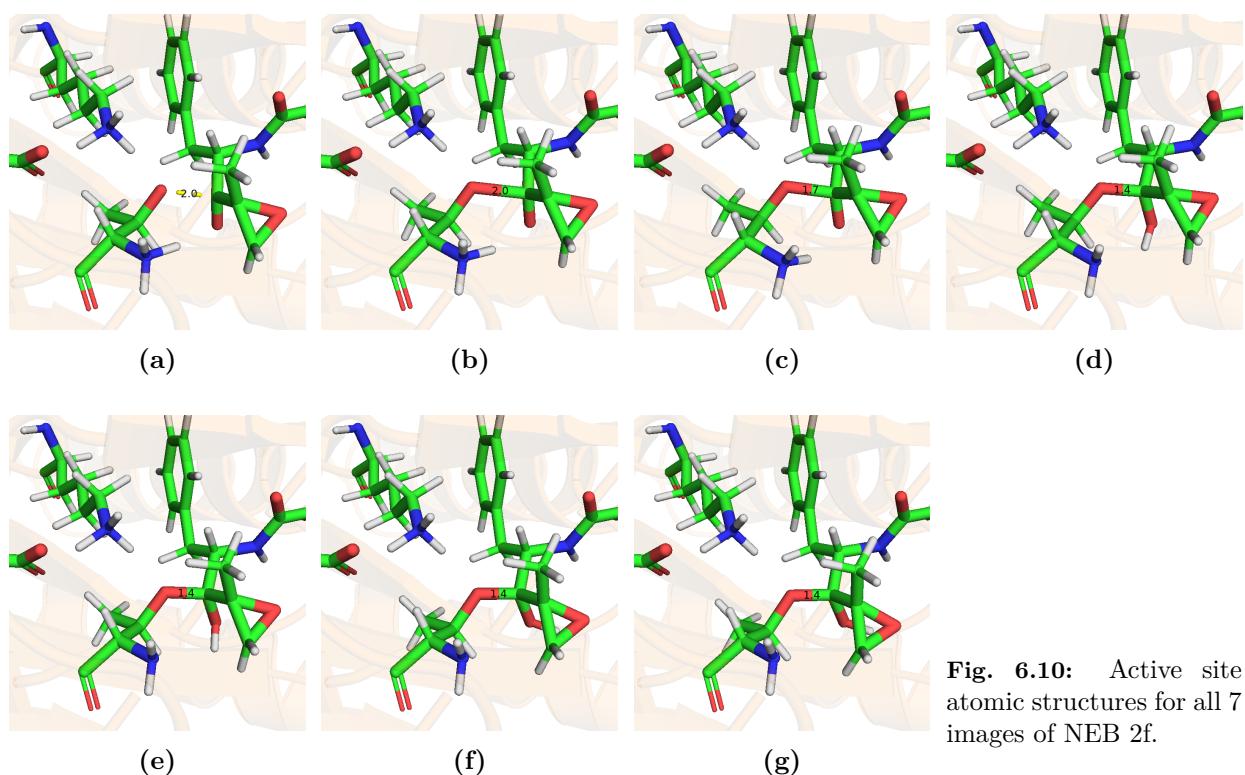


Fig. 6.10: Active site atomic structures for all 7 images of NEB 2f.

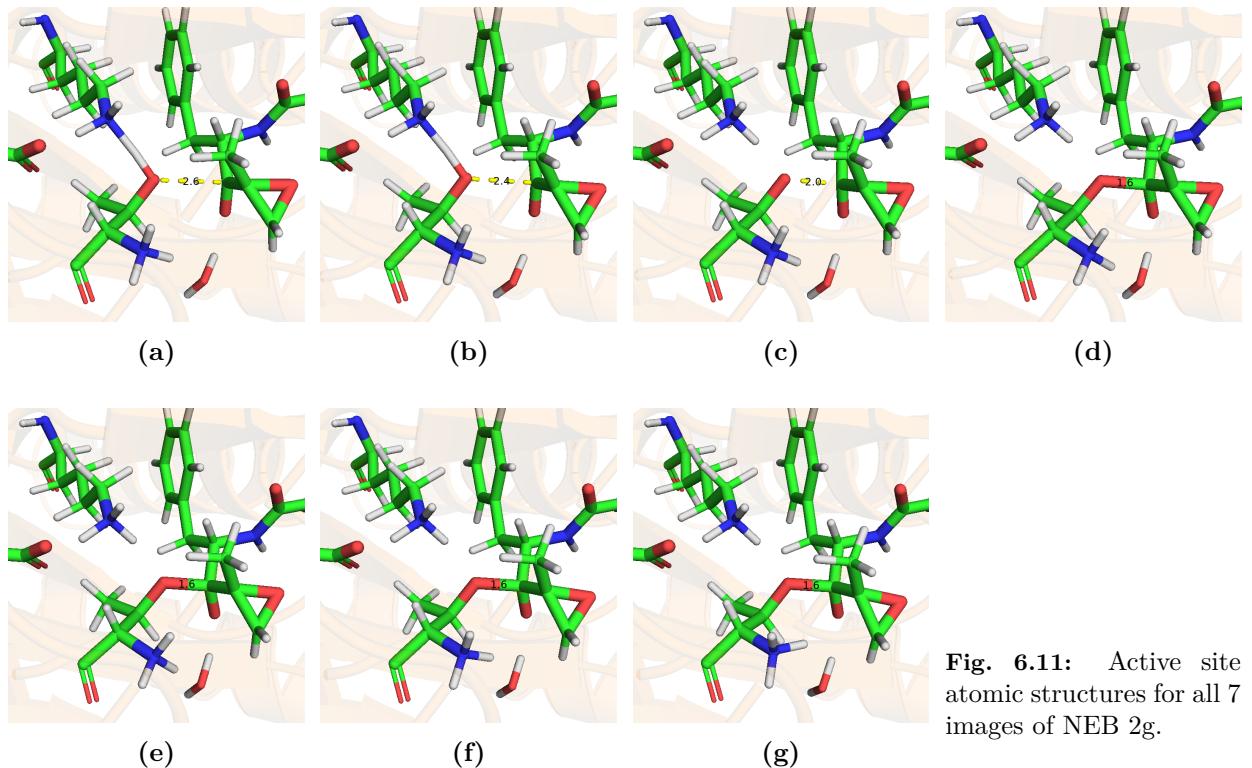


Fig. 6.11: Active site atomic structures for all 7 images of NEB 2g.

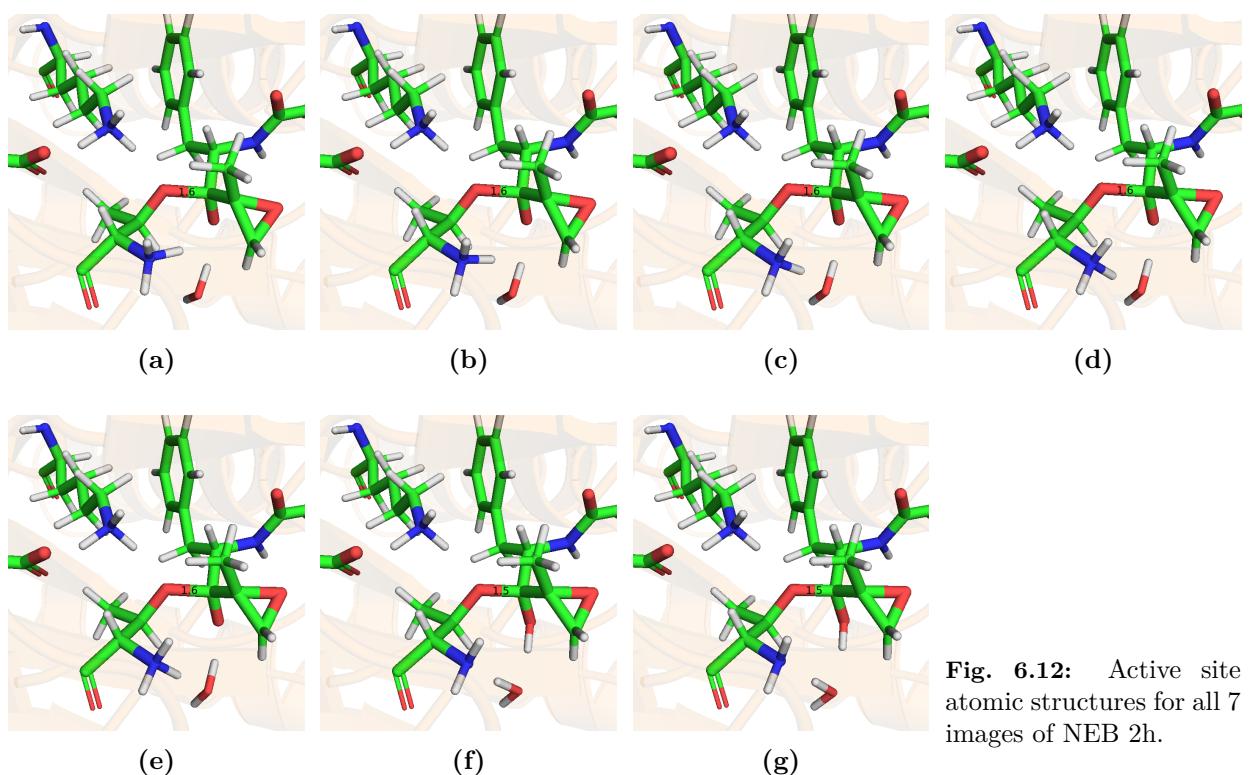


Fig. 6.12: Active site atomic structures for all 7 images of NEB 2h.

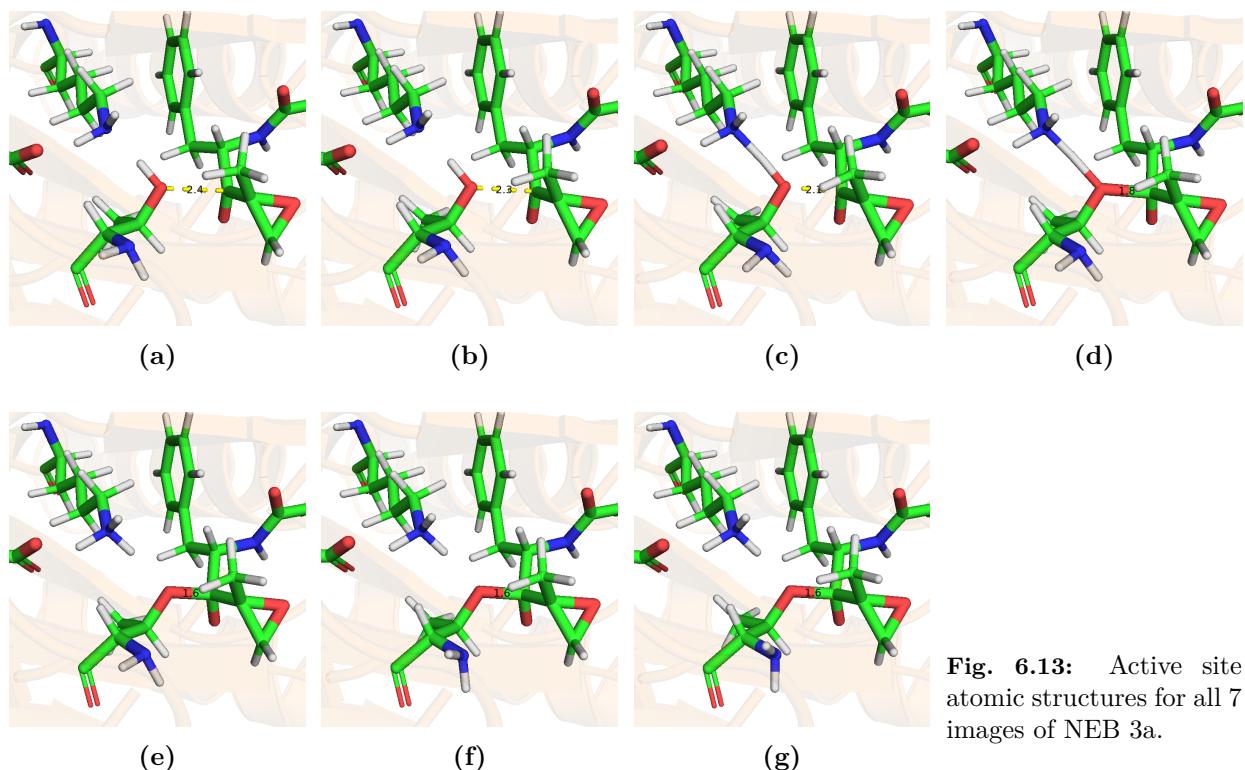


Fig. 6.13: Active site atomic structures for all 7 images of NEB 3a.

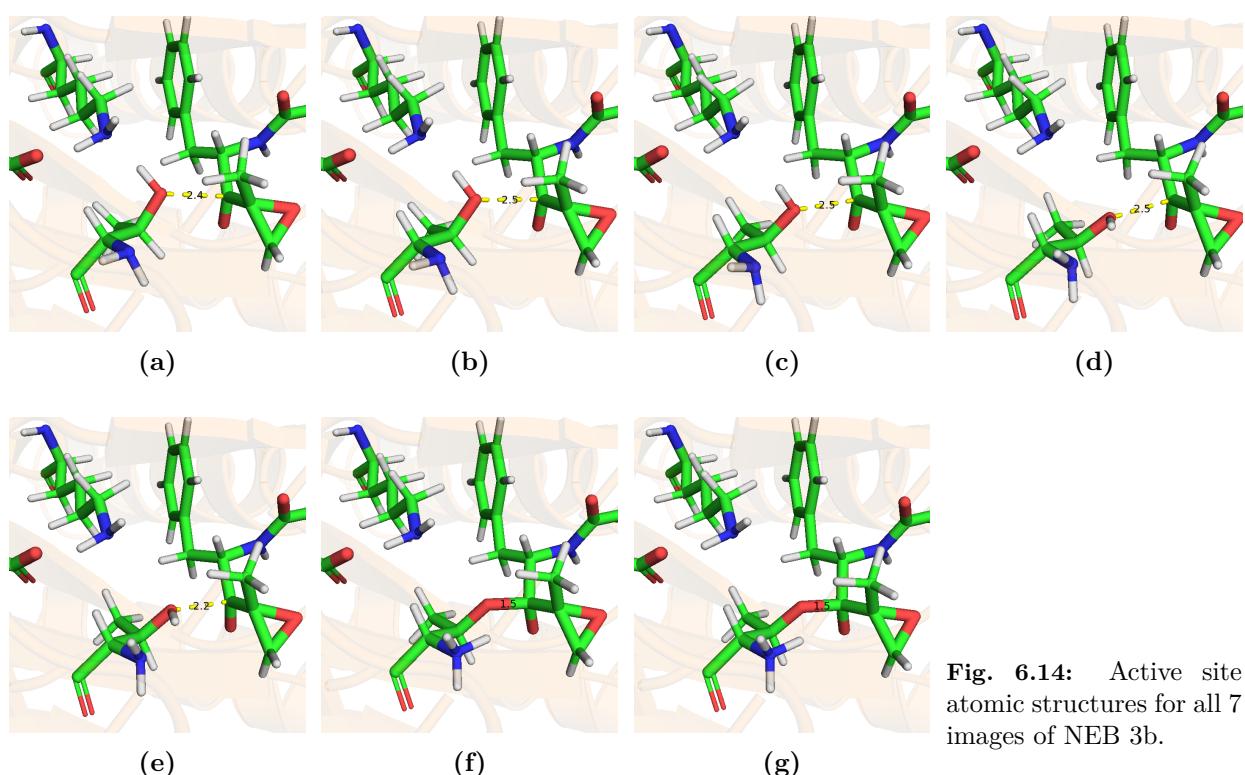


Fig. 6.14: Active site atomic structures for all 7 images of NEB 3b.

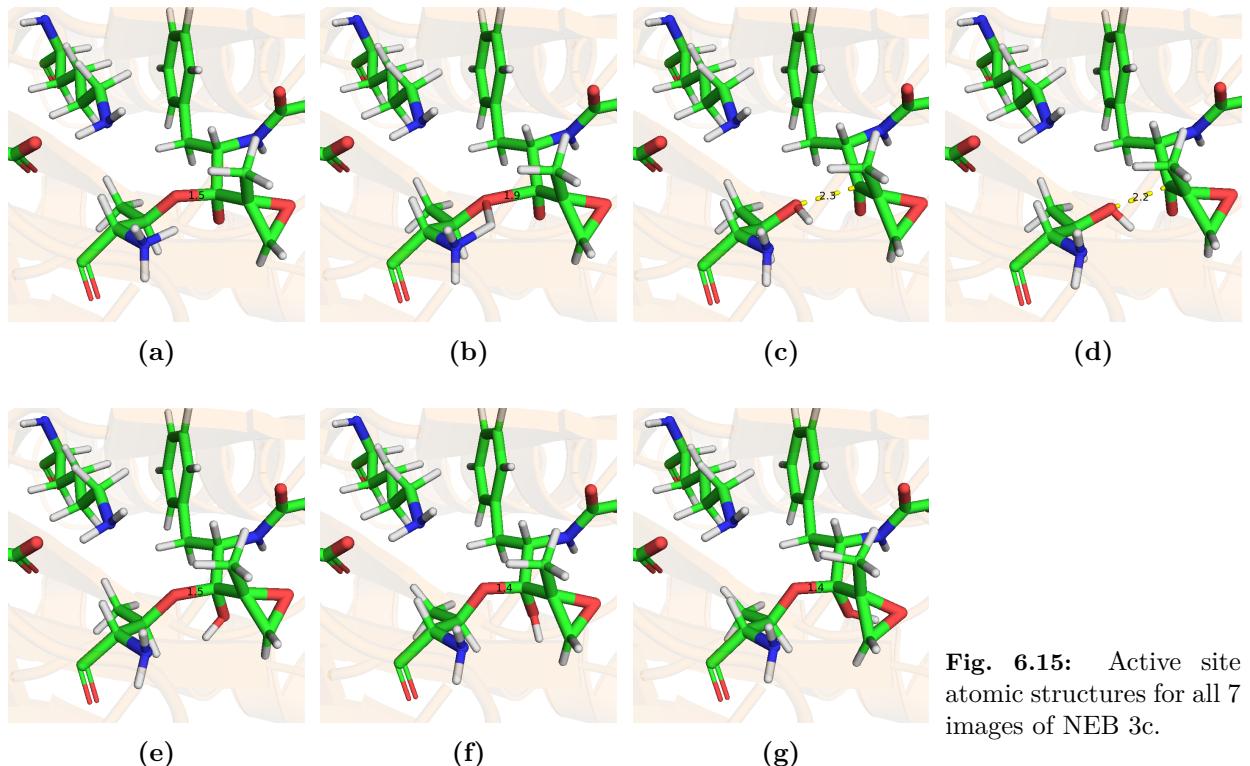


Fig. 6.15: Active site atomic structures for all 7 images of NEB 3c.

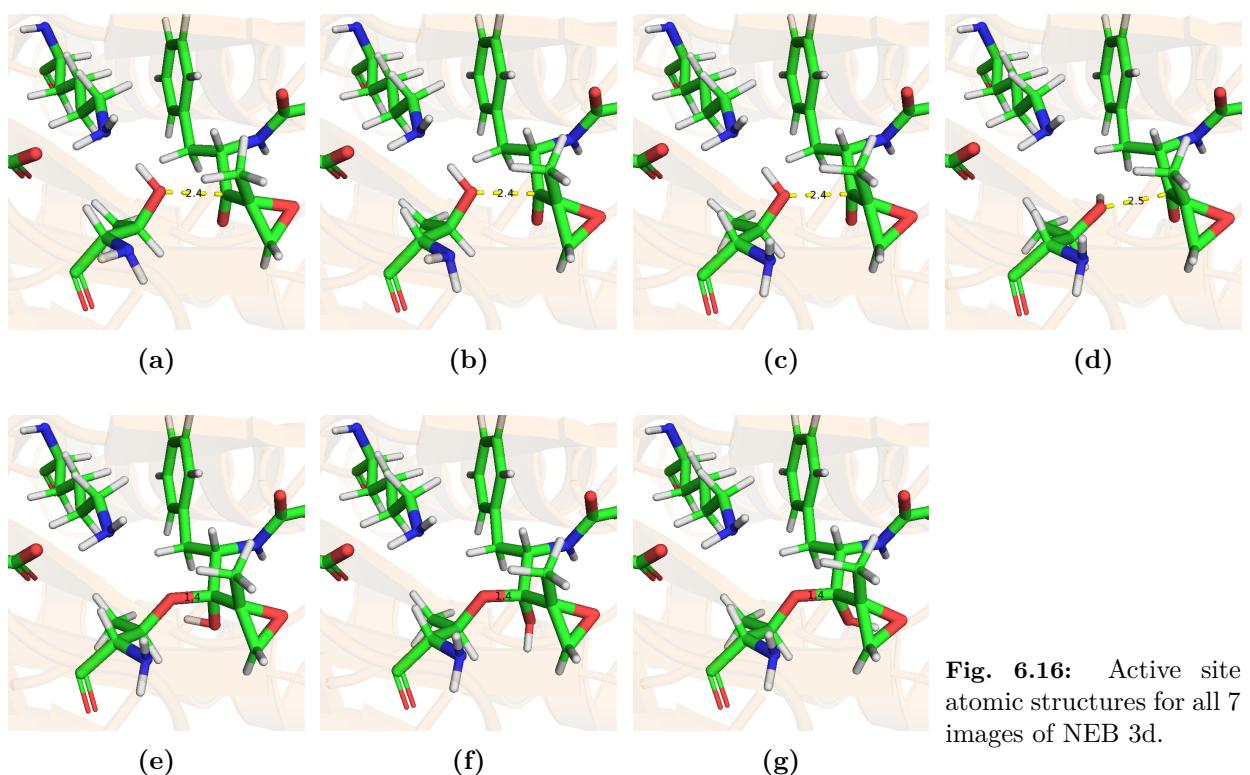


Fig. 6.16: Active site atomic structures for all 7 images of NEB 3d.

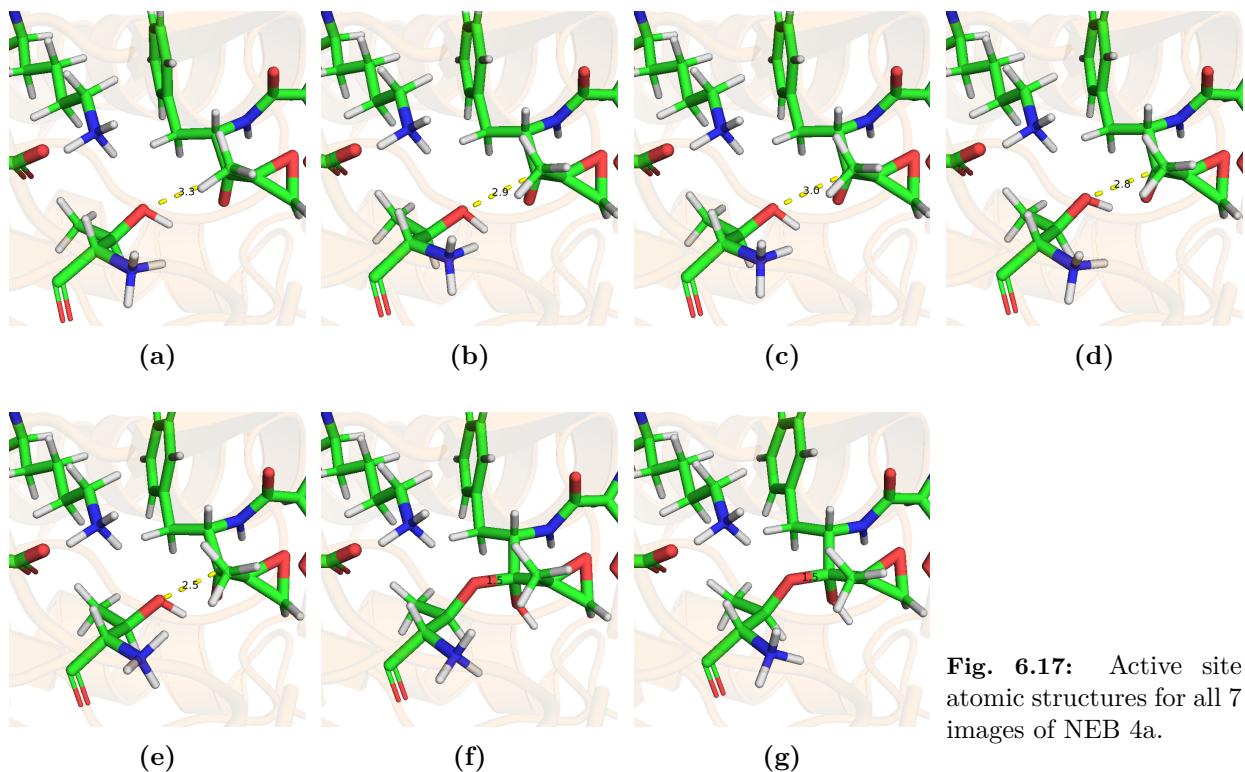


Fig. 6.17: Active site atomic structures for all 7 images of NEB 4a.

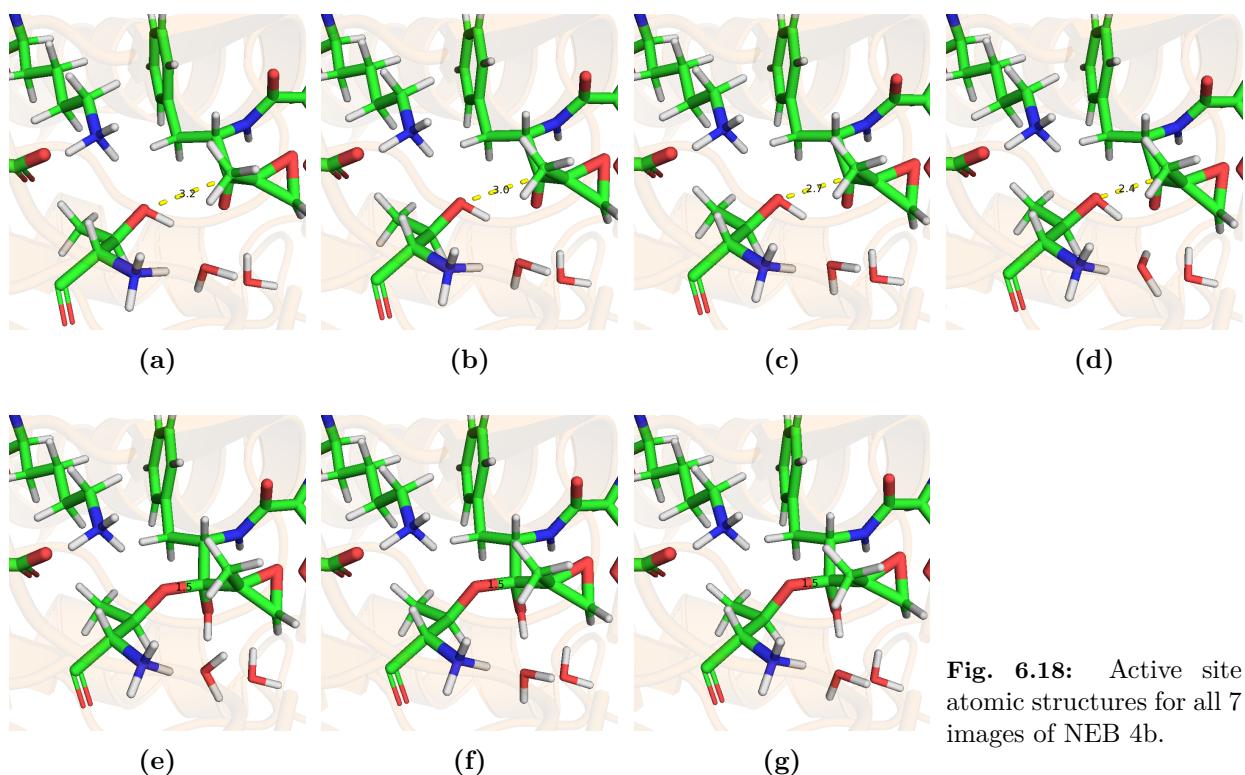


Fig. 6.18: Active site atomic structures for all 7 images of NEB 4b.

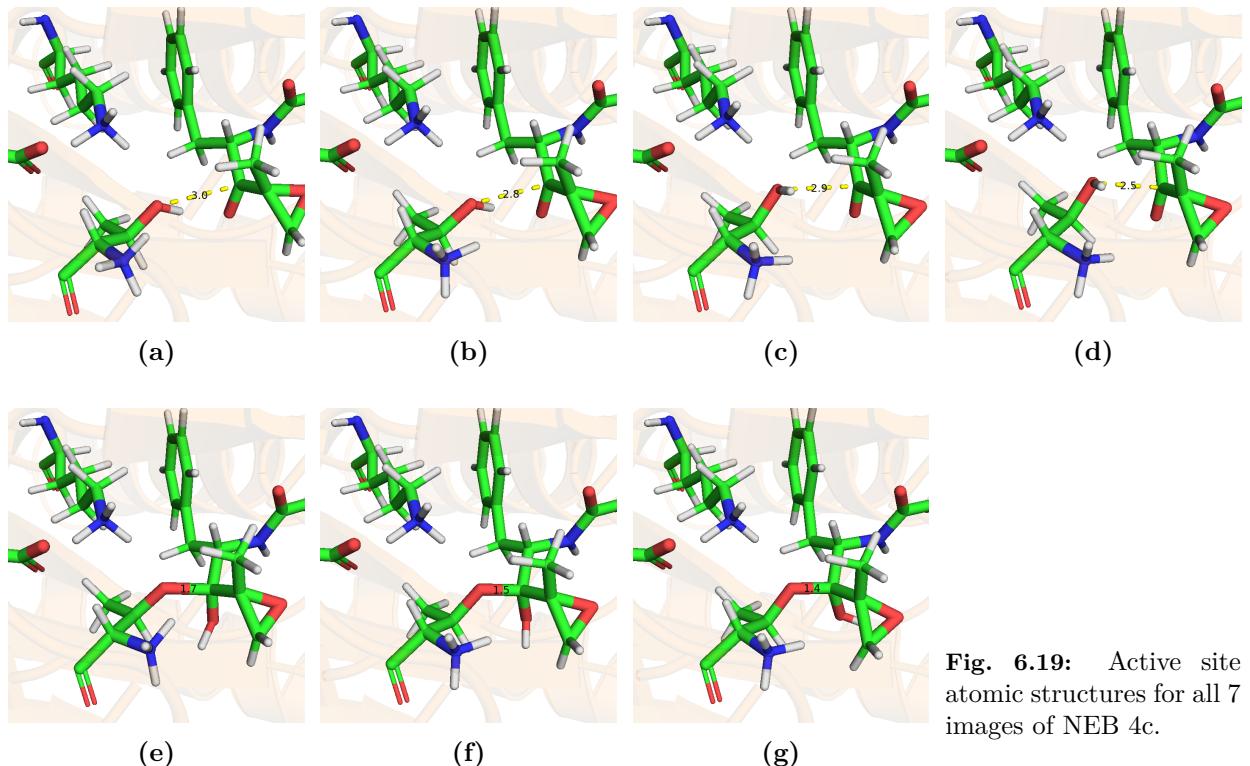


Fig. 6.19: Active site atomic structures for all 7 images of NEB 4c.

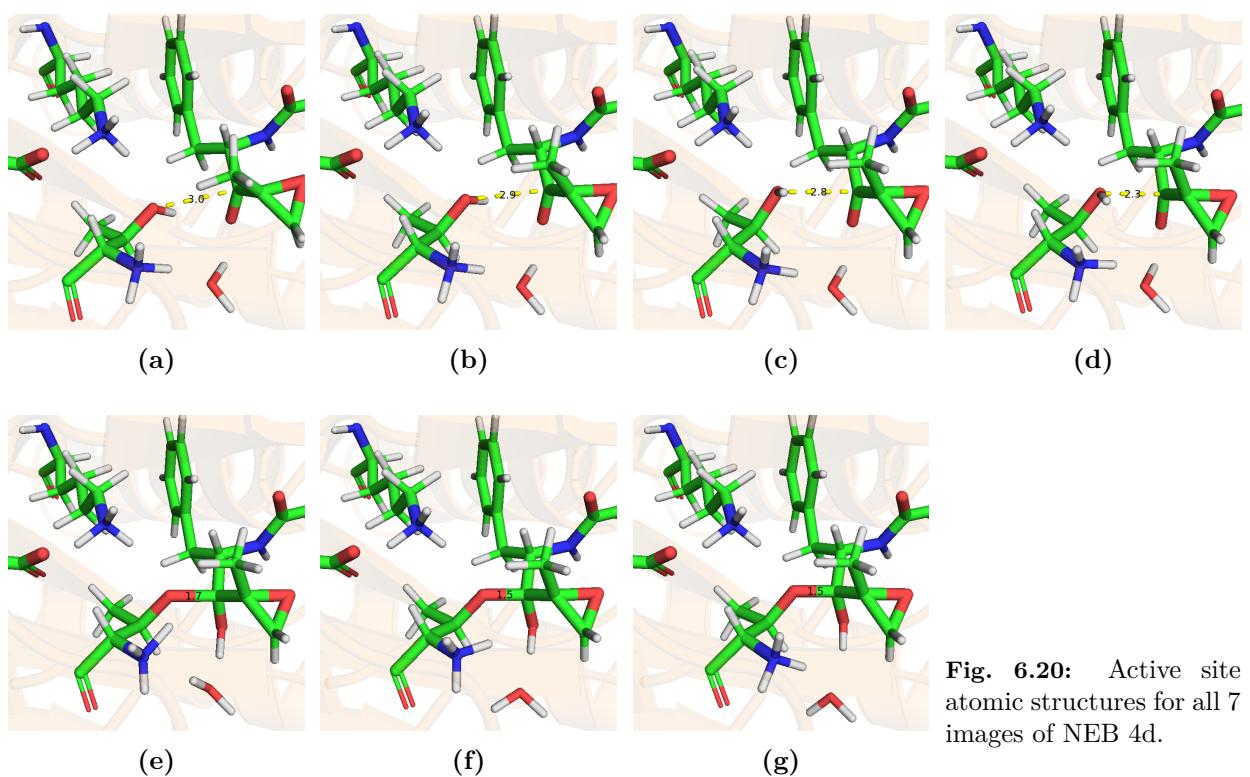


Fig. 6.20: Active site atomic structures for all 7 images of NEB 4d.

Bibliography

- [1] K. Tanaka, “The proteasome: overview of structure and functions,” *Proceedings of the Japan Academy, Series B*, vol. 85, no. 1, pp. 12–36, 2009.
- [2] D. Voges, P. Zwickl, and W. Baumeister, “The 26S proteasome: a molecular machine designed for controlled proteolysis,” *Annual review of biochemistry*, vol. 68, no. 1, pp. 1015–1068, 1999.
- [3] J. Adams, V. J. Palombella, E. A. Sausville, J. Johnson, A. Destree, D. D. Lazarus, J. Maas, C. S. Pien, S. Prakash, and P. J. Elliott, “Proteasome inhibitors: a novel class of potent and effective antitumor agents,” *Cancer research*, vol. 59, no. 11, pp. 2615–2622, 1999.
- [4] J. Adams, V. J. Palombella, and P. J. Elliott, “Proteasome inhibition: a new strategy in cancer treatment,” *Investigational new drugs*, vol. 18, no. 2, pp. 109–121, 2000.
- [5] A. L. Goldberg, “Protein degradation and protection against misfolded or damaged proteins,” *Nature*, vol. 426, no. 6968, pp. 895–899, 2003.
- [6] D. Nandi, P. Tahiliani, A. Kumar, and D. Chandu, “The ubiquitin-proteasome system,” *Journal of biosciences*, vol. 31, no. 1, pp. 137–155, 2006.
- [7] D. Wei, L. Fang, M. Tang, and C.-G. Zhan, “Fundamental reaction pathway for peptide metabolism by proteasome: insights from first-principles quantum mechanical/molecular mechanical free energy calculations,” *The Journal of Physical Chemistry B*, vol. 117, no. 43, pp. 13418–13434, 2013.
- [8] S. Ray and A. S. Murkin, “New electrophiles and strategies for mechanism-based and targeted covalent inhibitor design,” *Biochemistry*, vol. 58, no. 52, pp. 5234–5244, 2019. PMID: 30990686.
- [9] P. Hari, J. V. Matous, P. M. Voorhees, K. H. Shain, M. Obreja, J. Frye, H. Fujii, A. J. Jakubowiak, D. Rossi, and P. Sonneveld, “Oprozomib in patients with newly diagnosed multiple myeloma,” *Blood cancer journal*, vol. 9, no. 9, pp. 1–4, 2019.
- [10] J. Schrader, F. Henneberg, R. A. Mata, K. Tittmann, T. R. Schneider, H. Stark, G. Bourenkov, and A. Chari, “The inhibition mechanism of human 20S proteasomes enables next-generation inhibitor design,” *Science*, vol. 353, no. 6299, pp. 594–598, 2016.
- [11] D. Wei, B. Lei, M. Tang, and C.-G. Zhan, “Fundamental reaction pathway and free energy profile for inhibition of proteasome by epoxomicin,” *Journal of the American Chemical Society*, vol. 134, no. 25, pp. 10436–10450, 2012.

- [12] N. Serrano-Aparicio, K. Swiderek, and V. Moliner, “Theoretical study of the inhibition mechanism of human 20S proteasome by dihydroeponemycin,” *European journal of medicinal chemistry*, vol. 164, pp. 399–407, 2019.
- [13] C. Gu, I. Kolodziejek, J. Misas-Villamil, T. Shindo, T. Colby, M. Verdoes, K. H. Richau, J. Schmidt, H. S. Overkleeft, and R. A. van der Hoorn, “Proteasome activity profiling: a simple, robust and versatile method revealing subunit-selective inhibitors and cytoplasmic, defense-induced proteasome activities,” *The Plant Journal*, vol. 62, no. 1, pp. 160–170, 2010.
- [14] N. L. Doltsinis, “Nonadiabatic dynamics: mean-field and surface hopping,” *Quantum simulations of complex many-body systems: from theory to algorithms*, vol. 10, pp. 377–397, 2002.
- [15] J. B. Anderson, “Predicting rare events in molecular dynamics,” *Advances in Chemical Physics*, vol. 91, pp. 381–432, 1995.
- [16] C. A. Ullrich and Z.-h. Yang, “A brief compendium of time-dependent density functional theory,” *Brazilian Journal of Physics*, vol. 44, no. 1, pp. 154–188, 2014.
- [17] C. D. Sherrill, “An introduction to configuration interaction theory,” *School of Chemistry and Biochemistry, Georgia Institute of Technology*, 1995.
- [18] C. Møller and M. S. Plesset, “Note on an approximation treatment for many-electron systems,” *Physical review*, vol. 46, no. 7, p. 618, 1934.
- [19] M. Saitow, U. Becker, C. Ripplinger, E. F. Valeev, and F. Neese, “A new near-linear scaling, efficient and accurate, open-shell domain-based local pair natural orbital coupled cluster singles and doubles theory,” *The Journal of chemical physics*, vol. 146, no. 16, p. 164105, 2017.
- [20] A. Szabo and N. S. Ostlund, *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Corporation, 2012.
- [21] C. Ripplinger and F. Neese, “An efficient and near linear scaling pair natural orbital based local coupled cluster method,” *The Journal of chemical physics*, vol. 138, no. 3, p. 034106, 2013.
- [22] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, “A fifth-order perturbation comparison of electron correlation theories,” *Chemical Physics Letters*, vol. 157, no. 6, pp. 479–483, 1989.
- [23] S. Ahmadi, L. Barrios Herrera, M. Chehelamirani, J. Hostaš, S. Jalife, and D. R. Salahub, “Multiscale modeling of enzymes: QM-cluster, QM/MM, and QM/MM/MD: A tutorial review,” *International Journal of Quantum Chemistry*, vol. 118, no. 9, p. e25558, 2018.
- [24] A. R. Leach and A. R. Leach, *Molecular modelling: principles and applications*. Pearson education, 2001.

- [25] H. M. Senn and W. Thiel, “QM/MM methods for biomolecular systems,” *Angewandte Chemie International Edition*, vol. 48, no. 7, pp. 1198–1229, 2009.
- [26] D. Case, I. Ben-Shalom, S. Brozell, D. Cerutti, I. T.E. Cheatham, V. Cruzeiro, T. Darden, R. Duke, D. Ghoreishi, M. Gilson, H. Gohlke, A. Goetz, D. Greene, R. Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C. Simmerlinga, J. Smith, R. Salomon-Ferrer, J. Swails, R. Walker, J. Wang, H. Wei, R. Wolf, X. Wu, L. Xiao, D. York, and P. Kollman, “Amber 2018 reference manual,” 2018.
- [27] D. Sheppard, R. Terrell, and G. Henkelman, “Optimization methods for finding minimum energy paths,” *The Journal of chemical physics*, vol. 128, no. 13, p. 134106, 2008.
- [28] G. Henkelman and H. Jónsson, “Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points,” *The Journal of chemical physics*, vol. 113, no. 22, pp. 9978–9985, 2000.
- [29] E. L. Kolsbjerg, M. N. Groves, and B. Hammer, “An automated nudged elastic band method,” *The Journal of chemical physics*, vol. 145, no. 9, p. 094107, 2016.
- [30] J. Kastner, J. M. Carr, T. W. Keal, W. Thiel, A. Wander, and P. Sherwood, “DL-FIND: an open-source geometry optimizer for atomistic simulations,” *The journal of physical chemistry A*, vol. 113, no. 43, pp. 11856–11865, 2009.
- [31] G. Henkelman, B. P. Uberuaga, and H. Jónsson, “A climbing image nudged elastic band method for finding saddle points and minimum energy paths,” *The Journal of chemical physics*, vol. 113, no. 22, pp. 9901–9904, 2000.
- [32] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin, “UCSF Chimera – A visualization system for exploratory research and analysis,” *Journal of computational chemistry*, vol. 25, no. 13, pp. 1605–1612, 2004.
- [33] C. R. Søndergaard, M. H. Olsson, M. Rostkowski, and J. H. Jensen, “Improved treatment of ligands and coupling effects in empirical calculation and rationalization of pK_a values,” *Journal of chemical theory and computation*, vol. 7, no. 7, pp. 2284–2295, 2011.
- [34] M. H. Olsson, C. R. Søndergaard, M. Rostkowski, and J. H. Jensen, “PROPKA3: Consistent treatment of internal and surface residues in empirical pK_a predictions,” *Journal of chemical theory and computation*, vol. 7, no. 2, pp. 525–537, 2011.
- [35] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, “Comparison of simple potential functions for simulating liquid water,” *The Journal of chemical physics*, vol. 79, no. 2, pp. 926–935, 1983.
- [36] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, and C. Simmerling, “ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB,” *Journal of chemical theory and computation*, vol. 11, no. 8, pp. 3696–3713, 2015.

- [37] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, “Development and testing of a general amber force field,” *Journal of computational chemistry*, vol. 25, no. 9, pp. 1157–1174, 2004.
- [38] S. Adelman and J. Doll, “Generalized Langevin equation approach for atom/solid-surface scattering: General formulation for classical scattering off harmonic solids,” *The Journal of chemical physics*, vol. 64, no. 6, pp. 2375–2388, 1976.
- [39] H. J. Berendsen, J. v. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, “Molecular dynamics with coupling to an external bath,” *The Journal of chemical physics*, vol. 81, no. 8, pp. 3684–3690, 1984.
- [40] H. G. Petersen, “Accuracy and efficiency of the particle mesh Ewald method,” *The Journal of chemical physics*, vol. 103, no. 9, pp. 3668–3679, 1995.
- [41] P. Sherwood, A. H. de Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, *et al.*, “QUASI: A general purpose implementation of the QM/MM approach and its application to problems in catalysis,” *Journal of Molecular Structure: THEOCHEM*, vol. 632, no. 1-3, pp. 1–28, 2003.
- [42] S. Metz, J. Kästner, A. A. Sokol, T. W. Keal, and P. Sherwood, “ChemShell – A modular software package for QM/MM simulations,” *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 4, no. 2, pp. 101–110, 2014.
- [43] F. Spiegelman, N. Tarrat, J. Cuny, L. Dontot, E. Posenitskiy, C. Martí, A. Simon, and M. Rapacioli, “Density-functional tight-binding: basic concepts and applications to molecules and clusters,” *Advances in physics: X*, vol. 5, no. 1, p. 1710252, 2020.
- [44] I. T. Todorov, W. Smith, K. Trachenko, and M. T. Dove, “DL_POLY_3: new dimensions in molecular dynamics simulations via massive parallelism,” *Journal of Materials Chemistry*, vol. 16, no. 20, pp. 1911–1918, 2006.
- [45] F. Neese, “The ORCA program system,” *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 2, no. 1, pp. 73–78, 2012.
- [46] F. Neese, “Software update: the ORCA program system, version 4.0,” *Wiley Interdisciplinary Reviews: Computational Molecular Science*, vol. 8, no. 1, p. e1327, 2018.
- [47] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, “A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu,” *The Journal of chemical physics*, vol. 132, no. 15, p. 154104, 2010.
- [48] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, “Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields,” *The Journal of physical chemistry*, vol. 98, no. 45, pp. 11623–11627, 1994.

- [49] F. Weigend and R. Ahlrichs, “Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy,” *Physical Chemistry Chemical Physics*, vol. 7, no. 18, pp. 3297–3305, 2005.
- [50] D. C. Liu and J. Nocedal, “On the limited memory BFGS method for large scale optimization,” *Mathematical programming*, vol. 45, no. 1, pp. 503–528, 1989.
- [51] T. H. Dunning Jr, “Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen,” *The Journal of chemical physics*, vol. 90, no. 2, pp. 1007–1023, 1989.
- [52] M. Head-Gordon, J. A. Pople, and M. J. Frisch, “MP2 energy evaluation by direct methods,” *Chemical physics letters*, vol. 153, no. 6, pp. 503–506, 1988.
- [53] S. Kossmann and F. Neese, “Comparison of two efficient approximate Hartree-Fock approaches,” *Chemical Physics Letters*, vol. 481, no. 4-6, pp. 240–243, 2009.
- [54] M. Groll, C. R. Berkers, H. L. Ploegh, and H. Ovaa, “Crystal structure of the boronic acid-based proteasome inhibitor bortezomib in complex with the yeast 20s proteasome,” *Structure*, vol. 14, no. 3, pp. 451–456, 2006.
- [55] W. W. Bachovchin, “Contributions of NMR spectroscopy to the study of hydrogen bonds in serine protease active sites,” *Magnetic resonance in chemistry*, vol. 39, no. S1, pp. 199–213, 2001.
- [56] G. Dodson and A. Wlodawer, “Catalytic triads and their relatives,” *Trends in biochemical sciences*, vol. 23, no. 9, pp. 347–352, 1998.
- [57] J. Myung, K. B. Kim, and C. M. Crews, “The ubiquitin-proteasome pathway and proteasome inhibitors,” *Medicinal research reviews*, vol. 21, no. 4, pp. 245–273, 2001.
- [58] E. Genin, M. Reboud-Ravaux, and J. Vidal, “Proteasome inhibitors: recent advances and new perspectives in medicinal chemistry,” *Current topics in medicinal chemistry*, vol. 10, no. 3, pp. 232–256, 2010.
- [59] A. F. Kisselk, Z. Songyang, and A. L. Goldberg, “Why does threonine, and not serine, function as the active site nucleophile in proteasomes?,” *Journal of Biological Chemistry*, vol. 275, no. 20, pp. 14831–14837, 2000.
- [60] E. M. Huber, W. Heinemeyer, X. Li, C. S. Arendt, M. Hochstrasser, and M. Groll, “A unified mechanism for proteolysis and autocatalytic activation in the 20s proteasome,” *Nature Communications*, vol. 7, no. 1, pp. 1–10, 2016.
- [61] A. Saha, G. Oanca, D. Mondal, and A. Warshel, “Exploring the proteolysis mechanism of the proteasomes,” *The Journal of Physical Chemistry B*, vol. 124, no. 27, pp. 5626–5635, 2020.
- [62] J. L. Tymoczko, J. M. Berg, and L. Stryer, *Biochemistry: a short course*. Macmillan, 2011.

- [63] N. Agmon, “The Grotthuss mechanism,” *Chemical Physics Letters*, vol. 244, no. 5-6, pp. 456–462, 1995.
- [64] L. Xie, H. Liu, and W. Yang, “Adapting the nudged elastic band method for determining minimum-energy paths of chemical reactions in enzymes,” *The Journal of chemical physics*, vol. 120, no. 17, pp. 8039–8052, 2004.
- [65] W. Jia, H. Wang, M. Chen, D. Lu, J. Liu, L. Lin, R. Car, L. Zhang, *et al.*, “Pushing the limit of molecular dynamics with ab initio accuracy to 100 million atoms with machine learning,” *Proceedings of the International Conference for High Performance Computing, Network, Storage and Analysis*, no. 5, pp. 1–14, 2020.
- [66] G. Bistoni, I. Polyak, M. Sparta, W. Thiel, and F. Neese, “Toward accurate QM/MM reaction barriers with large QM regions using domain based pair natural orbital coupled cluster theory,” *Journal of chemical theory and computation*, vol. 14, no. 7, pp. 3524–3531, 2018.

Selbständigkeitserklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit ohne fremde Hilfe selbstständig verfasst und nur die von mir angegebenen Quellen und Hilfsmittel verwendet habe. Wörtlich oder sinngemäß aus anderen Werken entnommene Stellen habe ich unter Angabe der Quellen kenntlich gemacht. Die Richtlinien zur Sicherung der guten wissenschaftlichen Praxis an der Universität Göttingen wurden von mir beachtet. Eine gegebenenfalls eingereichte digitale Version stimmt mit der schriftlichen Fassung überein. Mir ist bewusst, dass bei Verstoß gegen diese Grundsätze die Prüfung mit "Nicht Bestanden" bewertet wird.

Göttingen, den 07.02.2021

David Prekel

Acknowledgements

First I would like to thank Prof. Dr. Ricardo Mata for acting as supervisor during the bachelor thesis and the previous internship, as well as for providing lots of support. The continuous work on problems of computational chemistry and biochemistry has significantly expanded my knowledge in this area. I would also like to thank Prof. Dr. Kai Tittmann for agreeing to be second examiner. Special thanks go to Dr. Jon Uranga for the incredible helpfulness and support over the course of this work. Last but not least, I would also like to thank my family for their support.