UiT The Arctic University of Norway

Faculty of Science and Technology
Department of Chemistry

Computational Studies of Selected Mutants of Shrimp Alkaline Phosphatase

Magnus Burkow
KJE-3900 Master thesis in Molecular Science, June 2020

Abstract

Drug discovery and refinement for use in the industry is an important and lucrative field within chemistry. In this regard, computational chemistry has shown to be a valuable tool in assisting drug discovery. A set of intriguing proteins used in drug discovery are cold-adapted enzymes. These enzymes have a lower activation point which makes them of interest in industrial use. In this study, the cold-adapted shrimp alkaline phosphatase's (SAP) affinity towards the immunoassay substrates PNPP and AMPPD has been studied. With free energy calculations using the linear interaction energy (LIE) and free energy perturbation (FEP) methods, SAP's affinity towards the substrates has been determined. In addition, putative mutations of SAP have been carried out by modelling SAP upon the already catalytic effective calf intestinal alkaline phosphatase (CIP) to increase its affinity towards PNPP and AMPPD. It was found that both substrates bound favourably to SAP and all its different mutations. Specificity towards PNPP and AMPPD changed depending on the mutations, where G102R would bind PNPP over AMPPD, while H149D would bind AMPPD over PNPP. The most successful mutation was a metal exchange of Zn^{2+} in the M3 metal site to Mg^{2+} according to the LIE method. The FEP method on the other hand revealed a decrease in binding free energies when transforming Zn^{2+} to Mg^{2+}. One possible solution given is that a key water molecule coordinated poorly in the FEP simulations compared to LIE simulations. However, even though literature shows that Mg^{2+} increases the catalytic effect, this does not necessarily imply that the proteins affinity towards the substrates need to increase.

Acknowledgment

I would like to extend my deepest appreciations to my supervisor Bjørn Olav Brandsdal for all his help and guidance throughout this master project.

My co-supervisor Geir Villy Isaksen deserves appreciation as well, without his help I would not have managed to get the calculations done in time. Additional thanks to Tor Arne Heim Andberg who helped me with the force field parameters.

I am also grateful to my family for supporting me throughout this endeavour. Special thanks to Ivan Cato Burkow, whose chemical knowledge and writing skills has been of impeccable help.

Finally, many thanks to all the wonderful bachelors, masters and PhD students at UiT, who has made the time even better than one could have hoped for.

Contents

1 Introduction

1.1 Proteins Importance

Of the four macromolecules that is present in our bodies, that being lipids, carbohydrates, nucleotides and proteins, none has probably been as widely investigated as proteins. Proteins are made up of the monomer amino acids. The tasks proteins perform in any living organism are plentiful, from signalling to transport, from structural integrity of the cell to storage [1]. Perhaps the most intriguing of all the proteins are enzymes. Enzymes are biological catalysers, which means they accelerates chemical reactions without itself being used up. Enzymes are essential for any living organism. The reason for this is the fact that most life sustaining chemical reactions that take place in any living organism, do not happen under normal circumstances. However, with the presence of enzymes, these chemical reactions can occur on a biological relevant timescale, and hence essential for survival [1].

Enzymes work by binding a target substrate, and through a chemical reaction converts the substrate to a product. The exact mechanism behind enzymatic reactions are not fully understood, but several models have been devised to describe them. The most used model to understand enzymatic reactions is the “induced fit” model, devised by Daniel Koshland in 1958 [2], which is an evolution of the old “lock and key” model. The “lock and key” model was proposed by Emil Fischer in 1894, stating that substrates and enzymes pose specific complementary geometric shapes which fits exactly to each other [3]. What this model fails to take into account, which the “induced fit” does, is the enzymes ability to change conformation. The “induced fit” model says that the substrate does not simply bind to the enzymes, but that the amino acids side chains will be moulded by the substrate, allowing it to fit perfectly in the active site of the enzyme. Enzymes catalyse a wide range of different reactions, and because of this, enzymes are divided into seven subgroups [4]. These are oxidoreductases, transferases, hydrolases, lyases, isomerases, ligases and translocases.

The study of how enzymes binds substrates, and their activity are essential for researchers wanting to understand the mechanism behind enzymes. As well as how to manipulate them. Exploration of enzymes interaction with different substrates, is also important in drug design.

1.1.1 Shrimp Alkaline Phosphatase, SAP

Alkaline phosphatases are a subcategory of hydrolases that dephosphorylates compounds, i.e. removes a PO_4^{3-} group. They are found in all kinds of organisms, from prokaryotic to eukaryotic cells. Phosphatases carry out an important function in the cells, as large amounts of cellular proteins are phosphorylated. It is estimated that between 30 – 65 % of proteins in the human genome are phosphorylated[5]. Alkaline phosphatases carry out the same function regardless of species, with structural differences depending on what kind of environment it is working in. Because alkaline phosphatases are present in human cells, especially in the liver, bile duct, kidney, bone and placenta, it is used as a biomarker in blood tests to help diagnose conditions such as hepatitis and osteomalacia[6]. The product, PO_4^{3-} is a natural inhibitor for phosphatases [7].

Alkaline phosphatases are homodimeric, where each monomer contains five cysteines, and has a metallic triad located in the active site. The metallic triad usually consists of two Zn^{2+} ions, and one Mg^{2+} ion, placed in the metal sites M1 through M3 respectively (Figure 2). It has been proposed that the metal ion in position M3 serves as a catalytic regulator, where the presence of Mg^{2+} increases activity [7]. A suggested mechanism for alkaline phosphatases was published in 1992[8], with a revised mechanism published in 2000[9] (Figure 1).

As mention above there exists countless different kinds of alkaline phosphates. One such alkaline phosphate is shrimp alkaline phosphate, abbreviated SAP. As the name suggest, it is an alkaline phosphatase found in the Arctic shrimp *Pandalus borealis*. SAP is a cold-adapted enzyme, which are enzymes often found in organisms living in cold environments.

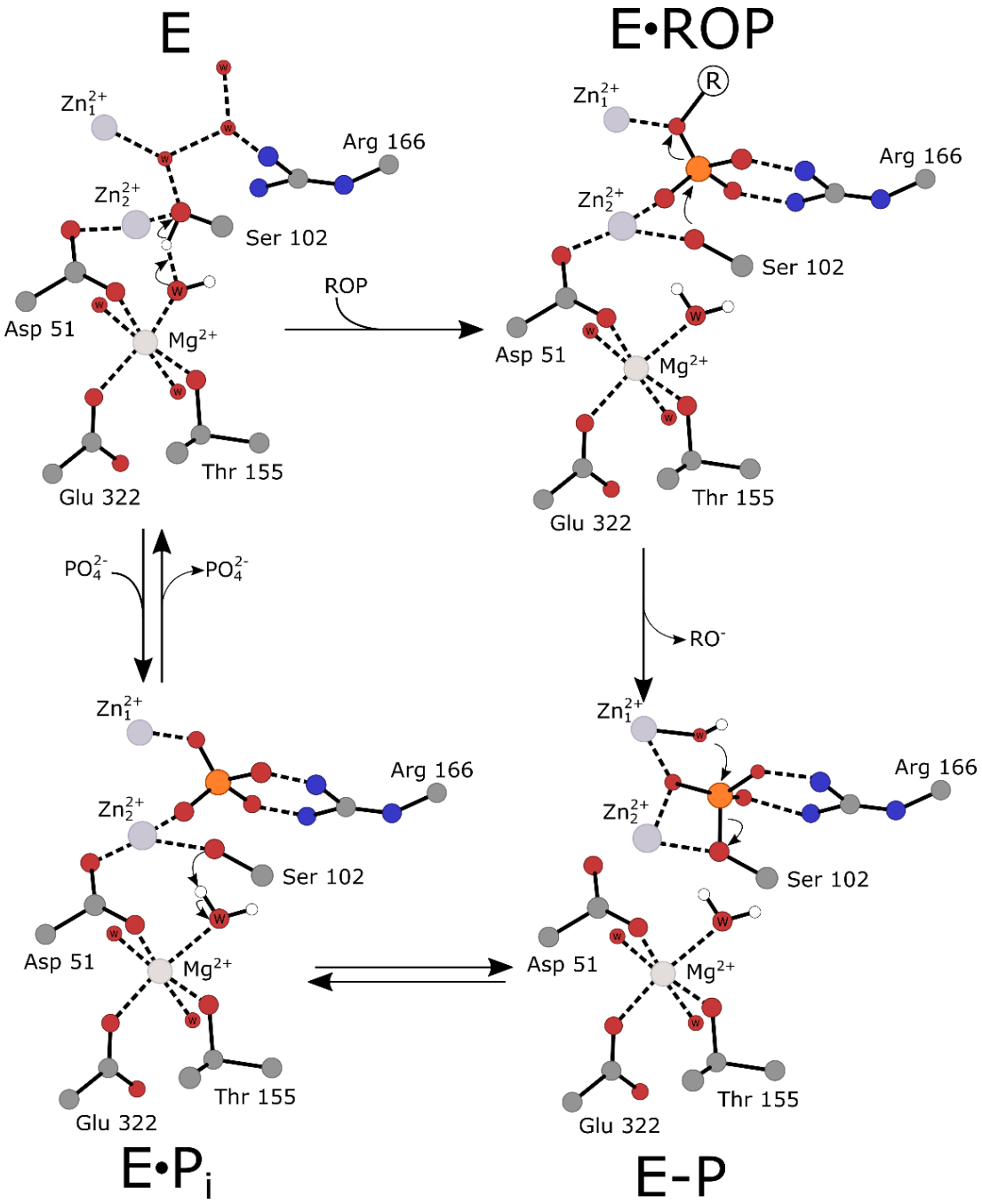

Figure 1. Reaction mechanism of an alkaline phosphatase. All hydrogens have been removed, except ones relevant for the catalytic reaction. Substrate is stabilized within the active site by Arg166 and the Zn^{2+} ions in position M1 and M2. A nucleophilic attack by Ser102 upon the central phosphate then takes place, resulting in the loss of the RO^- leaving group and the formation of a covalent enzymes-phosphate intermediate. After which a nucleophilic attack upon the covalent enzymes-phosphate complex is carried out by a hydroxide ion coordinated to Zn^{2+} in M1, resulting in a non-covalent enzymes phosphate complex. Then a water molecule coordinated to Mg donates a proton to Ser102, whereas the phosphate leaves the enzyme. Figure is based on *A revised mechanism for the alkaline phosphatase reaction involving three metal ions* [9].

These enzymes are of interests among other things for their high activity at lower temperature, where warm adapted enzymes would otherwise cease to function. The main characteristic of cold-adapted enzymes is a reduced activation enthalpy and a more negative activation entropy when compared to their warm-active counterparts [10]. SAP was first crystallized in 2002[11], where an interesting discovery was made, namely that the metallic triad in SAP consisted entirely of Zn^{2+} ions. A metal exchange study was carried out in 2004[7] where the Zn^{2+} ion in the M3 site was successfully replaced with a Mg^{2+} ion. When it comes to substrate binding, SAP binds the phosphate group by coordinating the oxygens with ARG 162 and the Zn^{2+} ion in the M1 site (Figure 2). A recombinant version of SAP is being produced and sold by several prominent pharmaceutical companies for use in immunoassays [12].

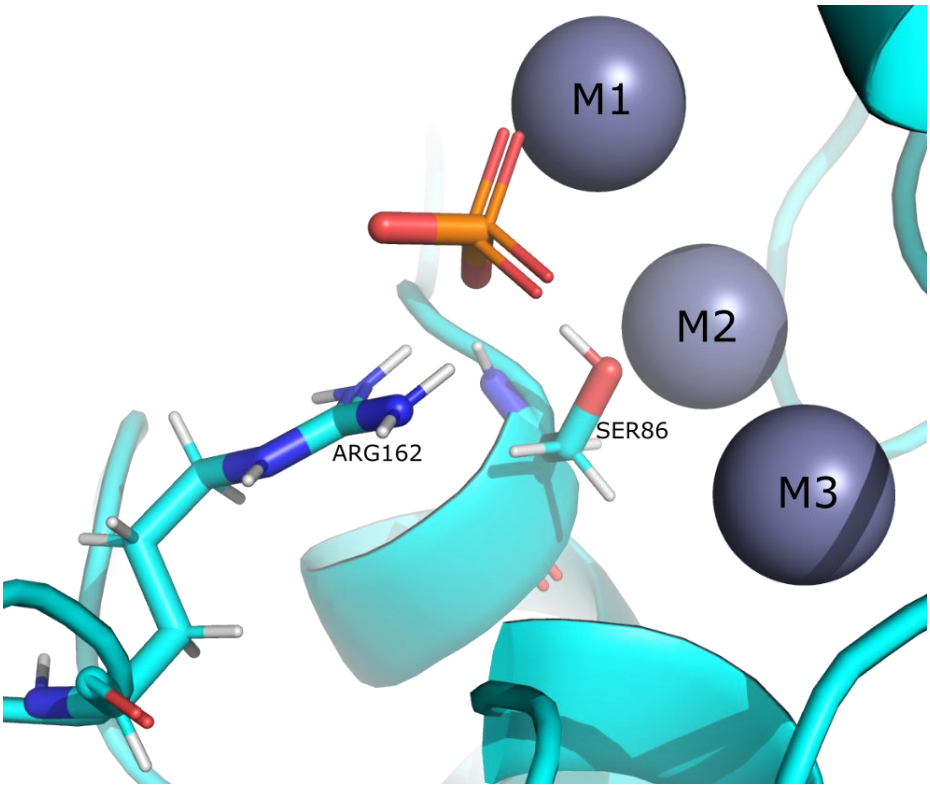

Figure 2. Snapshot of SAP's monomers active site. Zn^{2+} is present in all three metal sites.

1.1.2 The Substrates PNPP and AMPPD

A commonly used analytical biochemistry assay is the enzyme-linked immunosorbent assay (ELISA), which was described in 1971 by Engvall and Perlmann [13]. ELISA uses a solid-phase enzyme immunoassay to detect the presence of a given ligand. These ligands range from different structures, but they are often proteins. Because of this, antibodies specific to that protein can be used in the assay. It has seen wide use in medicine, plant pathology, biotechnology and quality control in various industries. Two well documented substrates used in ELISA are para-Nitrophenylphosphate (PNPP) [14, 15] and 3-(2'-spiroadamantyl)-4-methoxy-4-(3”-phosphoryloxy)-phenyl-1,2-dioxetane (AMPPD) [16] (Figure 3).

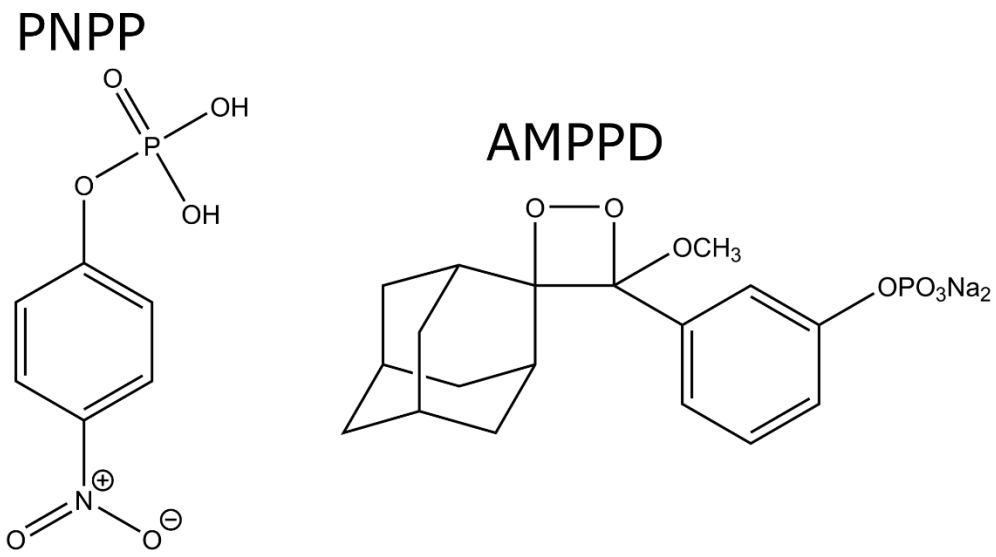

Figure 3. PNPP and AMPPD's chemical structures.

PNPP is a chromogenic substrate, meaning it is initially colourless, but through a chemical reaction can change its colour. PNPP is hydrolysed by alkaline phosphatases, removing the inorganic phosphate group, resulting in a yellow phenolate. Because of this is PNPP suited for colorimetric analysis. AMPPD on the other hand is a chemiluminescent substrate that reacts with alkaline phosphatases, and with the help of a trigger solution produces an ester moiety in a high excited state. When the ester moiety goes to ground state, light is emitted that can be detected. PNPP and AMPPD are both sold by well-known chemical providers.

1.2 Computational Chemistry

With the advent of computers, and the rise of their raw calculative power, it was inevitable that chemists would want to harness this resource to help assist solving chemical problems. As early as 1927, the first theoretical calculations in chemistry was done by Walter Heitler and Fritz London[17]. However, it was not until the 1950s that the real potential of computational chemistry was discovered, notably with the first simulation of a liquid (1953, [18]) , and the first *ab initio* Hartree-Fock method calculation on a diatomic system performed at MIT (1956, [19]). In the 1960s, the dream of an *in silico* chemical lab came to fruition by the first generations of "black box" computer programs such as POLYATOM[20] and Gaussian70, Gaussian16[21] is still being used today by computational chemists worldwide. During the same period, molecular graphics gained traction as both a research field within chemistry, but also as an invaluable tool for analysing data from both experimental research as well as computational research. By the start of the 1980s, the first publication of the *Journal of Computational Chemistry* was published[22]. In 1998 the field of computational chemistry was celebrated by the Nobel Prizes. Walter Kohn for his development of the density-function theory and John Pople for his development of computational methods in quantum chemistry. By the turn of the millennial, computational chemistry had segmented itself as a young but promising research field, which has only been growing the past decades.

1.2.1 Molecular Mechanics

The methods used in computational chemistry can broadly be divided into two groups, quantum mechanics (QM) [23] and molecular mechanics (MM) [24]. QM simulations considers the nuclei and electrons as individual quantities that can be treated separately. This approximation is called the Born-Oppenheimer approximation [25], and allows the electronic Schrodinger equation to be solved. QM is a computationally heavy simulation, while on the other hand MM simulations have a more classical approach. In MM, atoms are simulated as particles, each with a van der Waals radius, a constant atomic charge and if the force field (FF) allows it, polarizability [26]. The atomic charge is usually obtained either experimentally or from QM calculations. The bonds are characterised as springs, with an equilibrium distance equal to the experimental or calculated bond lengths. The potential energy of an MM system is the sum of the bonded and non-bonded interactions (Equation 1).

$$E = E_{bonded} + E_{non-bonded}$$ **Equation 1**

Where the bonded interactions consist of four components, bond stretching, bond angle bending, bond torsion and improper bond torsion (Equation 2).

$$E_{bonded} = E_{stretching} + E_{bending} + E_{torsion} + E_{improper}$$ **Equation 2**

The non-bonded interactions consist of two components, electrostatic interactions and van der Waals interactions (Equation 3).

$$E_{non-bonded} = E_{electrostatic} + E_{van\ der\ Waals}$$ **Equation 3**

A common functional form to describe a forcefield can be written (Equation 4) [27].

$$U(r^N) = \sum_{bond} \frac{k_i}{2}\left(l_i - l_{i,0}\right)^2 + \sum_{angle} \frac{k_i}{2}\left(\theta_i - \theta_{i,0}\right)^2 + \sum_{tors} \frac{V_n}{2}(1 + \cos(n\omega - \gamma)) + \sum_{im.tors} \frac{k_i}{2}\left(\zeta_i - \zeta_{i,0}\right)^2 + \sum_{i=1}^{N}\sum_{j=i+1}^{N}\left(4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}\right)$$ **Equation 4**

As described earlier, the bonded interactions can be divided into bond stretching, angle bending and torsional rotation. The energy that arises from bond stretching can be defined using a simple harmonic potential (Equation 5)

$$E_{bond}(l) = \frac{k}{2}(l - l_0)^2$$ **Equation 5**

where k is the force constant, l is the bond length and l_0 is the reference bond length. The force constant k is a constant that determines the stiffness of the harmonic oscillator. This is a simple model, whenever l is different from l_0, a quadratic increase in energy can be seen, resulting in energies moving towards infinity when the bond length increase or decrease. In reality the energies should not increase towards infinity as the bond length increase, rather it

should move towards the dissociation energy. A simple yet effective solution to this problem is the Morse potential [28], where the dissociation energy D is accounted for in the equation (Equation 6).

$$E_{Morse}(l) = D\left(1 - e^{-\alpha(l-l_0)}\right)^2, \alpha = \sqrt{\frac{k}{2D}} \quad \textbf{Equation 6}$$

The energy arising from angle bending can be modelled with a harmonic potential approach, similar to the bond stretching energy (Equation 7)

$$E_{angle}(\theta) = \frac{k}{2}(\theta - \theta_0)^2 \quad \textbf{Equation 7}$$

where θ is the angle, and θ_0 is the reference angle. The angle bending energy can also be described using a Taylor expansion terminated at the second order.

Torsional energy, energy related to the rotation of the central bond formed by four atoms, can be calculated using a Fourier series (Equation 8)

$$E_{tors}(\omega) = \sum_{n=1} V_n cos(n\omega) \quad \textbf{Equation 8}$$

where V_n is a constant that effects the energy barrier of rotation. As the angle increases from 0° to 360°, n gives the number of energy minima by describing the multiplicity of the function.

To ensure the forcefield handles planar arrangements correctly, such as aromatic rings, improper torsions are introduced. An improper torsion term is introduced between four atoms that are not connected in a standard way. Instead of having an atom sequence of 1-2-3-4, an improper torsion term would be introduced in a sequence of 1-2-4-3. A functional form of this would be a torsional potential that maintains angles at either 0° or 180° (Equation 9).

$$E_{im.tors}(\omega) = k(1 - \cos 2\omega) \quad \textbf{Equation 9}$$

The non-bonded interactions are, as described earlier, divided into electrostatic and van der Waals interactions. For the electrostatic interactions, the Coulomb's law is used to describe point charge electrostatics between pairs of non-bonded particles (Equation 10).

$$E_{elec}(r_{ij}) = \sum \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad \textbf{Equation 10}$$

q_i and q_j are the partial charges assigned to particle i and j, ϵ_o is the electric constant of a vacuum and r_{ij} is the distance between the respected particles.

The van der Waals interactions are described using the Lennard-Jones potential [29] which describes the interaction between two neutral particles, using a relatively simple mathematical model (Equation 11).

$$E_{vdW}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \quad \textbf{Equation 11}$$

ε is the potential well depth, σ is the distance at which the intermolecular potential equals zero and r is the distance of separation between both particles. The equation consists of two elements, Pauli-Repulsion and London dispersion, respectively the first and second part of the equation. This results in both repulsive and attractive forces being considered. When plotting the potential energy against distance between two atoms (Figure 4), one can clearly see the equilibrium distance for a system, in addition to the repulsive force acted upon the system.

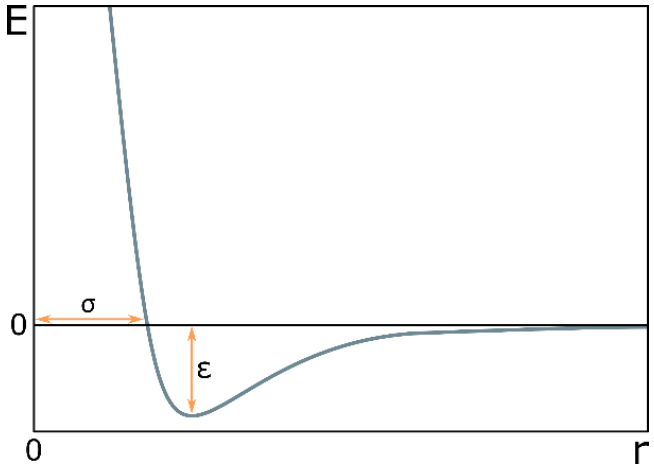

Figure 4. Lennard-Jones plot, σ is the distance at which the intermolecular potential equals zero, and ε is the potential well depth.

A set of parameters are used to characterise a chemical system, called a force field. A force field is a potential energy surface (PES) that contains parameters such as bond lengths, angles torsion etc. which has been obtained empirically through experimental work. There exist several different kinds of force fields, each tailored for different tasks. One such force field is OPLS (Optimized Potentials for Liquid Simulations) developed in the group of Prof. William L. Jorgensen, which is a popular forcefield for simulating biomolecules [30].

Because of MM`s use of classical mechanics, the size of the system simulated can range from small molecules, to large complex biological systems such as proteins and membranes. MM has however some drawbacks such as the fact that most FFs does not allow for bond breakage or formation. This can be solved by using a Morse potential [28] which is a more accurate model for bond breakage than the harmonic potential that most FF uses. Another drawback of MM is that the FFs are derived from ground states systems, meaning that excited states are not accounted for.

Molecular mechanics are most often applied either in molecular dynamics (MD) simulations or in Monte Carlo simulations. Running an MD simulation of a system lets the atoms and molecules interact with each other for a fixed period of time, resulting in a dynamic view of the system. By solving Newton`s equation of motion, trajectories of atoms and molecules can be calculated. Combining the potential energies calculated from the forcefield, and the trajectories, it is possible to simulate any given chemical system with time scales compatible with biological processes [31](Figure 5).

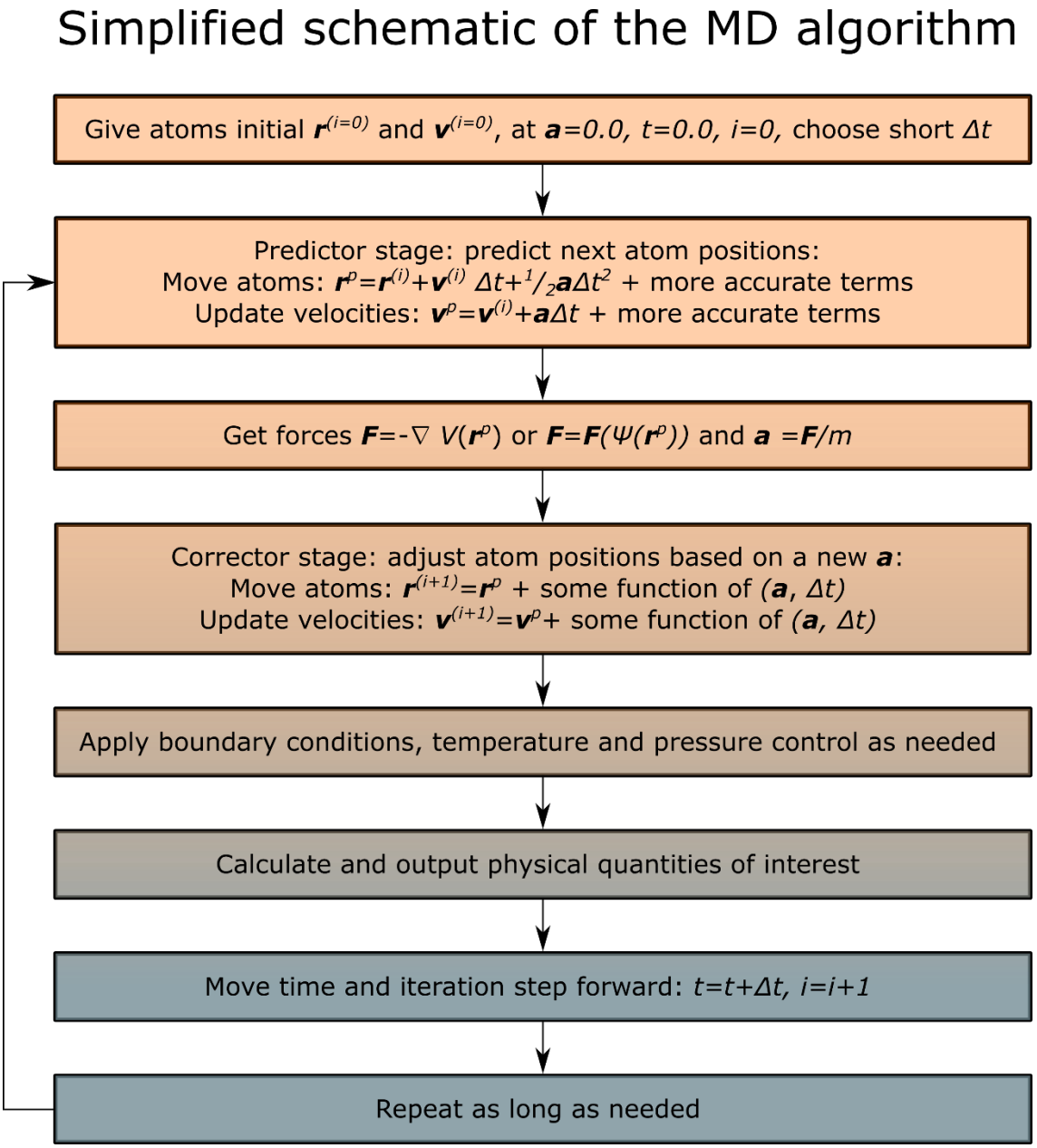

Figure 5. Simple schematic of the MD algorithm.

MD simulations use numerical methods to determine properties of complex systems since an analytical approach is impossible because of the complexity of the system. Because of this, long MD simulations were back in the day ill advised, seeing as they generated cumulative errors in numerical integration [31]. Nowadays progress has been made, making it possible to run simulations that last for several nano seconds without accumulating significant errors [32]. Simulations lasting over a milli second has also been achieved by the supercomputer Anton, which is purpose built for MD simulations [33].

1.2.2 Molecular Docking

In the field of structural based drug design (SBDD), molecular docking is extensively used. SBDD is a core approach to designing drugs, using structural data of molecular targets to improve different aspects related to ligand binding. Molecular docking aims to predict the most viable three-dimensional structure of a given ligand within a target binding site. In addition to giving a 3D structure, quantitative projections of binding energies are also calculated[34]. These calculated binding energies are often called docking score. Molecular docking can be divided into two separate processes, first predicting conformations of the ligand in the binding site, while the second is calculation of docking score for each predicted conformation of the ligand[35, 36].

By altering the ligands structural parameters such as dihedral angles, and translational and rotational degrees of freedom, the conformational space of the ligand can be explored. Prediction of the ligand's conformation are broadly achieved by using two different methods, systematic and stochastic search[37]. Systematic search explores the energy landscape of the conformational space, locating the energy minimum by altering the structural parameters gradually[36]. This is an effective method, but the search often yields a local minimum instead of the global minimum. This can be combatted by running several searches with different starting conformations. On the other hand, stochastic search alters the ligand structure parameters randomly to explore the energy landscape. Because stochastic search generates a wide range of solutions and explores a wider range of conformations, it avoids the local minimum and generates, more often than not, the global minimum.

In addition to predicting optimal ligand binding conformation, molecular docking also calculates binding energy, the so-called docking score. There are mainly three methods used to calculate the binding energy, force-field-based, empirical and knowledge-based functions[38]. The binding energy consists of two parameters, the binding constant and Gibbs free energy. Force-field-based methods uses *ab initio* calculations and classical mechanics to calculate these parameters, while empirical method uses sets of protein ligands complexes with known binding affinities. Knowledge-based method generates a general equation to calculate the energies, from known ligand-receptor complexes[39].

In summary, molecular docking methods predicts conformations of ligands at a reasonable accuracy rapidly, but estimation of desolvation contributions, entropic effects and protein flexibility are poorly handled [40, 41].

1.2.3 Linear Interaction Energy (LIE)

The Linear Interaction Energy method (LIE) is a computational method for calculating ligand-binding affinities[42]. It is an attractive method because of its good compromise between speed and accuracy. When it comes to accuracy, LIE calculations often have a root-mean-squared (RMS) error less than 1 kcal/mol when calculating binding free energies[43]. In contrast, the most common scoring functions, have an RMS error value around 2-2.5 kcal/mol[44].

Ligand binding can be considered as a process where a ligand (l) goes from being solvated in water (f), to bound in a solvated macromolecular target (b). This leads to the fact that for calculating free binding energies, the initial state of solvated ligand must be considered as a reference state. This is the foundation that LIE is built upon (Equation 12).

$$\Delta G_{bind}(l) = \Delta G^{b}_{sol}(l) - \Delta G^{f}_{sol}(l) \qquad \textbf{Equation 12}$$

To be able to calculate the free binding energy using the two states as mentioned earlier, a thermodynamic cycle must be made (Figure 6).

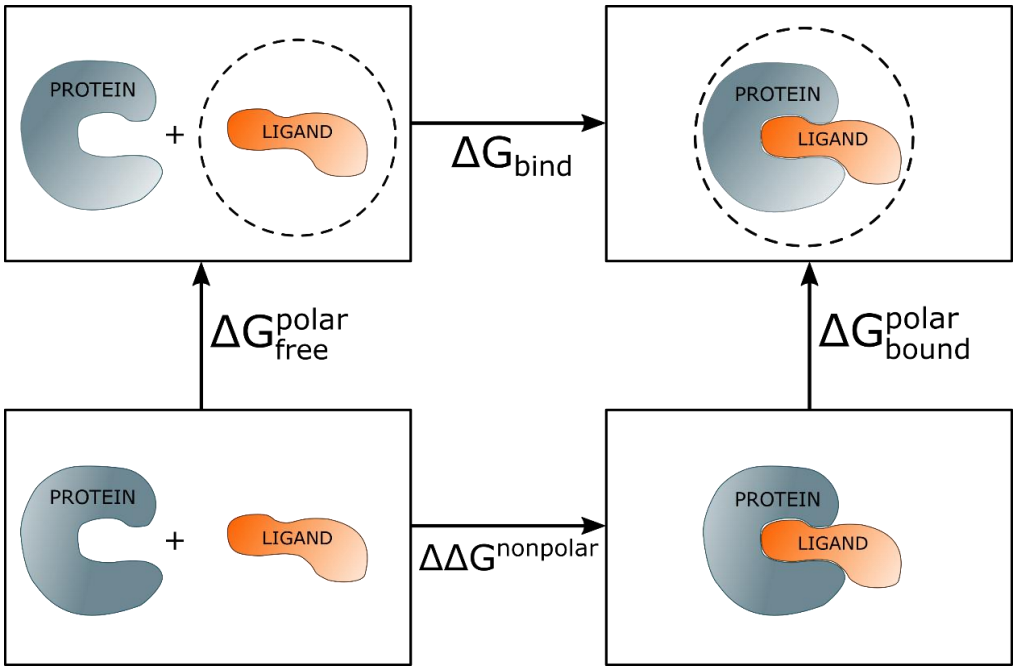

Figure 6. The thermodynamic cycle used to estimate binding free energies with the linear interaction energy method.

The top left and right corners are the unbound and bound states, while the bottom left and right are unphysical states where there are no intermolecular electrostatic interactions. By solving the thermodynamic cycle, the free binding energy can be calculated (Equation 13).

$$\Delta G_{bind} = \left(\Delta G_{bound}^{polar} - \Delta G_{free}^{polar}\right) + \Delta\Delta G_{bind}^{nonpolar}$$
$$= \Delta\Delta G_{bind}^{polar} + \Delta\Delta G_{bind}^{nonpolar}$$

Equation 13

Because the entropic term is implicitly incorporated into the nonpolar term, the binding free energy can be expressed as a sum of the polar and nonpolar interactions. This is very convenient, since molecular forcefields already split the nonbonded potential energy term into electrostatic and nonelectrostatic components. By using linear response theory for electrostatic forces[45], an empirical relationship for the polar contribution can be made (Equation 14).

$$\Delta G_{bind}^{polar} = \beta\left(\langle U_{l-s}^{el}\rangle_b - \langle U_{l-s}^{el}\rangle_f\right) = \beta\Delta\langle U_{l-s}^{el}\rangle$$

Equation 14

where ⟨⟩ are thermodynamic averages of ligand-surrounding, l-s are interaction energies calculated using standard force-field molecular dynamics and β is the polar scaling factor.

Using Equation 14, the nonpolar component of the free energy can be analogously estimated (Equation 15).

$$\Delta G_{bind}^{nonpolar} = \alpha\left(\langle U_{l-s}^{vdW}\rangle_b - \langle U_{l-s}^{vdW}\rangle_f\right) = \alpha\Delta\langle U_{l-s}^{vdW}\rangle + \gamma \qquad \textbf{Equation 15}$$

α is the empirically derived nonpolar scaling factor, and γ is a constant. Combining Equation 14 and Equation 15, the full LIE equation can be written (Equation 16).

$$\Delta G_{bind} = \alpha\Delta\langle U_{l-s}^{vdW}\rangle + \beta\Delta\langle U_{l-s}^{el}\rangle + \gamma \qquad \textbf{Equation 16}$$

The polar scaling factor β was determined by Åqvist and Hansson, on the basis of several FEP calculations done using varying chemical systems[46]. They determined that the most optimal value for the polar scaling factor, β, was 0.5 for any charged compounds. While the nonpolar scaling factor, α, was obtained using an empirical approach by using a set of 18 protein-ligand complexes, which resulted in a scaling factor of 0.18[46]. The γ constant is used as an offset parameter and is not required when calculating relative binding affinities. γ has been proposed to be connected with the hydrophobicity of the binding site[47].

In summary, the LIE method is a computationally light simulation with empirically sound results, making it an attractive method for calculating free binding energies.

3.2.4 Free Energy Perturbation (FEP)

Free energy perturbation, also known as FEP, is one of the most well-known methods for calculating free energy differences from simulations. Another name used for FEP is the Zwanzig relationship [48]. Zwanzig introduced the method in 1954 and theorized by doing an alchemical transformation with two analogous systems, one can find the difference in free energies. An example of such an alchemical transformation would be changing the ligand in a

ligand-binding reaction. The Zwanzig equation is used to calculate the free energy difference going from state A to state B (Equation 17).

$$\Delta G(A \rightarrow B) = G_B - G_A = -k_B T ln \left\langle exp\left[\frac{-(E_B - E_A)}{k_B T}\right]\right\rangle_A \quad \textbf{Equation 17}$$

T is temperature, k_B is the Boltzmann's constant, and $\langle\rangle$ denotes an average calculated over a simulation run for state A. By running a simulation for state A, the energy for state B is calculated using structures from state A. For this to be true, the two compounds must have a reasonable phase space overlap. By making a thermodynamic cycle (Figure 7) involving two different states, as well as a reaction in pure solvent, it is possible to calculate the difference in free energy using Gibs free energy (Equation 18).

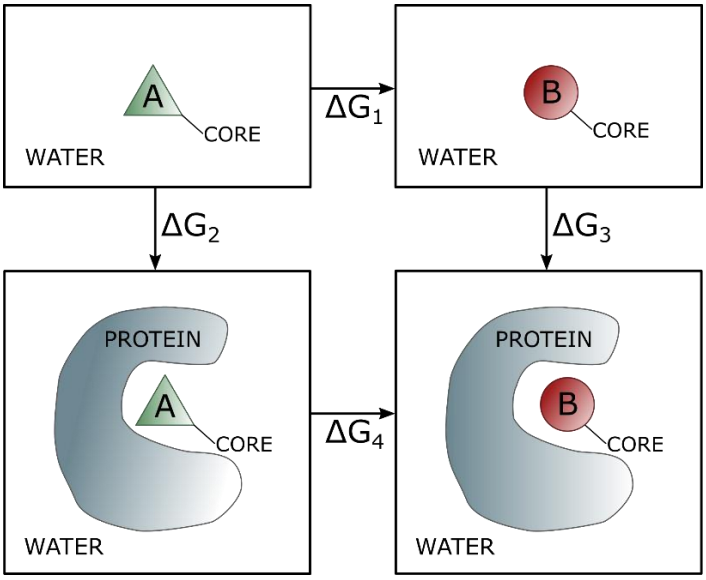

Figure 7. The thermodynamic cycle between two analogue compounds A and B. They both bind to the same protein.

$$\Delta\Delta G = \Delta G_3 - \Delta G_2 = \Delta G_4 - \Delta G_1 \quad \textbf{Equation 18}$$

ΔG_4 and ΔG_1 are calculated using the Zwanzig formula (Equation 17). When doing a FEP simulation, it is common to define a morph parameter, lambda (λ), that introduces an artificial

intermediate state. λ=0 is the initial state A, and λ=1 is the final state B. Utilizing this, the internal energy of a system can be calculated (Equation 19).

$$U_m = (1 - \lambda_m)U_A + \lambda_m U_B = U_A + \lambda_m \Delta U \, , \Delta U = U_B - U_A \qquad \textbf{Equation 19}$$

By summing the energies over the intermediate states, the binding free energy between the two states can be obtained (Equation 20).

$$\Delta G = G_B - G_A = -\beta^{-1} \sum_{m=1}^{n-1} ln \langle exp[-\beta(U_{m+1} - U_m)] \rangle_m \qquad \textbf{Equation 20}$$

1.2.5 Modelling of Ions in Classical Simulations

In biological systems, the presence of metal ions is ubiquitous. Especially for catalytic metal centres, which are found in all different classes of enzymes. They account for 36% of isomerases, 36% of lyases, 39% of hydrolases, 40% of transferases, 44% of oxidoreductases and 59% of ligases [49]. It is therefore important to be able to simulate metal ions in a satisfactory way. Some advice that such simulations are to be carried out using QM [50], while others want to use models that mimic real life more closely (i.e. larger systems). The problem with using a QM or QM/MM approach is the fact that QM is and will be for the foreseeable future too computationally demanding for large biosystems. In addition to computational cost, QM is not a perfect representation of real life either due to its size limitations [51]. There exist mainly three approaches to treat metal ions in a classical forcefield system, these are nonbonded soft-sphere model, the covalent bond approach, and the dummy-model.

Probably the most common of these three is the nonbonded soft-sphere model, which also is the simplest model, where metal-ligand interactions are simply described through electrostatic and van der Waals potentials. This model works fine for simple simulations and alkali and alkaline-earth metal ions, but struggles with more complex systems that involve multinuclear metal centres with metal ions in close proximity to each other. This is because of the model's inability to accurately calculate the solvation energy and coordinate water correctly.

The covalent bond approach suffers from a different kind of problem. That is the fact that because the model include predefined covalent bonds between the ligands and the metal, no ligand exchange can occur as well as no interconversion between different geometries [52].

The third approach, the dummy model, tries to solve the problems inherent in the previous two approaches. By describing the metal centre as a set of cationic dummy atoms, connected around a central atom, a more accurate representation of the metal ions can be achieved (Figure 8). The fact that the dummy atoms can adapt to its surrounding other than that which was originally intended increases its accuracy. There exists several different dummy models, though arguably the most accurate one is the octahedral dummy model that was originally presented by Åqvist and Warshel [53, 54]. Parameters for each dummy model for certain metal ions can vary, though the parameters presented by Duarte in 2014 have shown to be of good quality [55].

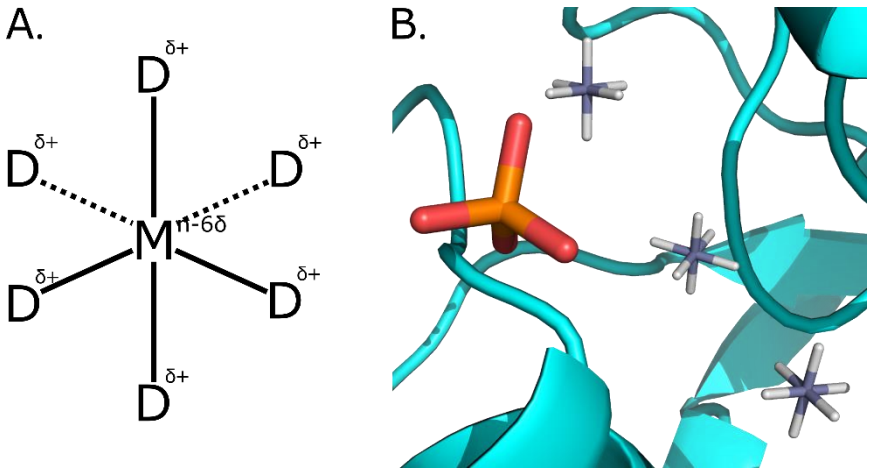

Figure 8. A. Schematic depiction of the dummy model. M is the central metal atom, while D are the dummy atoms, with charge divided between each other. B. Dummy model in use in the Zn^{2+} metllic triad in SAP.

Duarte et al. [55] refined parameters for a set of different metal ions by fitting simulation properties to known experimental values. By changing the van der Waals parameters, mainly the attractive and repulsive factor, and then validating the parameters through simulations done in Q with experimental values, parameters where refined. Since the dummy model uses a limited parameter set, there exists more than one combination of parameters that can reproduce the experimental observations. This is counteracted by iteratively refining the parameters over a range of relevant observables. Their results are among the most accurate parameters for metal ions used in MD simulations as of today.

1.3 Q – Software Package for MD simulations

Several different computer software packages exist for running MD simulations. One such package is Q, developed by Åqvist and his research group [56]. Q has been in constant development since 1998, and the latest release is Q6 [57]. In comparison to other MD simulation software, Q stands out by the fact that from its inception, the focus was on three main methods, those being empirical valence bond (EVB), free energy perturbation (EVB) and linear interaction energy (LIE). Q does not provide any force field, but instead allows users to choose force fields such as CHARMM, AMBER, OPLS and GROMOS. The software consists of one program for dynamic simulations, called qdyn, and several sub programs that handles different tasks. The workflow for Q starts with creation of a topology for a given structure, followed by creation of an input file, ending with the results needed to be processed by external means (Figure 9). Topology generation creates a PDB structural file where the given structure is surrounded by water, in addition to a topology file that contains information that defines which atoms are connected to each other, such as angles, torsions, dihedrals and impropers. The input file describes what kind of MD simulation is going to be run, that being EVB, FEP or LIE.

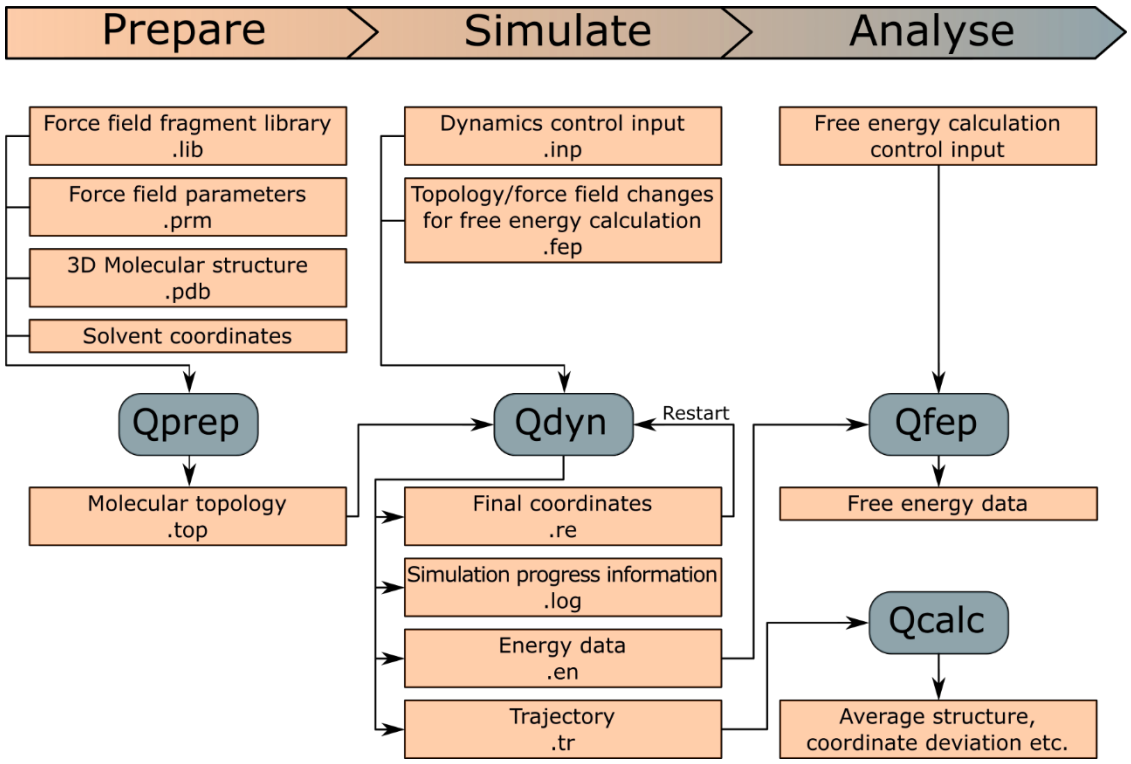

Figure 9. General workflow for the software Q. The orange boxes represent files, while the blue boxes are each of Q's modules. Based on *Manual for the Molecular Dynamics Package Q v 5.06* [58].

Because of the abstract nature of working with Q from a terminal, and the inherent problems of having to produce input files manually, as well as postprocessing the results, a GUI (graphical user interface) has been developed for Q in the group of Prof. Brandsdal, called Qgui [59]. Through Qgui it is possible to set up a workflow that handles creation of forcefield parameters, topology, input files for either MD, LIE, FEP or EVB, in addition to analysing the results from these simulations. Qgui's features can be seen in Figure 10.

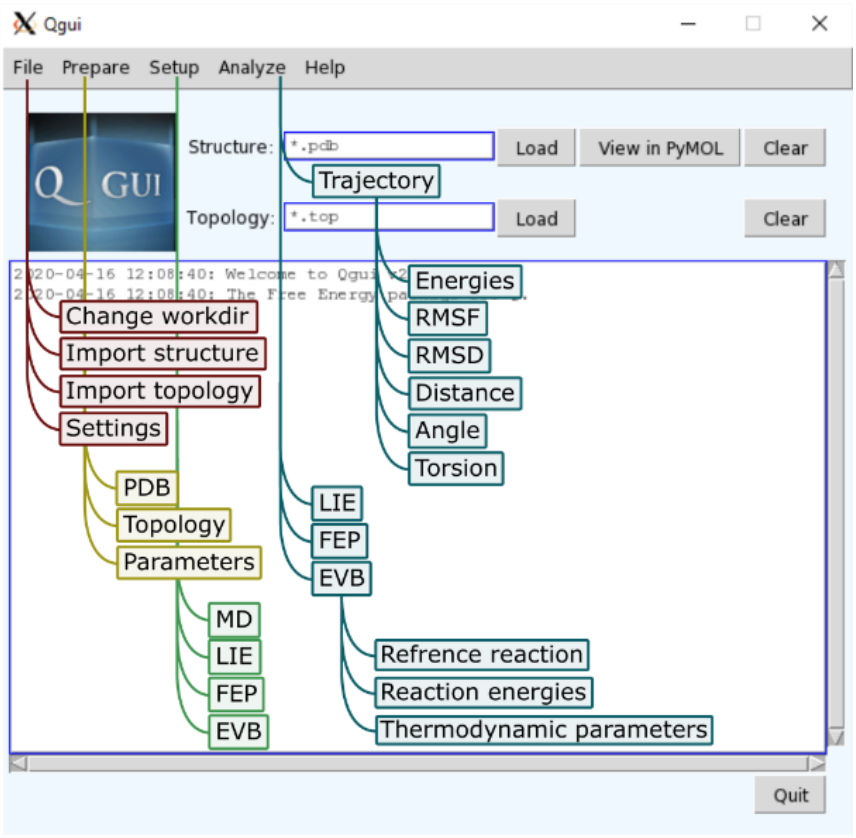

Figure 10. Main window of Qgui, illustrating the content of the drop-down menus. Based on *Qgui: A high-throughput interface for automated setup and analysis of free energy calculations and empirical valence bond simulations in biological systems* [59].

1.4 Bioinformatics

The field of bioinformatics is an interdisciplinary field that combines biology, computer science, information engineering, mathematics and statistics to analyse and understand biological data. The term “bioinformatics” was coined by Paulien Hogeweg and Ben Hesper in the 70s, elevating the field to that of biochemistry [60]. By using computers, bioinformatics aims to solve biological problems *in silico*, from predicting potential proteins in DNA strands, to predicting the functionality of unknown proteins. Also included in bioinformatics is the creation of homology models for proteins. There exist several tools that bioinformatics uses, two such tools that are of great importance is BLAST, for doing sequence queries, and SWISS-MODEL for making homology models.

1.4.1 Basic Local Alignment Search Tool (BLAST)

Basic Local Alignment Search Tool, also known as BLAST, is a bioinformatic tool that compares a subject protein or nucleotide sequence, called a query, with a library or database of sequences [61]. By comparing nucleotide and amino acid sequences, BLAST can identify library sequences that resembles the query sequence above a certain threshold. BLAST has a wide range of uses, ranging from identifying species, locating certain domains in proteins, establishing phylogeny, DNA mapping and comparing gene composition of different species.

1.4.2 Homology Model

In the last decades a huge number of sequences has been determined, thanks to massive developments in high throughput genome sequencing. For a large part of these, the 3D-structure is not yet available. By using a homology model, it is possible to obtain a three-dimensional structure of the protein, by modelling the polypeptide on an already existing solved structure. Studies have shown that the three-dimensional protein structure has a higher level of conservation than one would expect on the basis of sequence conservation alone [62]. The quality of a homology model is dependent on the sequence alignment and quality of the structure’s templates. A popular and well documented tool for making homology models is SWISS-MODEL [63].

1.5 Goal

The goal of this thesis is to learn computational methods such as molecular docking, molecular dynamics simulation, LIE and FEP. In addition, to use these tools to study shrimp alkaline phosphatase (SAP), and potentially increase its affinity towards two common targets within immunoassays, para-Nitrophenylphosphate (PNPP) and 3-(2'-spiroadamantyl)-4-methoxy-4-(3”-phosphoryloxy)-phenyl-1,2-dioxetane (AMPPD). Affinity increasing mutations will be explored, using calf intestine alkaline phosphatase (CIP) as a base. CIP is chosen because of its already proven use in the industry. A homology model of CIP will be made since no crystalized structure of CIP is published

2 Method

This chapter is divided into three sections: Molecular systems, pre-processing and simulations. Computational programs used in this thesis were Maestro [64], PyMol [65], Q5 [58] and Qgui [59]. All figures were made using PyMol [65], ChemDraw [66] and InkScape [67].

2.1 Molecular Systems

The structure of SAP was obtained from the PDB bank with the code 1SHN. This structure was solved in 2004 [7] by De Backer's team. There exist several solved structures of SAP, the reason for choosing 1SHN is the fact it was crystalized with a PO_4^{3-} molecule residing in the active site.

When it comes to the two ligands of interest, para-Nitrophenylphosphate (PNPP) and 3-(2'-spiroadamantyl)-4-methoxy-4-(3''-phosphoryloxy)-phenyl-1,2-dioxetane (AMPPD), both structures were provided by ArcticZymes [68] and built using Maestro's molecule builder.

To increase SAP's affinity towards PNPP and AMPPD, calf intestinal alkaline phosphatase (CIP) was chosen to model after. CIP has seen wide use in the industry and is known for its reliability. However, there exists no crystalized solved structure of CIP. To obtain a structure, a homology model needed to be created. The CIP sequence was obtained using BLAST, by querying SAP's amino acid sequence against *Bos taurus* taxonomy group. After which, a homology model was created in SWISSmodel. Using human phosphatase (1EW2) as a template to model CIP upon, a satisfactory homology model was created.

2.2 Pre-process

2.2.1 Protein Preparation and Docking

Since SAP is a homodimeric alkaline phosphatase with an active site in each monomer, one of the monomers was removed and its accompanying water molecules. The residues N-acetyl-D-glucosamine (NAG) was removed from the sequence, as it was used for crystallization[11]. Using the protein preparation tool in Maestro [69], hydrogens were added, H-bonds were

optimized using PROPKA [70] with a predicted pH of 7.0. Water molecules further than 3Å from heavy atoms were removed, as well as a restrained minimization using OPLS3e.

The ligands PNPP and AMPPD was created using Maestro's molecule builder. Structure optimization of the ligand was done using MacroModel [71] utilizing the OPLS3e force field with water as solvent.

Since SAP was already crystalized with PO_4^{3-} in the active site, there was no need to dock PO_4^{3-}. However, both PNPP and AMPPD needed to be docked to SAP. This was done using GLIDE [72], a docking module present in Maestro. GLIDE uses a grid-based docking method, meaning that a grid mesh needs to be generated for the protein. The GLIDE grid is a search space which contains force-field information surrounding the receptor protein. Using PO_4^{3-} as a focal point, GLIDE made a grid mesh of SAP with a van der Waals radius scaling factor of 1 with a partial charge cut-off of 0.25 as well as metal coordinate constraints, ensuring the ligand would interact with Zn^{2+} in position M1. Both PNPP and AMPPD was then flexibly docked using XP (extra precision) with a van der Waals radii scaling factor of 0.8 and a partial charge cut-off of 0.15. The reason for using XP rather than SP (standard precision) is that XP does more extensive sampling, weeding out false positives.

2.2.2 Mutation Study

As mentioned earlier, a homology model of CIP was created for the purpose of investigating potential mutations. Using PyMol, CIP was superimposed upon SAP. Visually investigating the differences between CIP`s and SAP`s active sites, revealed three promising mutations of SAP. Based on these three possible mutations, three modified structures of SAP were created, each docked with the three ligands, PO_4^{3-}, PNPP and AMPPD. The amino acid mutations of SAP were done using PyMol's mutagenesis wizard. All mutated SAP structures were optimized in Maestro using the protein preparation tool.

2.2.3 Parameterisation of Ligands and Ions

Running MD simulations require the use of a forcefield. For the natural amino acids in the protein, as well as water, OPLS2015 was used. Forcefield parameters for the ligands was generated through Qgui, using the Maestro ffld-server to create OPLS based forcefields. In addition, all ligands had charge groups made manually by hand. Because of the presents of metal ions in SAP, forcefield parameters and library files for Zn^{2+} and Mg^{2+} was created manually using Duarte's parameters [55].

2.2.4 Topology Creation

Topologies were created using Q for each variants of SAP, that is the mutants with the three different ligands, PO_4^{3-}, PNPP and AMPPD. In addition to this, topology of just the ligands alone was also created. Simulation sphere was set to 22Å for both protein complex (SAP + ligand) as well as the ligands. For the complex, SER86 was chosen as centre. Charges was turned on, but because LIE were being used, the charges between complex and ligand needed to be identical. This was solved by turning off the charges on amino acids that were not directly affecting the active site, ASP433, ASP436, ASP182 and ASP173. This resulted in a net charge of minus three for the complexes bound to PO_4^{3-}, and minus two for the complexes bound to PNPP and AMPPD. Water molecules were added using the TIP3P model [73]. The same topologies were used for LIE and FEP simulations.

2.3 Simulation

The LIE and FEP method were used to calculate the free binding energies between SAP and the three different ligands. LIE was utilized for calculating the free binding energies between all different versions of SAP and the three different ligands, as well as calculating free binding energies for SAP where Zn^{2+} in M3 was changed to Mg^{2+}. FEP was used to calculate the difference in free binding energies between SAP and the three ligands, where Zn^{2+} in M3 changed to Mg^{2+}. All simulations were carried out using Q5.

2.3.1 Linear Interaction Energy (LIE) Parameters

The LIE simulations were carried out on four different versions of SAP bound to three different ligands. The four versions of SAP were native SAP, G102R, H149D and $Zn^{2+}3Mg^{2+}$, and as mentioned earlier, the three ligands were PO_4^{3-}, PNPP and AMPPD. For both complexes and ligands, a set of MD simulations were prepared and executed using Q5. For each complex and ligand, an initial heating of the system from 1K to 300K using short simulations of 51ps, followed by 1ns equilibration at 300K was done. Total simulation time of 10ns divided over four simulations with 2 500 000 steps and 1fs stepsize for each of the parallels was prepared. Simulations ran at 300K coupled to a temperature bath with a relaxation time of 10fs. The solvents angles and bonds were constrained using SHAKE [74]. Cut-offs for non-bonded interactions between solute-solute, solvent-solvent and solute-solvent were all set at 10Å. No cut-off was applied for Q-atoms (i.e. ligands). A shell of 18.7Å was used, with a shell force of 10 kcal/(mol·Å^2). The ligand in both complex and ligand simulation was chosen as Q-atoms. In the pure ligand simulation, a restraint was put on the ligand to prevent it from leaving the simulation sphere. When analysing the LIE results, α was set to 0.18, β was set to 0.5 and γ was set to 0. Experimental data is needed to do fitting of LIE's results, therefor α, β and γ remained unaltered.

2.3.2 Free Energy Perturbation (FEP) Parameters

The FEP simulations were carried out on SAP bound to PO_4^{3-}, PNPP and AMPPD, where Zn^{2+} in the M3 metal site was transformed to a Mg^{2+}. A thermodynamic cycle was created, where Zn^{2+} is transformed to Mg^{2+} in solvent, and Zn^{2+} in the M3 metal site in SAP bound to ligand is transmuted to Mg^{2+}. A FEP file was created, denoting the atomic change of Zn^{2+} to Mg^{2+}, reflecting state A and state B. MD simulations were done at 300K coupled to a temperature bath with a relaxation time of 100fs. For each ligand, four parallels were simulated, each containing 51 000 steps, resulting in 0.51ns simulation time for each parallel. The solvents angles and bonds were constrained using SHAKE [74]. Cut-offs for non-bonded interactions between solute-solute, solvent-solvent and solute-solvent were all set at 10Å. No cut-off was used for the Q-atoms. A shell of 18.7Å was used, with a shell force of 10 kcal/(mol·Å^2).

3 Results and discussion

Before any kind of simulations can be conducted, the system in question must be configured. Therefor the first task was to obtain a 3D structure to work with. In this case, the structure of SAP (1SHN) has previously been obtained through X-ray crystallography, which is a common way of determining protein structures. A homology model of CIP was created to search for potential mutations. In addition, these mutations were studied using the LIE and FEP method.

This chapter will start with an investigation of CIP's homology model acquired using BLAST and SWISSmodel. Followed by determination of possible mutations in the active site. Next, results from docking the ligands PO_4^{3-}, PNPP and AMPPD to different mutants of SAP will be discussed. After which binding free energies calculated for each mutant of SAP with the three ligands using the LIE method will be presented and discussed. Finally, free binding energies calculated using the FEP method for the transformation of Zn^{2+} to Mg^{2+} will be investigated.

3.1 Mutations

The sequence identity between SAP and CIP when doing BLAST was 44% (Figure 11 A). This would be enough to create a homology model, since sequence identity as low as 30% yields structures that are comparable to a low-resolution X-ray structures [75]. The amount of identity required for a successful homology model is linked to the alignment length. If the alignment length is over a 100, less than 30% sequence identity is required. However, if the alignment length is as low as 20, a sequence identity over 55% is needed to yield a successful homology model [76].

The homology model of CIP was created using SWISS-MODEL (Figure 11 B) and used human phosphatase (1EW2) as a template instead of SAP. The sequence identity between CIP and 1EW2 were 77.85%, which was why this structure was chosen as a template. Besides, The active site of alkaline phosphates are highly conserved, meaning the largest differences between them would be in the peripheral regions of the protein [77].

A.

intestinal-type alkaline phosphatase [Bos taurus]

Sequence ID: XP_002685648.1 **Length:** 528 **Number of Matches:** 1

See 1 more title(s)

Range 1: 25 to 491 GenPept Graphics ▼ Next Match ▲ Previous

Score	Expect	Method	Identities	Positives	Gaps
363 bits(932)	2e-120	Compositional matrix adjust.	210/480(44%)	285/480(59%)	18/480(3%)

```
Query  1    EEDKAYWNKDAQDALDKQLGIKLREKQAKNVIFFLGDGMSLSTVTAARIYKGGLTGKFER  60
            EED A+WN+ A  ALD    ++  +  AKNVI  FLGDGM +STVTAARI KG + GK
Sbjct  25   EEDPAFWNRQAAQALDVAKKLQPIQTAAKNVILFLGDGMGVSTVTAARILKGQMAGKPGP  84

Query  61   E-KISWEEFDFAALSKTYNTDKQVTDSAASATAYLTGVKTNQGVIGLDANTVRTNCSYQL  119
            E  ++ ++F + ALSKTYN D+ V DSA + TAYL GVKTN   IG+ A      C+
Sbjct  85   ETPLAMDQFPYLALSKTYNVDRDVPDSAGTTTAYLCGVKTNMRTIGVSAAARFDQCNTTR  144

Query  120  DESLFTYSIAHWFQEAGRSTGVVTSTRVTHATPAGTYAHVADRDWENDSDVVHDREDPEI  179
               +   S+ +  ++AG+S GVVT+TRV  A+PAG YAH  +RDW +D+D+  D +
Sbjct  145  GNEV--TSVINRAKKAGKSVGVVTTTRVQDASPAGAYAHTVNRDWFSDADLPPDAQT-YG  201

Query  180  CDDIAEQLVFREPGKNFKVIMGGGRRGFFPEEALDIEDGIP---GEREDGKHLITDWLDD  236
            C DIA QLV+     +  VI+GGGR+  FP    D E   P   G R+D ++L+ +W
Sbjct  202  CLDIATQLVYN---MDIDVILGGGRKYMFPAGTPDPE--YPDDNGVRKDKRNLVQEW--Q  254

Query  237  KASQGATASYVWNRDDLL-AVDIRNTDYLMGLFSYTHLDTVLTRDAEMDPTLPEMTKVAI  295
               QGA   YVWNR +LL A D  N  +LMGLF    +  + RD   DP+L EMT+ A+
Sbjct  255  AKYQGA--QYVWNRTELLKAADDSNVTHLMGLFQPGEMAYEIFRDHTTDPSLEEMTEAAL  312

Query  296  EMLTKDENGFFLLVEGGRIDHMHHANQIRQSLAETLDMEEAVSMALSMTDPEETIILVTA  355
             +L+++  GFFL VEGGRIDH HHAN    +L ET+  + A++ A  +T   +T+ LVTA
Sbjct  313  RVLSRNPRGFFLFVEGGRIDHGHHANTAYWALNETIMFDNAIAKASQLTSEADTLTLVTA  372

Query  356  DHGHTLTITGYADRNTDILDFAGISDLDDRRYTILDYGSGPGYHITEDGKRYEPTEEDLK  415
            DH H  T  GY  R T I   A     D + YT L YG+GPGY + + G R +  E++
Sbjct  373  DHSHVFTFGGYPLRGTSIFGLADGKAKDGKSYTSLLYGNGPGYRL-DVGPRPDVNEKEST  431

Query  416  DINFRYASAAPKHSVTHDGTDVGIWVNGPFAHLFTGVYEENYIPHALAYAACVGTGRTFC  475
            D  ++  +A P  S TH G DV ++  GP+AHL  GV E+ ++ H +A+AACV    T C
Sbjct  432  DPEYQQQAAVPLDSETHAGEDVAVFARGPWAHLVHGVQEQTFVAHVMAFAACVEPYTTDC  491
```

B.

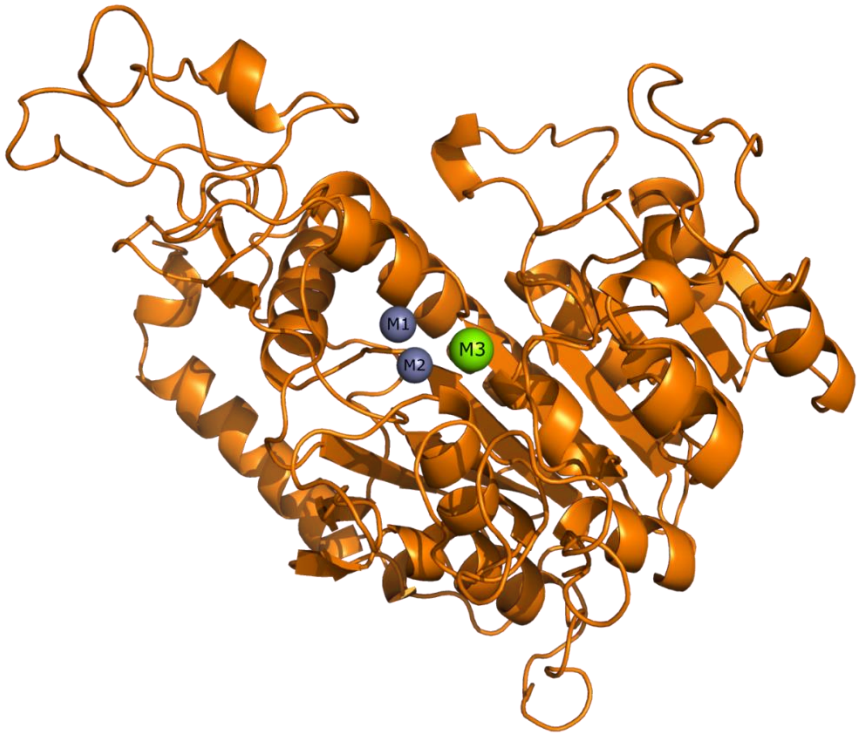

Figure 11. A: CIP's amino acid sequence found using BLAST, by querying SAP`s amino acid sequence against *Bos taurus*. B: Homology model of CIP created using SWISS-MODEL.

Upon further investigation of CIP, superimposing CIP on SAP (Figure 12) yielded an RMSD value of 0.698Å. By visually inspecting the active site of the superimposed proteins, potential mutations were discovered. That being GLY 102 substituted with ARG (SAP G102R), HIS 149 substituted with ASP (SAP H149D) and Zn^{2+} in M3 substituted with Mg^{2+} (SAP $Zn^{2+}3Mg^{2+}$). The arginine in position 102 and asparagine in position 149 could potentially coordinate with the ligand and provide favourable interactions. In addition, the M3 metal site has been proposed to be a regulatory site, where Mg^{2+} is said to increase activity of the protein compared to Zn^{2+} [7].

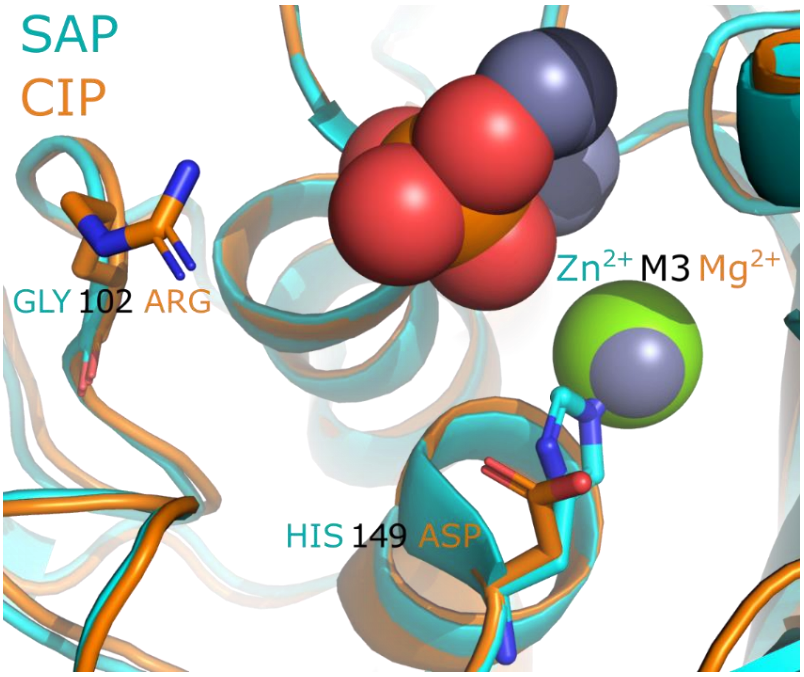

Figure 12. CIP superimposed upon SAP, possible mutations shown to be GLY 102 ARG, HIS 149 ASP and Zn^{2+} M3 Mg^{2+}.

3.2 Docking

Both ligands, PNPP and AMPPD, were docked successfully to SAP and its mutants using XP in GLIDE. Docking scores of SAP and its mutant with PNPP and AMPPD are shown in Table 1.

Table 1. Docking scores (kcal/mol) for SAP, SAP G102R, SAP H149D and SAP $Zn^{2+}3Mg^{2+}$ with ligand PNPP and AMPPD using XP in GLIDE.

Protein	Ligand	Docking Score
SAP	PNPP	-7.5
	AMPPD	-8.9
SAP G102R	PNPP	-8.6
	AMPPD	-8.6
SAP H149D	PNPP	-7.8
	AMPPD	-8.0
SAP $Zn^{2+}3Mg^{2+}$	PNPP	-8.4
	AMPPD	-9.5

The docking conformations (Figure 13) showed that the phosphate group of each ligand docked in a similar fashion as PO_4^{3-}. This was expected as the initial GLIDE grid was modelled upon the already existing PO_4^{3-} bound to the active site. In general, molecular docking suffers from the lack of conformational sampling of the protein, and one alternative strategy is to use MD simulations to incorporate conformational sampling.

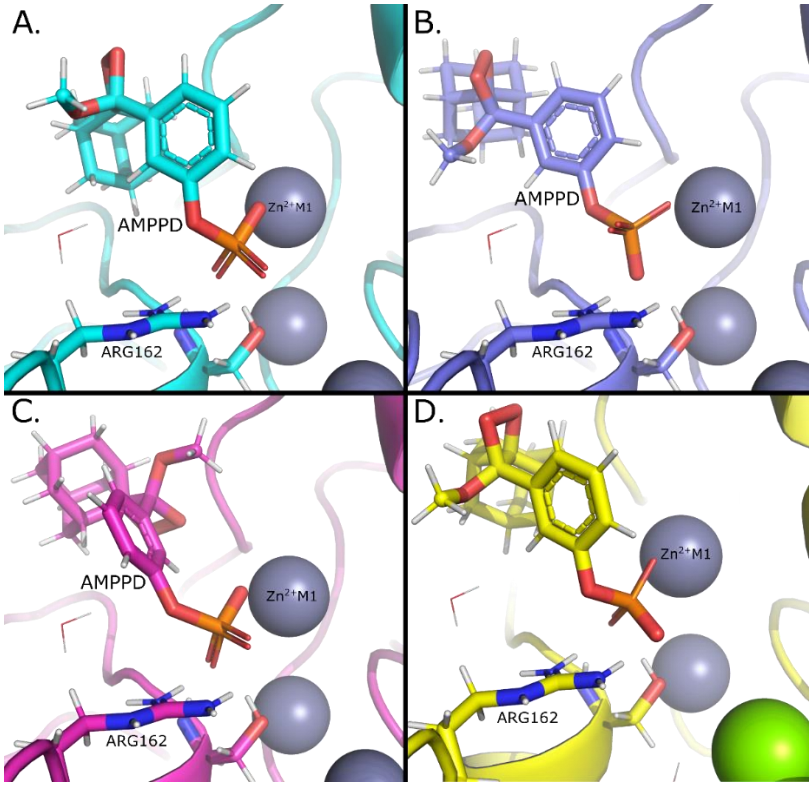

Figure 13. A: AMPPD docked with native SAP. B: AMPPD docked with SAP G102R. C: AMPPD docked with SAP H149D. D: AMPPD docked with SAP $Zn^{2+}3Mg^{2+}$.

3.3 Linear Interaction Energy (LIE)

3.3.1 SAP

The results from the LIE calculations are presented in Table 2. Investigating the binding free energy results from the LIE simulations of native SAP one can clearly see that PO_4^{3-} binds the strongest. This is as expected, as PO_4^{3-} is a natural inhibitor for all alkaline phosphatases.

Table 2. Free binding energies calculated using LIE for SAP binding PO_4^{3-}, PNPP and AMPPD. Change in energy, as well as electrostatic and van der Walls energies are shown.

		SAP		
Ligand	$\Delta G_{kcal/mol}$	**Std.err.**	$\Delta E^{el}_{kcal/mol}$	$\Delta E^{vdw}_{kcal/mol}$
PO_4^{3-}	-87.7	±10.6	-173.1	-6.5
PNPP	-75.2	±5.5	-145.1	-14.5
AMPPD	-76.2	±2.8	-144.3	-22.3

PNPP and AMPPD binds favourably to native SAP, which was expected considering their use in immunoassays. On closer inspection, one can see that the largest contribution towards the binding free energies for all three ligands is the electrostatic interactions. This makes sense as the ligands have a negative charge, and the active site is filled with positively charged zinc ions resulting in strong electrostatic interactions. In addition, ARG162 coordinates with the phosphate group, adding to the electrostatic interactions.

The ligand with the most favourable electrostatic contribution, (PO_4^{3-}), is also the ligand with the smallest van der Waals contribution. This is explained by the fact that PO_4^{3-} has a larger negative charge than PNPP and AMPPD. In addition, PO_4^{3-} has no hydrocarbon chain that can contribute towards van der Waals interactions. While AMPPD, which has the largest hydrocarbon chain, also has the largest van der Waals contribution. This is as expected since the non-polar contribution in LIE is size-dependent. It is important to mention that the binding free energies calculated are too high. In reality they would likely lie between 5kcal/mol and 10kcal/mol. However, when comparing the difference between each ΔG value, it is clear these are comparable values. The difference in ΔG and the difference in docking scores (Table 1) are similar, supporting this claim.

Investigating the electrostatic and van der Waals interactions between PNPP and AMPPD, shows that the binding free energies are almost identical, despite their van der Waals interactions varying highly. A possible reason is that within the LIE equation, the electrostatic interaction is more weighted than the van der Waals interaction. This can be seen by comparing the non-polar and polar scaling factor, α and β respectively. α is 0.18 while β is 0.5 which are the default values [46]. Adding to this, the electrostatic interactions for PNPP and AMPPD are overly large, overshadowing the van der Waals interactions.

Since no experimental data has been published, a good assessment of the accuracy of the simulations is not entirely possible. However, since the structure that is being used originates from a crystal structure, one can calculate the RMSD value of the MD simulations and use it to gauge the structural stability. The RMSD values for native SAP simulations were between 0.40 and 0.75 (Figure 18 A), which means the simulations does not diverge greatly from the crystal structure, implying that they are stable.

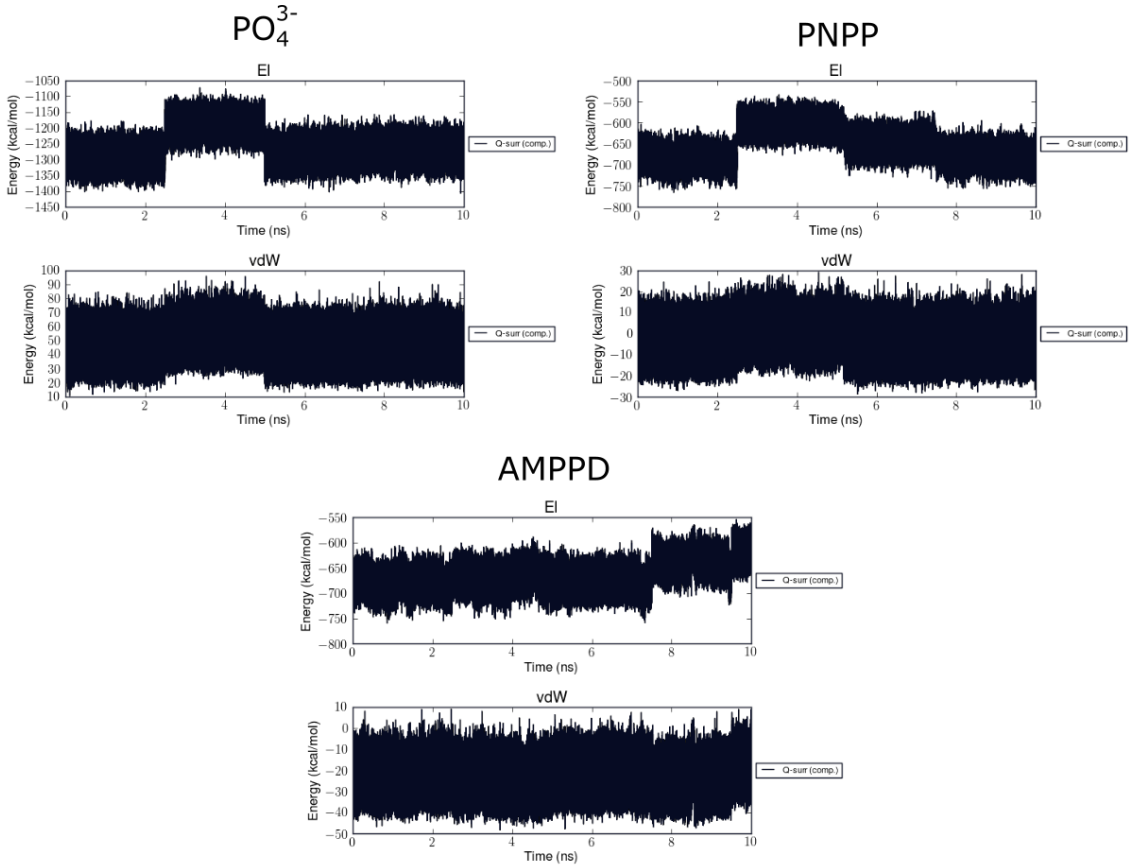

Figure 14 Energy fluctuation over time in MD simulations for SAP bound to PO_4^{3-}, PNPP and AMPPD.

When inspecting the stability of the simulations (Figure 14), despite energy fluctuations are all simulations stable. Each simulation is divided into four segments of 2.5ns, meaning the jumps in energy that are visible are between each parallel, not within a single parallel. This indicates that there is more than one stable conformation for the ligands within SAP. These different conformations are clearly visible in Figure 15. The most stable conformation for the phosphate group is when all four oxygens are coordinated, two with ARG162 and two with zinc's dummy atoms. This is the case for all three of the ligands across all different mutations of SAP.

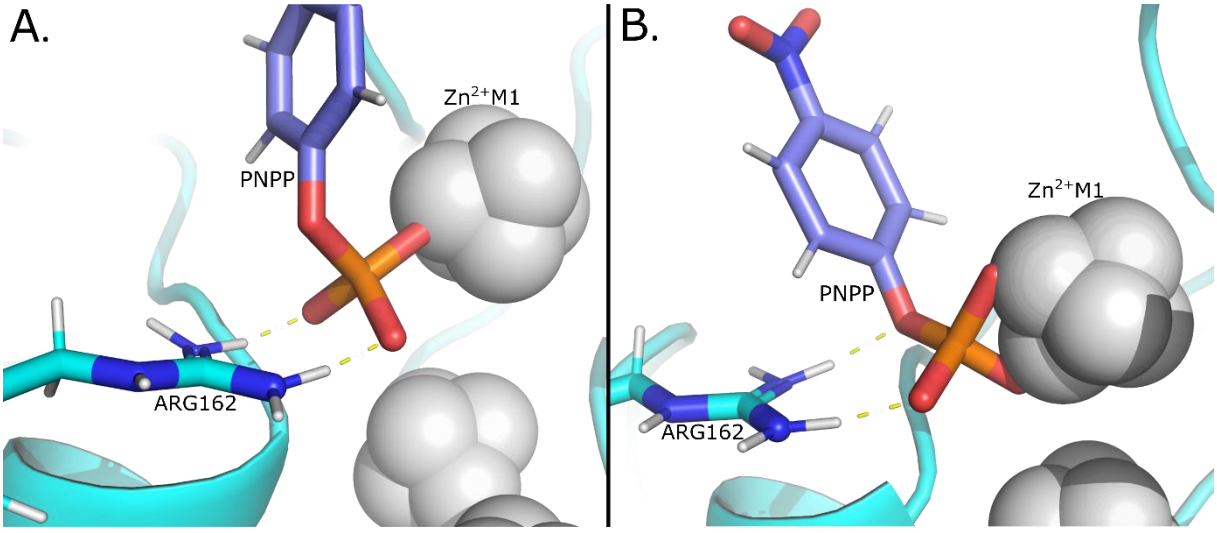

Figure 15 A. PNPP's high energetic conformation in SAP. B. PNPP's low energetic conformation in SAP. Made using PyMol.

3.3.2 SAP G102R

The binding free energy results from LIE simulations for SAP G102R, shows that just as with native SAP, SAP G102R binds PO_4^{3-} most favourably (Table 3).

Table 3. Free binding energies calculated using LIE for SAP G102R binding PO_4^{3-}, PNPP and AMPPD. Change in energy, electrostatic and van der Walls energies, as well as difference between native SAP are shown.

			SAP G102R		
Ligand	$\Delta G_{kcal/mol}$	**Std.err.**	$\Delta E^{el}_{kcal/mol}$	$\Delta E^{vdw}_{kcal/mol}$	$\Delta\Delta G_{SAP}$
PO_4^{3-}	-65.5	±8.9	-130.0	-2.8	22.2
PNPP	-58.9	±8.2	-112.8	-14.0	16.2
AMPPD	-49.2	±9.0	-92.7	-16.2	26.9

SAP G102R does bind PNPP and AMPPD favourably as well. Interestingly enough, PNPP is bound better than AMPPD, which indicates that the G102R mutation changes SAP`s specificity. Compared to native SAP the binding energies are poor, meaning the mutation G102R is potentially a weak one regarding binding affinity. One reason for this could be that arginine is a bulky amino acid compared to glycine, resulting in a cramped active site (Figure 16). Upon closer inspection of the active site with AMPPD, one can see that AMPPD is forced further out of the active site. The amino acids GLN101 and ARG112 are also forced away to make room for ARG102. Since AMPPD is forced further out of the active site, it does not coordinate as well with Zn^{2+} in M1 as it would have otherwise. The guanidine group of ARG102 is bent away from the active site and the phosphate group of the ligand. Because of this, any electrostatic interactions with the ligand that could have arisen from ARG102 positive charge is thereby lost.

AMPPD's van der Waals interactions are also weakened by the fact that many of the non-polar amino acids around AMPPD are forced away by the mutation.

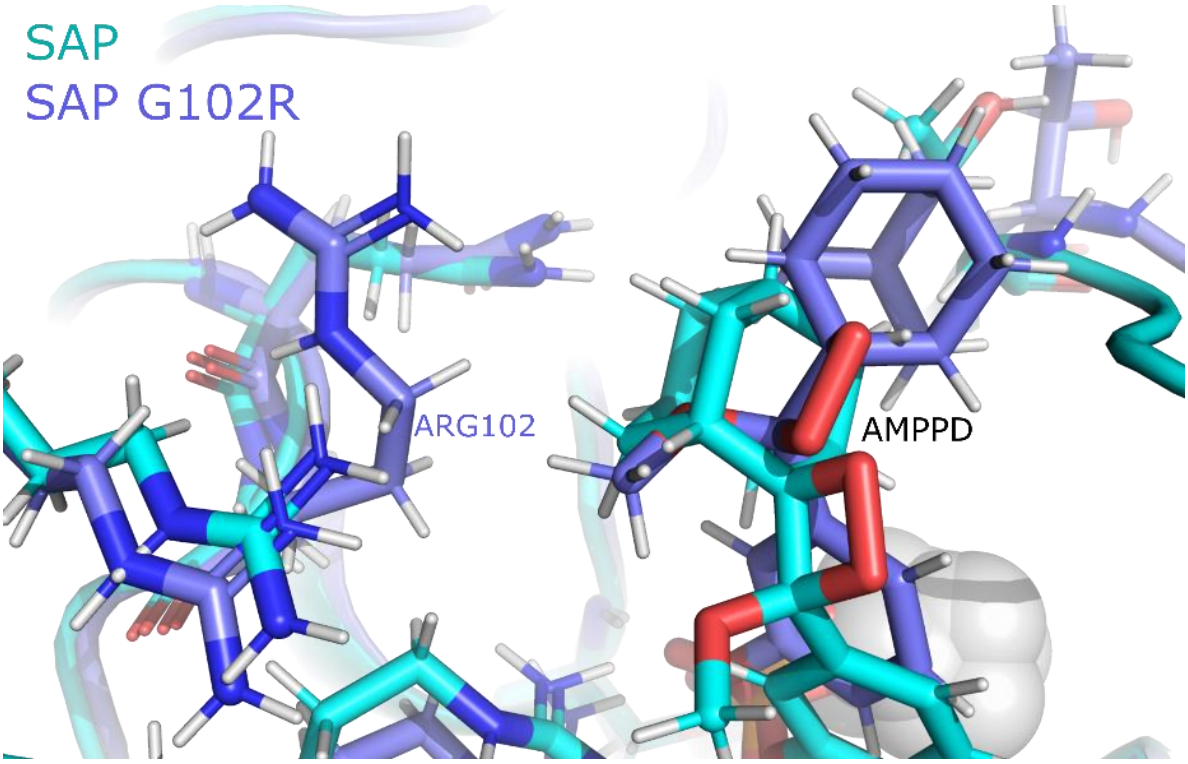

Figure 16. SAP G102R and native SAP superimposed on each other, both bound to AMPPD.

When it comes to PNPP, the van der Waals interactions are almost identical as in native SAP, although the electrostatic interactions are weaker. This might mean that the electrostatic interactions around the phosphate group has been weakened as a result of the mutation. As with AMPPD, PNPP is not as firmly placed in the active site because ARG102 is taking up a large amount of space. This results in PNPP coordinating poorly with Zn^{2+} in M1 and ARG162. However, PNPP is not as bulky as AMPPD, which might be why SAP G102R binds PNPP better than AMPPD.

With PO_4^{3-}, as with AMPPD and PNPP, ARG102 takes up space, forcing amino acids around it to bend away. That said, no direct loss of electrostatic interactions towards PO_4^{3-} can be easily observed. This might lead to the conclusion that the G102R mutation indirectly destabilizes the active site, weakening the electrostatic interactions, resulting in poorer binding affinities for all ligands.

The RMSD values for the simulation lay between 0.50Å and 0.75Å (Figure 18 B), indicating that, as with native SAP, these are somewhat stable simulations. A similar pattern as seen for native SAP can be observed (Figure 17), that being the presence of two stable conformations that each parallel simulation adapt. The more oxygens that are coordinated, the more energetically favourable the conformation is.

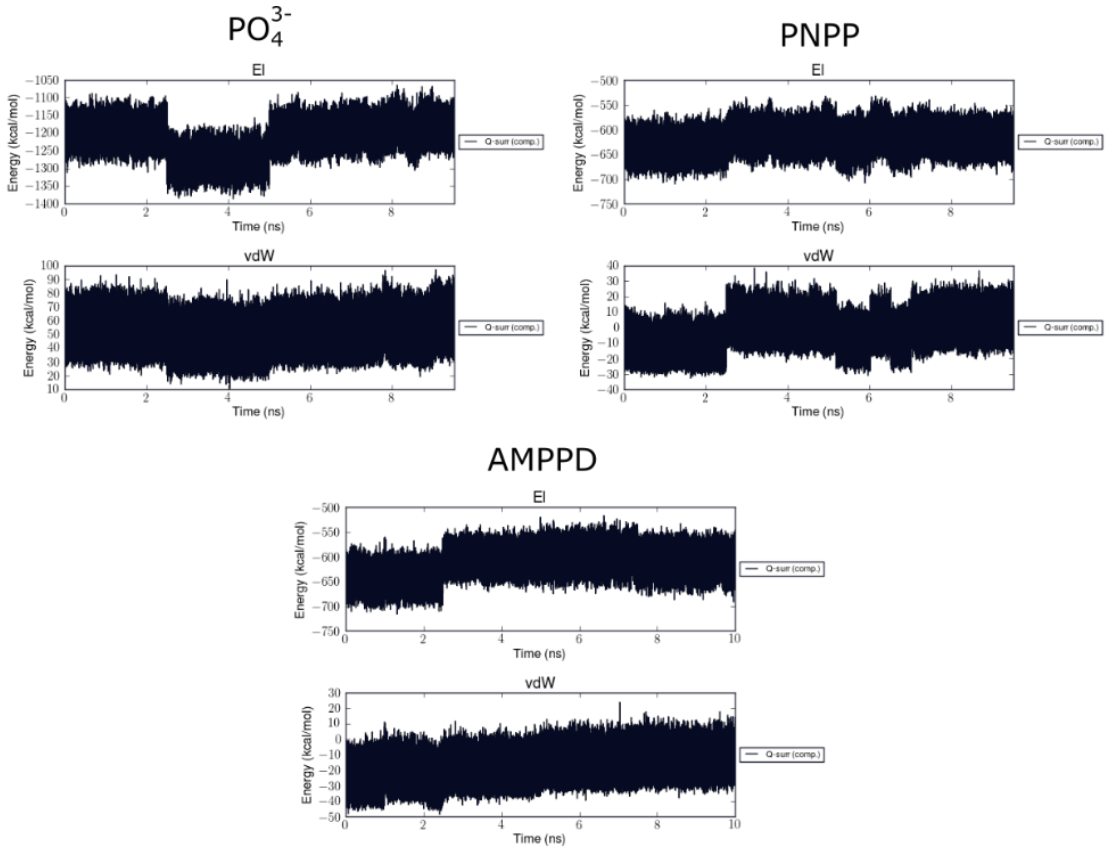

Figure 17. Energy fluctuation over time in MD simulations for SAP G102R bound to PO_4^{3-}, PNPP and AMPPD.

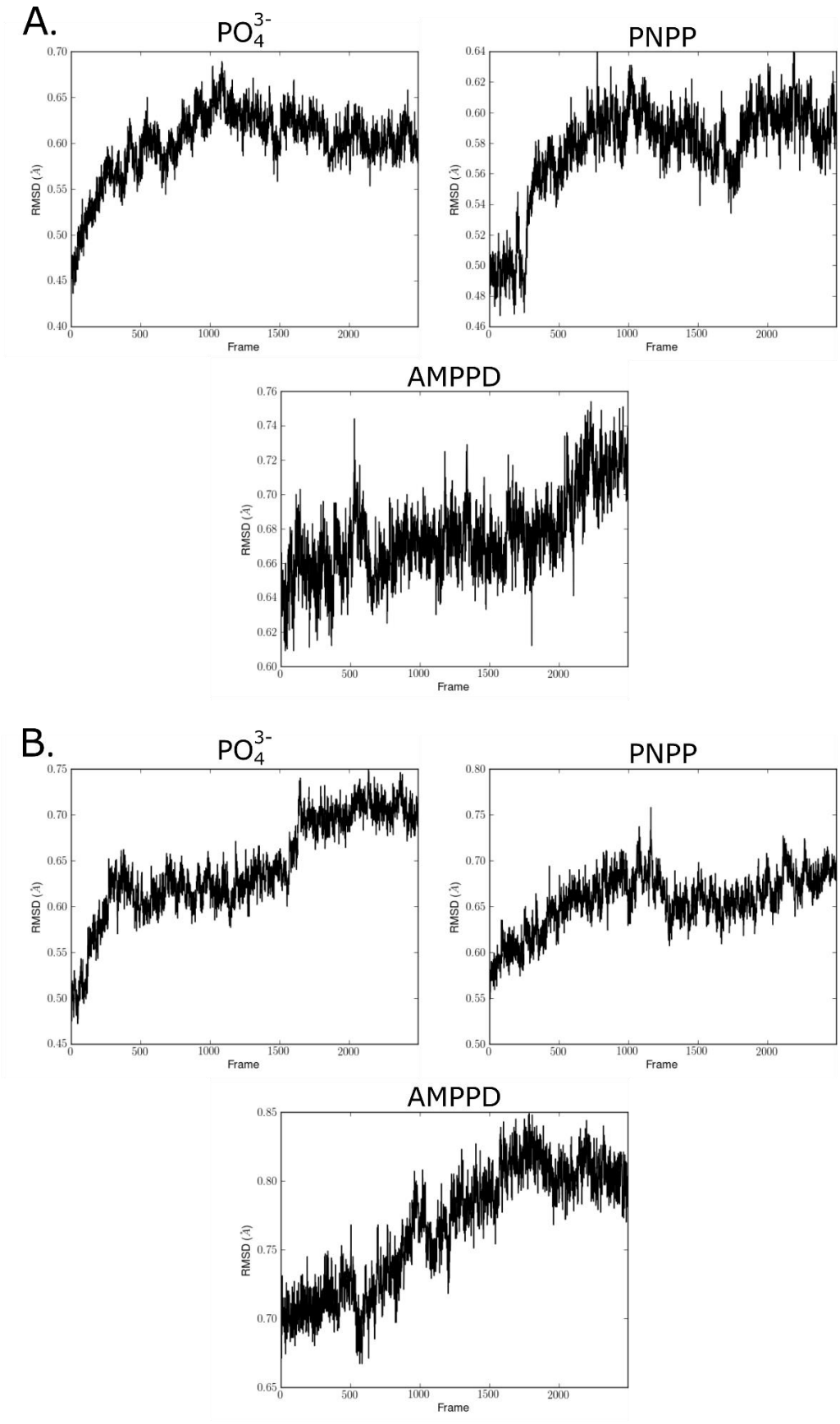

Figure 18. A: RMSD values calculated for the first parallel of each LIE simulations of native SAP bound to PO_4^{3-}, PNPP and AMPPD. B: RMSD values calculated for the first parallel of each LIE simulations of SAP G102R bound to PO_4^{3-}, PNPP and AMPPD.

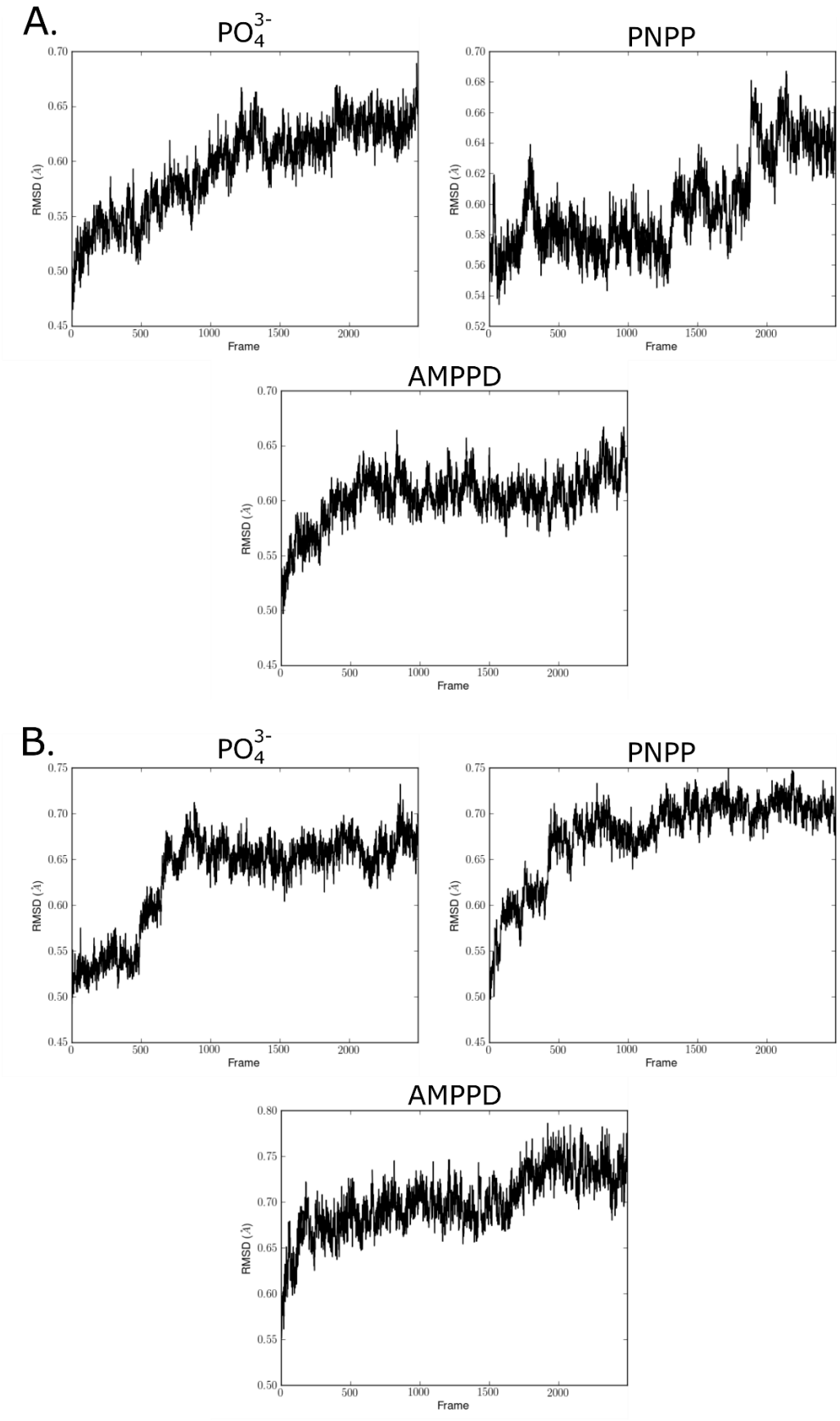

Figure 19. A: RMSD values calculated for the first parallel of each LIE simulations of SAP H149D bound to PO_4^{3-}, PNPP and AMPPD. B: RMSD values calculated for the first parallel of each LIE simulations of SAP $Zn^{2+}3Mg^{2+}$ bound to PO_4^{3-}, PNPP and AMPPD.

3.3.3 SAP H149D

The results from simulations of SAP H149D show that the mutant bound AMPPD better than native SAP (Table 4). The other two ligands, PO_4^{3-} and PNPP both bound worse than in native SAP, with electrostatic interactions being the main difference. For AMPPD, the electrostatic interactions were stronger than in native SAP. This mutation changes SAP's specificity towards PNPP and AMMPD, seemingly favouring AMPPD over PNPP. When inspecting the active site around the H149D mutation for all three ligands, no radical changes to the conformation could be observed, leaving the question, why does AMPPD bind more favourably. It is possible that since HIS149 coordinates with Zn^{2+} in M3, changing it to an ASP results in a more favourable coordination and stabilization of the active site. The question then becomes, why did not the affinity for PO_4^{3-} and PNPP increase as well. A possible solution to this question might lie in the multiple stable conformations of the ligands that are observed in the active site (Figure 20).

Table 4. Free binding energies calculated using LIE for SAP H149DR binding PO_4^{3-}, PNPP and AMPPD. Change in energy, electrostatic and van der Walls energies, as well as difference between native SAP are shown.

SAP H149D

Ligand	$\Delta G_{kcal/mol}$	**Std.err.**	$\Delta E^{el}_{kcal/mol}$	$\Delta E^{vdw}_{kcal/mol}$	$\Delta\Delta G_{SAP}$
PO_4^{3-}	-64.7	±10.3	-106.3	-2.0	23.0
PNPP	-69.9	±3.8	-135.9	-11.0	5.2
AMPPD	-84.7	±3.0	-161.4	-22.3	-8.5

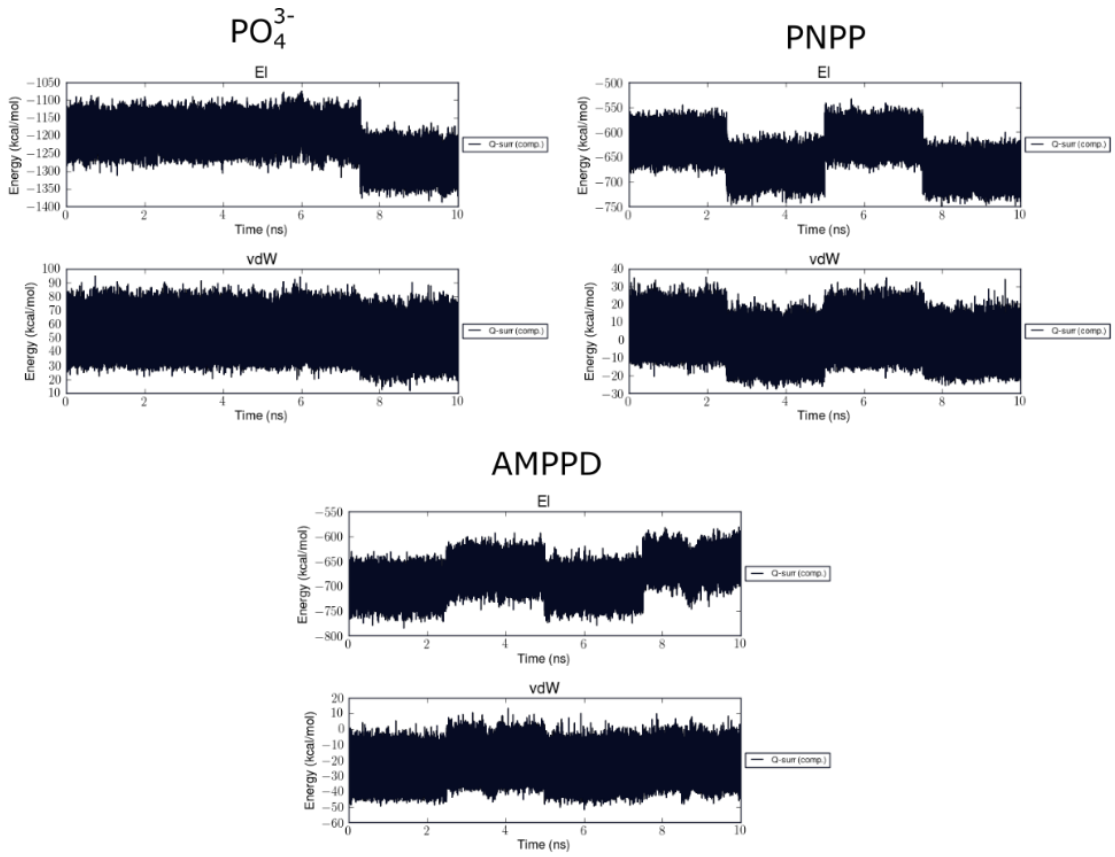

Figure 20. Energy fluctuation over time in MD simulations for SAP H149D bound to PO_4^{3-}, PNPP and AMPPD.

Since all ligands has at least two well stabilized structures when bound to the active site, with highly varying energies, it is possible that the energies that are calculated is not a true average. Because only four parallels of each simulation are run, a true distribution of conformations might not be reached. For instance, when looking at PO_4^{3-}'s electrostatic interactions in all simulations, one can see that when bound to native SAP, three of the four parallels reaches a more favourable energy level. While in SAP G102R and SAP H149D, only one of the parallels reaches the most energetically favourable electrostatic level. This is also true for the van der Waals contribution, especially with PNPP across all the different systems. The parallels end up being distributed among different energetic yet stable conformations. This results in contributions to the free binding energies that are skewed one way or another because of its uneven distribution.

The RMSD values for the simulation lay between 0.50Å and 0.75Å (Figure 19 A). As with the previous two sets of LIE simulations, this indicates that the simulations are structurally stable.

3.3.4 SAP $Zn^{2+}3Mg^{2+}$

Inspecting the last set of LIE simulations, SAP $Zn^{2+}3Mg^{2+}$, reveals an interesting yet plausible result. SAP $Zn^{2+}3Mg^{2+}$ binds all three ligands better than native SAP (Table 5).

Table 5. Free binding energies calculated using LIE for SAP $Zn^{2+}3Mg^{2+}$ binding PO_4^{3-}, PNPP and AMPPD. Change in energy, electrostatic and van der Walls energies, as well as difference between native SAP are shown.

			SAP $Zn^{2+}3Mg^{2+}$		
Ligand	$\Delta G_{kcal/mol}$	**Std.err.**	$\Delta E^{el}_{kcal/mol}$	$\Delta E^{vdw}_{kcal/mol}$	$\Delta\Delta G_{SAP}$
PO_4^{3-}	-105.0	±5.33	-206.9	-8.6	-17.3
PNPP	-89.3	±1.8	-172.6	-16.6	-14.1
AMPPD	-80.9	±8.2	-155.7	-17.1	-4.8

It is postulated that the M3 metal site in alkaline phosphatases is a regulatory site [7]. By exchanging Zn^{2+} with Mg^{2+} the overall catalytic affinity of the protein will increase. The LIE results support this theory as SAP $Zn^{2+}3Mg^{2+}$ binds all three ligands better than native SAP. Changing the metal ion from Zn^{2+} to Mg^{2+} increases the electrostatic interactions among the tested ligands, while maintaining the same van der Waals interactions. A possible reason for the increase in binding affinity may lie in the fact that the simulations that contained Mg^{2+} almost always ended up coordinating the four oxygens in the phosphate group better, resulting in a more stable conformation.

When it comes to the stability of the simulations (Figure 21), it is clear that all three ligands reach a more stable conformation. For the three previous LIE simulations, native SAP, SAP G102R and H149D, the ligands alternate between two stable, yet energetically different conformations. While for this particular mutant, it seems the most favourable conformation is reached the majority of the time.

The RMSD values for the simulation lay between 0.50Å and 0.80Å (Figure 19 B). As with the previous three sets of LIE simulations, this indicates that the simulations are structurally stable.

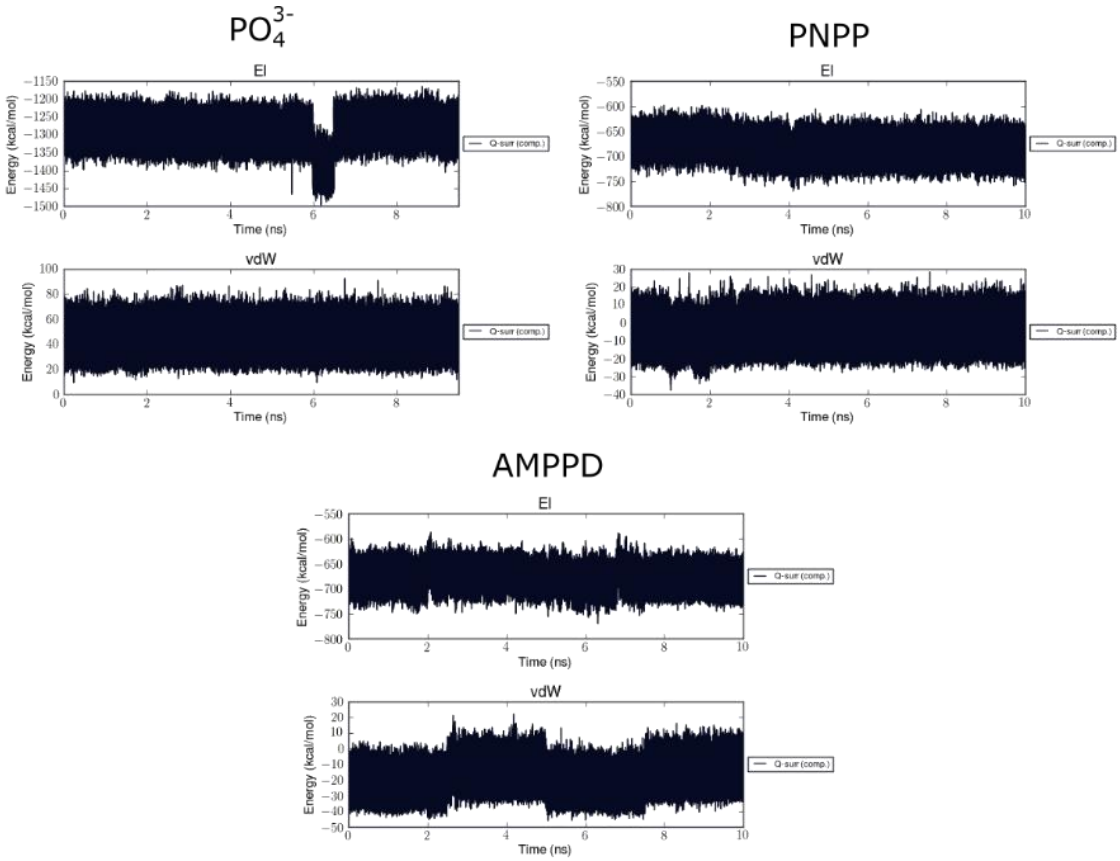

Figure 21. Energy fluctuation over time in MD simulations for SAP $Zn^{2+}3Mg^{2+}$ bound to PO_4^{3-}, PNPP and AMPPD.

3.4 Free Energy Perturbation (FEP)

Investigating the difference in binding free energies for SAP $Zn^{2+}3Mg^{2+}$ with the three ligands PO_4^{3-}, PNPP and AMPPD, calculated using FEP is presented in Table 6. Both the literature [7] and the previous LIE simulations indicate for the $Zn^{2+}3Mg^{2+}$ mutation of SAP an increased affinity. In contrast, the FEP results show the opposite trend. The FEP results indicate that the mutation will substantially decrease the free binding energies.

Table 6. Difference in free binding energies for SAP $Zn^{2+}3Mg^{2+}$ binding PO_4^{3-}, PNPP and AMPPD, calculated using FEP.

	SAP $Zn^{2+}3Mg^{2+}$		
Ligand	$\Delta G_{kcal/mol}$	**Std.err.**	$\Delta\Delta G_{kcal/mol}$
H_2O	48.9	±0.01	X
PO_4^{3-}	59.7	±1.3	10.7
PNPP	61.8	±0.1	12.9
AMPPD	61.0	±0.5	12.0

The solvation energy of Zn^{2+} transformed into Mg^{2+} calculated using the parameters from Duarte [55] are in the same range as published values. The experimental difference should be around 57.3 kcal/mol [78], meaning the solvation free energy difference calculated is within an acceptable range. Since there are indications that the binding free energy difference should be negative, it is plausible that the protein simulations contains structural errors. No obvious explanation was found in the MD simulations. Though when compared to the successful LIE simulations, an interesting discovery was made. A water molecule that was present in the LIE simulation, was not present in the FEP simulations (Figure 22). This water molecule coordinates the metal ion in the M3 site, with two oxygen atoms on the phosphate group of the ligand.

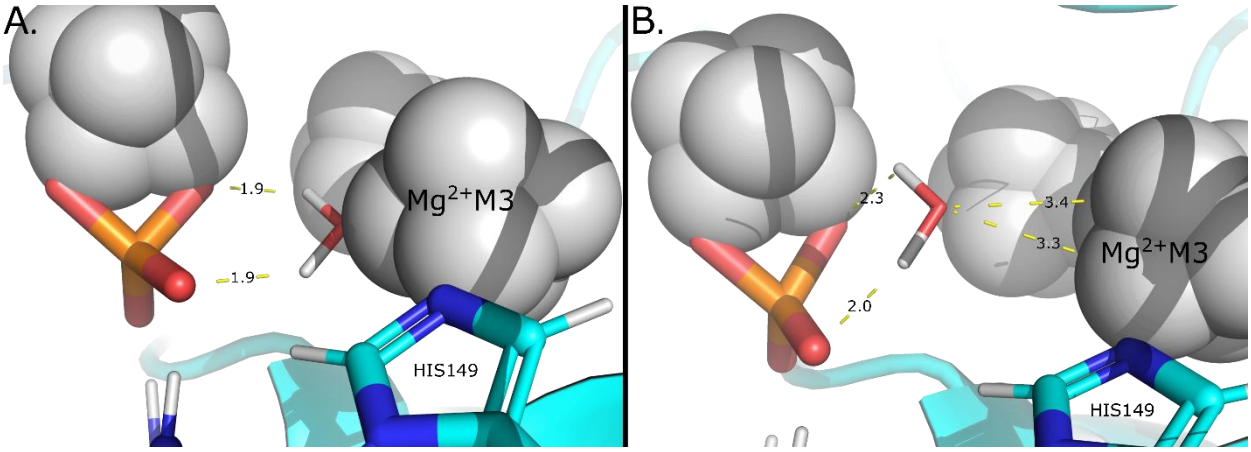

Figure 22. A: LIE simulation of SAP $Zn^{2+}3Mg^{2+}$ bound to PO_4^{3-} with focus on a coordinating water molecule. B: FEP simulation of SAP $Zn^{2+}3Mg^{2+}$ bound PO_4^{3-} with focus on a coordinating water molecule. Made using PyMol.

This water molecule is not present in the FEP simulation, reason for this can be a simple fact that each step in the FEP simulations are too short for this molecule to get into position. And by not getting into position, helping with coordinating the ligand, the stability of the active site is slightly reduced, enough to give a positive ΔΔG.

Although the literature postulates an increased catalytic effect by changing the metal ion in M3 from Zn^{2+} to a Mg^{2+}, does not mean a decrease in affinity is not possible. It is possible to have a situation where the affinity towards certain substrates are decreased, while the catalytic effect overall for the protein is increased. The free binding energies that has been calculated is difficult to comment on the catalytic effect, apart from the fact that the ligands do still bind favourably.

4 Concluding remarks

In this project, the value of bioinformatic tools and computational tools such as MD, LIE and FEP has been assessed. This was achieved by increasing SAP's affinity towards the ligands PNPP and AMPPD. Using BLAST and SWISSmodel, a homology model of CIP was created, that formed the basis for potential mutations in SAP. By superimposing CIP on SAP, several possible mutations were discovered, where a set of three mutations where chosen to investigate further. These were SAP G102R, H149D and SAP $Zn^{2+}3Mg^{2+}$. Binding free energies were calculated for all four versions of SAP, binding three different ligands, PO_4^{3-}, PNPP and AMPPD. For all four versions of SAP, the free binding energy were negative for all ligands, meaning all ligands bound favourably. A general trend in binding free energies were discovered, PO_4^{3-} bound the strongest while AMPPD the weakest, with PNPP right in the middle. This is not surprising as PO_4^{3-} is a natural inhibitor for SAP and alkaline phosphatases in general. SAP H149D bound AMPPD the strongest, and PO_4^{3-} the weakest. Specificity towards PNPP and AMMPD also changed depending on the mutation, where G102R would bind PNPP over AMPPD, while H149D would bind AMPPD over PNPP. Out of the three mutations that were investigated, the most successful one was the SAP $Zn^{2+}3Mg^{2+}$. The differences in binding free energies from native SAP were -17.3kcal/mol for PO_4^{3-}, -14.1kcal/mol for PNPP and -4.8 kcal/mol for AMPPD. This strengthens the theory that the M3 metal site in alkaline phosphates is a regulatory site, where Mg^{2+} increases the catalytic effect compared to Zn^{2+}. Although, LIE only calculates the binding free energies, which is not a direct compression to catalytic effect, one can suggest that doing a metal exchange on SAP will increase affinity towards PNPP and AMPPD.

The two other mutants, SAP G102R and SAP H149D, showed to decrease the affinity for all ligands, except SAP H149D which bound AMPPD better than native SAP. Several possible solutions to this have been discussed, one being the overall decrease of electrostatic interactions in the active site. The other is related to the two distinct stable conformations that the ligands ends up in. For each ligand, two stable, yet energetically different conformations are seen in all the parallel LIE simulations. This is a result from the number of oxygen atoms in the phosphate group that are coordinated in the active site. SAP H149D had three parallels of AMPPD where the ligands had the most stable conformation, while native SAP only had one parallel with the most stable conformation. It is possible that this changes

the average so that it seems that H149D binds stronger, when in fact the difference might be negligible. Regarding the G102R mutant, an overall loss inn affinity towards the ligands can be observed. Again, different solutions have been discussed, and perhaps the most reasonable one being the fact that the large arginine displaces the ligands in the active site, as well as the surrounding ligands, decreasing the interactions substantially. Despite an overall decrease in binding affinity towards the ligand PNPP and AMPPD, these mutations could be useful in relation to ligand specificity.

On closer inspection of the LIE results, it is clear that the largest contribution to the free binding energy is the electrostatic interactions. This makes sense since the active site is highly positively charged from all the metal ions, and the ligands are highly negatively charged. However, it is possible that the polar scaling factor β should have been calibrated lower for the metal ions, yielding lower contribution. As the calculations are done now, the van der Waals interaction has almost no impact on the binding free energies.

FEP simulation of SAP was also done, where Zn^{2+} in the M3 site was transmuted to Mg^{2+}. The solvation energy calculated for Zn^{2+} to $Mg^{2+}$was 48.9kcal/mol which coincided with values from the literature. However, when calculating the difference in free binding energies with this transformation, all ligands were positive, meaning this transformation lowered the affinity. A possible solution to this were the fact that a water molecule that coordinated the metal ion in M3 and two oxygens in the phosphate group had not the same position in the FEP simulations as in the LIE simulations. Another solution might be that this metal exchange actually decreases the affinity of SAP towards PO_4^{3-}, PNPP and AMPPD. Since free binding energy calculation does not tell us directly anything regarding the catalytic effect, it is possible to have a situation where the affinity is decreased, while the catalytic effect is increased.

Regardless, through this project, possible changes to SAP has been explored that can see use in the industry, from increased binding affinity, to increased specificity.

4.1 Future work

The overall goal of this project was to increase SAPs affinity towards PNPP and AMPPD. While interesting discoveries were made, primely that the metal ion in the M3 site plays a vital role in determining the proteins function, additional avenues can and should be explored. A larger set of LIE simulations could be run to clarify that the results are indeed valid. Incorporate a possible distance restraint on the phosphate group, preventing it from organizing in two different conformations should also be considered. Dampening the electrostatic interactions within the LIE equation by changing the polar scaling factor β might also be advisable. Several more mutations can be explored, especially mutations that does not directly affect the active site. Doing multiple mutations at the same time might yield interesting results as well. When it comes to FEP, it would go without saying that doing resFEP [79] on all the possible amino acid mutations would be advisable to validate the LIE results. Since there exists a proposed reaction mechanism for alkaline phosphatases, DFT (density field theory) calculations could be carried out, which then could be used in an EVB (empirical valence bond) study. Doing a proper EVB study on the system would tell us if the mutations can indeed increase the catalytical effect. Lastly, VR is a new and exciting tool that is slowly being refined, and in regards to chemistry there already exists promising software[80]. Preliminary work utilizing VR has been carried out, but not part of this project. Initial findings are positive to the degree that looking at proteins in VR can yield a greater understanding of the proteins. For SAPs case, more work is needed to incorporate the protein into the VR simulations. The reason for this is the problems that arise from the metal ions in the active site.

References

1. Nelson, D.L., A.L. Lehninger, and M.M. Cox, *Lehninger Principles of Biochemistry*. 2008: W. H. Freeman.
2. Koshland, D.E., *Application of a Theory of Enzyme Specificity to Protein Synthesis.* Proceedings of the National Academy of Sciences, 1958. **44**(2): p. 98-104.
3. Fischer, E., *Einfluss der Configuration auf die Wirkung der Enzyme. II.* 1894. **27**(3): p. 3479-3483.
4. McDonald, A.G., S. Boyce, and K.F. Tipton, *ExplorEnz: the primary source of the IUBMB enzyme list.* Nucleic Acids Research, 2009. **37**(Database): p. D593-D597.
5. Cohen, P., *The origins of protein phosphorylation.* Nature Cell Biology, 2002. **4**(5): p. E127-E130.
6. Test, A.A.P.L. *ALP (Alkaline Phosphatase Level) Test*. [cited 2020 20.03.2020]; Available from: https://www.healthline.com/health/alp?m=0.
7. De Backer, M.M.E., S. McSweeney, P.F. Lindley, and E. Hough, *Ligand-binding and metal-exchange crystallographic studies on shrimp alkaline phosphatase.* 2004. **60**(9): p. 1555-1561.
8. Coleman, J.E., *Structure and Mechanism of Alkaline Phosphatase.* Annual Review of Biophysics and Biomolecular Structure, 1992. **21**(1): p. 441-483.
9. Stec, B., K.M. Holtz, and E.R. Kantrowitz, *A revised mechanism for the alkaline phosphatase reaction involving three metal ions.* Journal of Molecular Biology, 2000. **299**(5): p. 1303-1311.
10. Siddiqui, K.S. and R. Cavicchioli, *Cold-Adapted Enzymes.* Annual Review of Biochemistry, 2006. **75**(1): p. 403-433.
11. de Backer, M., S. McSweeney, H.B. Rasmussen, B.W. Riise, P. Lindley, and E. Hough, *The 1.9 A crystal structure of heat-labile shrimp alkaline phosphatase.* J Mol Biol, 2002. **318**(5): p. 1265-74.
12. *Shrimp Alkaline Phosphatase (SAP)* 2020 09.04.2020]; Available from: https://www.thermofisher.com/order/catalog/product/783905000UN#/783905000UN.
13. Engvall, E. and P. Perlmann, *Enzyme-Linked Immunosorbent Assay, Elisa.* III. Quantitation of Specific Antibodies by Enzyme-Labeled Anti-Immunoglobulin in Antigen-Coated Tubes, 1972. **109**(1): p. 129-135.
14. Lorenz, U., *Protein tyrosine phosphatase assays.* Current protocols in immunology, 2011. **Chapter 11**: p. Unit-11.7.
15. Sancenon, V., W.H. Goh, A. Sundaram, K.S. Er, N. Johal, S. Mukhina, G. Carr, and S. Dhakshinamoorthy, *Development, validation and quantitative assessment of an enzymatic assay suitable for small molecule screening and profiling: A case-study.* Biomolecular detection and quantification, 2015. **4**: p. 1-9.
16. Yue, L. and Y.J. Liu, *Mechanism of AMPPD Chemiluminescence in a Different Voice.* J Chem Theory Comput, 2013. **9**(5): p. 2300-12.
17. Heitler, W. and F. London, *Wechselwirkung neutraler Atome und höomopolare Bindung nach der Quantenmechanik.* Zeitschrift für Physik, 1927. **44**(6-7): p. 455-472.
18. Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, *Equation of State Calculations by Fast Computing Machines.* The Journal of Chemical Physics, 1953. **21**(6): p. 1087-1092.
19. *Quantum chemistry; the development of ab initio methods in molecular electronic structure theory. (reprint, 1984)*. 2004, Ringgold Inc: Portland. p. n/a.
20. Barnett, M.P., *Mechanized Molecular Calculations---The POLYATOM System.* Reviews of Modern Physics, 1963. **35**(3): p. 571-572.
21. Frisch, M.J., G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, G.A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A.V. Marenich, J. Bloino, B.G. Janesko, R. Gomperts, B. Mennucci, H.P. Hratchian, J.V. Ortiz, A.F. Izmaylov, J.L. Sonnenberg, Williams, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V.G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M.

Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M.J. Bearpark, J.J. Heyd, E.N. Brothers, K.N. Kudin, V.N. Staroverov, T.A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A.P. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, J.M. Millam, M. Klene, C. Adamo, R. Cammi, J.W. Ochterski, R.L. Martin, K. Morokuma, O. Farkas, J.B. Foresman, and D.J. Fox, *Gaussian 16 Rev. C.01*. 2016: Wallingford, CT.

22. *Masthead.* 1980. **1**(1): p. fmi-fmi.
23. Atkins, P.W. and R.S. Friedman, *Molecular Quantum Mechanics*. 2011: OUP Oxford.
24. Vanommeslaeghe, K., O. Guvench, and A.D. MacKerell, Jr., *Molecular mechanics.* Current pharmaceutical design, 2014. **20**(20): p. 3281-3292.
25. Born, M. and R. Oppenheimer, *Zur Quantentheorie der Molekeln.* Annalen der Physik, 1927. **389**(20): p. 457-484.
26. Antila, H.S. and E. Salonen, *Polarizable force fields.* Methods Mol Biol, 2013. **924**: p. 215-41.
27. Leach, A.R., *Molecular modelling : principles and applications*. 2001, Harlow, England; New York: Prentice Hall.
28. Morse, P.M., *Diatomic Molecules According to the Wave Mechanics. II. Vibrational Levels.* Physical Review, 1929. **34**(1): p. 57-64.
29. Jones, J.E., *On the determination of molecular fields. —II. From the equation of state of a gas.* Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 1924. **106**(738): p. 463-477.
30. Jorgensen, W.L. and J. Tirado-Rives, *The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin.* J Am Chem Soc, 1988. **110**(6): p. 1657-66.
31. Gelpi, J., A. Hospital, R. Goñi, and M. Orozco, *Molecular dynamics simulations: advances and applications.* Advances and Applications in Bioinformatics and Chemistry, 2015: p. 37.
32. Lopes, P.E.M., O. Guvench, and A.D. Mackerell, *Current Status of Protein Force Fields for Molecular Dynamics Simulations*. 2015, Springer New York. p. 47-71.
33. Shaw, D.E., P. Maragakis, K. Lindorff-Larsen, S. Piana, R.O. Dror, M.P. Eastwood, J.A. Bank, J.M. Jumper, J.K. Salmon, Y. Shan, and W. Wriggers, *Atomic-Level Characterization of the Structural Dynamics of Proteins.* Science, 2010. **330**(6002): p. 341-346.
34. Meng, X.-Y., H.-X. Zhang, M. Mezei, and M. Cui, *Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery.* Current Computer Aided-Drug Design, 2011. **7**(2): p. 146-157.
35. Kitchen, D.B., H. Decornez, J.R. Furr, and J. Bajorath, *Docking and scoring in virtual screening for drug discovery: methods and applications.* Nature Reviews Drug Discovery, 2004. **3**(11): p. 935-949.
36. Ferreira, L., R. Dos Santos, G. Oliva, and A. Andricopulo, *Molecular Docking and Structure-Based Drug Design Strategies.* Molecules, 2015. **20**(7): p. 13384-13421.
37. Yuriev, E., M. Agostino, and P.A. Ramsland, *Challenges and advances in computational docking: 2009 in review.* Journal of Molecular Recognition, 2011. **24**(2): p. 149-164.
38. Huang, S.Y., S.Z. Grinter, and X. Zou, *Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions.* Phys Chem Chem Phys, 2010. **12**(40): p. 12899-908.
39. Huang, S.-Y. and X. Zou, *An iterative knowledge-based scoring function to predict protein–ligand interactions: I. Derivation of interaction potentials.* 2006. **27**(15): p. 1866-1875.
40. Mysinger, M.M. and B.K. Shoichet, *Rapid context-dependent ligand desolvation in molecular docking.* J Chem Inf Model, 2010. **50**(9): p. 1561-73.
41. Ruvinsky, A.M., *Role of binding entropy in the refinement of protein-ligand docking predictions: analysis based on the use of 11 scoring functions.* J Comput Chem, 2007. **28**(8): p. 1364-72.

42. Åqvist, J., C. Medina, and J.-E. Samuelsson, *A new method for predicting binding affinity in computer-aided drug design.* "Protein Engineering, Design and Selection", 1994. **7**(3): p. 385-391.
43. Aqvist, J. and J. Marelius, *The linear interaction energy method for predicting ligand binding free energies.* Comb Chem High Throughput Screen, 2001. **4**(8): p. 613-26.
44. Wang, R., L. Lai, and S. Wang, *Further development and validation of empirical scoring functions for structure-based binding affinity prediction.* J Comput Aided Mol Des, 2002. **16**(1): p. 11-26.
45. Marcus, R.A., *Chemical and Electrochemical Electron-Transfer Theory.* Annual Review of Physical Chemistry, 1964. **15**(1): p. 155-196.
46. Aqvist, J. and T. Hansson, *On the Validity of Electrostatic Linear Response in Polar Solvents.* The Journal of Physical Chemistry, 1996. **100**.
47. Almlof, M., B.O. Brandsdal, and J. Aqvist, *Binding affinity prediction with different force fields: examination of the linear interaction energy method.* J Comput Chem, 2004. **25**(10): p. 1242-54.
48. Zwanzig, R.W., *High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases.* The Journal of Chemical Physics, 1954. **22**(8): p. 1420-1426.
49. Andreini, C., I. Bertini, G. Cavallaro, G.L. Holliday, and J.M. Thornton, *Metal ions in biological catalysis: from enzyme databases to general principles.* J Biol Inorg Chem, 2008. **13**(8): p. 1205-18.
50. Siegbahn, P.E., *The performance of hybrid DFT for mechanisms involving transition metal complexes in enzymes.* J Biol Inorg Chem, 2006. **11**(6): p. 695-701.
51. Harvey, J.N., *On the accuracy of density functional theory in transition metal chemistry.* 2006. **102**: p. 203.
52. Lin, F. and R. Wang, *Systematic Derivation of AMBER Force Field Parameters Applicable to Zinc-Containing Systems.* J Chem Theory Comput, 2010. **6**(6): p. 1852-70.
53. Aqvist, J. and A. Warshel, *Computer simulation of the initial proton transfer step in human carbonic anhydrase I.* J Mol Biol, 1992. **224**(1): p. 7-14.
54. Aaqvist, J. and A. Warshel, *Free energy relationships in metalloenzyme-catalyzed reactions. Calculations of the effects of metal ion substitutions in staphylococcal nuclease.* Journal of the American Chemical Society, 1990. **112**(8): p. 2860-2868.
55. Duarte, F., P. Bauer, A. Barrozo, B.A. Amrein, M. Purg, J. Åqvist, and S.C.L. Kamerlin, *Force Field Independent Metal Parameters Using a Nonbonded Dummy Model.* The Journal of Physical Chemistry B, 2014. **118**(16): p. 4351-4362.
56. Marelius, J., K. Kolmodin, I. Feierberg, and J. Åqvist, *Q: a molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems.* Journal of Molecular Graphics and Modelling, 1998. **16**(4-6): p. 213-225.
57. Bauer, P., A. Barrozo, M. Purg, B.A. Amrein, M. Esguerra, Philippe, D.T. Major, J. Åqvist, and S.C.L. Kamerlin, *Q6: A comprehensive toolkit for empirical valence bond and related free energy calculations.* SoftwareX, 2018.
58. *Manual for the Molecular Dynamics Package Q v 5.06*. 2004; Available from: http://xray.bmc.uu.se/~aqwww/q/documents/qman5.pdf.
59. Isaksen, G.V., T.A.H. Andberg, J. Åqvist, and B.O. Brandsdal, *Qgui: A high-throughput interface for automated setup and analysis of free energy calculations and empirical valence bond simulations in biological systems.* 2015. **60**: p. 15-23.
60. Hogeweg, P., *The Roots of Bioinformatics in Theoretical Biology.* PLoS Computational Biology, 2011. **7**(3): p. e1002021.
61. Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman, *Basic local alignment search tool.* J Mol Biol, 1990. **215**(3): p. 403-10.
62. Kaczanowski, S. and P. Zielenkiewicz, *Why similar protein sequences encode similar three-dimensional structures?* Theoretical Chemistry Accounts, 2010. **125**(3-6): p. 643-650.

63. Schwede, T., *SWISS-MODEL: an automated protein homology-modeling server.* Nucleic Acids Research, 2003. **31**(13): p. 3381-3385.
64. *Schrödinger Release 2019-3: Maestro, Schrödinger, LLC, New York, NY, 2020.*
65. *The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.*
66. *ChemDraw 19.1, PerkinElmer Informatics.*
67. TheInkscapeProject. *InkScape 1.0*. 2020; Available from: https://inkscape.org/.
68. *ArcticZymes*. 2020; Available from: https://arcticzymes.com/.
69. Sastry, G.M., M. Adzhigirey, T. Day, R. Annabhimoju, and W. Sherman, *Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments.* J Comput Aided Mol Des, 2013. **27**(3): p. 221-34.
70. Olsson, M.H.M., C.R. Søndergaard, M. Rostkowski, and J.H. Jensen, *PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions.* Journal of Chemical Theory and Computation, 2011. **7**(2): p. 525-537.
71. 2019-2, S.R., *MacroModel.* 2019.
72. Friesner, R.A., R.B. Murphy, M.P. Repasky, L.L. Frye, J.R. Greenwood, T.A. Halgren, P.C. Sanschagrin, and D.T. Mainz, *Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes.* 2006. **49**(21): p. 6177-6196.
73. Jorgensen, W.L., 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, 1983. **79**(2): p. 926-935.
74. Ryckaert, J.-P., G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.* Journal of Computational Physics, 1977. **23**(3): p. 327-341.
75. Xiang, Z., *Advances in homology protein structure modeling.* Current protein & peptide science, 2006. **7**(3): p. 217-227.
76. Sander, C. and R. Schneider, *Database of homology-derived protein structures and the structural meaning of sequence alignment.* Proteins: Structure, Function, and Genetics, 1991. **9**(1): p. 56-68.
77. Galperin, M.Y. and M.J. Jedrzejas, *Conserved core structure and active site residues in alkaline phosphatase superfamily enzymes.* Proteins, 2001. **45**(4): p. 318-24.
78. Marcus, Y., *A simple empirical model describing the thermodynamics of hydration of ions of widely varying charges, sizes, and shapes.* Biophysical Chemistry, 1994. **51**(2): p. 111-127.
79. Jespers, W., G.V. Isaksen, T.A.H. Andberg, S. Vasile, A. van Veen, J. Åqvist, B.O. Brandsdal, and H. Gutiérrez-de-Terán, *QresFEP: An Automated Protocol for Free Energy Calculations of Protein Mutations in Q.* Journal of Chemical Theory and Computation, 2019. **15**(10): p. 5461-5473.
80. O'Connor, M.B., S.J. Bennie, H.M. Deeks, A. Jamieson-Binnie, A.J. Jones, R.J. Shannon, R. Walters, T.J. Mitchell, A.J. Mulholland, and D.R. Glowacki, *Interactive molecular dynamics in virtual reality from quantum chemistry to drug binding: An open-source multi-person framework.* The Journal of Chemical Physics, 2019. **150**(22): p. 220901.