UiT
THE ARCTIC
UNIVERSITY
OF NORWAY

Faculty of Science and Technology

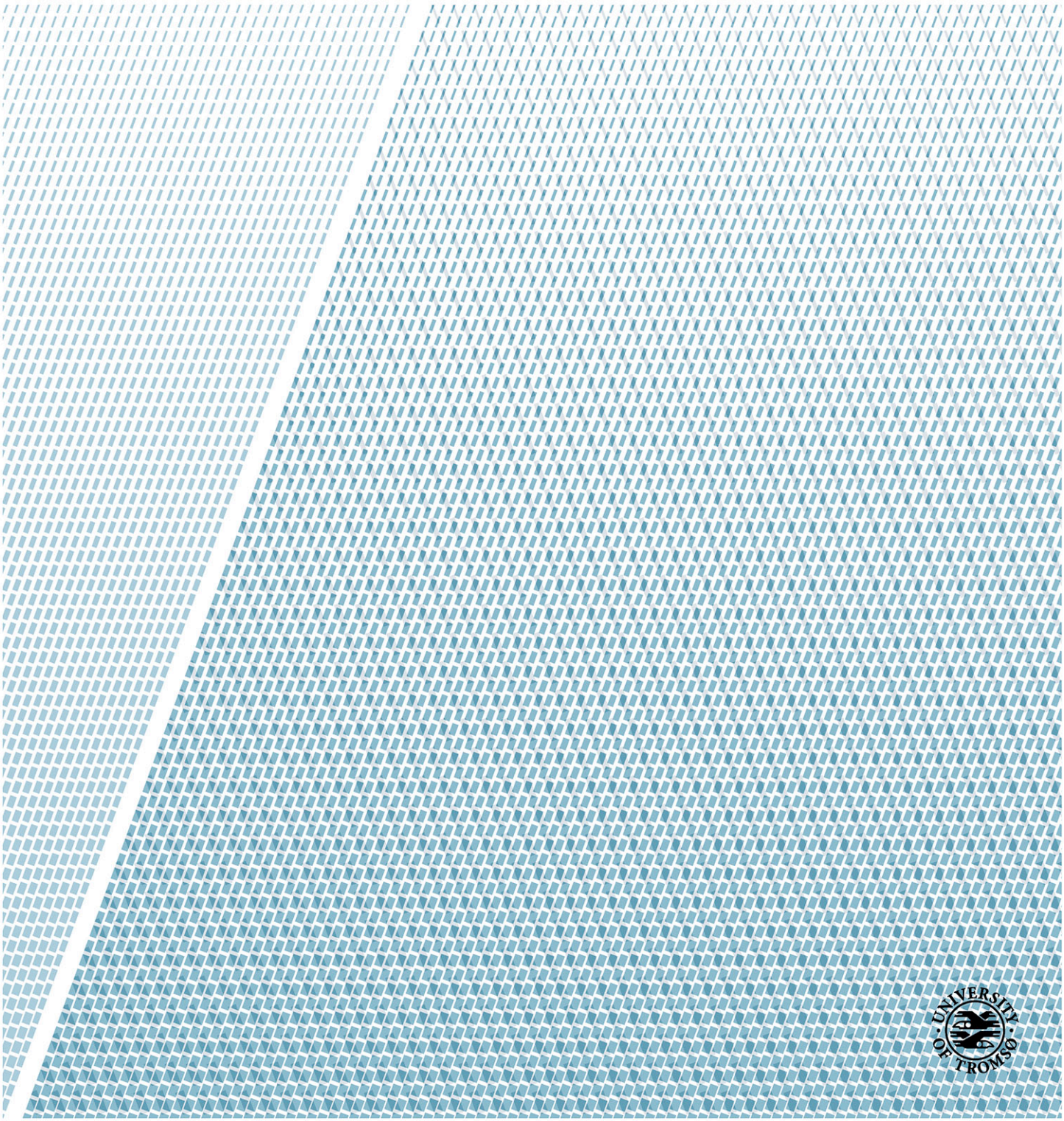
Department of Chemistry

Target characterization and ligand design directed at deoxyuridine 5'-triphospate nucleotidohydrolase

—

Laura Elina Liikanen

KJE-3900 Master's Thesis in Chemistry, July 2014



Abstract

dUTPase is an important enzyme in DNA metabolism. It can be thought of as a gate keeper for apoptosis and is therefore an attractive target when trying to kill malign cells which cause disease in the human body. dUTPase has been found to be an important drug target in diseases such as cancer and malaria, to mention two. Due to problems with drug resistance in existing treatments, the search for new and more efficient inhibitors against dUTPase is very relevant in present-day drug design.

Computational methods play an important role in the development of novel inhibitors and can reduce the time and cost of the drug design process. These methods are applied to sample the binding modes of ligands to the receptor. To make the search base for ligands smaller, one must be able to rank the ligands with respect to the binding affinities. However, in order to integrate computational methods into the drug design process, suitable computational tools are needed, that can efficiently carry out and conduct ligand binding free energy calculations. As a part of this work, a graphical user interface for the simulation tools was developed.

Molecular dynamics simulations and docking have been combined with the free energy calculations to predict the binding affinity for inhibitors against the human dUTPase using Qgui. A LIE model was first constructed to reproduce the experimental binding free energies for a training set of ligands. Judged by the coefficient of determination (R^2), good agreement between the experimental and theoretical binding affinities was obtained. The model was tested for selectivity with a set of ligands targeted at *Plasmodium falsiparum* dUTPase showing the pivotal importance of good starting structure. The binding free energies of two novel ligands were measured with the LIE model and the calculations resulted in moderate binding affinities.

Acknowledgements

I would like to express my very great appreciation to my supervisor Professor Bjørn Olav Brandsdal for providing invaluable guidance throughout this thesis.

I would also like to thank PhD student Geir Isaksen for the technical support and troubleshooting my system. I will also use this opportunity to thank scientist Johan Isaksson for providing valuable advice.

My thanks goes to the fellow students which I shared office with.

Last, but not the least, a special thanks go to my fiancé Martin Ernstsen for believing in me and supporting my choices in life.

Contents

List of Figures

List of Tables

List of Abbreviations

*Pf*dUTPase	*Plasmodium falciparum* dUTPase.
5-FU	5-fluorouracil.
COD	coefficient of determination.
dTMP	deoxythymine monophosphate.
dTTP	deoxythymine triphosphate.
dUMP	deoxyuridine monophosphate.
dUTP	deoxyuridine triphosphate.
dUTPase	deoxyuridine triphosphate nucleotidohydrolase.
EL	electrostatic.
EVB	empirical valence bond.
FEP	free energy perturbation.
GUI	graphical user interface.
LIE	linear interaction energy.
MC	monte carlo.
MD	molecular dynamics.
MM-PBSA	molecular mechanics/Poisson-Boltzmann/surface area.
pdb	Protein Data Bank.

Qgui	Q graphical user interface.
RMSD	root mean square deviation.
SSE	sum of squared errors.
TS	thymidylate synthase.
vdW	van der Waals.

Chapter 1

Introduction

This study is a part of a larger drug design project at the University of Tromsø targeted at deoxyuridine triphosphate nucleotidohydrolase (dUTPase) enzyme . The project aims at increasing the knowledge of the dUTPase characteristics especially its role in malaria, tuberculosis and cancer, and also to develop novel inhibitors.

1.1 dUTPase

1.1.1 Human dUTPase

dUTPase is an important enzyme in DNA metabolism. The enzyme hydrolyzes deoxyuridine triphosphate (dUTP) to deoxyuridine monophosphate (dUMP) as shown in Equation 1.1.

$$dUTP + H_2O \rightarrow dUMP + diphosphate \tag{1.1}$$

dUTPase keeps the cellular dUTP:dTTP ratio low, which is essential in maintaining DNA integrity [1]. Since most DNA polymerases cannot distinguish between dUTP and deoxythymine triphosphate (dTTP) a low ratio of dUTP:dTTP is crucial. If uracil is incorporated into DNA, it is subjected to uracil-excision repair mechanism. Too high rate of uracil integration in the DNA can lead to DNA fragmentation and eventually cell death [2], and

the availability of dUTP as a DNA precursor must therefore be restricted. dUTPase has a dual role in the cell since it also provides dUMP, a dTTP precursor [2]. Recently dUTPases has been linked to other cellular processes like the transfer of mobile genetic elements, autoimmunity or apoptosis and the regulation of the immune system [3]. These findings suggest that dUTPases are involved in many cell regulatory functions.

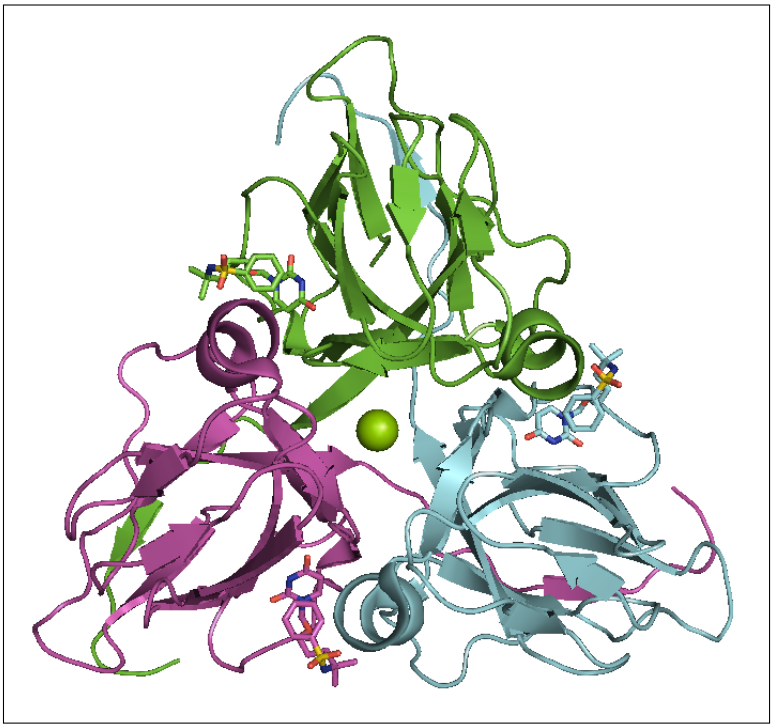


Figure 1.1: Crystal structure of the human dUTPase co-crystallized with an inhibitor ligand (pdb:3ARN). The figure illustrates the water channel through the trimer, the Mg^{2+} ion in the channel and three inhibitors bound at the active sites.

dUTPases are divided into monomeric, homodimeric and homotrimeric enzymes. The human dUTPase biological unit is composed of three identical subunits which together form a homotrimer (Figure 1.1). The first detailed crystal structure of human dUTPase was published in 1996 by Mol *et al* [1]. According to this structure each subunit in the dUTPase trimer is made of an eight-stranded jelly-roll β barrel. Each subunit has a long floppy C-terminal β strand that is organized at the surface of a neighboring subunits active site. There are three active sites in each biological unit of dUTPase. The C-terminal β strand is incorporated to the neighboring surface by main chain interaction with the N-terminal β strand. The C-terminal strand is close to

the active site (formed between subunits) and makes contact with the bound substrate at the active site. The substrate interacts with all three subunits. In human dUTPase a Mg^{2+} ion is located in a water tunnel formed between the three subunits [1, 4].

dUTPase has a high specificity to uracil and does not hydrolyze other similar compounds. This is achieved by a β hairpin motif in two ways [1]. Firstly, the binding site provides very effective steric hindrance against purines, thymine, ribose, and secondly the hydrogen bonding pattern in the active site suites the uracil ring well. The hydrogen bonding to uracil is caused by main chain atoms, which makes the main chain folding important for uracil recognition [2]. When dUTPase binds dUTP a Mg^{2+} ion plays an important role of coordinating the phosphates chain of dUTP for a nucleophilic attack from a water molecule [5].

DNA and RNA have three bases, adenine, guanine and cytosine, which are incorporated in both. In addition, thymine and uracil are only found in DNA and RNA, respectively. However, uracil can be incorporated into DNA in two ways: either through DNA polymerase, which will lead to U:A base pairing, or by DNA cytosine deamination, which in turn will lead to U:G mismatch base pairing [2]. Both cases of uracil incorporation will be detected by the DNA repair system and excised. It is believed that uracil is discriminated from DNA because it would be difficult to tell the difference between uracil mutated from cytosine or an uracil that originates from DNA polymerase activity [2]. Uracil is capable of base pairing with all of the bases found in DNA and RNA, including itself. The methyl group of thymine makes it more hydrophobic than uracil and therefore the placement of thymine will be more restricted than the placement of uracil, since the major part of the DNA is hydrophilic. The methyl group of thymine make the DNA strand more stable by protecting it from environmental threats. While uracil incorporation is generally prevented by the DNA repair mechanism, recent studies show that this is not always the case [6, 7, 8]. One such example is in the chromosome of B-cells, where uracil is tolerated, thereby promoting diversity in antibodies [7].

The dUTPase gene *dut* is found everywhere in eukaryotes, eubacteria and archaea [9]. It is also present in some retroviruses and DNA viruses [9]. The human *dut* gene encodes for two dUTPase isoenzymes which are generated by alternative promoters [10]. One of the isoenzymes is localized in the nucleus and one in mitochondria. The nuclear isoform concentration is controlled by the cell cycle. The mitochondrial isoform concentration is constitutive and

not linked to the stage of the cell cycle [10].

1.1.2 Human dUTPases Role in Cancer

Human dUTPase has been suggested to protect cancer cells from elevated dUTP levels and anti-cancer drugs [11, 12]. Therefore dUTPase is a drug target in cancer treatment [13]. If dUTPase is inhibited the levels of dUTP will increase and promote uracil incorporation in DNA. In addition, since the dUMP levels are decreased, thymidylate synthase (TS) lacks its substrate - dUMP - to make deoxythymine monophosphate (dTMP), which in turn is the precursor of thymine.

TS catalyses the dUMP conversion to dTMP. This reaction produces *de novo* source of thymidylate for DNA replication and repair [14]. Inhibition of TS will disturb the reaction and can cause an inbalance in the deoxynucleotide pool. This is thought to disturb the DNA synthesis and repair mechanism and can lead to serious DNA damage due to lack of dTMP [14]. TS inhibition also leads to accumulation of dUMP which in turn can lead to elevated dUTP levels [15, 16]. The cancer drug 5-fluorouracil (5-FU) inhibits TS in cancer cells to induce cell death [14]. Unfortunately 5-FU treatment is subjected to drug resistance in a large percent of tumors [17]. Inhibiting the dUTPase in combination with 5-FU treatment has been shown to decrease drug resistance [18].

1.1.3 *Plasmodium falciparum* dUTPase and Malaria

Plasmodium falciparum is a parasite which causes malaria. It is carried by a mosquito which infects humans with the parasite. Malaria is the cause of death of estimated 627 000 people in 2012 in sub-Saharan Africa [19]. There are drugs developed for malaria, but unfortunately *Plasmodium falciparum* has evolved resistance to common malaria chemotherapy treatments [20]. The inhibition of *Plasmodium falciparum* dUTPase (*Pf*dUTPase) is a promising anti-malaria drug target [21]. *Pf*dUTPase has a low sequence similarity with the human dUTPase (28% identity), which makes *Pf*dUTPase a suitable drug target in malaria [21]. There has been reported inhibitors against *Pf*dUTPase which a have strong selectivity for the malaria enzyme over the human counterpart [20].

1.2 Force Fields

Computer simulations have become very important in chemistry, enabling the researcher to model and predict the behavior of chemical systems before performing time-consuming and expensive wet-lab experiments.

Different computational models can be used in the simulations. Quantum mechanical methods can be used to study atoms and small molecules, but they are computationally expensive, and do not presently scale to larger molecules such as proteins. For the larger molecules, force field methods are more suitable. These methods ignore electron motion, and consider the system as a function of the nuclear positions.

Equation 1.2 is a basic force field equation that considers bond stretching, angle bending, angle rotation (torsion) and non-bonded interactions. The potential energy U of the system is a function of the position,r , of N particles:

$$\begin{aligned} U(r^N) = & \sum_{bonds} \frac{k_i}{2}(l_i - l_{i,0})^2 + \sum_{angles} \frac{k_i}{2}(\theta_i - \theta_{i,0})^2 \\ & + \sum_{torsions} \frac{V_n}{2}(1 + cos(n\omega - \gamma)) \\ & + \sum_{i=1}^{N} \sum_{j=i+1}^{N} (4\epsilon_{ij}[(\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^6] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}) \end{aligned} \tag{1.2}$$

The first term in Equation 1.2 represents bond stretching, and provides the energy difference from an ideal bond length between two covalently bonded atoms. In the first term, k_i is the force constant of the bond between the two atoms, l_i is the length of the bond between the atoms and $l_{i,0}$ is the reference bond length between the atoms. The stretching and compression of a bond is best described by a Morse potential. Calculating the Morse potential is computationally intensive due to its exponential form [22]. Instead Hooke's law, which is a harmonic potential, is used as an effective approximation. Hooke's law is accurate for small oscillations in the bond length, but is not accurate for large deviations from the reference bond length.

The second term in Equation 1.2 represents angle bending between three covalently bonded atoms. Angle bending follows Hooke's law near the equilibrium angle. Again, k_i denotes the force constant for the angle bending, θ_i is the angle between the two bonds and $\theta_{i,0}$ is the reference bond angle. In

general, angle requires less energy to be distorted from the reference state compared to the bond [22].

The third term in Equation 1.2 represents the torsional energy barriers. The torsional term is expressed as a cosine series expansion. V_n is the dihedral force constant, n is the number of minimum energy points when the bond is rotated around 360 °, ω is the dihedral angle and γ is the phase factor, which determines the phase shift. Changes in the torsions of a molecule are closely related to the non-bonded contributions and therefore not all force fields use torsional potentials.

The last term in Equation 1.2 represents the non-bonded interactions in the system. The non-bonded term is divided into van der Waals (vdW) and electrostatic (EL) contributions. The vdW interactions are modeled with a 6-12 Lennard-Jones potential. In Equation 1.2 this term, σ_{ij}, denotes the collision diameter between atoms i and j, ϵ_{ij} is the potential well depth and r_{ij} is the separation between atoms i and j. The Lennard-Jones potential has a term for repulsion and a term for attraction. The repulsive part varies with r^{-12} and the attractive part varies with r^{-6}. Consequently the repulsive part of the Lennard-Jones potential converges quicker to zero when the distance between two atoms increases. The repulsive term has its origin in Pauli principle, which states that two electrons can not have the same set of quantum numbers. The attractive force is also called the London force. The London force occurs when the electron cloud fluctuates and influences the near by electrons and creates instantaneous dipoles in neighboring atoms which last for a very short period of time. In general, the larger the molecule, the more important is the London force.

The EL contributions are normally modeled with a Coulomb potential. In Equation 1.2 q_i and q_j are the charges for atoms i and j, and ϵ_0 is the permittivity of free space. Electronegative elements attract electrons and creates unequal distribution of electrons in the molecule. This gives the rise to electrostatic interactions in the system which are a very important part of the forces that drive the changes in chemical systems.

There are additional terms which can be added to the basic force field equation. If necessary, out-of-plane bending term can be used to keep, for example, the oxygen of cyclobutane in the plane of the ring. The out-of-plane term can be modeled with torsional potential to only allow certain improper angles. Another additional term is the coupling of internal coordinates with cross terms. A simple example of internal coordinate coupling is when decreasing an angle between two atoms causes an increased repulsion between

atoms. This effect can be compensated by adding length to the bonds. The most important cross terms are believed to be stretch-stretch, stretch-bend, bend-bend, stretch-torsion and bend-bend-torsion [22]. The cross terms can be modeled as bond stretching and angle bending potentials, and non-bonded interactions.

Parametrization of a force field is time consuming and involves methods like IR-spectroscopy and quantum mechanical calculations. Parameters are often derived for small molecules, however transferability makes it possible to share parameters developed for small systems with large systems.

1.3 Molecular Dynamics

molecular dynamics (MD) is a way of simulating molecular motions, based on Newton's laws of motion. These laws are:

1. A particle continues to move in a straight line at constant velocity (or remains at rest) if there is no external force, $\Sigma\mathbf{F}$, acting upon it:
$$\Sigma\mathbf{F} = 0 \Rightarrow \frac{d\mathbf{v}}{dt} = 0 \tag{1.3}$$
where $\mathbf{v}$ is the velocity of the particle.

2. The net force acting on a particle is equal to the rate of change of its linear momentum:
$$\Sigma\mathbf{F} = \frac{d\mathbf{p}}{dt} = \frac{d(m\mathbf{v})}{dt} \tag{1.4}$$
where $\mathbf{p}$ is the linear momentum of the particle.

3. To every force, $\mathbf{F}_A$, acting on a particle, there exists an equal and opposite force, $\mathbf{F}_B$, acting on a different particle:
$$\mathbf{F}_A = -\mathbf{F}_B \tag{1.5}$$

MD was developed in the late 1950's [23, 24]. These algorithms used hard-sphere calculations where no force exist between the particles until they collide. The basic steps in the algorithm were:

1. find the pair of spheres that will be the next to collide and predict when this will happen

2. at the time of collision calculate the position of all spheres
3. calculate the new velocities for the spheres that collided

This procedure is repeated until the simulation is finished.

The algorithms in use today use a continuous potential model, such as the Lennard-Jones [25]. This means that the force between particles vary continuously as their separation changes. Using a continuous potential couples the motion of all the particles in the system. Generally, it is not possible to solve the resulting equations analytically, and numerical methods, such as finite difference methods must be used.

The integration is divided into time steps and the force acting on each particle at time t is calculated as the vector sum of the particles interactions with other particles. Since we now have the force it is possible to determine the accelerations of the particles. The positions and velocities at time t are combined with the accelerations and force to calculate the positions and velocities at time $t + \delta$t.

All algorithms that are used to solve the finite difference method for the equations of motion assume that the positions, velocities, accelerations, etc. can be approximated as a Taylor series expansions:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \delta t\mathbf{v}(t) + \frac{1}{2}\delta t^2\mathbf{a}(t) + \frac{1}{6}\delta t^3\mathbf{b}t + \frac{1}{24}\delta t^4\mathbf{c}(t) + \cdots \tag{1.6a}$$

$$\mathbf{v}(t+\delta t) = \mathbf{v}(t) + \delta t\mathbf{a}(t) + \frac{1}{2}\delta t^2\mathbf{b}(t) + \frac{1}{6}\delta t^3\mathbf{c}(t) + \cdots \tag{1.6b}$$

$$\mathbf{a}(t+\delta t) = \mathbf{a}(t) + \delta t\mathbf{b}(t) + \frac{1}{2}\delta t^2\mathbf{c}(t) + \cdots \tag{1.6c}$$

$$\mathbf{v}(t+\delta t) = \mathbf{b}(t) + \delta t\mathbf{c}(t) + \cdots \tag{1.6d}$$

During the time interval the forces acting on atoms are considered to be constant. The length of the time step in the above equation should be adapted to the system. With a too small time step, the phase space is not sampled efficiently. With a too large time step, the energy of the system can change too rapidly, causing atoms to overlap and resulting in too high potential energy.

It is suggested that the time step for flexible molecules should be less than the fastest motions in the system [22]. The fastest motion is often the vibration

of hydrogen bonds which in protein simulations is not the interesting part of the system. These fast motions can be taken away from consideration by using constraints like the SHAKE [26] algorithm which reduces the fastest movement of hydrogens and makes it possible to use longer time steps to cover more phase space.

The simulation can be done in vacuum or in solvent. When done in solvent the simulation becomes more computer intensive because of the added interactions of solvent molecules. This problem can be partially solved with setting cutoff boundaries for interactions so that atoms interact only inside a cutoff distance or introducing a neighbor list. Common thermodynamic ensembles for molecular dynamics are microcanonical (NVE, with constant number of particles, volume and total energy), canonical (NVT, with constant number of particles, volume and temperature), and isothermal-isobaric (NPT, with constant number of particles, pressure and temperature) [22].

The MD methods have enabled the study of molecular motion since the late 1950's. The first protein simulation appeared in 1977 [27] . Even though the simulations were primitive, they made the foundation for the later appearing methods such as free energy perturbation (FEP), linear interaction energy (LIE) and molecular mechanics/Poisson-Boltzmann/surface area (MM-PBSA).

1.4 Conformation Sampling

Computational methods, such as docking, LIE, MM-PBSA and FEP, are utilized in the search for information about the binding free energies in ligand-protein binding. At one end of the spectrum there is docking, which is a quick and time efficient semi-empirical method, while at the other end there are more computationally expensive and precise methods such as the FEP. These different methods are applied in different parts of a drug design project. Docking is used in the first steps of the computational part to search through large ligand libraries for new candidates and decrease the search base for new ligands, while for example the LIE method is applied for selected compounds to conduct a closer analysis of the structure and the binding affinity to rule out false positives. Even though docking is not a conformation sampling method is it included in this chapter since we utilize it in our working protocol. In the following chapter we will look more closely at the techniques mentioned above.

1.4.1 Docking and Scoring

Docking is a widely used method to predict the binding mode of a ligand and a receptor (protein). The interactions between a ligand and the target protein are crucial pieces of information in drug design. Methods like X-ray crystallography and NMR-studies are important in the analysis of molecular interactions. These methods give invaluable insight to the molecular structure for computational drug design. However, they have their limitations, especially when it comes to the search for novel ligands due to the time consuming nature of experimental work. The screening of thousands of novel ligands for a single protein is not feasible with experimental techniques. Computer based screening methods, like docking, have reached an important place next to the experimental methods. Docking offers the possibility to go through large ligand libraries in the search for good binding qualities. The pioneering work for docking programs was done in the 1980s [28]. Since then there has been major improvements in the speed and accuracy of docking programs. Some well known docking programs are GOLD [29], FlexX [30], DOCK [31] and Glide [32].

Docking has been developed for rapid and precise prediction of binding modes between molecules. The "docking problem" includes the generation and evaluation of relevant ligand-receptor poses. Docking structures can be generated with different degrees of freedom. The simplest way is to have six degrees of translational and rotational freedom of one molecule in relation to the other, meaning that the molecules are rigid during the fitting. If the ligand is allowed to change conformation the conformational degrees of freedom arise and makes the search more complex. Preferably the protein should also be flexible in some areas (side-chain). In general, the more degrees of freedom are allowed, the more laborious is the search. Common algorithms to generate conformations are MD, Monte Carlo and fragment based, to mention a few [33].

Evaluation of the poses (ligand-receptor conformation) is done with scoring functions. Scoring functions serve three purposes in docking [33]:

1. Resolve the binding mode of the ligand in the protein.

2. Predict the binding affinity for the ligand binding.

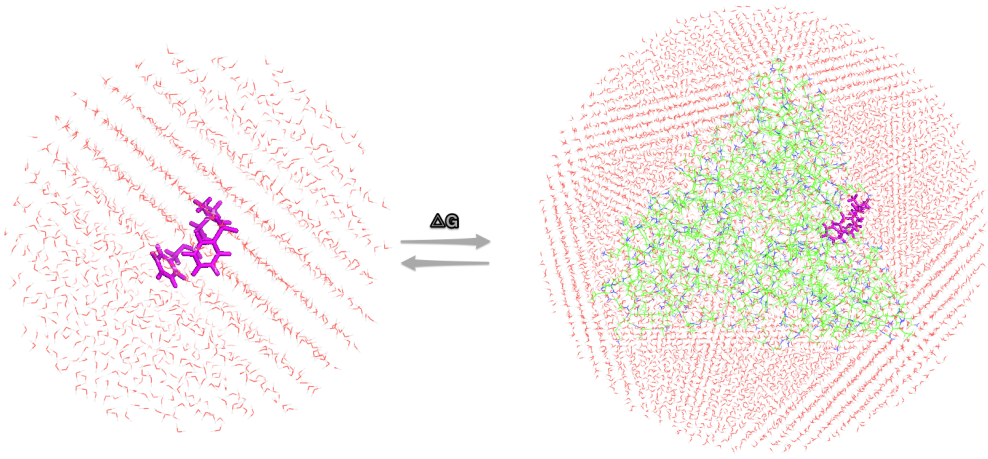
3. Identifying potential lead compounds for drug design.

Popular scoring functions in docking are force-field, empirical and knowledge-based scoring [33].

While docking and scoring can give estimates of the binding free energies and conformations, more precise methods are required to confirm the results, such as the LIE method.

1.4.2 Linear Interaction Energy

The LIE method is a semi-empirical method developed to estimate absolute binding free energies [34]. It was originally designed for estimating ligand-protein binding affinities. The idea behind a LIE calculation is to sample the physically important states of the system, rather than sampling confirmations that have no real relevance, like done in FEP [35].



(a) (b)

Figure 1.2: The LIE method calculates the difference in the binding free energy from two states, the: a) ligand free in water and b) ligand bound to the protein.

When the LIE method was developed the fundamental idea was to measure the absolute binding free energy of a ligand as a the change in free energy when a ligand is transferred from water to the receptor protein (Figure 1.2), which would correspond to free and bound state of the ligand [35]. The idea is summed up in the following equation:

$$\Delta G_{bind} = \Delta G^{p}_{sol} - \Delta G^{w}_{sol} \tag{1.7}$$

where p denotes protein, w denotes water and sol stands for solvation. Solvation of the ligand means the process of transferring the ligand from vacuum to the solvent environment. This is done in two steps where first the ligand is transferred to a van der Waals cavity in the solute and then the electrostatic interactions between the solute and the solvent are turned on. The electrostatic interactions can be approximated by the linear response method [34] and results in:

$$\Delta G_{el} = 1/2\{\langle U^{el}_{l-s}\rangle_{on} + \langle U^{el}_{l-s}\rangle_{off}\} \tag{1.8}$$

where the first term is the averaged electrostatic term of the solvation energy when the electrostatic interactions between the ligand and the surroundings are turned on and the second term is when the interactions are turned off. Since in the second term the electrostatic interactions are turned off the ligand is not experiencing any electrostatic interactions. This justifies neglecting of the off term and yields a simplified version of Equation 1.8.

$$\Delta G_{el} = 1/2\{\langle U^{el}_{l-s}\rangle_{on}\} \tag{1.9}$$

This is considered to be the electrostatic contribution to the solvation free energy of a given system and is the second term in Equation 1.11.

The nonpolar contribution to the free energy is given in Equation 1.10.

$$\Delta G_{vdw} = \alpha\{\langle U^{vdw}_{l-s}\rangle\} \tag{1.10}$$

In Equation 1.10 α is an empirically derived scaling factor for the nonpolar part of the binding free energy. The difference in the solute-solvent van der Waals energy between the free and the bound states of the ligand is connected to a corresponding nonpolar binding contribution by α. This approximation is based on the observed linear dependence of the solvation free energy on the size for nonpolar compounds [36].

The interactions in the LIE equation 1.11 [35] for ΔG are divided into electrostatic and van der Waals components (Equation 1.11).

$$\Delta G_{bind} = \alpha \Delta \langle U_{l-s}^{vdW} \rangle + \beta \Delta \langle U_{l-s}^{el} \rangle + \gamma \tag{1.11}$$

where 1.11 $\langle\rangle$ indicate ensemble averages, α and β are weight constants and γ is an additional constant. The l-s denotes ligand interacting with its surroundings. The polar contributions, $\Delta \langle U_{l-s}^{el} \rangle$, comes from the original version of LIE where the electrostatic forces for solvation/binding free energies were estimated by the linear response approximation [35]. As mentioned earlier the value of the constant $\beta = 1/2$ is derived from the approximation, although the value can vary, depending on the system. Compounds can be divided into four classes of β values based on the chemical nature: charged, neutral with no hydroxyl groups, neutral with one hydroxyl group and neutral with two or more hydroxyl groups. These four groups respond to β values of 0.50, 0.43, 0.37 and 0.33, respectively [37].

The non-polar part of equation 1.11, $\langle U_{l-s}^{vdW} \rangle$, responds to the size of the ligand and should not be interpreted as a result of linear response towards forces in the Lennard-Jones potential. $\langle U_{l-s}^{vdW} \rangle$ is a measure of the density of heavy-atoms surrounding the ligand/solute, which can make the value of $\langle U_{l-s}^{vdW} \rangle$ more negative for a ligand when it is surrounded by a protein than when it is free in water [35]. The constant α is determined empirically to scale the non-bonded contribution to the free energy of binding and its value depends on the functional groups of the ligand. In the initial LIE model the value of α was 0.16, but later it was estimated to 0.18 [38]. The γ is added to the LIE equation if the binding free energy is dominated by vdW interactions [38].

The Δ's' in equation 1.11 denote the difference between the bound and free state. The free energy terms of binding are retrieved from simulations where the ligand is free in water and when it is bound to the receptor, as follows:

$$\begin{aligned} \Delta G_{bind} = \alpha(\langle U_{complex}^{vdW} \rangle - \langle U_{wat}^{vdW} \rangle) \\ +\beta(\langle U_{complex}^{el} \rangle - \langle U_{wat}^{el} \rangle) \end{aligned} \tag{1.12}$$

The simulation used in LIE can either be Molecular Dynamics or Monte Carlo.

The original, electrostatic linear response approximation, parametrization reproduced binding energies with good results for many systems [39], [40],

[41], [42]. Later the parametrization was revised for deviations from the linear response theorem and exhibited the ability to predict the relative binding free energies for chemically different ligands [43]. The LIE method has also indicated the ability to predict absolute binding energies in addition to relative ones, in contrast to FEP, which only is capable of resolving the relative energies between two ligands [35].

1.4.3 Molecular Mechanics/Poisson-Boltzmann/Surface Area

An alternative to the LIE method, which also uses conformation sampling, is the MM-PBSA [44]. It was originally applied for studying the stability of DNA and RNA fragments [45]. The MM-PBSA method analyses molecular dynamics trajectories with a continuum solvent approach and estimates the free energy state of the system as a trajectory average of the molecular mechanics energy, solvation free and nonpolar energies. In addition the solute entropy is included in the equation:

$$\Delta G_{bind} = \langle E_{MM} \rangle + \langle E_{PBSA} \rangle - T\langle S_{MM} \rangle \tag{1.13}$$

where $\langle E_{MM} \rangle$ is the average molecular mechanical energy, which usually includes bond, angle, torsion, van der Waals, and electrostatic term from a regular force field. These terms are evaluated with the absence of nonbonded term cutoffs. The $\langle E_{PBSA} \rangle$ term consists of the solvation free enegies calculated by a numerical solution of the Poisson-Boltzmann equation, and a surface-area based estimate of the nonpolar free energy. The last term $T\langle S_{MM} \rangle$, the solute entropy, can be for example approximated by a quasi-harmonic analysis of the trajectory [45].

There are two ways to apply the MM-PBSA to ligand binding free energies. In the first option the terms in Equation 1.13 are evaluated from separate directories for the complex, receptor and the ligand. The different contributions are then added together to form ΔG_{bind}:

$$\Delta G_{bind} = \langle G_{complex} \rangle - \langle G_{receptor} \rangle - \langle G_{ligand} \rangle \tag{1.14}$$

The second alternative includes evaluating the terms in Equation 1.14 from snapshots of a trajectory of the complex. The $\langle G_{receptor} \rangle$ and $\langle G_{ligand} \rangle$ are estimated by subtracting one of the molecules from the trajectory. The second option is mainly used since the $\langle E_{MM} \rangle$ term for the complex or the receptor will not converge within normal computational times.

The second option assumes that the conformation of the receptor and the ligand do not change upon binding since the intramolecular term in the protein and the ligand are not taken into account. This can be problematic because then it is assumed that the protein and the ligand do not change conformation upon binding. In the case of flexible ligands the ligand conformation can be quite different from free in solvent compared to when it is bound to the protein. This is especially important for large flexible hydrophobic ligands which can go through a "hydrophobic collapse", meaning they rearrange them selves to minimize the expose to water molecules [43]. This can lead to decreased binding affinities if the free state becomes more favorable for the ligand. Therefore using a simulation only for the protein might not result in correct energies.

1.4.4 Free Energy Perturbation

One of the most rigorous molecular dynamics binding free energy methods is the FEP method. The FEP method can be thought of as the mutation of ligand L to ligand L', free in water, and measuring the free energies, ΔG^w_{mut}, between these two states (Figure 1.3). The same procedure is done with the ligands bound to the protein and retrieve ΔG^p_{mut}. $\Delta G_{bind}(L)$ and $\Delta G_{bind}(L')$ are the binding free energies for both of the ligands L and L', respectively, between the free and bound state. The relative free energy of binding between the two ligands L and L' can be deduced from the thermodynamical cycle in Figure 1.3 and yields Equation 1.15:

$$\Delta\Delta G_{bind} = \Delta G_{bind}(L') - \Delta G_{bind}(L) = \Delta G^p_{mut} - \Delta G^w_{mut} \tag{1.15}$$

which states that the relative free energy of binding can be measured from the unphysical path when mutating L to L'.

The free energy difference between two states A and B is given by Zwanzig's formula [46]:

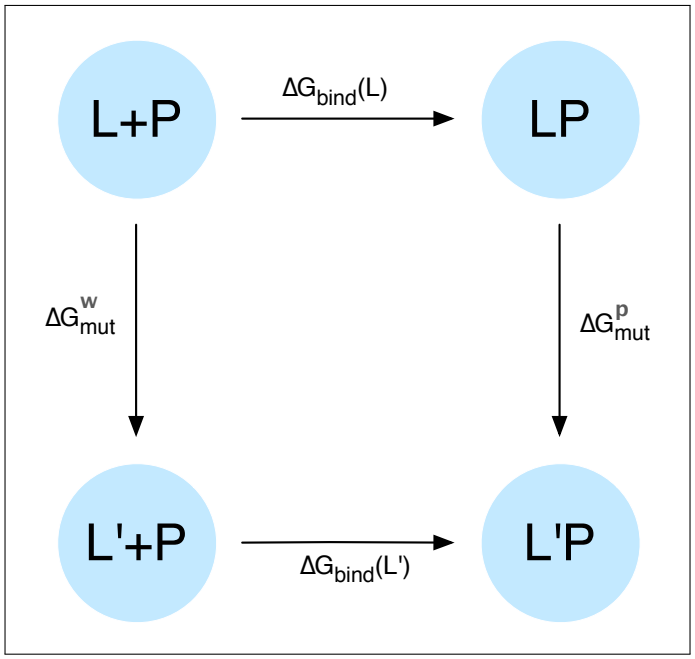


Figure 1.3: The thermodynamical cycle which represents the binding of ligands L and L' to the protein P, in terms of binding free energies and the free energy differences between the ligands.

$$\Delta G = G_B - G_A = -\beta^{-1}\langle exp(-\beta\Delta V)\rangle_A \tag{1.16}$$

where $\beta = 1/kT$ and $\langle\rangle$ is the ensemble average of ΔV sampled with the V_A potential from MD or monte carlo (MC) simulations. For the binding free energy measurements the conformation sampling is done with isothermal-isobaric conditions. The kinetic energy contribution to the binding free energy difference is not considered since it cancels out due to the equipartition theorem.

To be able to use Equation 1.16 the thermally accessible regions of the potential V_A and V_B should overlap. To achieve this the change from state A to state B is done step wise as a linear combination of the first and the last state of the change:

$$V_m = (1 - \lambda_m)V_A + \lambda_m V_B \tag{1.17}$$

where λ_m can have values between 0 and 1, and m=1,...,n. The total free

energy change, consisting of n intermediate states, is the sum of the intermediate states specified by the λ variables:

$$\Delta G = G_B - G_A = -\beta^{-1} \sum_{m=1}^{n-1} ln\langle exp[-\beta(V_{m+1} - V_m)]\rangle_m \tag{1.18}$$

The linear combination of state A and B is used for all terms in the potential energy equation: bond stretching, bond bending, improper torsions, non-bonded interactions etc. The simulations are done both with the ligands solvated in water and bound to the receptor solvated in water. The suggested number of λ steps for a FEP simulation is 50-100 [43].

When the FEP method is applied to ligand binding studies there are some important modifications which can be made to decrease instabilities in the simulations due to appearing and vanishing atoms:

- change the Lennard-Jones potential in the non-bonded energy term to a softer variant [47], [48]
- make the bond length of vanishing atoms shorter until they are pulled inside the vdW radius of atoms they are connected to
- increase the amount of λ sampling points near the end-points

Even though the FEP method is accurate in measuring the relative binding free energies in ligand binding it requires large amounts of computing time. Methods such as docking and LIE are more efficient in the field of drug design when estimating the binding free energies of ligand-protein complexes.

1.5 Q Graphical User Interface

Q [49] is a software package for doing free energy calculations. Q has modules to do FEP simulations [50, 51], empirical valence bond (EVB) calculations of reaction free energies [52] and LIE calculations [34] for ligand-receptor binding free energies.

Q consists of multiple applications (Figure 1.4 [53]): Qprep, Qdyn, Qfep and Qcalc. All interactions with Q are done through a command-line interface. To use Q effectively, the user must be proficient at writing job submission scripts, navigating the file systems, executing applications and edit data and configuration files.

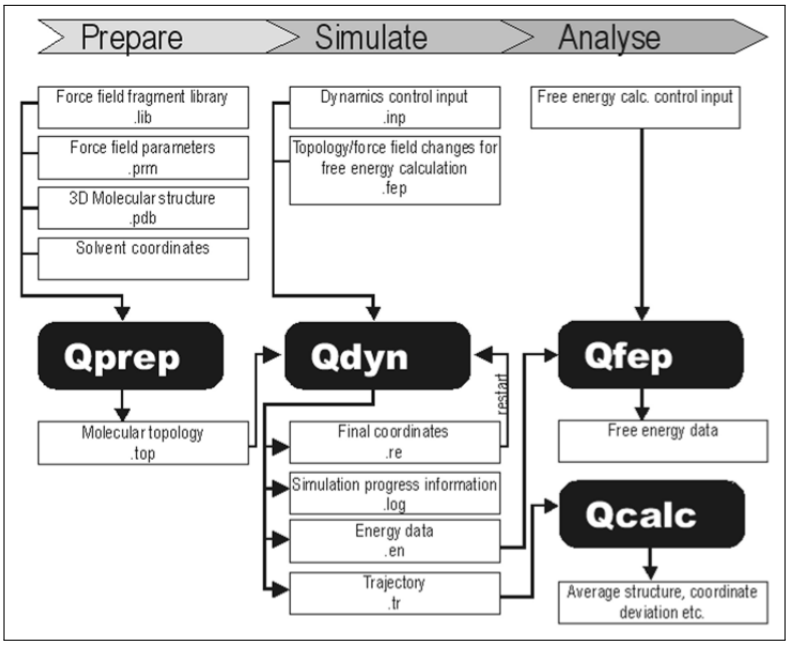


Figure 1.4: Overview of the work flow in Q. The white boxes represent the files and the black boxes the programs.

For a novice computational chemist the topology generation, setting up MD, LIE, FEP and EVB jobs, in addition to analyze the simulations can be an overwhelming task:

- Keeping track of available user settings is difficult.
- It is hard to keep track of all files involved.

- It is cumbersome and error prone to transform data using ad-hoc scripts or sed commands. These tools often require knowledge of regular expressions.
- Detecting defects in settings and data transformations is difficult, when the settings are spread over multiple files and scripts and the data transformations may affect only small parts of large data files.

In an attempt to solve these problems the idea of a graphical user interface came alive.

Q graphical user interface (Qgui) has its origin in a project lead by Geir Villy Isaksen (a PhD student at CTCC, Department of Chemistry, University of Tromsø). The main goal with Qgui has been to make Q user friendly for users not familiar with command-line interfaces and making the use of Q faster and more efficient for both novice and experienced users. New users of Q are already used to interact with applications through graphical user interfaces (GUIs). All settings are visible in the various windows and user dialogs. Some of the major advantages of using Q with Qgui is the efficiency of submitting large jobs and analyzing the simulations with just a few clicks, in contrast to the amount of effort it would take when using Q manually.

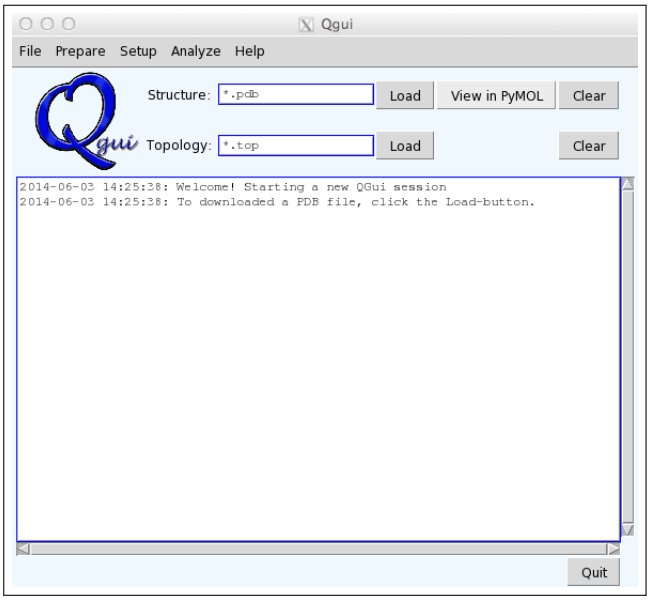


Figure 1.5: The Qgui start window.

The GUI can performs task the same way every time, making errors less

likely. This is convenient for the novice user since they do not have to have the skills to write scripts for data manipulation. Also the experienced user of Q will find this feature practical since keeping track of old and new scripts is difficult, not to mention to remember the purpose and function of these scripts. The GUI will also alert the user of a missing input or settings before starting the calculations, therefore saving time. This is a huge improvement form Q, since Q does not have a good feedback system.

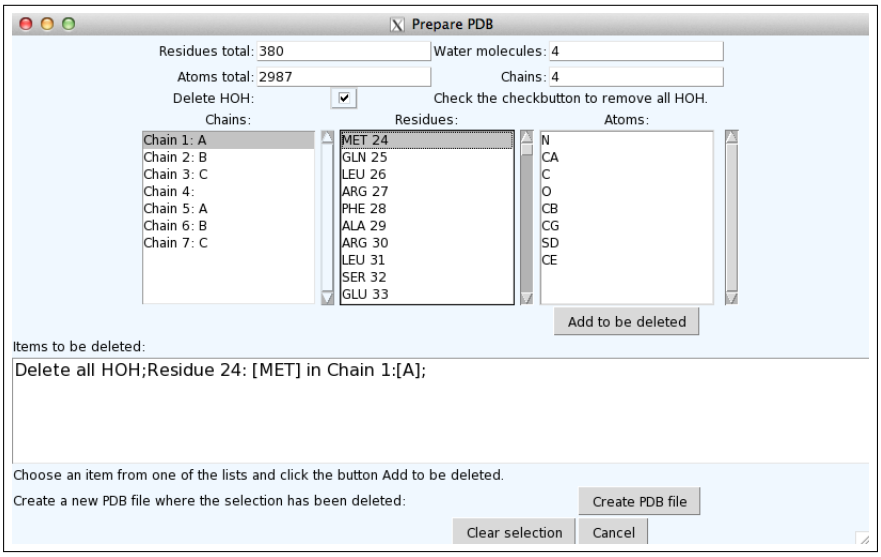


Figure 1.6: The Qgui PDB Prepare window.

Qgui is implemented in Python using object oriented design principles. It is based on the TkInter GUI library. Qgui is built around a main start window (Figure 1.5). In this window you can load structure and topology files and have access to the different modules of Qgui. A direct download of structures files from Protein Data Bank (pdb) is also possible. A monitor on the main window displays status information and messages. In Prepare PDB it is possible to inspect and modify your structure files (Figure 1.6). In Prepare Topology you can prepare your topology file (Figure 1.7).

In Topology Prepare there are several options to customize the molecular topology, it is possible to:

- choose the size of the solvation sphere
- change the simulation center (it is defined as the center of the system by default)

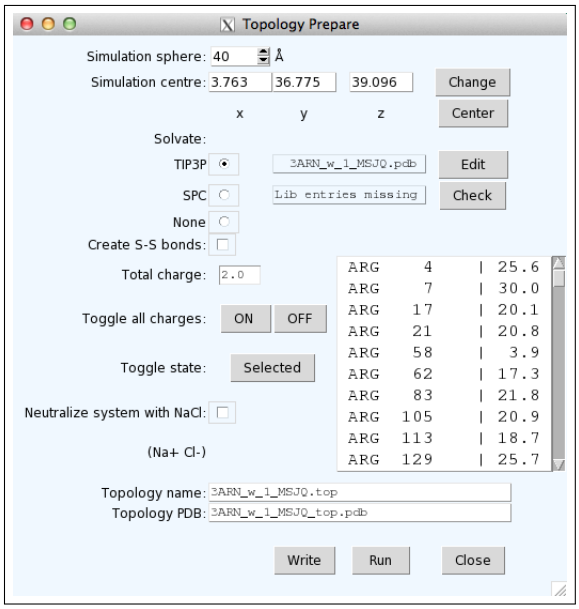


Figure 1.7: The Qgui Topology Prepare window.

- type of solvation
- create S-S bonds
- check the total charge of the system
- turn charges on/off
- specify C and N terminal residues
- customize the topology name
- inspect and modify the molecular structure file

It takes very little effort to prepare a topology in this way.

With Qgui it is possible to set up MD, LIE, FEP and EVB calculations with great ease compared to doing it manually. In Figure 1.8 and 1.9 the set up scene of LIE simulations is displayed. In Setup LIE the setup for a LIE simulation is specified by loading structure and topology for the complex and the ligand. It is also possible to only load the structure for the complex and use the functionality in Setup LIE to prepare the complex topology and, the ligand structure and topology. Through Configure MD (Figure 1.9) the MD simulation set up is chosen. After completing the MD set up and choosing

the number of runs in the setup LIE window there is an opportunity to either write input files, or submit the simulations. All the input files are generated automatically and divided into corresponding directories for the complex and the ligand.

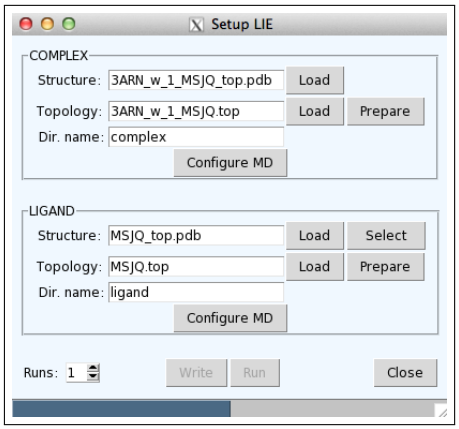


Figure 1.8: The Qgui LIE set up window.

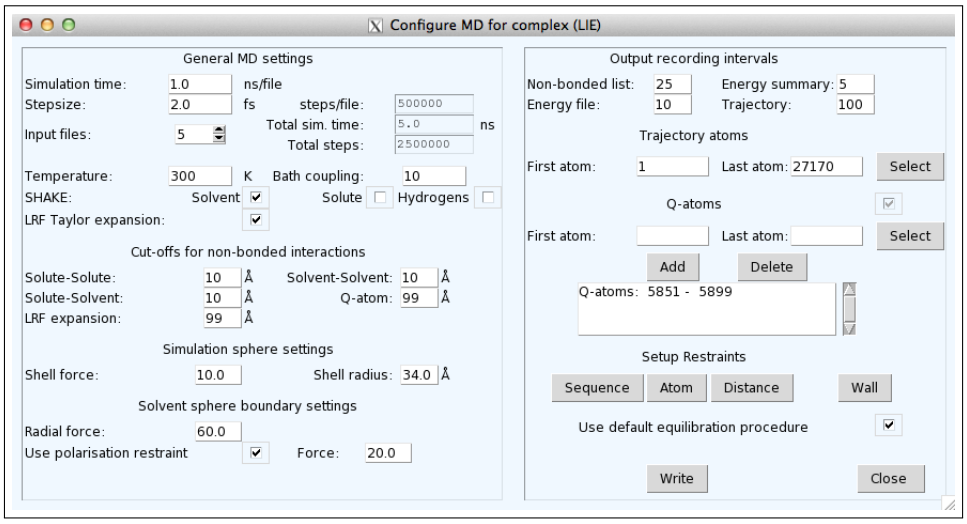


Figure 1.9: The Qgui LIE MD simulation set up window.

1.6 Aims of Study

dUTPase is an important drug target in various types of cancer. The development of new inhibitors targeting human dUTPase in cancerous cells can benefit from thorough computational studies of the binding site of the enzyme. One step in this direction is to utilize existing experimental data and build a linear interaction energy model which can predict binding free energies for novel inhibitors. The specific goals of this project are:

- develop a graphical user interface for the molecular dynamics software Q, and apply it for the running simulations and analyzing them
- develop a deeper understanding of molecular dynamics methods and apply them for binding free energy calculations
- become proficient in setting up and analyzing molecular dynamics simulations for the LIE method
- build a protocol for predicting inhibitor binding free energies in human dUTPase
- test the selectivity of the LIE model
- predict binding free energies for novel ligands with the LIE model

Should the project be successful then the LIE model could be implemented in a drug design pipeline to detect potent novel inhibitors.

Chapter 2

Methods

2.1 Protein Preparation

The PDB file of the crystal structure of 3ARN [18] was prepared in Maestro [54] using the Protein Preparation Wizard [55, 56]. The default settings in Protein Preparation Wizard were used to preprocess the protein and the ligand.

The preprocess option:

- assigns bond orders if a bond order is not given in the structure
- adds hydrogens
- breaks bonds to metals and creates zero-order bonds to nearby atoms
- corrects formal charges to metals and nearby atoms
- creates disulfide bonds if it is possible
- deletes all waters that are further away than 5 Å from non-standard residues

Automatic optimization was used to optimize hydroxyl, Asn, Gln and His states. A restrained minimization was executed for the complex, with the OPLS2005 force field [57], where the heavy atoms are restrained and hydrogens are unrestrained.

The biological unit of the enzyme was included in the model. The active sites of the enzyme are located at the intersections of two subunits, at three separate locations. A ligand, a Mg^{2+} ion and eight structural waters were included in the model. The Mg^{2+} ion and a water coordinated to it were considered to be structurally important for the trimer [1, 4]. In addition to the magnesium coordinated water, three waters that were reported to be structurally important in the active site [18] and four additional waters were included in the model.

Three sets of ligands were built manually in Maestro from the structure of 3ARN. The first set was built from ligands 9-26 [18] by substituting the hydrogen atoms of the phenyl ring by components listed in Table 3.1. In the first set, also referred to as modification 1, the substitutions on the phenyl ring were made at carbons 15, 16 and 17. Set two was built from ligands 9-26, but the substitutions on the phenyl ring was made at carbons 17, 18 and 19. Set two is also referred to as modification 2. The third and last set of the ligands were made by including the ligands from modification 2 and by changing a Lys 318 rotamer at the enzyme.

To study the selectivity of the human dUTPase model, selected ligands from an inhibitor serie [21] designed for *Pf*dUTPase, with poor binding affinity to the human counterpart, were prepared. The *Pf*dUTPase ligands were built in the human enzyme active site by using the initial coordinates of the uracil ring of compound 9 as a base. The rest of the ligand conformation was built to mimic the crystallographic structure of 1VQY [58].

Two novel ligands for human dUTPase were built manually in Maestro and docked to the 3ARN crystalstructure to sample conformations for LIE simulations.

2.2 Docking

The selected *Pf*dUTPase ligands [21], as well as other novel ligands, were docked into the crystallographic structure of the human dUTPase [18]. The Glide (version 6.3) [32] program was used for the docking. A receptor grid was generated for the human dUTPase [18] crystallographic structure. The vdW radius the for nonpolar parts of the receptor was scaled with a scaling factor of 1.0, while the partial charge cutoff was scaled with a factor of 0.25. A SP(standard precision) docking was performed with flexible ligand sam-

pling. Nitrogen inversions and ring conformations were sampled. Nonplanar conformations for amides were penalized and epik state penalties were added to the docking score. vdW radii for the nonpolar parts of the ligand was scaled with the a factor of 0.8 and the partial charge cutoff with a factor of 0.15.

In the docking panel the maximum amount of ligands allowed generated for each ligand can be limited to a certain number. First all ligands were docked with this setting set to 10. Then ligands 3a, 3b, 4, 5a, 30, 7a, 7b, 7c, 14b, 14c, 14d and 14e were docked with the number set to 40. A post-docking minimization was run with a 0.5 kcal/mol threshold for rejecting the minimized pose.

Two novel ligands for human dUTPase were docked to the crystallographic structure 3ARN. The docking protocol followed the same speficications as for ligands 1a-14e, with maximum 10 poses allowed per ligand. The three highest scoring poses from each docking run was selected for the LIE simulations.

2.3 Topology Preparation

Topologies for the LIE MD simulations were prepared with the Prepare Topology module in Qgui. The ligand and Mg^{2+} ion parameters were generated with the Prepare Parameter module in Qgui according to the OPLS-AA force field [59]. The enzyme was solvated in a 40 Å water sphere and the ligand in a 20 Å water sphere, with TIP3P [60] water grid. The charges in the charged residues were turned on. The enzyme N and C ends were marked as N- and C-terminals.

2.4 Linear Interaction Energy Setup

The LIE calculations were run with the program Q [49]. The automated LIE setup was used in Qgui, where the ligand free energy simulation in water and in protein are configured simultaneously.

Equilibration was run for all structures before the MD simulations. The equilibration process was executed in five steps. In the four first steps the system was gradually heated up with restraints for all coordinates in the

system. Each of the four steps had 10000 simulation steps. The restraints were gradually decreased. The last part of the equilibration was 100000 unrestrained steps at the final temperature.

The MD simulations for the LIE method were run with 1.0 ns simulation time per file, with 2.0 fs time step and five input files per run. This results in 2500000 simulation steps and 5 ns of simulation per run. Each run was repeated 10 times with different random seeds. This resulted in 50 ns of simulation time per compound. The temperature was set to 300 K. System was coupled to a temperature bath which was had a relaxation time of 10 fs. The SHAKE [26] algorithm was applied for solvent bonds and angles. LRF Taylor Expansion [61] was applied for long range non-bonded interactions, with no cutoff (99 Å). A 10 Å cut-off was applied for both solute-solute, solute-solvent and solvent-solvent non-bonded interactions. An exception in the non-bonded interaction cut-offs was the Q-atoms (the ligand), where no cut-off was applied (99 Å). The simulation sphere radius was 34 Å for the protein and 17 Å for the ligand. The outermost layer of the simulation sphere is the buffer zone and is applied a constant 10 kcal*mol$^{-1}\times$Å^{-2} restraining force. The boundary of the solvent sphere was applied a radial force of 60 kcal*mol$^{-1}\times$Å^{-2}. The polarization restraints were enabled in the solvent boundary with polarization force of 20 kcal*mol$^{-1} \times red^{-2}$. The non-bonded energy list was updated every 25 simulation steps, the energy of the system was recorded every 5 simulation steps, the recorded energy was written to a output file every 10 simulation steps and the coordinates of the system were written in a trajectory file every 100 simulation steps. Positional restraints were applied on the free ligand solvated in water to keep the ligand from moving to the solvation sphere boundaries. The parameters were assigned according to the OPLS-AA [59] force field.

2.5 Analysis of the Linear Interaction Energy Simulations

Root mean square deviation (RMSD) analysis of the MD simulations was performed with the Analyze Trajectory module of Qgui. Only the backbone of the protein was included in the analysis. The analysis of the LIE model was performed with the Qgui Analyze LIE module. The fitting of the binding free energies was done i Qgui with the "Fit" option. The fitting is performed with a standard least squares calculation. The LIE fitting will give you the

opportunity to change the α, β and γ parameters. For our models, only the β parameter was changed from its initial value of 0.50 to 0.43.

The LIE fitting also provides measures of the goodness-of-fit for the model. sum of squared errors (SSE) is the sum of squares of residuals:

$$SSE = \sum_{i=1}^{n} (y_i - f(x_i))^2, \tag{2.1}$$

where y_i is the i^{th} value of the computational binding free energy, and $f(x_i)$ is the value of the experimental energy.

SSE is a measure of the difference between the computational binding free energies and the experimental model. If the value of SSE is close to 0 it indicates that the fitted model has a small random error component and that the model can be useful for prediction.

The coefficient of determination (COD), or R^2, indicates how well the computational binding free energies fit the LIE model:

$$COD = 1 - \frac{SS_{res}}{SS_{tot}} \tag{2.2}$$

where the SS_{res} is the residual sum of squares:

$$SS_{res} = \sum_{i} (y_i - f_i)^2 \tag{2.3}$$

and SS_{tot} is the total sum of squares:

$$SS_{\text{tot}} = \sum_{i} (y_i - \bar{y})^2 \tag{2.4}$$

COD can have any value between 0 and 1. The closer COD is to 1, the bigger the portion of the variance from the average can be explained by the model. If the value of COD is 1, then the computational binding free energies are in complete agreement with the experimental energies.

2.6 Design and Implementation of Qgui

The Qgui project was initiated to make the Q molecular dynamics software more user friendly.

Qgui was implemented in the Python programming language using the Tk-Inter GUI library. The organization of the source code in Qgui follows object oriented software design principles. The architecture of Qgui consists of three layers (Figure 2.1): the GUI layer, the controller layer and the Q layer.

The GUI layer contains the different windows and user dialogs, for example, the main application window, the prepare PDB window and the prepare topology window. It displays information to the user, and forwards user events, such as the click of a button, to the controller layer below.

The controller layer includes the Qgui main controller class which can be thought of as the engine of Qgui. The main controller class handles user generated events from the GUI layer. These events can include for example the loading of an input file, opening up a new dialog window, or executing a Q application. The controller layer also updates the GUI with new data, such as log messages or data from files. Finally, the controller is responsible for executing external scripts that generate the Q input files, and for executing the different Q applications in the Q layer.

The Q layer consists of the different modules of the Q molecular dynamics software. Qgui generates the input files to the Q modules, execute the Q module, and also use the output files for further processing or analysis.

After the initial design and implementation of Qgui for this project, much work has been done by Geir Isaksen to get Qgui to the state where it is today. Qgui now includes many components that are not show in Figure 2.1, such as more windows and dialogs, various setup protocols for molecular dynamics simulations and result analysis tools. Geir Isaksen has also made the data transformation scripts used by Qgui.

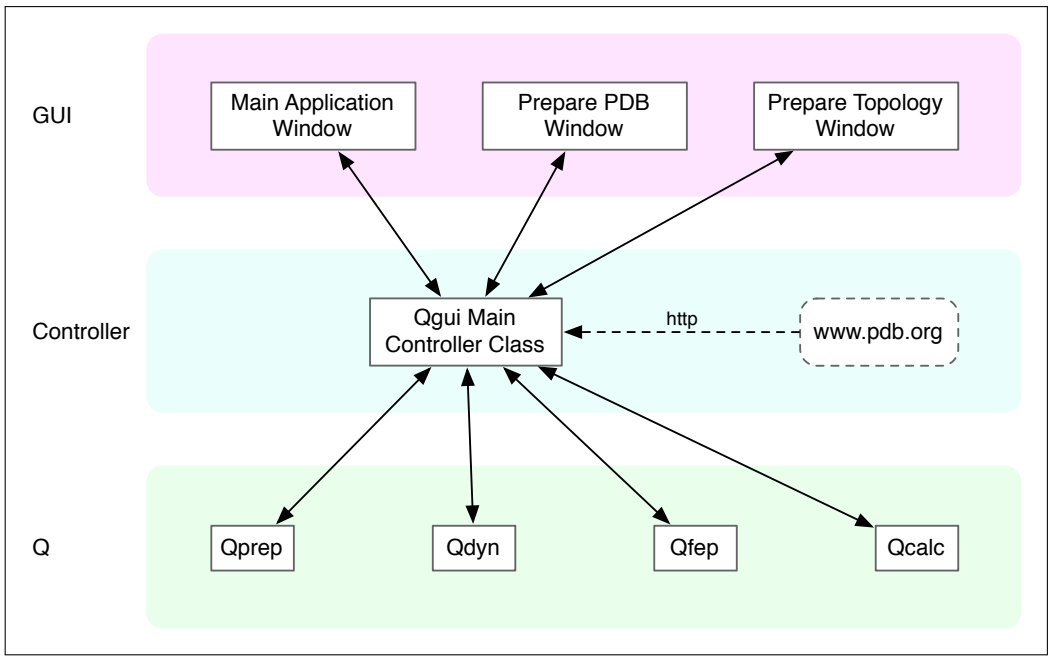


Figure 2.1: The architecture of Qgui represented as layers. Qgui is centered around the Qgui Main Controller Class which communicates with the other components: Main Application Window, Prepare PDB Window and Prepare Topology Window. The controller class retrieves PDB structure files from pdb.org. The different Q modules get input files from Qgui and sends back output files.

Chapter 3

Results and Discussion

dUTPase plays a crucial role in DNA metabolism by converting dUTP to dUMP. The dUTP:dUMP ratio is kept low in the cell to prevent the incorporation of uracil into DNA. An accumulation of dUTP can lead to degeneration of the DNA double strand due to excessive excision repair mechanism evocation and ultimately to thymineless cell death [2]. dUTPase is a drug target in the treatment of diseases such as malaria, tuberculosis and cancer, where it has a role of keeping the malign cell alive [58, 62, 13].

TS is also a drug target in cancer treatment [14]. 5-FU derivates have succesfully been used to inhibit TS in a variety of cancer types [14]. The inhibtion of TS results in the reduction of dTTP concentration in the cell, and can therefore also lead to thymineless cell death. Unfortunately drug resistance is found in a large percentage of tumors [17]. The inhibition of human dUTPase in combination with 5-FU has been shown to be a promising strategy in cancer treatment and to address the resistance issue of 5-FU treatment [18].

New dUTPase inhibitors with low nanomolar IC_{50} and EC_{50} values were recently reported for cancer treatment [18]. This makes the time right to use the accumulated knowledge about dUTPase and its inhibition to explore new ligands to be used in dUTPase inhibition.

The present study tries to develop a computational model to predict binding free energies based on available experimental data [18]. Compounds 9-26 exhibit medium to strong binding to the dUTPase enzyme and compound 26 shows high in vivo activity when combined with 5-FU treatment [18].

This chapter will first discuss the ligands constituting the training set, then the molecular dynamics simulations, and the free energy calculations. Finally, the quality and selectivity of the model, in addition to the novel ligands, are discussed.

3.1 Ligands Examined

The ligands examined in our case were compounds 9, 10, 11, 12, 13, 14, 15, 16, 24, 25R, 25S and 26 [18] (Table 3.1 and Figure 3.1). Ligands 9-16 were built by substituting hydrogen atoms from the phenyl ring at ortho, meta and/or para positions. The substituents were chlor, methyl, methoxy and cyclopropylmethoxy. Ligands 24-26 were built based on the structure of compound 16. The link structure in ligands 24-26 was modified by changing the place of the sulphur dioxide group. The diversity in compounds 24-26 was created by substituting one of the methyl groups preceding the phenyl ring with a hydrogen, and also adding a fluor atom at the phenyl ring at para position.

Table 3.1: Substitution elements and IC_{50} values for compounds 9, 10, 11, 12, 14, 15, 16, 24, 25S, 25R and 26.

Compound	R_1 [a]	R_2	chirality	X	$IC_{50}(\mu M)$ [b]
9	H				3.9
10	o-CH_3O, p-Me				19.5
11	m-CH_3O				1.2
12	p-CH_3				18.2
13	o-Cl				9.4
14	m-Cl				4.5
15	p-Cl				5.7
16	m-cyclopropylmethoxy				0.035
24	Me	Me		H	0.34
25S	H	Me	S	H	2.5
25R	Me	H	R	H	0.040
26	Me	H	R	F	0.021

[a] o, m and p denotes ortho, meta and para, respectively.
[b] [18]

In compounds 10, 11, 13, 14, 16, 24, 25R, 25S and 26 the substitutions at ortho or meta positions on the phenyl ring could be made at carbon 15 and

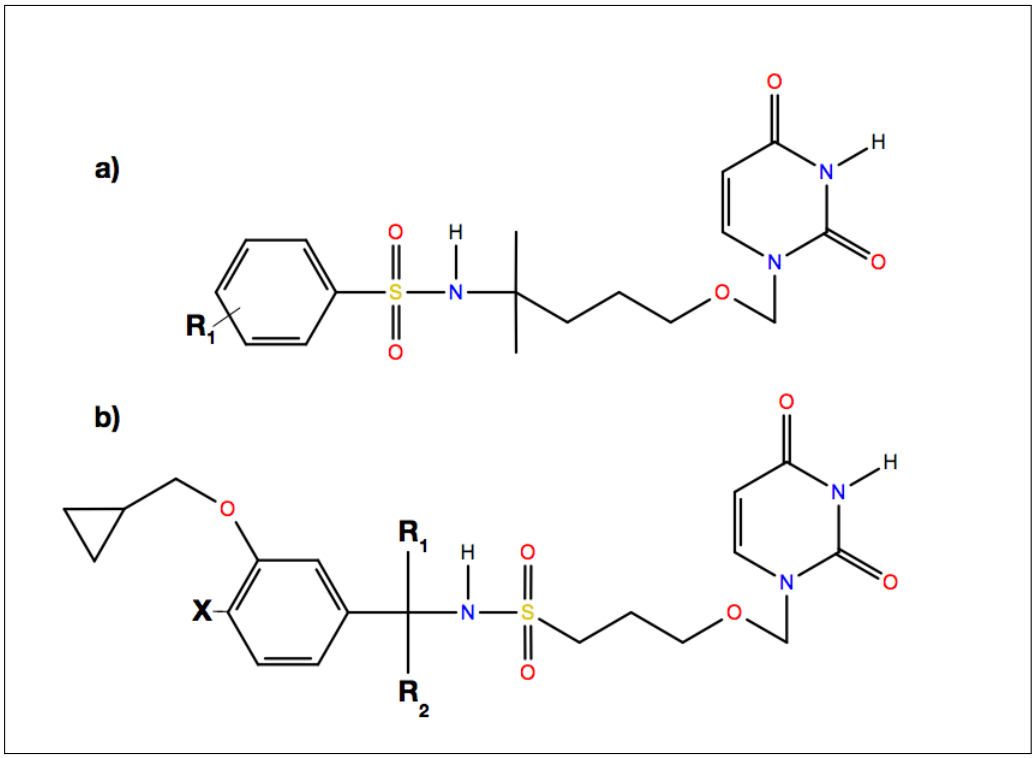


Figure 3.1: The chemical structures for a) ligands 9-16 and b) ligands 24-26. The labels R_1, R_2 and X are the substitution sites for new chemical groups on the ligand, which are listed in Table 3.1

16, or at carbon 18 and 19 (Figure 3.2).

3.2 Linear Interaction Energy

The LIE method was used to calculate the binding free energy of the ligands, at the enzyme active site. This calculation involves doing MD simulations for the ligand in water and in the enzyme active site, finding the electrostatic and the van der Waals energies. The difference in these energies is then used to estimate the binding free energy. This is summarized in equation 3.1[35].

$$\Delta G_{bind} = \alpha \Delta \langle V_{l-s}^{vdW} \rangle + \beta \Delta \langle V_{l-s}^{el} \rangle + \gamma \tag{3.1}$$

where $\langle\rangle$ denotes ensemble averages and l-s denotes ligand-surrounding.

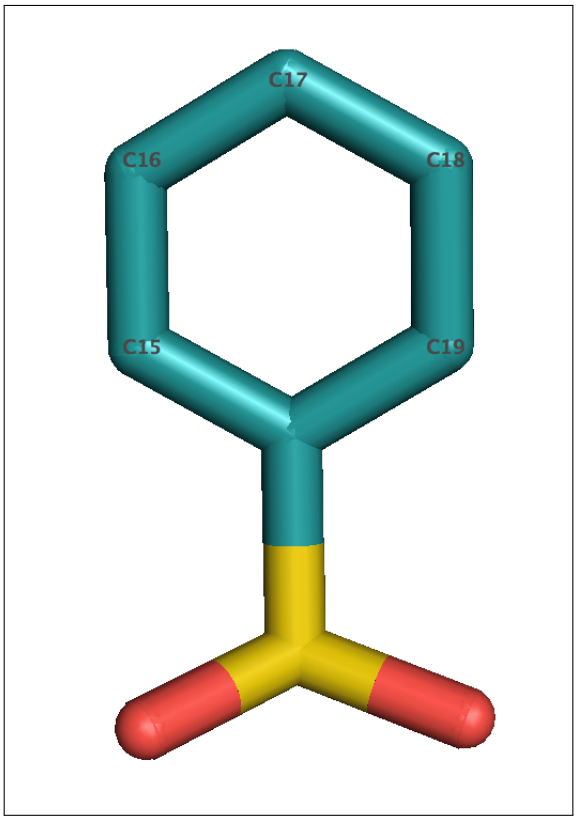


Figure 3.2: The numbering of carbons on the phenyl ring in compound number 9.

The first step was to create a protocol to calculate the binding free energies for compounds 9-26 with the LIE method.

3.2.1 Comparison of Experimental and Calculated Binding Free Energies

Three different LIE sets were constructed for the same series of ligands, by substituting new elements:

1. at the phenyl ring on the side facing away from the enzyme (substitutions at carbon 15, 16 and 17 in Figure 3.2)

2. at the phenyl ring on the side facing the enzyme (substitutions at carbon 17, 18 and 19 in Figure 3.2)

3. at the same position as in set 2, but changing the Lys 318 rotamer in the enzyme

These three variations in the ligand series are referred to as sets 1, 2, and 3, respectively. The parameters used for the ligand series 9-26 were α=0.18 and β=0.43. Previous estimations of β indicate that values lower than 0.5 is a good approximations for ligands which deviate from the linear response approximation [63]. Ligands are divided into four classes with regards to the β value: charged(0.5), neutral (0.43), neutral with one hydroxyl group (0.37) and neutral with two or more hydroxyl groups (0.33). The deviations due to the addition of a hydroxyl group reflect the strong hydrogen bonds the hydroxyl groups are able the create with the surroundings. To improve the predicted binding free energies it was necessary to add the γ parameter to the LIE equation. Slightly different values of γ was generated for each of the sets: γ_1=-7.48, γ_2=-7.30 and $\gamma_3 = -7.24$ for modifications 1, 2 and 3, respectively. Similar γ values for each of the sets indicates that the LIE model used is robust with respect to the parametrization.

Table 3.2: The EL and vdW energies, and the predicted and experimental binding free energies for compounds 9-26 given in kcal/mol. For every compound there are three different LIE calculations which are variations of the ligand structure and/or the enzyme structure.

Compound	ΔEL	$\beta\Delta EL$	ΔvdW	$\alpha\Delta vdW$	ΔG_{calc} [a]	ΔG_{exp} [b] [18]
9_1	7.45	3.20	-19.08	-3.43	-7.71	-7.68
9_2	7.45	3.20	-19.08	-3.43	-7.53	-7.68
9_3	7.32	3.15	-18.23	-3.28	-7.38	-7.68
10_1	7.87	3.38	-19.62	-3.53	-7.63	-6.68
10_2	9.57	4.12	-20.71	-3.73	-6.91	-6.68
10_3	7.42	3.19	-21.60	-3.89	-7.94	-6.68
11_1	7.16	3.08	-19.99	-3.60	-8.00	-8.40
11_2	7.02	3.02	-18.56	-3.34	-7.62	-8.40
11_3	7.23	3.11	-19.91	-3.58	-7.72	-8.40
12_1	5.85	2.52	-18.59	-3.35	-8.31	-6.73
12_2	5.85	2.52	-18.59	-3.35	-8.13	-6.73
12_3	7.11	3.06	-20.13	-3.62	-7.81	-6.73
13_1	8.04	3.46	-18.65	-3.36	-7.38	-7.13
13_2	7.82	3.36	-20.64	-3.72	-7.65	-7.13
13_3	6.09	2.62	-19.30	-3.47	-8.10	-7.13
14_1	7.19	3.09	-18.96	-3.41	-7.81	-7.58
14_2	6.13	2.64	-20.04	-3.61	-8.27	-7.58
14_3	7.92	3.41	-19.78	-3.56	-7.40	-7.58
15_1	5.62	2.84	-20.24	-3.64	-8.71	-7.44
15_2	5.62	2.84	-20.24	-3.64	-8.52	-7.44
15_3	5.61	2.41	-20.29	-3.65	-8.47	-7.44
16_1	9.09	3.91	-23.07	-4.15	-7.73	-10.58
16_2	5.69	2.45	-21.29	-3.83	-8.68	-10.58
16_3	4.47	1.92	-23.84	-4.29	-9.61	-10.58
24_1	5.92	2.55	-19.95	-3.59	-8.53	-9.18
24_2	6.43	2.76	-23.60	-4.25	-8.78	-9.18
24_3	5.38	2.31	-23.48	-4.23	-9.16	-9.18
$25S_1$	3.05	1.31	-20.89	-3.76	-9.93	-7.95
$25S_2$	5.45	2.34	-22.27	-4.01	-8.96	-7.95
$25S_3$	7.53	3.24	-21.78	-3.92	-7.93	-7.95
$25R_1$	4.77	2.05	-21.93	-3.95	-9.38	-10.50
$25R_2$	3.45	1.48	-23.01	-4.14	-9.96	-10.50
$25R_3$	3.74	1.61	-23.96	-4.31	-9.95	-10.50
26_1	4.44	1.91	-22.41	-4.03	-9.61	-10.89
26_2	4.23	1.82	-23.59	-4.25	-9.73	-10.89
26_3	5.30	2.28	-24.06	-4.33	-9.29	-10.89

[a] The free binding energy is calculated with the LIE equation $\Delta G_{calc} = \alpha\Delta\langle V_{l-s}^{vdW}\rangle + \beta\Delta\langle V_{l-s}^{el}\rangle + \gamma$. Parameters used in the LIE fitting are $\alpha = 0.18$, $\beta = 0.43$ [63], $\gamma_1 = -7.48$, $\gamma_2 = -7.30$ and $\gamma_3 = -7.24$.

[b] $\Delta G_{exp} = RTlnK_i$ where $K_i \approx IC_{50}$ [18].

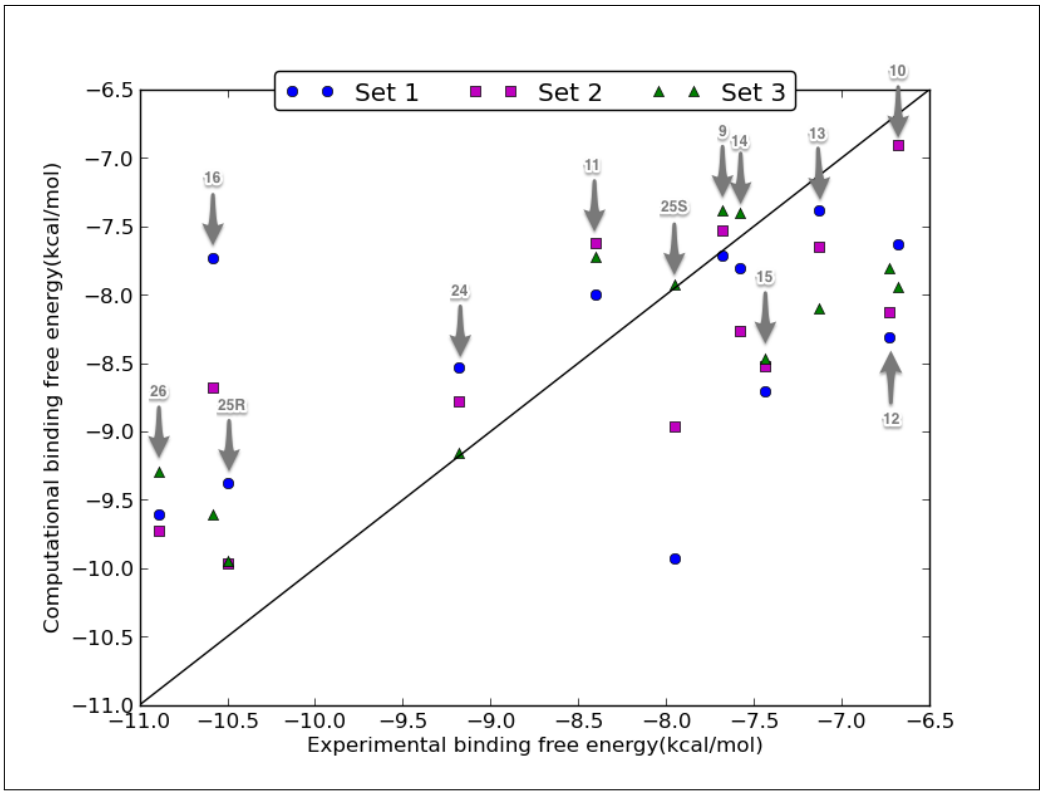


Figure 3.3: The distribution of binding free energies from the LIE calculations for modifications 1, 2 and 3. The gray arrows point out every three sets of simulations for compounds 9-26.

The computational binding free energies were generally in close agreement with the experimental binding free energies (Table 3.2 and Figure 3.3). When the computational binding free energies differ with less then 1.5 kcal/mol from the experimental energies, the agreement is considered to be good. All binding affinities were dominated by vdW contributions. The vdW contributions ranged from -5.3 kcal/mol to -3.3 kcal/mol and the EL energies ranged from 1.3 kcal/mol to 4.1 kcal/mol. It can be noted that the vdW energies decrease as the ligand increases in size and the EL energies became less unfavorable for the best ligands.

Modification 2 of the ligands turned out to create a better agreement than modification 1 with the experimental binding data for 8 out of 12 ligands (Table 3.2). The substitutions on the phenyl ring in modification 2 are placed to fit in a cavity on the surface of the protein (Figure 3.4) which might be a part of the reason for the better binding energies. Modification 3 of the ligands was created by taking variation 2 ligand structure and changing Lys

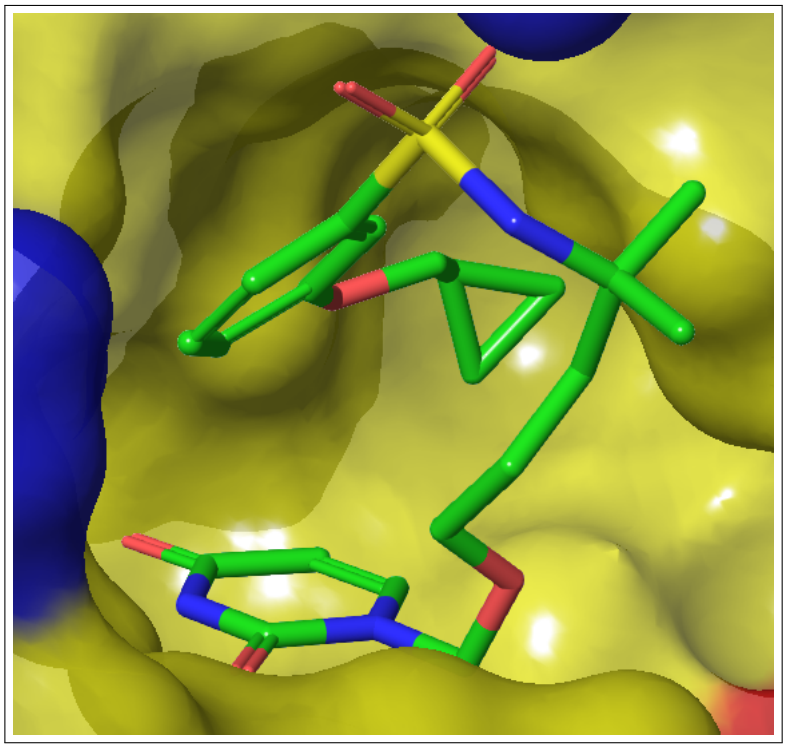


Figure 3.4: 16_2 *shown with the cavity where the side group of the phenyl ring fits well.*

318 rotamer in the enzyme. This modification resulted in the best agreement with the experimental data in 6 out of 12 compounds. The magnitude of the difference between the calculated and experimental energies varied between a minimum of 0.02 kcal/mol ($25S_3$) and a maximum of 2.85 kcal/mol (16_1).

The LIE simulations successfully estimated the binding free energy for ligands 9, 10, 11, 13, 14, 24 and 25R. All three modifications of the structures gave good agreement with the experimental binding data. This indicates that the structure of these compounds were able to capture the binding modes for these ligands.

Compound 9 had the best agreement with the experimental binding data. The LIE simulations for compound 9 yielded binding free energies of -7.7 kcal/mol, -7.5 kcal/mol and -7.4 kcal/mol for 1, 2 and 3, respectively. The experimental binding free energy is -7.7 kcal/mol. The conformation of compound 9 stays approximately the same throughout the simulations (Figure 3.5), and it is therefore not surprising that the LIE method was able to reproduce binding energies for 9_1, 9_2 and 9_3.

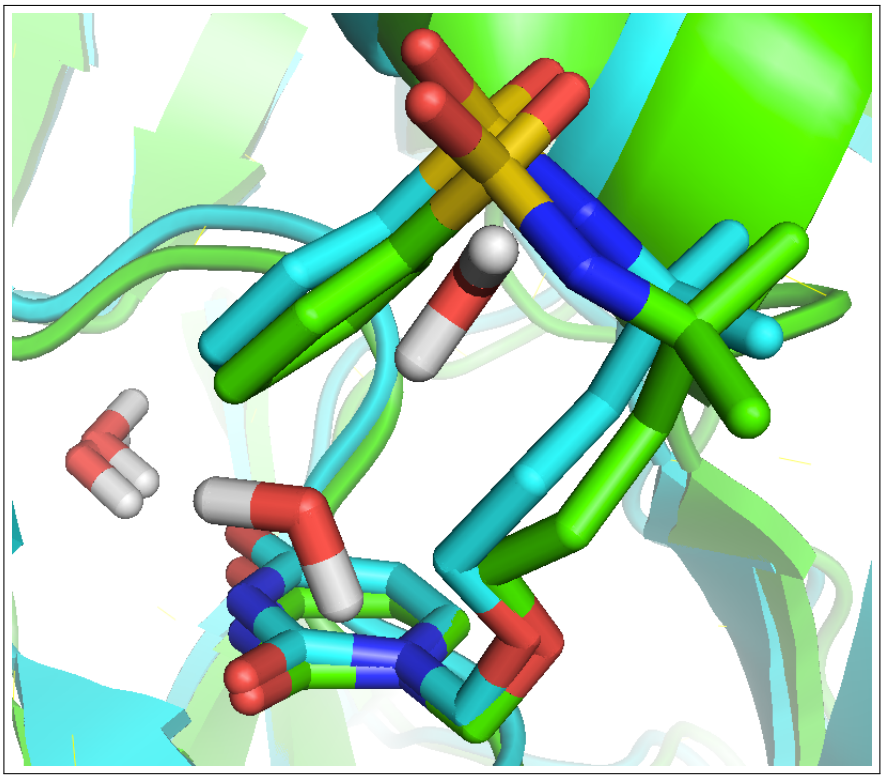


Figure 3.5: Compound 9 ligand structure from the crystal structure 3ARN (cyan) and a MD simulation snapshot (green).

The LIE method did not produce good binding energies for all or some of the modifications of compounds 12, 15, 16, 25S and 26. It is reasonable to assume that this can be related to the binding mode of the ligand in the complex.

LIE calculations for compound 16 yielded the following binding free energies for modifications 1, 2 and 3: -7.7 kcal/mol, -8.7 kcal/mol and -9.6 kcal/mol, respectively. The experimental binding free energy is -10.6 kcal/mol. The energies improved and became less positive with modification 1, 2 and 3. The electrostatic contributions to the the binding free energy (Table 3.2) changes from 3.9 kcal/mol to 1.9 kcal/mol, when progressing from 9_1 to 9_3. This indicates that there is either more EL attraction in 9_3 or that there is more EL repulsion in 9_1, or a combination of both.

Figure 3.6 illustrates that there is a prominent repulsion from Lys 44 in 16_1 with the magnitude of 1.1 kcal/mol. The same repulsion is present in 16_3 with a smaller value 0.3 kcal/mol. The orientation of the ligand and

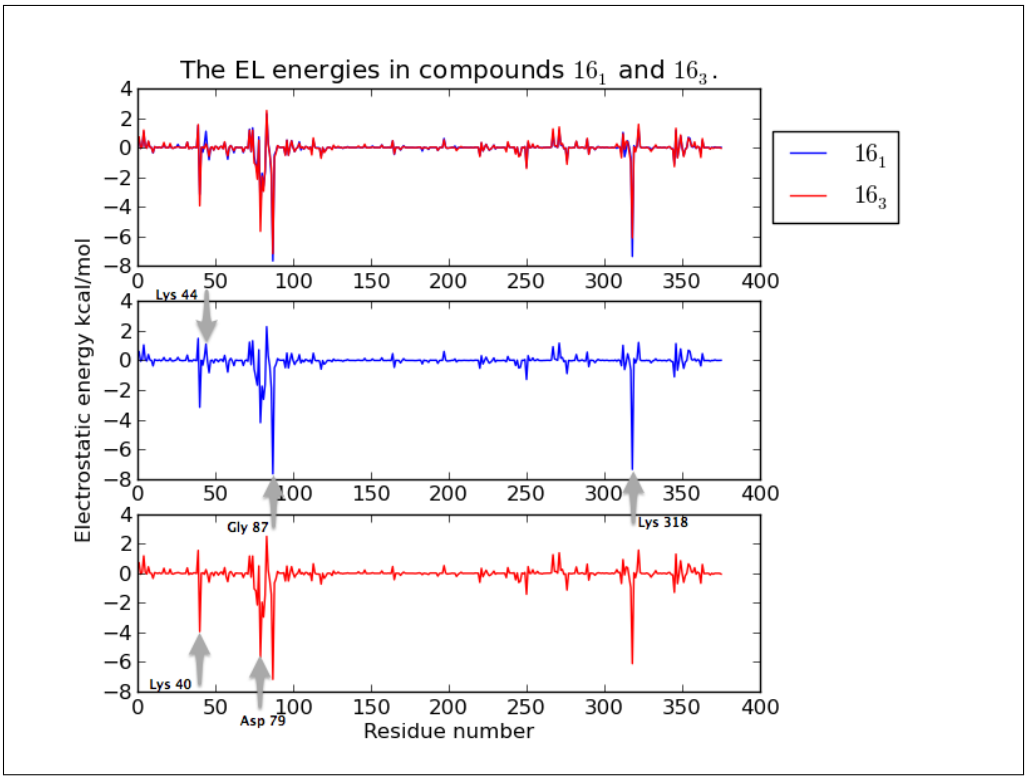


Figure 3.6: The electrostatic contributions to the binding free energy from each residue in compounds 16_1 *(blue) and* 16_3 *(red).*

Lys 44 in the active site illustrates that the methoxy group of the ligand changes conformation during the simulation (Figure 3.7 (a)). Lys 44 points further away from the uracil ring, changes orientation and seeks away from the methoxy group. The amino group of Lys 44 can be the cause of the repulsion against the hydrophobic tail of the methoxy group. The repulsion from Lys 44 diminishes as the methoxy group is changed to the other side of the phenyl ring.

Other residues that might explain the difference in electrostatic energies between 16_1 and 16_3 are Lys 40 and Asp 79. Lys 40 contributed with -3.7 kcal/mol and -8.9 kcal/mol for the polar part of the binding free energy for 16_1 and 16_3. Asp 79 contributed with -4.2 kcal/mol to 16_1 electrostatic energy and with -5.7 kcal/mol to 16_3 electrostatic energy. 16_1 has more favorable EL interactions with Gly 87 and Lys 318, but these contributions do not compensate the EL contributions that favor modification 16_3.

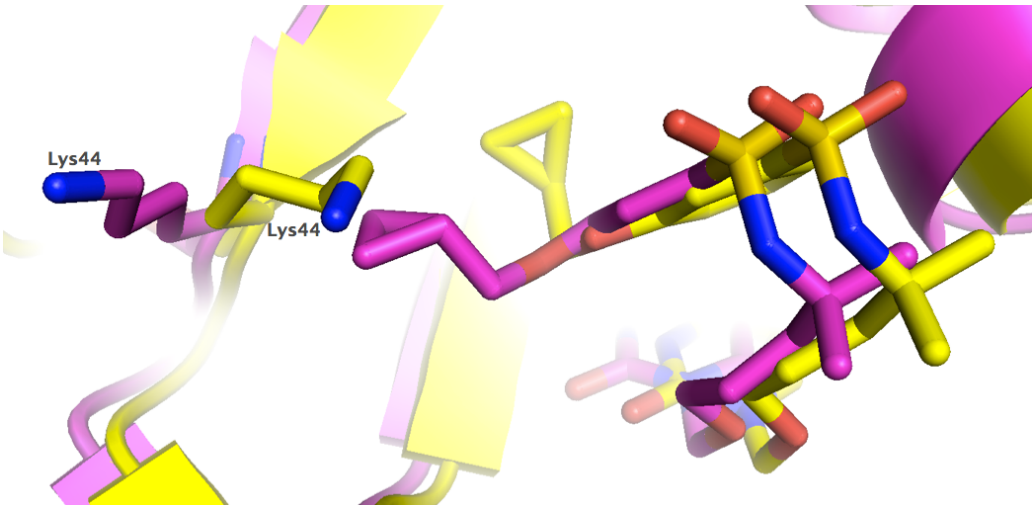
The predicted starting structure for modification 16_3 in the LIE simulations is very similar to the observed conformation later in the MD trajectory (Figure 3.7 (b)). Based on the conformational similarity during the simulation and previous observations it seems that the effort made to improve the initial structure of compound 16 from modifications 1 and 2 was successful.

Compound 26 was the leading compound for Miyahara *et al.* [18] with the best binding affinity to human dUTPase and greatest in vitro potency in the study. The LIE calculations for compound 26 gave binding free energies of -9.6 kcal/mol, -9.7 kcal/mol and -9.3 kcal/mol for variation 1, 2 and 3, respectively. The experimental binding free energy is -10.9 kcal/mol. All of the variations underpredict the binding energy, and the best result was obtained with 26_2.

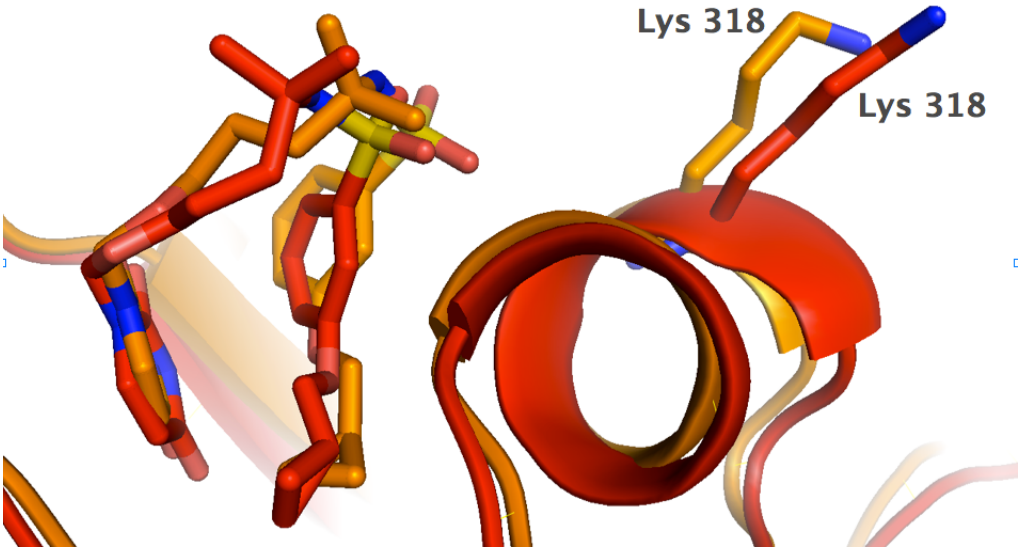
Free energy calculations for compound $25R$ resulted in energies of -9.4 kcal/-mol, -10.0 kcal/mol and -10.0 kcal/mol for modifications 1, 2 and 3, respectively, which compares rather well with the experimental binding free energy of -10.5 kcal/mol.

Experimentally, compound 26 has stronger binding affinity compared to $25R$. In the calculated binding affinities 26 was a better binder in set 1. The only difference between $25R$ and 26 is a fluor atom at para position on the phenyl ring in 26. According to the experimental data, the fluor atom appears to enhance the binding affinity. Yet, the opposite effect is observed in 26_2 and 26_3. The 26_1 ligand experiences more repulsion from residues such as Lys 44 and Asp 322 compared to $25R_1$ ligand (Figure B.1). The electrostatic interaction energy was 2.1 kcal/mol and 0.6 kcal/mol for Lys 44 and Asp 322, respectively, with the $25R_1$ ligand. For the 26_1 ligand the electrostatic interactions were 3.2 kcal/mol and 1.5 kcal/mol with Lys 44 and Asp 322, respectively. Despite the higher electrostatic repulsions, 26_1 has stronger favorable electrostatic interactions compared to $25R_1$ and results in 26_1 being a better binder (Figure B.1).

$25R_2$ resulted in higher binding affinity in comparison with 26_2 from the LIE simulations. Change in the position of cyclopropylmethoxy substitution in $25R_2$ resulted in an increase of electrostatic attractions from residues Asp 79 and Lys 318 (Figure 3.8). The electrostatic contributions from Asp 79 changed from -2.4 kcal/mol to -5.8 kcal/mol for $25R$ when going from modification 1 to 2. Lys 318 electrostatic interaction changed from -2.2 kcal/mol to -4.2 kcal/mol, respectively, for $25R$ in modification 1 and 2. Repulsion from Lys 44 was reduced for both modification $25R_2$ and 26_2 (Figures 3.8 and 3.9). The electrostatic interaction from Lys 44 was 0.8 kcal/mol and 1.3



(a)



(b)

Figure 3.7: (a)Ligand and Lys 44 orientation in the starting structure (yellow) and in a snap shot of the MD simulation (magenta), for compound 16_1*. (b) Ligand and Lys 318 orientation in the starting structure (orange) and in a snapshot of the MD simulation (red), for compound* 16_3*.*

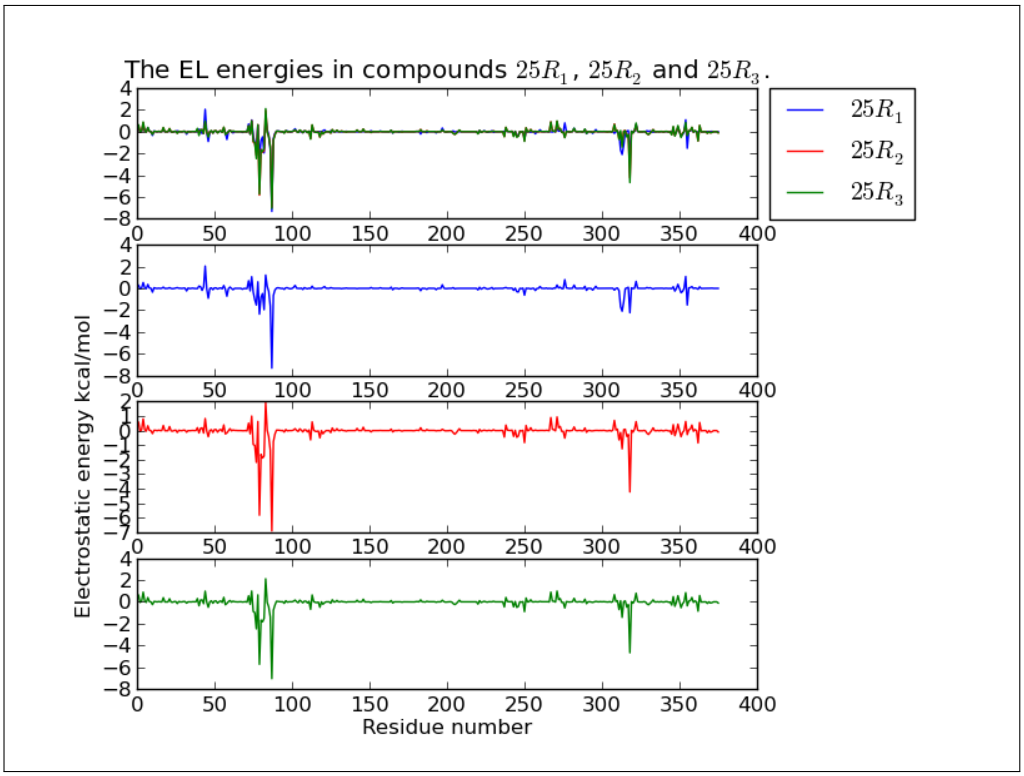


Figure 3.8: A comparison of the electrostatic energies from the LIE simulations for $25R_1$, $25R_2$ and $25R_3$.

kcal/mol, respectively, in $25R_2$ and 26_2. The attractive electrostatic forces from Lys 318 seem to weaken in modification 26_2 when compared to 26_1 (Figure 3.9). The electrostatic interaction from Lys 318 was -6.1 kcal/mol and -3.4 kcal/mol, respectively, in 26_1 and 26_2. Added steric hindrance from the cyclopropylmethoxy group can be the cause of this effect.

The LIE simulations yielded lower binding free energy for $25R_3$ than for 26_3. The change in the rotamer of Lys 318 gave less EL attraction for modification 26_3. Lys 318 EL interactions were more favorable in modification $25R_3$ when compared to 26_3 (Figure B.3). The electrostatic contribution from Lys 318 was -4.7 kcal/mol and -2.6 kcal/mol for $25R_3$ and 26_3, respectively. The loss of Lys 318 interactions in modification 26_3 resulted in less favorable binding free energy. There was no change in the binding energy of 25_3.

The modifications of $25R$ and 26 were not able to capture the true binding modes of the ligands. This may be due to the orientation of ligand 26, since the fluor atom fails to enhance the binding affinity over compound $25R$.

There is also a possibility that the force field parameters for the fluor are not optimal and results in too low attractive forces.

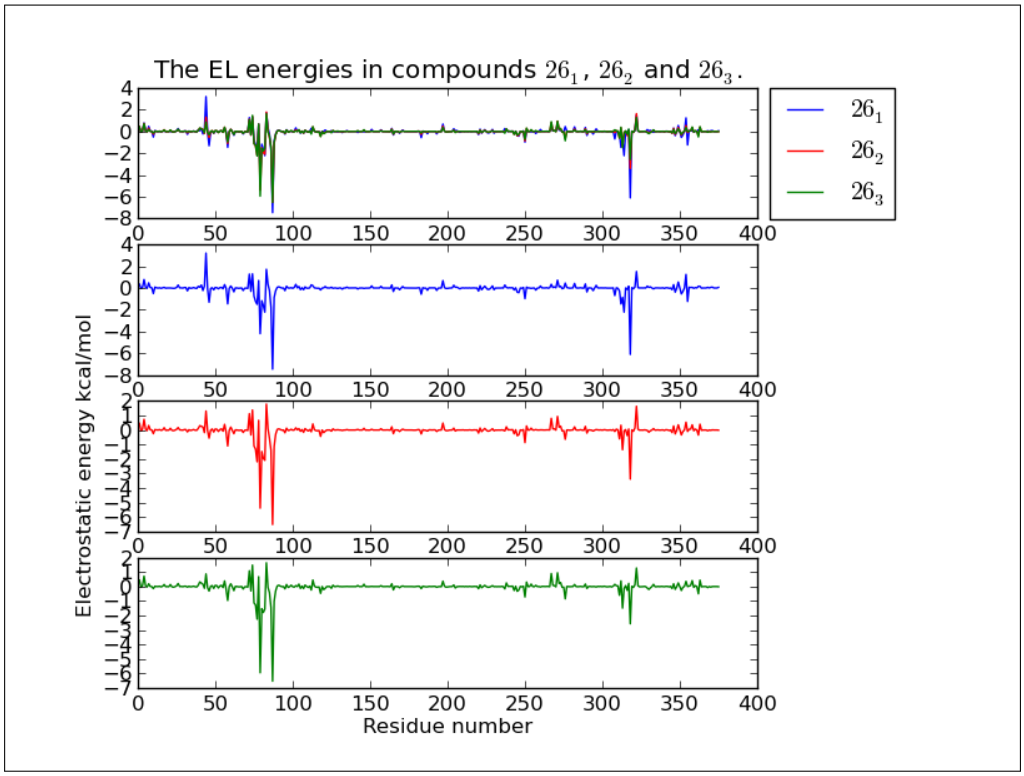
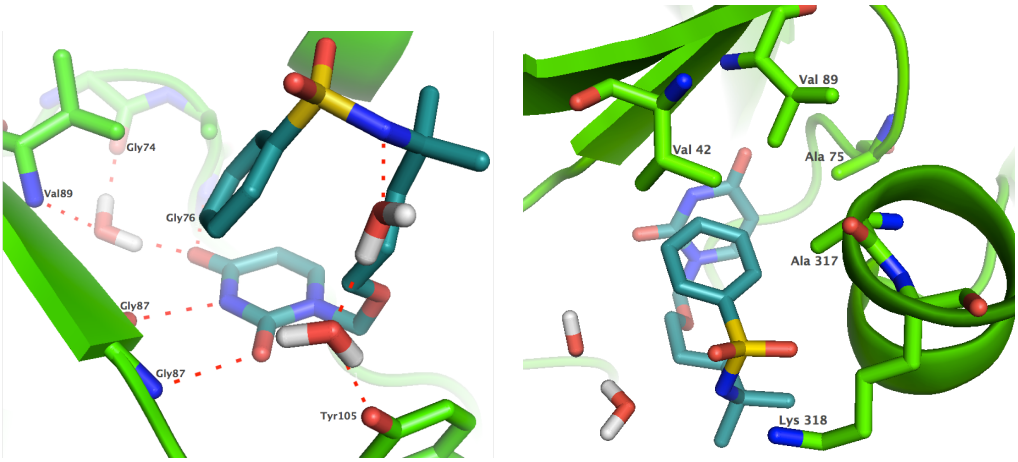


Figure 3.9: A comparison of the electrostatic energies from the LIE simulations for 26_1, 26_2 *and* 26_3.

3.2.2 Main Contributors to the Difference in Binding Free Energies Between the Structures

The binding mode of compound 9 in the human dUTPase active site was obtained from the crystallographic model 3ARN [18]. The hydrogen bonds to the ligand originate from residues Gly 74 (mediated by HOH 378), Gly 76, Tyr 82 (mediated by HOH 379 and HOH 382), Gly 87 and Val 89 (mediated by HOH 378) (Figure 3.10 (a)). In the crystallographic structure there is also a hydrophobic region consisting of residues Val 42, Ala 75, Val 89 and Ala 317 (Figure 3.10 (b)) which interact with the phenyl ring of the ligand.

The hydrogen bonding pattern was to a large extent captured in the MD simulations. The EL interactions appear stable during the simulations (Table



(a) (b)

Figure 3.10: (a)The hydrogen bonding pattern in the active site of 3ARN with compound 9 bound to it (ligand in blue), and (b) residues contributing to a hydrophobic region near the ligand phenyl ring. Lys 318 is also displayed.

C.1). However, the simulations failed to reproduce the two water mediated hydrogen bonds between the hydroxyl group of Tyr 82 and the sulfonamide group of the ligand. HOH 379 and 382 quickly disappeared in the MD simulations from the ligand proximity in all compounds. There was no permanent substitution from other waters.

vdW interactions from residues Val 42 and 89, and Ala 75 and 317 were present in the MD simulations (Table 3.3). There were only small changes in the vdW energies and most of these can be explained from the changes made in the ligand structures. In compound $25R$ the change of the position of the cyclopropylmethoxy on the phenyl ring made a difference in the vdW energies. Val 42, for example, is situated in a position that it is closer to the cyclopropylmethoxy group of the ligand in modification 1 compared with modification 2. This is reflected in the vdW energies which change from -2.5 kcal/mol to -1.5 kcal/mol for $25R_1$ and $25R_2$, respectively.

Table 3.3: The van der Waals energies for selected residues from the LIE simulations with ligands 9-26

Compound	Val 42	Ala 75	Val 89	Ala 317
9_1	-1.79	-2.18	-1.86	-1.49
9_3	-1.85	-1.68	-1.85	-1.64
10_1	-1.99	-1.71	-1.94	-1.57
10_2	-2.61	-1.90	-1.73	-1.39
10_3	-2.40	-1.91	-2.04	-1.36
11_1	-2.11	-1.76	-1.85	-1.53
11_2	-2.17	-1.79	-1.76	-1.65
11_3	-2.15	-2.00	-1.85	-1.55
12_1	-1.76	-1.68	-1.72	-1.26
12_3	-1.71	-2.13	-2.07	-1.38
13_1	-1.93	-1.59	-1.76	-1.69
13_2	-2.23	-1.99	-2.04	-1.58
13_3	-2.37	-1.23	-2.05	-1.52
14_1	-2.05	-1.39	-1.87	-1.58
14_2	-1.70	-1.81	-1.69	-1.70
14_3	-1.80	-1.75	-1.69	-1.77
15_1	-1.74	-1.89	-1.99	-1.53
15_3	-1.84	-1.92	-2.02	-1.55
16_1	-2.69	-1.99	-1.94	-1.74
16_2	-1.24	-3.04	-1.99	-2.20
16_3	-1.16	-3.64	-2.01	-2.34
24_1	-2.32	-1.92	-1.84	-1.48
24_2	-1.79	-3.02	-2.30	-2.57
24_3	-1.88	-3.21	-2.34	-2.42
$25S_1$	-2.39	-1.02	-1.81	-1.50
$25S_2$	-1.32	-3.21	-2.08	-2.10
$25S_3$	-1.46	-2.99	-2.24	-2.16
$25R_1$	-2.47	-1.68	-1.89	-1.64
$25R_2$	-1.54	-3.45	-2.04	-2.14
$25R_3$	-1.74	-3.53	-2.15	-2.24
26_1	-2.54	-1.73	-2.03	-1.79
26_2	-1.43	-3.41	-1.94	-2.27
26_3	-1.35	-3.36	-1.94	-2.24

There were also prominent contributions to the electrostatic energy from other residues that were not pointed out from 3ARN crystal structure [18] (Table 3.4). Asp 79 and Lys 318 stand out with large contributions to the electrostatic energy (Figure D.1). The side chain nitrogen of lysine was interacting with the SO_2 oxygens of the ligand. The EL attractions from Lys 318 were weaker in structures 24-26. This is most likely due to the ligand backbone having a different conformation and the SO_2 oxygens were placed further away from the positively charged nitrogen. There is uncertainties with respect to the validity of the Lys 318 interactions. The Lys 318 rotamer was changed so that it pointed away from the ligand and new LIE simulations were run with the altered structures.

3.2.3 The Effect of Changing the Lysine 318 Rotamer State on the Electrostatic Contribution to the Binding Free Energy

Lys 318 was found to be one of the major contributors to the EL part of the binding energy. Since the contribution was not described as important for the ligand binding [18] there was a question of its importance. The electron density was poor for the Lys 318 side chain (Figure 3.11). This gave us reason to doubt the orientation of Lys 318 side chain in the crystal structure. Crystal packing can also have an effect on how the Lys 318 would orient in the crystal and might disturb the real orientation. A new series of LIE simulations was run with Lys 318 in a different rotameric state.

In the new simulations, for 6 out of the 12 compounds, the interactions with Lys 318 were less favorable (Table 3.4). This resulted in improved agreement for some of the compounds. For example, the binding free energy of compound 9 changed from -7.5 kcal/mol in modification 2, to -7.4 kcal/mol in modification 3. The EL Lys 318 interaction with the ligand changed from -6.5 kcal/mol in modification 2 to -6.2 kcal/mol in modification 3. Nitrogen and the sulfure oxide oxygens moved further away from each other after the rotamer modification. The distance between the nitrogen and $O4$ changed from 7.4 Å in modification 2 to 8.1 Å in modification 3. The same was observed for the distance between the nitrogen and $O5$ where the distance changed from 6.9 Å i modification 2 to 7.5 Å in modification 3. The change in the distance between the sulfur oxide oxygens and the side chain nitrogen appeared to correlate with the change in the binding free energies.

Table 3.4: EL energies for selected residues from the LIE simulations with ligands 9-26

Compound	Asp 79	Lys 318
9_1	-5.34	-6.47
9_3	-5.09	-6.22
10_1	-4.79	-10.59
10_2	-4.61	-6.73
10_3	-4.58	-8.34
11_1	-5.27	-7.03
11_2	-4.84	-3.73
11_3	-5.52	-4.98
12_1	-4.93	-6.52
12_3	-5.17	-6.92
13_1	-5.29	-6.94
13_2	-4.50	-4.83
13_3	-4.80	-3.71
14_1	-5.95	-5.00
14_2	-4.95	-7.15
14_3	-4.51	-7.53
15_1	-5.57	-8.89
15_3	-5.19	-6.78
16_1	-4.21	-7.37
16_2	-5.73	-6.89
16_3	-5.67	-6.11
24_1	-2.21	-1.67
24_2	-4.82	-3.30
24_3	-4.95	-2.96
$25S_1$	-3.29	-3.02
$25S_2$	-5.30	-3.11
$25S_3$	-5.22	-3.32
$25R_1$	-2.37	-2.24
$25R_2$	-5.83	-4.23
$25R_3$	-5.74	-4.68
26_1	-4.20	-6.11
26_2	-5.39	-3.39
26_3	-5.95	-2.57

Regardless of the change in the rotamer, Lys 318 did after a while move closer to the ligand in most of the structures. In fact, for compound $25S$ the opposite was observed compared to compound 9. The computational binding free energy of modification $25S_2$ was -9.0 kcal/mol and for modification $25S_2$ it reduced to -7.9 kcal/mol. The EL interactions with Lys 318 changed from -3.1 kcal/mol to -3.3 kcal/mol in $25S_2$ and $25S_3$, respectively. The distance N-O4 was reduced from 9.5 Å to 9.0 Å in $25S_2$ and $25S_3$, respectively. Similar change occurred in the N-O5 distance. Even though the binding free energy of modification 3 was reduced, the EL contributions became stronger from Lys 318 and the distance shortened between the sulfur oxide and nitrogen.

The cases discussed above show that the structures responded in different ways to the modification of the Lys 318 rotamer.

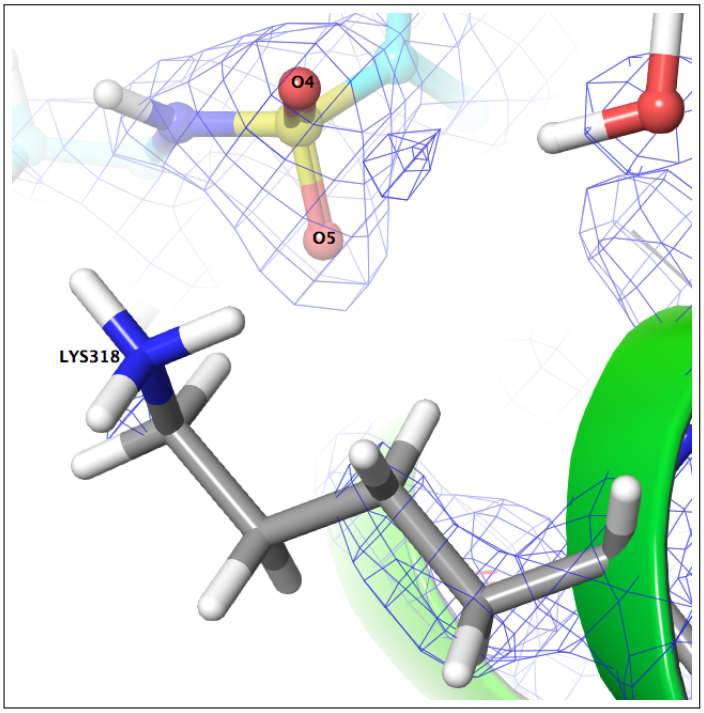
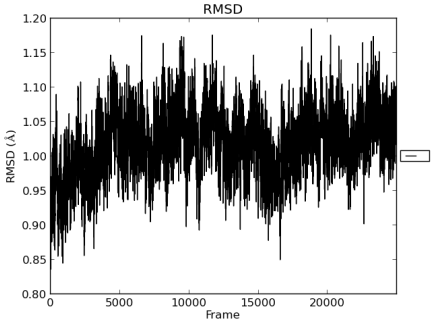


Figure 3.11: Electron density around Lys 318 in 3ARN crystal structure

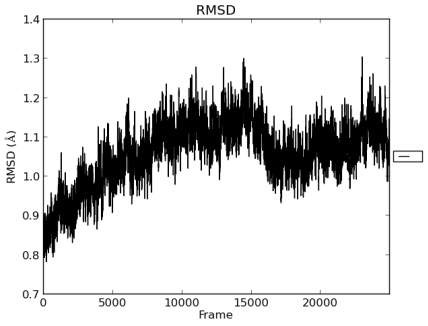
3.2.4 Stability of the MD Simulations

The quality of the LIE simulations was determined with RMSD calculations, statistical parameters from the fitting of the LIE model and the stability of the electrostatic and van der Waals energies during the MD simulations.

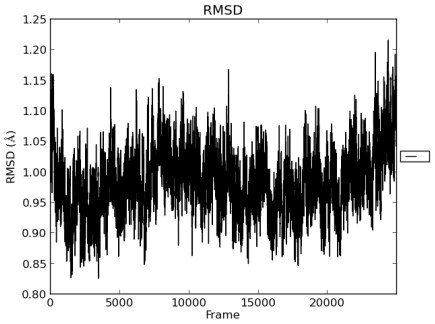
The RMSD gives the average deviation of the enzyme structure at any given frame in the MD directory from the initial structure. RMSD can be calculated from the whole system, the back bone or the side chains of the enzyme. Deviations of 2 Å or smaller are considered to be acceptable for a protein. The RMSD measured for 9_1, 16_1 and 26_1 show that the deviations in the protein bakc bone were less than 1.6 Å (Figures 3.12, E.1 and E.2). Therefore the stability of the MD simulations was considered good.



(a)



(b)



(c)

Figure 3.12: RMSD calculations for 9_1 *from simulations 1, 5 and 10.*

The LIE fitting was done with a lest squares linear regression. The fitting resulted in parameters SSE and COD. The parameters give a measure about the quality of the LIE model. COD gives a measure of the fit of the model. COD can have any value between 0 and 1, and the closer the value is to 1, the better the fit is between the experimental and the computational energies.

COD progressed from 0.6, 0.7 to 0.8 in sets 1, 2 and 3, respectively. The closer the SSE value is to zero the less random error there is in the model. SSE changed from 6.3, 3.5 to 2.5 in sets 1, 2 and 3, respectively. Gradually better values were attained for both parameters when the LIE calculations progressed from set 1 to 3. This change is in agreement with the analysis of the LIE results for sets 1, 2 and 3, where set number 3 created the best overall agreement with the experimental binding free energies.

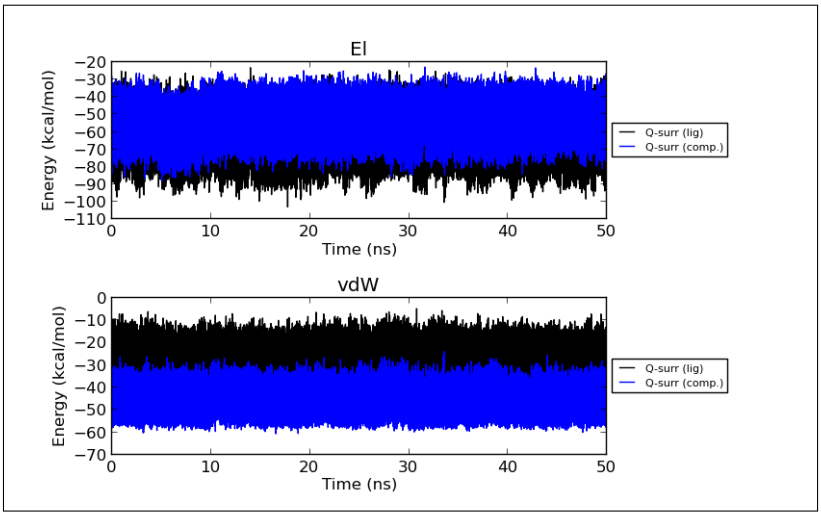


Figure 3.13: The electrostatic and van der Waals energies in MD simulations for the protein in 9_1.

The LIE simulations for every compound were run with 5 ns per simulation, and the simulations were repeated 10 times with an unique random seed. These simulations added up to 50 ns per compound. The electrostatic and van der Waals contributions fluctuate during the MD simulations. What is important though, is that the energies should stay at the same range, or level, through the simulation. In the MD simulations executed for our system both the electrostatic and van der Waals energies converge quickly and are stable as can be observed for ligands 9_1, 16_1 and 26_1 from Figures 3.13, 3.14 and 3.15.

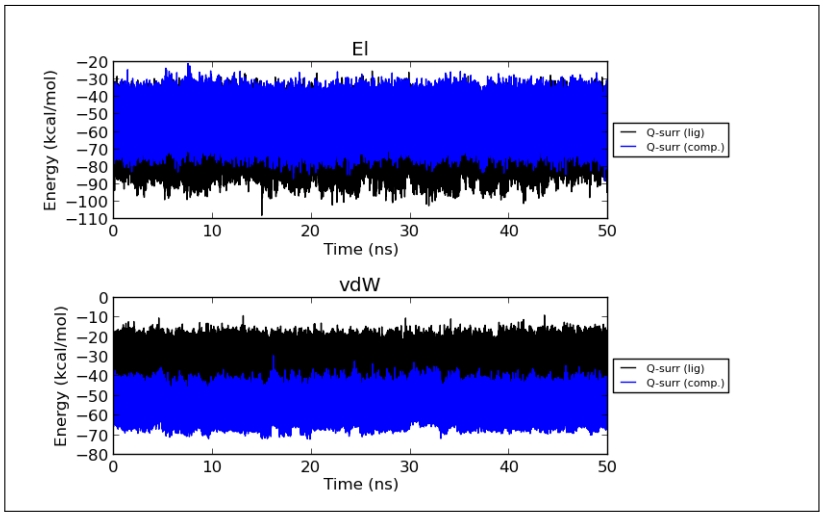


Figure 3.14: The electrostatic and van der Waals energies in MD simulations for the protein in 16_1.

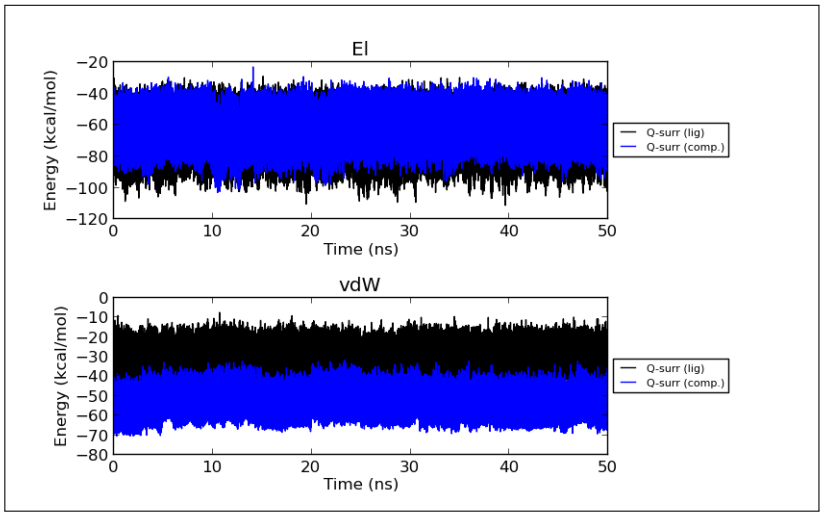


Figure 3.15: The electrostatic and van der Waals energies in MD simulations for the protein in 26_1.

3.3 Selectivity of the LIE Model

3.3.1 *Plasmodium falciparum* dUTPase

dUTPase from *Plasmodium falciparum* represents an ideal system to further test the LIE model derived for the human dUTPase, particularly with respect to selectivity. *Plasmodium falciparum* is a parasite causing malaria in humans. Although there is medication against malaria, there are problems with drug resistance [20]. *Pf*dUTPase has shown to be a good candidate for a drug target in malaria due to low sequence similarity with human dUTPase (28% identity)[21]. Therefore inhibitors that bind well to *Pf*dUTPase have shown poor binding affinities to the human counterpart [21].

Recent potential medicines developed against malaria show both enzymatic and cellular inhibition activity against the *Pf*dUTPase enzyme and *Plasmodium falciparum* culture [21]. At the same time, the same inhibitors show weak enzymatic and cellular inhibition against the human dUTPase and mammalian L6 cells [21]. These inhibitors are deoxyuridine derivatives with a triphenyl group. The selectivity is thought to origin from a difference in a hydrophobic region in the enzyme active site. In *Pf*dUTPase residues Phe 46 and Ile 117 are a part of a hydrophobic pocket and are thought to have interactions with the triphenylmethane group of the ligand. In the human counterpart these residues are replaced with a Val and a Gly. The hypothesis is that the residues in human dUTPase provide less hydrophobic interactions.

3.3.2 Ligands

The ligands 1a, 1b, 1c, 2a, 2b, 3a, 3b, 4, 5a, 5b, 7a, 7b, 14b, 14c, 14d and 14e (Table 3.5) were chosen for docking studies and LIE simulations [21]. The ligands were manually built into the active site of the crystallographic structure of human dUTPase [18] (3ARN). The position of the uracil ring of ligands 1a-14e was based on the crystallographic structure of *Pf*dUTPase accommodating a similar ligand [58] (1VYQ). The position of the uracil ring in the structures 3ARN and 1VYQ was almost identical. The remaining parts of the ligands were constructed to have structural resemblance with the ligand in 1VYQ. It was assumed that the ligands 1a-14e had a similar binding mode both in both enzymes. The most obvious difference between

ligands 1a-14e and 9-26 is the size of the ligands. The *Pf*dUTPase ligands have a large trityl moiety which is much bulkier compared to the phenyl group of the human dUTPase ligand 9. The substitution elements and patter for ligands 1a-14e are shown in Figure 3.16 and Table 3.5.

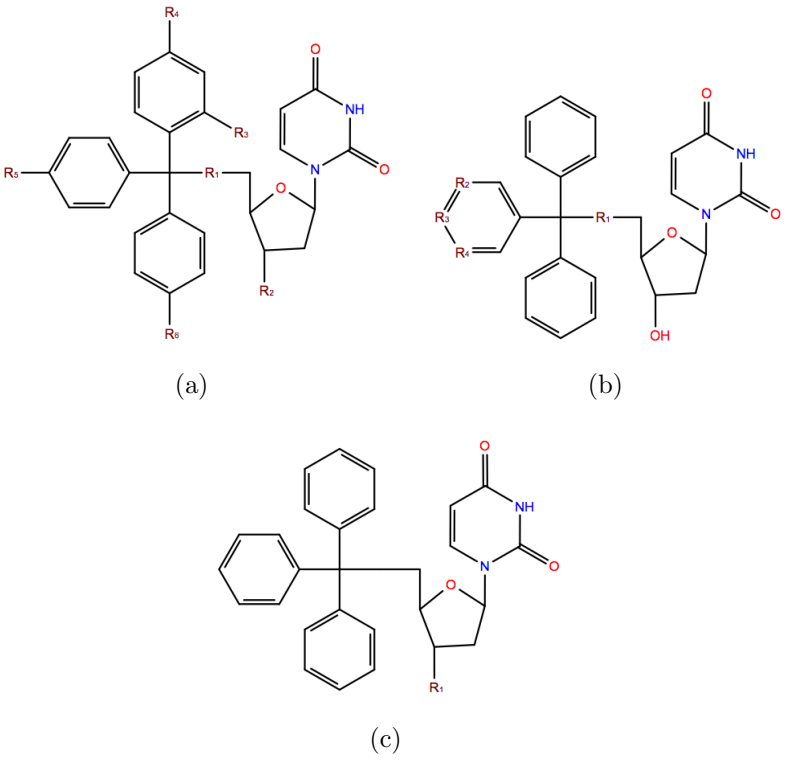


Figure 3.16: Chemical structure and substitution pattern of the malaria ligands. (a) 1a, 1b, 1c, 2a, 2b, 3b, 4, 5a and 5b, (b) 7a, 7b and 7c and c) 14b, 14c, 14d and 14e. Substitutes are given in Table 3.5.

Table 3.5: Substitution elements and K_i values for PfdUTPase ligands 1a-14e. The template of the ligands are shown in Figure 3.16(a), (b) and (c).

Figure (a)	ligand	R_1	R_2	R_3	R_4	R_5	R_6	K_i [a]	K_i [b]
	1a	O	OH	H	H	H	H	1.8	17.7
	1b	NH	OH	H	H	H	H	0.2	46.3
	1c	S	OH	H	H	H	H	4.5	>1000
	2a	O	OH	H	H	H	OMe	5.2	>1000
	2b	NH	OH	H	H	H	OMe	0.7	>374.3
	3a	O	OH	H	H	H	CN	1.1	816.0
	3b	NH	OH	H	H	H	CN	0.4	231.5
	4	O	OH	H	H	H	OMe	2.2	>1000
	5a	O	OH	Cl	H	H	H	2.6	>1000
	5b	NH	OH	Cl	H	H	H	3.4	238

Figure (b)	ligand	R_1	R_2	R_3	R_4	K_i [b]	K_i [c]
	7a	NH	CH	N	CH	0.23	>1000
	7b	O	N	CH	N	1.8	>1000
	7c	NH	N	CH	N	0.98	>1000

Figure (c)	ligand	R_1	K_i [b]	K_i [c]
	14b	N_3	48.1	>1000
	14c	NH_2	1.3	>1000
	14d	NHAc	1.4	>1000
	14e	F	5.0	457.0

[a] *Pf*dUTPase K_i [21]
[b] Human dUTPase K_i [18]

3.3.3 Linear Interaction Energy

The LIE model built for the human dUTPase was developed for an eventual drug design purpose. We wanted to test the selectivity of the human dUTPase LIE model and if it would discriminate the ligands designed for *Pf*dUTPase. The LIE method was applied two times on ligands 1a-14e. In

the first LIE run, LIE_1, the simulations were conducted with the ligands manually built in the active site of human dUTPase, and in the second run, LIE_2, the starting structures were chosen from the docking simulations. The parameters in the LIE calculations were adopted from the human dUTPase model, with the exception of electrostatic scaling constant that was adjusted according the chemical nature of the ligands ($\gamma = 0.37$).

Table 3.6: Binding free energies for PfdUTPase ligands 1a-14e. The energies are given in kcal/mol.

Ligand	ΔG_{exp} [a] [b]	ΔG_{LIE1} [c]	Docking	ΔG_{LIE2} [c]
1a	-6.7	-9.3	-3.6	-7.5
1b	-6.2	-9.2	-4.4	-6.7
1c	>-4.3	-8.5	-5.5	-6.3
2a	>-4.3	-9.0	-4.1	-8.8
2b	-4.9	-8.7	-3.8	-7.1
3a	-4.4	-7.1	-3.1	-7.8
3b	-5.2	-7.2	-4.7	-6.2
4	>-4.3	-8.7	-3.7	-9.0
5a	>-4.3	-9.0	-3.9	-7.1
5b	-5.0	-8.9	-4.4	-6.4
7a	>-4.3	-7.5	-4.0	-5.2
7b	>-4.3	-7.2	-3.7	-8.1
7c	>-4.3	-7.4	-3.5	-8.3
14b	>-4.3	-5.9	-1.9	-9.4
14c	>-4.3	-8.1	-2.4	-6.6
14d	>-4.3	-9.4	-6.4	-9.5
14e	-4.7	-9.6	-2.1	-8.1

[a] [21]
[b] The experimental binding free energy calculated from the equation $\Delta G_{exp} = RTlnK_i$.
[c] The parameters used for the binding free energy calculations:α=0.18, β=0.37 and γ=-7.24

In the first LIE run the binding free energies were in disagreement with the experimental binding data. The binding energies ranged from -5.9 kcal/mol to -9.6 kcal/mol. 16 out of the 17 ligands were overpredicted by 2.0-5.1 kcal/-mol (Figure 3.17, Table 3.6). The energies suggest that the LIE model might be system dependent or actually wrong. But there is also a possibility that the way the ligands were built in the active site did not represent the true

binding mode of these ligands. In fact, the protein might even need to undergo significant structural rearrangement at the active site to accommodate such ligands. When the ligands are built manually directly into the active site we assume that that specific binding mode is possible and the ligand might be forced into an artificial energy minimum. In addition, if there is a possibility for the ligand to bind in the active site in a other way the LIE simulations are most likely too short to capture any significant reorganization of the ligand. Due to the poor agreement with the experimental binding data the docking method was applied on the ligands to generate alternative binding modes.

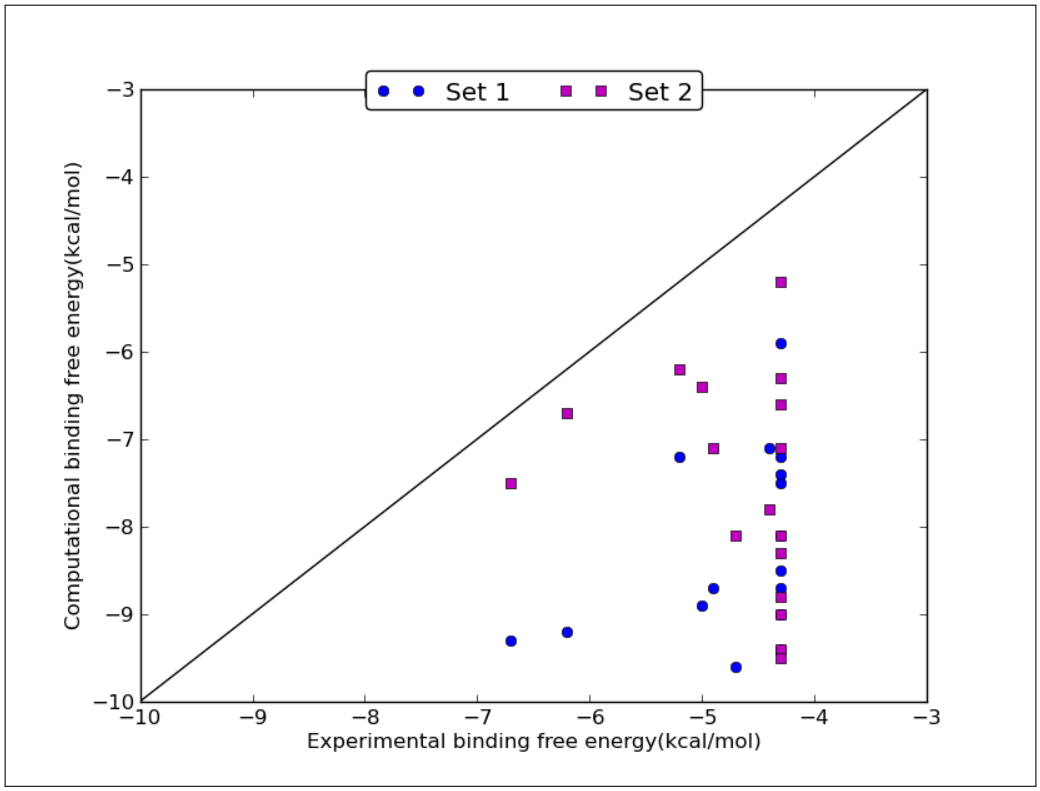
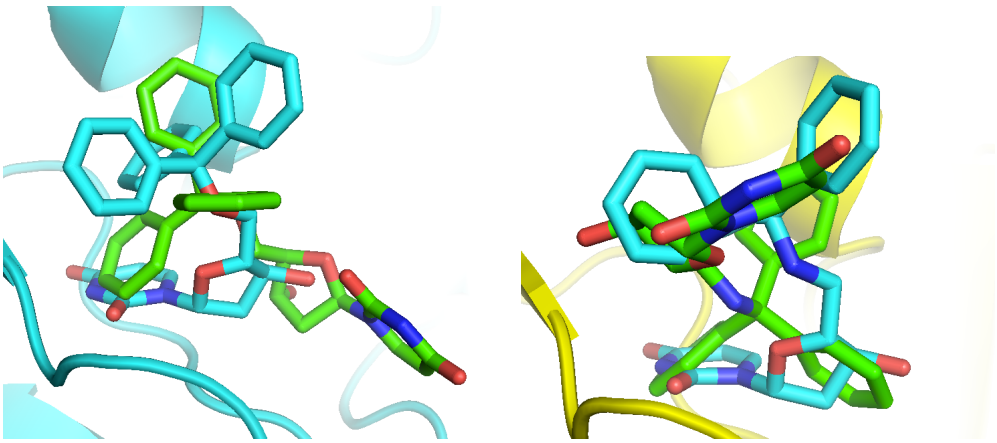


Figure 3.17: The distribution of the binding free energies for PfdUTPase ligands calculated by the LIE method. LIE_1 in blue and LIE_2 in magenta.

The docking of ligands 1a, 1b, 1c, 1a, 2b and 5b generated alternative positions for the uracil ring when the maximum number of poses were limited to 10. For the rest of the ligands, 10 was not enough to create alternative uracil conformations. The docking for these ligands was repeated with a treshold of 40, which gave in significantly different uracil positions. The poses chosen for further LIE simulations possessed both an alternative uracil binding

mode and a low docking score (>-6.5 kcal/mol). A comparison of the starting structure of ligand 1a when built manually into the active versus the conformation from the docking in Figure 3.18 (a) show the difference in the uracil ring position. Figure 3.18 also show the same comparison for ligand 7a.



(a) (b)

Figure 3.18: Ligand conformation in the binding site when built manually (cyan) and when chosen from the docking simulation (green) for ligand (a) 1a and (b) 7a.

The simulations based on poses from the docking improved the LIE binding free energies slightly compared to the first round of LIE simulations. The binding free energies ranged from -5.2 kcal/mol to -9.5 kcal/mol. In 6 out of the 17 ligands the binding free energies were in agreement with the experimental binding free energies. The difference from the experimental binding free energies was 0.8-5.2 kcal/mol. Ligands with 2 kcal/mol or smaller deviation from the experimental binding free energy were 1a, 1b, 1c, 3b, 5b and 7a, which is an improvement from the first LIE run. These results indicate that the docking of the ligands gave better starting structures for the LIE simulations.

In both LIE runs the binding free energies were dominated by van der Waals energies, which were negative. The electrostatic energies were positive through out. The large contribution from the non-polar interactions is not a surprise since the ligands have the large trityl moiety which is prone to van der Waals interactions. The LIE raw data is given in Tables F.1 and F.2.

3.4 Novel ligands

Two novel ligands were tested against the human dUTPase with the human dUTPase LIE model (Figure 3.20). The ligands follow a basic template which is drafted for the human dUTPase project. The ligand template (Figure 3.19) consists of:

- a uracil ring
- linker A
- a substituted heterocycle
- linker B
- a hydrophobic aromatic group

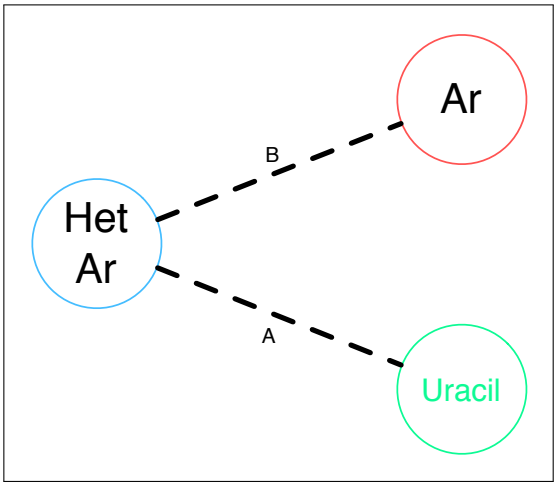
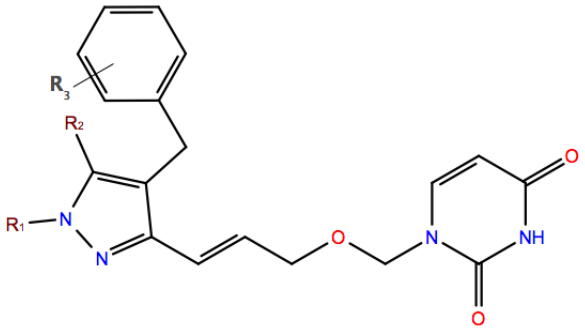


Figure 3.19: The template for the novel ligands in the human dUTPase drug design project.

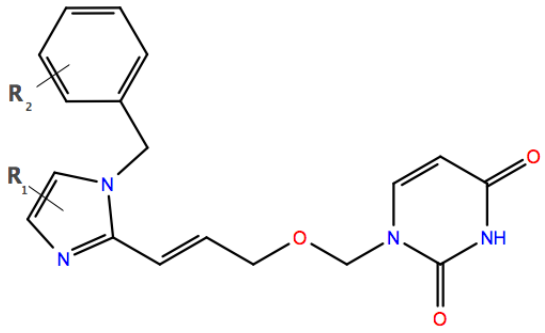
The drafted structure includes several points where the ligand can be modified. The aryl group has been hypothesized in previous works to bend in the active site towards the uracil ring [64]. In this way the ring would contribute to interactions with the hydrophobic pocket by π-stacking interactions. The heterocyclic group could enhance rigidity and directionality.

The novel ligands presented in this thesis are illustrated in Figure 3.20. Ligand 1 has a pyrazole ring at the linker structure preceding the phenyl ring,

while ligand 2 has a imidazole ring at the same position. Ligand 1 allows substitution at position R_1, R_2 and R_3 (Figure 3.20 (a)). Ligand 2 has possible locations of substitution at R_1 and R_2 (Figure 3.20 (b)).



(a)



(b)

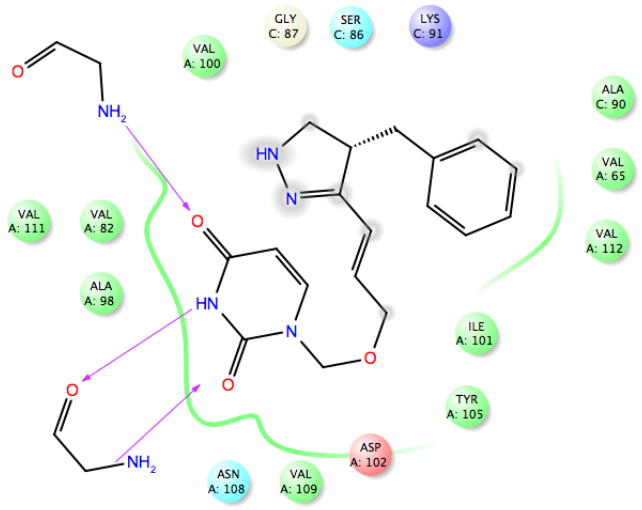
Figure 3.20: The novel ligand (a) 1 and (b) 2 structures. R_1, R_2 and R_3 mark possible positions of substitution elements.

3.4.1 Docking

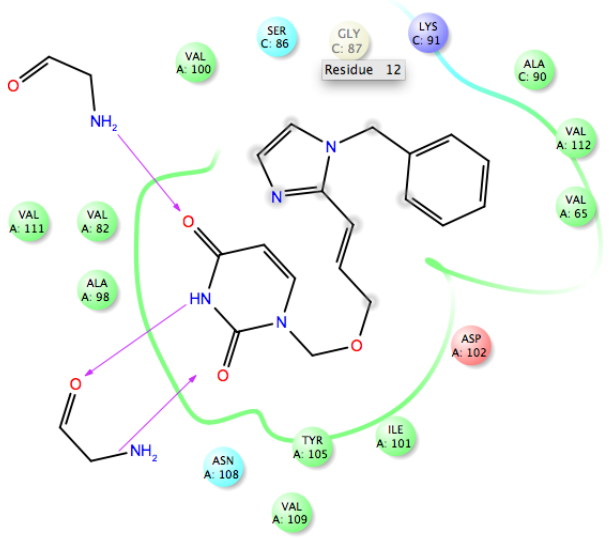
Ligands 1 and 2 were docked to the crystallographic structure of human dUTPase [18] (3ARN). The docking was conducted with maximum 10 poses per ligand. The docking yielded in 10 poses for ligand 1 and 9 poses for ligand 2. The docking score for ligand 1 ranged from -4.9 kcal/mol to -7.9 kcal/mol. The top scoring poses for ligand 1 possessed a conformation where the uracil ring is in the same position as in ligand 9 in the human dUTPase crystals structure. The phenyl ring was bent towards the uracil ring, most

likely interacting with the hydrophobic pocket of the binding site (Figure 3.21 (a)). The top three scoring poses from ligand 1 docking were selected for the LIE simulations. The ligands had docking scores -7.9 kcal/mol, -7.6 kcal/mol and -7.3 kcal/mol, and they were named 1_a, 1_b and 1_c, respectively (Table 3.7).

For ligand 2 the docking score varied from -3.9 kcal/mol to -7.8 kcal/mol. The position of the uracil ring was the same as in ligand 1_a for the top scoring poses and the phenyl ring was, similar to ligand 1_a, bent towards the uracil ring (Figure 3.21 (b). The top three ranked ligands from docking with ligand 2 were selected for the LIE simulations. The poses had docking scores of -7.8 kcal/mol, -7.2 kcal/mol and -7.5 kcal/mol and were named as 2_a, 2_b and 2_c, respectively (Table 3.7).



(a)



(b)

Figure 3.21: The ligand (a) 1_a and (b) 2_a interactions in the active site of human dUTPase. The figure represents the docked structures. The purple arrows show the ligand-protein backbone hydrogen bonding, the green circles are hydrophobic residues, the cyan circles are polar residues, the red circles are negatively charged residues, the purple circles are positively charged residues and the gray circles on the ligand are solvent exposed atoms. The figures were generated in Maestro.

3.4.2 Linear Interaction Energy

The docking of ligands 1 and 2 provided a binding poses which were further studied with the LIE method. The LIE method is more accurate at predicting binding free energies than docking. The LIE method was applied on ligands 1 and 2 with the human dUTPase LIE model. The top three scoring poses from the docking were chosen for each of the ligands as the starting structures in the LIE simulations.

The LIE method reproduced the docking scores to a large degree. In the LIE calculations the parameters from the human dUTPase model were used. The binding free energies varied from -7.1 kcal/mol to -8.4 kcal/mol. The binding energies were dominated by the vdW contributions, which varied from -3.0 kcal/mol to -3.1 kcal/mol. The EL contributions varied from 3.2 kcal/mol to 1.9 kcal/mol. The LIE raw data is given in Table G.1.

The LIE simulations resulted in binding free energies -7.8 kcal/mol, -8.4 kcal/mol and -7.3 kcal/mol for poses 1_a, 1_b and 1_c, respectively. The binding free energies from the LIE simulations for ligands 2_a, 2_b and 2_c were -7.3 kcal/mol, -7.12 kcal/mol and -8.23 kcal/mol, respectively.

For both of the ligands the binding free energies did not vary a lot. This is expected since the ligands conformations are quite similar with regards to both uracil ring placement and the orientation of the phenyl ring. The ligand binding energies should be tested experimentally to confirm the binding affinities. We need more experimental data on the novel ligands to judge the quality of the LIE model with respect to the ability to predict the binding affinities.

Table 3.7: Binding free energies for ligands 1 and 2. The energies are given in kcal/mol.

Ligand	Docking	ΔEL_{LIE}	ΔvdW_{LIE}	ΔG_{LIE} [a]
1_a	-7.87	2.53	-3.07	-7.77
1_b	-7.60	1.93	-3.12	-8.43
1_c	-7.34	2.90	-2.98	-7.32
2_a	-7.79	3.01	-3.10	-7.33
2_b	-7.20	3.24	-3.12	-7.12
2_c	-7.53	2.15	-3.13	-8.23

[a] The parameters used for the binding free energy calculations are: α=0.18, β=0.43 and γ=-7.24

Chapter 4

Conclusion

The LIE simulations and the analysis of the results were conducted with GUI developed for Q, and the Qgui turned out to work very well.

The LIE model constructed for the crystallographic structure of human dUTPase [18] was to a large degree able to predict the binding free energies for ligands 9-26 [18]. The LIE model was dependent on the addition of the γ scaling factor to predict the binding affinities. The γ has been found to be important in compounds where the hydrophobic interactions in the active site are important for the binding [65]. The best fit of the binding free energies was produced for compound 25S (modification 3), with a 0.02 kcal/-mol difference between the experimental and the computational binding free energies. The largest deviation between the computational and the experimental binding free energies was produced for compound 16 (modification 1).

The LIE simulations for the malaria ligands shows the importance of a good starting structure for the prediction of binding free energies. The first LIE simulation done for the ligands 1a-14e produced agreement with the experimental binding energies in 1 out of the 17 ligands. Due to the low agreement between the experimental and the computational binding affinities the docking method was applied to provide alternative starting structures for the malaria ligands. The second LIE simulations, run with these starting structures, presented binding free energies which were in better agreement with the experimental data: 6 out of the 17 ligands created a good agreement the experimental binding affinities. The lack of a good agreement between the experimental and the computational binding free energies for the malaria

ligands may also indicate system dependence of the LIE model, or in fact, the model might not be correct.

The LIE model was applied on two novel ligands and produced moderate binding free energies. The ligands should be tested with experimental techniques to be able to judge the LIE models ability to predict the binding affinities and conformations.

Chapter 5

Future Work

5.1 dUTPase

A goal of the dUTPase project is to develop new potent inhibitors against human dUTPase. New derivatives of the novel ligands with structural diversifications at sites R1-R3 could be constructed/created and tested with docking and the LIE model. One could start exploring the diversification with chemical compounds similar to the substitutions which are found at the human dUTPase ligands. Compounds such as the cyclopropylmethoxy and the fluor atom enhanced the binding qualities in the human dUTPase ligands. Similar compounds could be tested on the phenyl ring of the novel ligands.

The LIE model developed in this work could also be tested with more ligands targeted at the human dUTPase, to further test the models quality.

The experimental binding free energies produced for the human dUTPase ligands 9-26 could be tested with the FEP method with regards to the relative binding free energies.

In addition, the EL and vdW contributions to the binding free energies from the LIE calculations in the malaria and novel ligands should be examined closer to search for clues for further work.

5.2 Qgui

5.2.1 Replication support

With the growing use of computational methods in science there has been an increased focus on the difficulties involved in replicating scientific results [66]. According to Ref [67], the authors list specific guidelines for making computational research reproducible. The essence of the guidelines is that not only must any input data be stored and made available, but also the exact code and applications used for the computations. In order to meet the recommendations, it is possible to add a snapshot or working project feature to Qgui. This feature would make record all settings and input data and store at least the following:

- all input data, including files downloaded from pdb
- all settings
- version numbers of applications used
- version information of all scripts used
- all launch scripts used
- random number seeds used for the simulations

5.2.2 Help system

One of the goals of Qgui is to make Q more user friendly and accessible to new users. Currently, Qgui does not have a help system. A help system with tool tips and user instructions would make a new Qgui user even less dependent on user support.

Chapter 6

Concluding Remarks

This project has given me the opportunity to gain experience with the use of computational methods in chemistry. The learning curve has been steep and the process frustrating at times, but it is very satisfying when everything finally works as it should.

The process of setting up MD simulations for the LIE method was difficult due to instabilities in the molecular system and required several attempts before the system was cleared of structural errors. One of the reasons for this being very time consuming is that the molecular software Q provides very little feed back which can help you to find the error in your system. Another reason is that it is very hard to get an overview of the different tools provided with Q to analyze your system. Learning to use these tools required a lot of help from experienced users (Bjørn Olav Brandsdal and Geir Isaksen).

Also my computer skills have skyrocketed after I started on the Qgui project. I am still by no means a professional but have gained so much confidence and experience in computer programming, working with a supercomputer and writing various scripts to manage my jobs.

I have also experienced many surprises when it comes to my knowledge of the LIE method. When theory is applied in practice you learn a lot from the mistakes you make.

References

[1] C. D. Mol, J. M. Harris, E. M. McIntosh, and J. A. Tainer, "Human dutp pyrophosphatase: uracil recognition by a β hairpin and active sites formed by three separate subunits," *Structure*, vol. 4, no. 9, pp. 1077–1092, 1996.

[2] B. G. Vertessy and J. Tóth, "Keeping uracil out of dna: physiological role, structure and catalytic mechanism of dutpases," *Accounts of chemical research*, vol. 42, no. 1, pp. 97–106, 2008.

[3] J. R. Penadés, J. Donderis, M. García-Caballer, M. Á. Tormo-Más, and A. Marina, "dutpases, the unexplored family of signalling molecules," *Current opinion in microbiology*, 2013.

[4] J. Kovári, O. Barabás, E. Takács, A. Békési, Z. Dubrovay, V. Pongrácz, I. Zagyva, T. Imre, P. Szabó, and B. G. Vértessy, "Altered active site flexibility and a structural metal-binding site in eukaryotic dutpase kinetic characterization, folding, and crystallographic studies of the homotrimeric drosophila enzyme," *Journal of Biological Chemistry*, vol. 279, no. 17, pp. 17932–17944, 2004.

[5] B. Varga, O. Barabás, J. Kovári, J. Tóth, É. Hunyadi-Gulyás, É. Klement, K. F. Medzihradszky, F. Tölgyesi, J. Fidy, and B. G. Vértessy, "Active site closure facilitates juxtaposition of reactant atoms for initiation of catalysis by human dutpase," *FEBS letters*, vol. 581, no. 24, pp. 4783–4788, 2007.

[6] N. Yan, E. O'Day, L. A. Wheeler, A. Engelman, and J. Lieberman, "Hiv dna is heavily uracilated, which protects it from autointegration," *Proceedings of the National Academy of Sciences*, vol. 108, no. 22, pp. 9244–9249, 2011.

[7] R. W. Maul and P. J. Gearhart, "Chapter six-aid and somatic hypermutation," *Advances in immunology*, vol. 105, pp. 159–191, 2010.

[8] V. Muha, A. Horváth, A. Békési, M. Pukáncsik, B. Hodoscsek, G. Merényi, G. Róna, J. Batki, I. Kiss, F. Jankovics, *et al.*, "Uracil-containing dna in drosophila: stability, stage-specific accumulation, and developmental involvement," *PLoS genetics*, vol. 8, no. 6, p. e1002738, 2012.

[9] A. M. Baldo and M. A. McClure, "Evolution and horizontal transfer of dutpase-encoding genes in viruses and their hosts," *Journal of virology*, vol. 73, no. 9, pp. 7710–7721, 1999.

[10] R. D. Ladner and S. J. Caradonna, "The human dutpase gene encodes both nuclear and mitochondrial isoforms differential expression of the isoforms and characterization of a cdna encoding the mitochondrial species," *Journal of Biological Chemistry*, vol. 272, no. 30, pp. 19072–19080, 1997.

[11] P. M. Wilson, W. Fazzone, M. J. LaBonte, J. Deng, N. Neamati, and R. D. Ladner, "Novel opportunities for thymidylate metabolism as a therapeutic target," *Molecular cancer therapeutics*, vol. 7, no. 9, pp. 3029–3037, 2008.

[12] T. Chano, K. Mori, K. Scotlandi, S. Benini, C. Lapucci, M. C. Manara, M. Serra, P. Picci, H. Okabe, and N. Baldini, "Differentially expressed genes in multidrug resistant variants of u-2 os human osteosarcoma cells," *Oncology reports*, vol. 11, no. 6, pp. 1257–1264, 2004.

[13] W. Yano, S. Tsukioka, T. Yokogawa, S. Ito, K. Ishida, K. Yamamura, A. Fujioka, M. Fukuoka, K. Matsuo, K. Noguchi, *et al.*, "Abstract b88: Tas-114, a dutpase inhibitor, in combination with pemetrexed is a novel strategy for the treatment of nsclc.," *Molecular Cancer Therapeutics*, vol. 12, no. 11 Supplement, pp. B88–B88, 2013.

[14] D. B. Longley, D. P. Harkin, and P. G. Johnston, "5-fluorouracil: mechanisms of action and clinical strategies," *Nature Reviews Cancer*, vol. 3, no. 5, pp. 330–338, 2003.

[15] G. W. Aherne, A. Hardcastle, F. Raynaud, and A. L. Jackman, "Immunoreactive dump and ttp pools as an index of thymidylate synthase inhibition; effect of tomudex (zd1694) and a nonpolyglutamated quinazoline antifolate (cb30900) in l1210 mouse leukaemia cells," *Biochemical pharmacology*, vol. 51, no. 10, pp. 1293–1301, 1996.

[16] B. Mitrovski, J. Pressacco, S. Mandelbaum, and C. Erlichman, "Biochemical effects of folate-based inhibitors of thymidylate synthase in mgh-u1 cells," *Cancer chemotherapy and pharmacology*, vol. 35, no. 2, pp. 109–114, 1994.

[17] S. Popat, A. Matakidou, and R. S. Houlston, "Thymidylate synthase expression and prognosis in colorectal cancer: a systematic review and meta-analysis," *Journal of Clinical Oncology*, vol. 22, no. 3, pp. 529–536, 2004.

[18] S. Miyahara, H. Miyakoshi, T. Yokogawa, K. T. Chong, J. Taguchi, T. Muto, K. Endoh, W. Yano, T. Wakasa, H. Ueno, *et al.*, "Discovery of a novel class of potent human deoxyuridine triphosphatase inhibitors remarkably enhancing the antitumor activity of thymidylate synthase inhibitors," *Journal of medicinal chemistry*, vol. 55, no. 7, pp. 2970–2980, 2012.

[19] W. H. Organization, "Factsheet on the world malaria report 2013," June 2014.

[20] R. G. Ridley, "Medical need, scientific opportunity and the drive for antimalarial drugs," *Nature*, vol. 415, no. 6872, pp. 686–693, 2002.

[21] G. F. Ruda, C. Nguyen, P. Ziemkowski, K. Felczak, G. Kasinathan, A. Musso-Buendia, C. Sund, X. X. Zhou, M. Kaiser, L. M. Ruiz-Pérez, *et al.*, "Modified 5-trityl nucleosides as inhibitors of plasmodium falciparum dutpase," *ChemMedChem*, vol. 6, no. 2, pp. 309–320, 2011.

[22] A. R. Leach, *Molecular modelling: principles and applications*. Pearson Education, 2001.

[23] B. Alder and T. Wainwright, "Phase transition for a hard sphere system," *The Journal of Chemical Physics*, vol. 27, no. 5, pp. 1208–1209, 1957.

[24] B. J. Alder and T. Wainwright, "Studies in molecular dynamics. i. general method," *The Journal of Chemical Physics*, vol. 31, no. 2, pp. 459–466, 1959.

[25] J. E. Jones, "On the determination of molecular fields. i. from the variation of the viscosity of a gas with temperature," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, pp. 441–462, 1924.

[26] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of¡ i¿ n¡/i¿-alkanes," *Journal of Computational Physics*, vol. 23, no. 3, pp. 327–341, 1977.

[27] J. A. McCamm0n, "Dynamics of folded proteins," *Nature*, vol. 267, p. 16, 1977.

[28] I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, and T. E. Ferrin, "A geometric approach to macromolecule-ligand interactions," *Journal of molecular biology*, vol. 161, no. 2, pp. 269–288, 1982.

[29] G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor, "Development and validation of a genetic algorithm for flexible docking," *Journal of molecular biology*, vol. 267, no. 3, pp. 727–748, 1997.

[30] M. Rarey, B. Kramer, T. Lengauer, and G. Klebe, "A fast flexible docking method using an incremental construction algorithm," *Journal of molecular biology*, vol. 261, no. 3, pp. 470–489, 1996.

[31] G. L. Warren, C. W. Andrews, A.-M. Capelli, B. Clarke, J. LaLonde, M. H. Lambert, M. Lindvall, N. Nevins, S. F. Semus, S. Senger, *et al.*, "A critical assessment of docking programs and scoring functions," *Journal of medicinal chemistry*, vol. 49, no. 20, pp. 5912–5931, 2006.

[32] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, *et al.*, "Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy," *Journal of medicinal chemistry*, vol. 47, no. 7, pp. 1739–1749, 2004.

[33] A. Hernández-Santoyo, A. Y. Tenorio-Barajas, V. Altuzar, H. Vivanco-Cid, and C. Mendoza-Barrera, "Protein-protein and protein-ligand docking," 2013.

[34] J. Åqvist, C. Medina, and J.-E. Samuelsson, "A new method for predicting binding affinity in computer-aided drug design," *Protein engineering*, vol. 7, no. 3, pp. 385–391, 1994.

[35] J. Åqvist, V. B. Luzhkov, and B. O. Brandsdal, "Ligand binding affinities from md simulations," *Accounts of chemical research*, vol. 35, no. 6, pp. 358–365, 2002.

[36] K. A. Sharp, A. Nicholls, R. F. Fine, and B. Honig, "Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects," *Science*, vol. 252, no. 5002, pp. 106–109, 1991.

[37] J. Åqvist and T. Hansson, "On the validity of electrostatic linear response in polar solvents," *The Journal of Physical Chemistry*, vol. 100, no. 22, pp. 9512–9521, 1996.

[38] T. Hansson, J. Marelius, and J. Åqvist, "Ligand binding affinity prediction by linear interaction energy methods," *Journal of computer-aided molecular design*, vol. 12, no. 1, pp. 27–35, 1998.

[39] J. Aaqvist, "Comment on" transferability of ion models"," *The Journal of Physical Chemistry*, vol. 98, no. 33, pp. 8253–8255, 1994.

[40] T. Hansson and J. Åqvist, "Estimation of binding free energies for hiv proteinase inhibitors by molecular dynamics simulations," *Protein Engineering*, vol. 8, no. 11, pp. 1137–1144, 1995.

[41] J. ÅQvist, "Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions," *Journal of computational chemistry*, vol. 17, no. 14, pp. 1587–1597, 1996.

[42] J. Hultén, N. M. Bonham, U. Nillroth, T. Hansson, G. Zuccarello, A. Bouzide, J. Åqvist, B. Classon, U. H. Danielson, A. Karlén, *et al.*, "Cyclic hiv-1 protease inhibitors derived from mannitol: synthesis, inhibitory potencies, and computational predictions of binding affinities," *Journal of medicinal chemistry*, vol. 40, no. 6, pp. 885–897, 1997.

[43] B. O. Brandsdal, F. Osterberg, M. Almlof, I. Feierberg, V. B. Luzhkov, and J. Aqvist, "Free energy calculations and ligand binding," *Advances in protein chemistry*, vol. 66, pp. 123–158, 2003.

[44] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, *et al.*, "Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models," *Accounts of chemical research*, vol. 33, no. 12, pp. 889–897, 2000.

[45] J. Srinivasan, T. E. Cheatham, P. Cieplak, P. A. Kollman, and D. A. Case, "Continuum solvent studies of the stability of dna, rna, and phosphoramidate-dna helices," *Journal of the American Chemical Society*, vol. 120, no. 37, pp. 9401–9409, 1998.

[46] R. W. Zwanzig, "High-temperature equation of state by a perturbation method. i. nonpolar gases," *The Journal of Chemical Physics*, vol. 22, no. 8, pp. 1420–1426, 1954.

[47] T. Beutler and A. Mark, "Wf v. gunsteren. 1994. avoiding singularities and numerical instabilities in free-energy calculations based on molecular simulations," *Chem. Phys. Lett*, vol. 222, pp. 529–539.

[48] M. Zacharias, T. Straatsma, and J. McCammon, "Separation-shifted scaling, a new scaling method for lennard-jones interactions in thermodynamic integration," *The Journal of chemical physics*, vol. 100, no. 12, pp. 9025–9031, 1994.

[49] J. Marelius, 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*, vol. 16, no. 4, pp. 213–225, 1998.

[50] P. Kollman, "Free energy calculations: applications to chemical and biochemical phenomena," *Chemical Reviews*, vol. 93, no. 7, pp. 2395–2417, 1993.

[51] D. L. Beveridge and F. DiCapua, "Free energy via molecular simulation: applications to chemical and biomolecular systems," *Annual review of biophysics and biophysical chemistry*, vol. 18, no. 1, pp. 431–492, 1989.

[52] J. Åqvist and A. Warshel, "Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches," *Chemical reviews*, vol. 93, no. 7, pp. 2523–2544, 1993.

[53] *Manual for the Molecular Dynamics Package Q.*

[54] M. Schrödinger, "Suite 2012: Maestro, version 9.3, schrödinger, llc, new york, ny, 2012.," 2012.

[55] M. Schrödinger, "Suite 2012: Schrödinger suite 2012 protein preparation wizard; epik version 2.3, schrödinger, llc, new york, ny, 2012; impact version 5.8, schrödinger, llc, new york, ny, 2012; prime version 3.1, schrödinger, llc, new york, ny, 2012.," 2012.

[56] G. M. Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, and W. Sherman, "Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments," *Journal of computer-aided molecular design*, pp. 1–14, 2013.

[57] J. L. Banks, H. S. Beard, Y. Cao, A. E. Cho, W. Damm, R. Farid, A. K. Felts, T. A. Halgren, D. T. Mainz, J. R. Maple, *et al.*, "Integrated modeling program, applied chemical theory (impact)," *Journal of computational chemistry*, vol. 26, no. 16, pp. 1752–1780, 2005.

[58] J. L. Whittingham, I. Leal, C. Nguyen, G. Kasinathan, E. Bell, A. F. Jones, C. Berry, A. Benito, J. P. Turkenburg, E. J. Dodson, *et al.*, "dutpase as a platform for antimalarial drug design: structural basis for the selectivity of a class of nucleoside inhibitors," *Structure*, vol. 13, no. 2, pp. 329–338, 2005.

[59] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, "Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids," *Journal of the American Chemical Society*, vol. 118, no. 45, pp. 11225–11236, 1996.

[60] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *The Journal of chemical physics*, vol. 79, no. 2, pp. 926–935, 1983.

[61] F. S. Lee and A. Warshel, "A local reaction field method for fast evaluation of long-range electrostatic interactions in molecular simulations," *The Journal of chemical physics*, vol. 97, no. 5, pp. 3100–3107, 1992.

[62] S. Chan, B. Segelke, T. Lekin, H. Krupka, U. S. Cho, M.-y. Kim, M. So, C.-Y. Kim, C. M. Naranjo, Y. C. Rogers, *et al.*, "Crystal structure of the¡ i¿ mycobacterium tuberculosis¡/i¿ dutpase: Insights into the catalytic mechanism," *Journal of molecular biology*, vol. 341, no. 2, pp. 503–517, 2004.

[63] H. Gutiérrez-de Terán and J. Åqvist, "Lie: Method and applications in drug design.,"

[64] H. Miyakoshi, S. Miyahara, T. Yokogawa, K. T. Chong, J. Taguchi, K. Endoh, W. Yano, T. Wakasa, H. Ueno, Y. Takao, *et al.*, "Synthesis and discovery of n-carbonylpyrrolidine-or n-sulfonylpyrrolidine-containing uracil derivatives as potent human deoxyuridine triphosphatase inhibitors," *Journal of medicinal chemistry*, vol. 55, no. 7, pp. 2960–2969, 2012.

[65] M. Almlöf, B. O. Brandsdal, and J. Åqvist, "Binding affinity prediction with different force fields: examination of the linear interaction energy

method," *Journal of computational chemistry*, vol. 25, no. 10, pp. 1242–1254, 2004.

[66] R. D. Peng, "Reproducible research in computational science," *Science (New York, Ny)*, vol. 334, no. 6060, p. 1226, 2011.

[67] G. K. Sandve, A. Nekrutenko, J. Taylor, and E. Hovig, "Ten simple rules for reproducible computational research," *PLoS computational biology*, vol. 9, no. 10, p. e1003285, 2013.

Appendix A

LIE raw data human dUTPase ligands.

Table A.1: LIE raw data for human dUTPase ligands 9-26 with substitution pattern 1.

Compound		EL_w	vdW_w	EL_p	vdW_p
9	ligand	-61.48	-28.57	0.00	0.00
	complex	-29.91	-8.95	-24.12	-38.70
10	ligand	-62.07	-31.83	0.00	0.00
	complex	-26.91	-11.83	-27.29	-39.61
11	ligand	-62.10	-29.96	0.00	0.00
	complex	-29.87	-9.75	-25.07	-40.20
12	ligand	-63.58	-30.73	0.00	0.00
	complex	-34.01	-11.52	-23.72	-37.80
13	ligand	-59.77	-32.03	0.00	0.00
	complex	-26.51	-11.37	-25.22	-38.81
14	ligand	-61.45	-32.11	0.00	0.00
	complex	-30.14	-11.57	-24.12	-39.50
15	ligand	-60.96	-31.54	0.00	0.00
	complex	-29.73	-12.33	-25.61	-39.45
16	ligand	-63.29	-33.92	0.00	0.00
	complex	-29.69	-11.49	-24.51	-45.50
24	ligand	-64.44	-33.90	0.00	0.00
	complex	-38.79	-12.17	-19.73	-41.68
25S	complex	-62.75	-35.60	0.00	0.00
	ligand	-37.16	-14.49	-22.54	-42.00
25R	complex	-65.77	-35.20	0.00	0.00
	ligand	-41.24	-12.87	-19.76	-44.26
26	ligand	-66.80	-32.57	0.00	0.00
	complex	-37.24	-9.96	-25.12	-45.02

Table A.2: LIE raw data for human dUTPase ligands 9-26 with substitution pattern 2.

Compound		EL_w	vdW_w	EL_p	vdW_p
9	ligand	-61.48	-28.57	0.00	0.00
	complex	-29.91	-8.95	-24.12	-38.70
10	ligand	-68.83	-31.06	0.00	0.00
	complex	-32.50	-10.40	-23.76	-41.37
11	ligand	-62.86	-30.60	0.00	0.00
	complex	-33.48	-10.09	-22.36	-39.07
12	ligand	-63.58	-30.73	0.00	0.00
	complex	-34.01	-11.52	-23.72	-37.80
13	ligand	-62.76	-30.70	0.00	0.00
	complex	-32.62	-10.68	-22.32	-40.66
14	ligand	-61.40	-31.84	0.00	0.00
	complex	-30.92	-12.08	-24.35	-39.80
15	ligand	-60.96	-31.54	0.00	0.00
	complex	-29.73	-12.33	-25.61	-39.45
16	ligand	-62.77	-34.32	0.00	0.00
	complex	-30.73	-11.15	-26.35	-44.46
24	ligand	-67.07	-37.87	0.00	0.00
	complex	-37.75	-12.35	-22.89	-49.12
25S	ligand	-65.19	-36.04	0.00	0.00
	complex	-37.39	-11.10	-22.35	-47.21
25R	ligand	-67.36	-35.67	0.00	0.00
	complex	-41.33	-11.29	-22.58	-47.39
26	ligand	-68.31	-35.39	0.00	0.00
	complex	-42.93	-11.11	-21.87	-47.87

Table A.3: LIE raw data for human dUTPase ligands 9-26 with substitution pattern 3

Compound		EL_w	vdW_w	EL_p	vdW_p
9	ligand	-61.56	-31.02	0.00	0.00
	complex	-31.59	-10.92	-22.65	-38.33
10	ligand	-64.35	-33.55	0.00	0.00
	complex	-32.53	-11.92	-24.40	-43.23
11	ligand	-63.10	-32.91	0.00	0.00
	complex	-32.88	-11.23	-22.99	-41.59
12	ligand	-63.25	-33.18	0.00	0.00
	complex	-31.69	-12.62	-24.45	-40.69
13	ligand	-62.47	-30.35	0.00	0.00
	complex	-34.65	-11.94	-21.73	-37.71
14	ligand	-62.12	-32.18	0.00	0.00
	complex	-30.00	-11.72	-24.20	-40.24
15	ligand	-60.75	-31.38	0.00	0.00
	complex	-31.11	-12.15	-24.03	-39.52
16	ligand	-62.65	-37.15	0.00	0.00
	complex	-32.42	-12.02	-25.76	-48.97
24	ligand	-66.40	-37.61	0.00	0.00
	complex	-40.04	-11.75	-20.98	-49.34
25S	ligand	-66.46	-36.63	0.00	0.00
	complex	-35.91	-11.21	-23.02	-47.20
25R	ligand	-66.42	-35.49	0.00	0.00
	complex	-39.73	-10.50	-22.95	-48.95
26	ligand	-69.03	-35.56	0.00	0.00
	complex	-43.48	-11.05	-20.25	-48.57

Appendix B

Electrostatic energies in compounds 25R and 26.

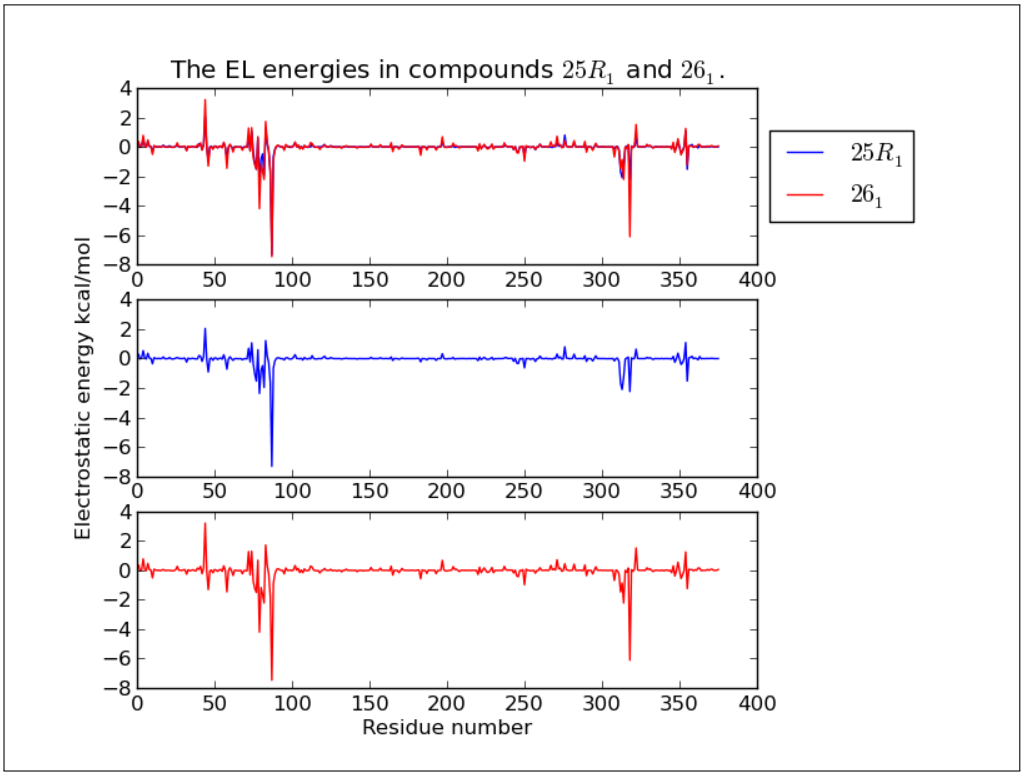


Figure B.1: A comparison of the electrostatic energies from the LIE simulations for $25R_1$ and 26_1.

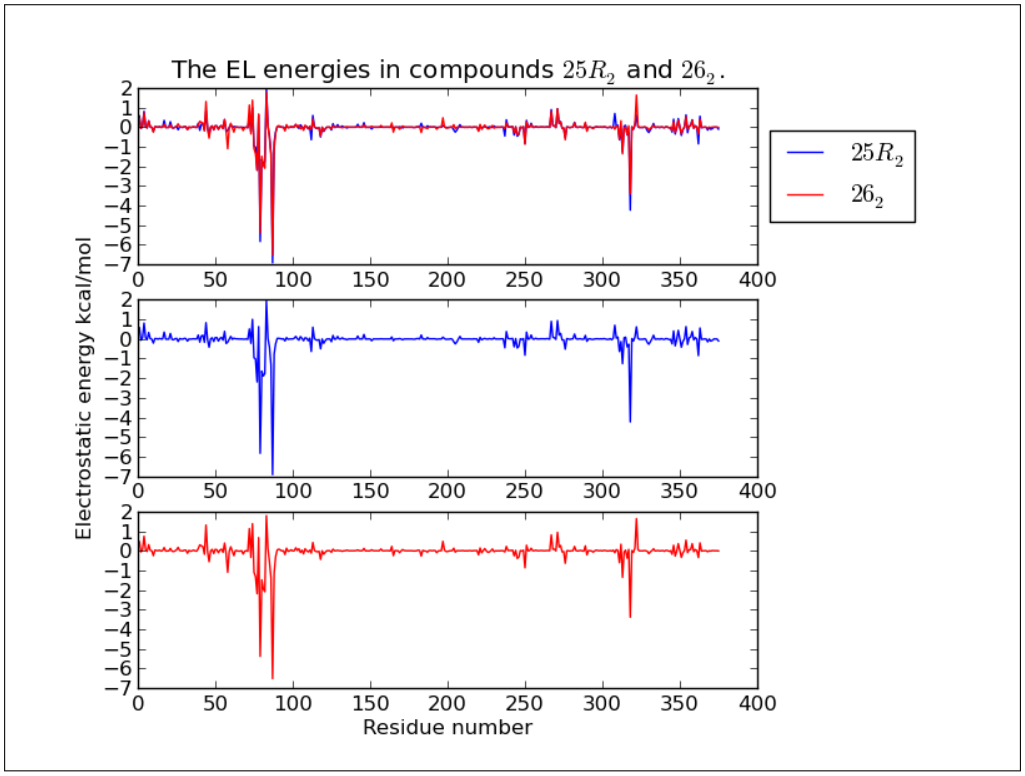


Figure B.2: A comparison of the electrostatic energies from the LIE simulations for $25R_2$ and 26_2.

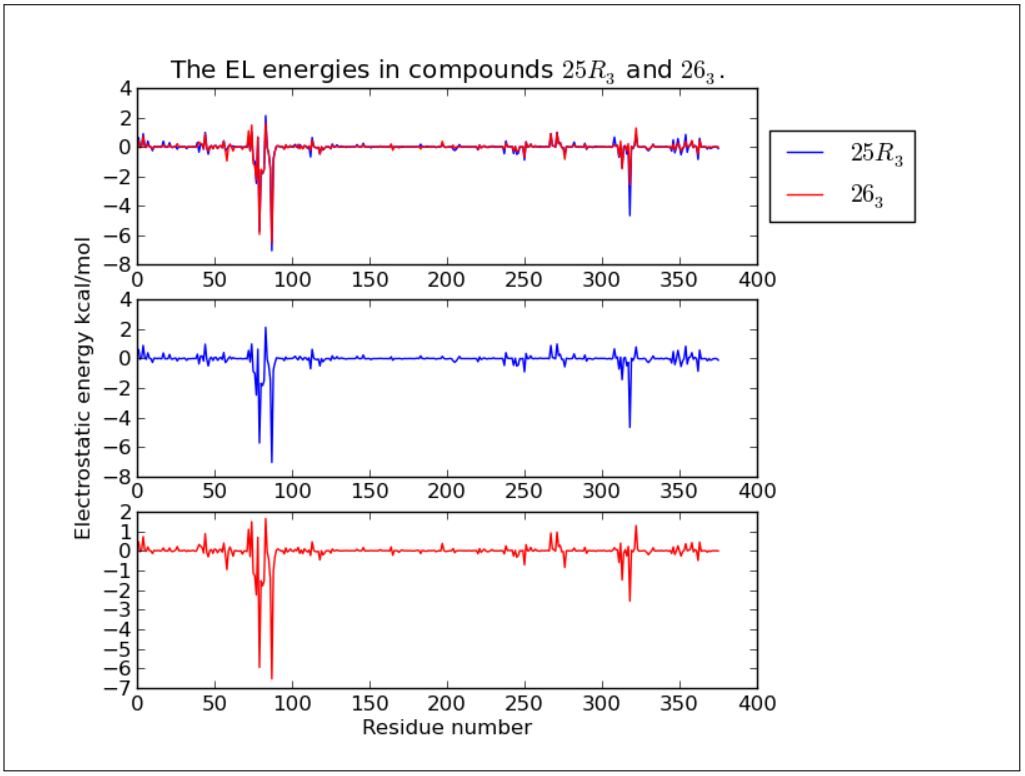


Figure B.3: A comparison of the electrostatic energies from the LIE simulations for $25R_3$ and 26_3.

Appendix C

The electrostatic energies from the MD simulations for interactions predicted from the 3ARN crystallographic structure.

Table C.1: Electrostatic energies for residues which were predicted to interact with the ligand [18]. Enzyme-ligand interactions are between a residue and the ligand, between a residue and a water molecule, and between the ligand and a water molecule.

Compound	Enzyme-Ligand		Enzyme-Water		Ligand-Water
	Gly 76	Gly 87	Gly 74 [a]	Val 89 [a]	HOH378
9_1	-1.17	-6.70	-6.20	-5.00	-5.13
9_3	-1.11	-6.61	-5.39	-4.99	-4.62
10_1	-1.03	-5.95	-5.93	-5.08	-4.98
10_2	-1.11	-6.54	-6.19	-5.05	-5.16
10_3	-1.18	-6.43	-6.04	-4.98	-5.02
11_1	-1.10	-7.43	-5.54	-5.14	-4.84
11_2	-0.96	-7.25	-5.65	-4.92	-4.70
11_3	-1.23	-7.09	-6.14	-4.97	-4.84
12_1	-1.05	-6.33	-5.75	-5.03	-4.57
12_3	-1.11	-6.30	-6.25	-4.90	-5.13
13_1	-0.96	-6.77	N/A	N/A	N/A
13_2	-1.30	-6.45	-6.16	-5.17	-5.42
13_3	-1.26	-6.45	-4.28	-4.36	-4.22
14_1	-1.02	-6.89	-4.28	-4.57	-4.46
14_2	-1.41	-6.50	-5.54	-4.66	-4.59
14_3	-1.40	-6.57	-5.74	-5.08	-4.97
15_1	-1.23	-6.96	-6.06	-4.80	-4.78
15_3	-1.19	-6.85	-6.18	-5.01	-4.96
16_1	-1.12	-7.67	-6.16	-5.01	-5.00
16_2	-1.13	-6.93	-5.97	-5.41	-4.61
16_3	-1.17	-7.17	-5.97	-5.31	-5.12
24_1	-0.87	-7.35	-5.95	-4.93	-4.77
24_2	-1.05	-6.88	-5.24	-5.02	-4.33
24_3	-0.88	-6.96	-5.20	-4.84	-4.54
$25S_1$	-1.10	-7.44	-4.13	-4.70	-4.11
$25S_2$	-1.00	-6.96	-5.58	-5.30	-4.96
$25S_3$	-0.93	-7.01	-4.53	-4.38	-4.20
$25R_1$	-1.05	-7.33	-5.01	-4.87	-4.75
$25R_2$	-1.02	-6.92	-5.74	-5.25	-4.84
$25R_3$	-1.00	-7.06	-5.68	-5.24	-4.94
26_1	-1.23	-7.47	-5.08	-4.98	-4.88
26_2	-1.32	-6.52	-5.74	-5.08	-5.17
26_3	-1.27	-6.53	-5.51	-5.05	-5.25

[a] Interactions with HOH378.

Appendix D

Distance between Lys 318 and ligand.

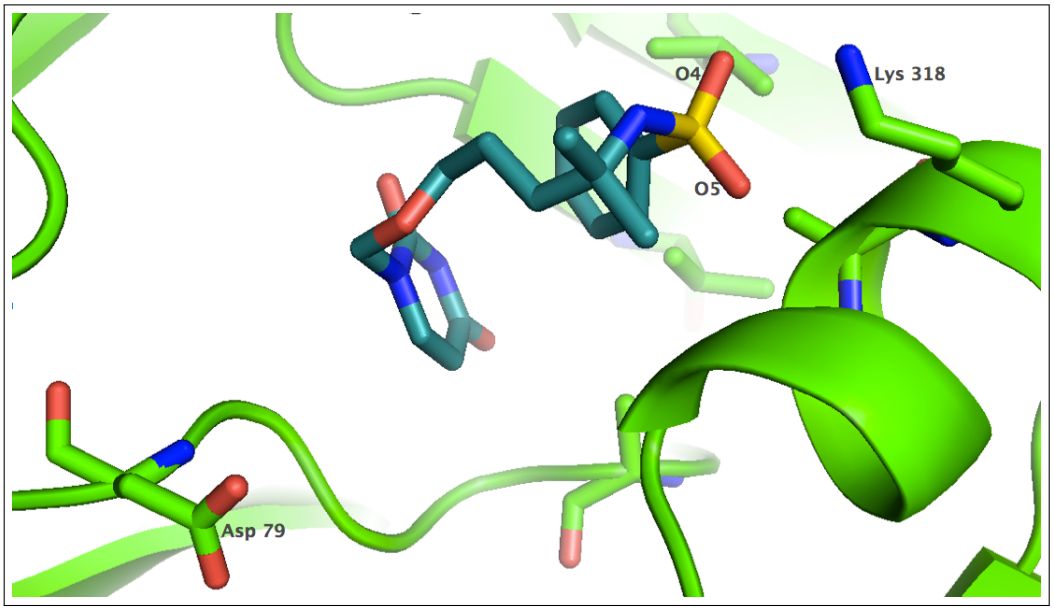


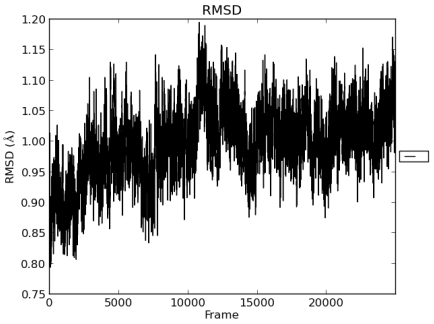
Figure D.1: Asp 79, Lys 318 and ligand 9 displayed in the active site of 3ARN.

Table D.1: The average distance (Å) between Lys 318 side chain nitrogen and ligand SO_2 oxygens in modifications 2 and 3.

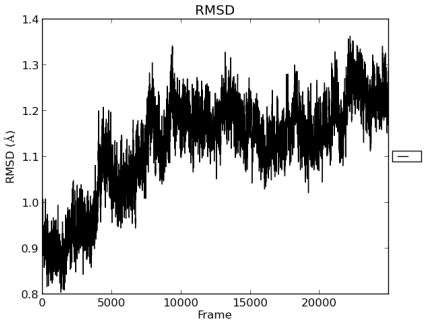
Compound	Distance N-O4	Distance N-O5
9_1	7.4	6.9
9_3	8.1	7.5
10_2	6.8	7.1
10_3	7.3	7.4
11_2	7.7	7.7
11_3	8.3	7.9
12_1	7.3	7.3
12_3	7.9	7.6
13_2	8.0	7.8
13_3	9.2	8.9
14_2	7.3	6.7
14_3	7.2	6.7
15_1	6.6	6.1
15_3	7.6	7.2
16_2	8.0	7.7
16_3	8.5	7.8
24_2	8.9	8.0
24_3	8.8	8.1
$25S_2$	9.5	8.5
$25S_3$	9.0	8.1
$25R_2$	9.6	8.4
$25R_3$	9.3	8.0
26_2	9.2	8.1
26_3	9.6	8.6

Appendix E

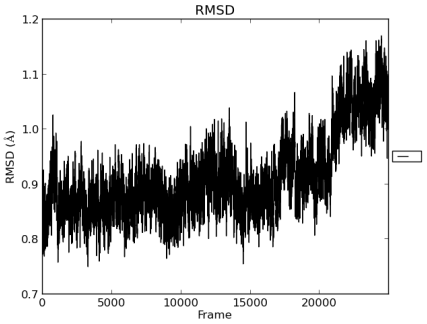
RMSD for molecular dynamics complex simulations for compound 16_1 and 26_1 .



(a)

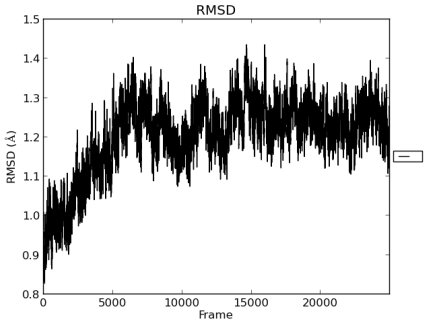


(b)

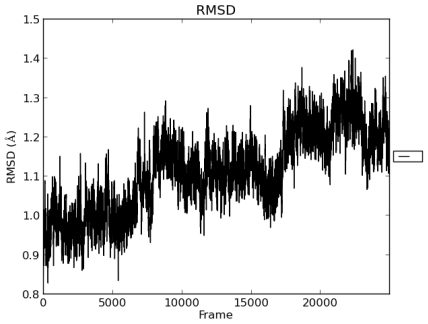


(c)

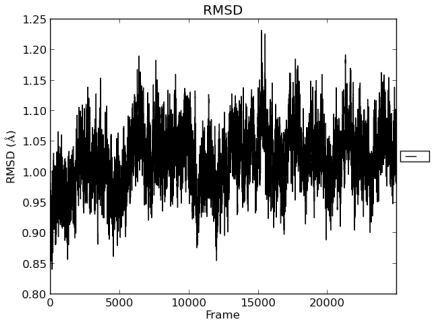
Figure E.1: RMSD calculations for 16_1 *from simulations 1, 5 and 10.*



(a)



(b)



(c)

Figure E.2: RMSD calculations for 26_1 *from simulations 1, 5 and 10.*

Appendix F

LIE raw data for *Pf*dUTPase ligands.

Table F.1: LIE_1 raw data for PfdUTPase ligands 1a-14e.

Ligand		EL_w	vdW_w	EL_p	vdW_p
1a	ligand	-47.03	-40.56	0.00	0.00
	complex	-29.27	-15.73	-13.69	-44.41
1b	ligand	-48.00	-40.91	0.00	0.00
	complex	-28.46	-15.22	-15.20	-45.25
1c	ligand	-47.63	-41.85	0.00	0.00
	complex	-25.83	-17.16	-16.52	-42.41
2a	ligand	-49.09	-43.61	0.00	0.00
	complex	-29.72	-17.41	-14.72	-45.56
2b	ligand	-49.53	-43.52	0.00	0.00
	complex	-30.00	-18.15	-14.30	-44.30
3a	ligand	-51.91	-42.95	0.00	0.00
	complex	-29.57	-17.60	-13.00	-43.98
3b	ligand	-52.92	-43.10	0.00	0.00
	complex	-30.98	-19.22	-13.25	-41.67
4	ligand	-51.19	-46.69	0.00	0.00
	complex	-28.63	-16.62	-15.92	-52.04
5a	ligand	-48.05	-41.76	0.00	0.00
	complex	-31.45	-16.66	-11.94	-44.38
5b	ligand	-48.92	-41.84	0.00	0.00
	complex	-31.43	-17.00	-12.48	-44.06
7a	ligand	-53.18	-39.80	0.00	0.00
	complex	-29.36	-16.18	-14.96	-43.52
7b	ligand	-55.76	-38.88	0.00	0.00
	complex	-31.33	-15.69	-14.68	-43.19
7c	ligand	-59.46	-39.14	0.00	0.00
	complex	-34.36	-15.71	-15.52	-43.93
14b	ligand	-96.76	-37.86	0.00	0.00
	complex	-52.53	-13.84	-30.80	-44.27
14c	ligand	-47.30	-41.17	0.00	0.00
	complex	-27.65	-16.80	-12.79	-43.10
14d	ligand	-51.10	-46.07	0.00	0.00
	complex	-26.87	-20.06	-19.36	-48.11
14e	ligand	-34.08	-42.09	0.00	0.00
	complex	-17.13	-17.48	-13.66	-44.21

Table F.2: LIE_2 raw data for PfdUTPase ligands 1a-14e.

Ligand		EL_w	vdW_w	EL_p	vdW_p
1a	ligand	-49.33	-40.42	0.00	0.00
	complex	-27.94	-21.43	-13.41	-37.06
1b	ligand	-50.92	-40.65	0.00	0.00
	complex	-18.82	-20.46	-18.70	-44.65
1c	ligand	-49.06	-41.66	0.00	0.00
	complex	-29.32	-18.81	-7.23	-43.44
2a	ligand	-50.25	-43.40	0.00	0.00
	complex	-28.91	-23.12	-15.53	-40.99
2b	ligand	-53.69	-43.80	0.00	0.00
	complex	-27.03	-20.92	-14.61	-46.68
3a	ligand	-50.88	-42.72	0.00	0.00
	complex	-33.70	-21.44	-10.11	-39.97
3b	ligand	-53.46	-42.95	0.00	0.00
	complex	-24.61	-28.87	-16.63	-33.40
4	ligand	-50.78	-46.40	0.00	0.00
	complex	-35.30	-21.50	-9.45	-47.01
5a	ligand	-47.93	-41.43	0.00	0.00
	complex	-32.25	-19.91	-4.91	-42.92
5b	ligand	-49.48	-41.81	0.00	0.00
	complex	-31.33	-20.13	-4.85	-44.28
7a	ligand	-52.98	-39.73	0.00	0.00
	complex	-26.41	-17.57	-10.50	-44.09
7b	ligand	-55.96	-38.87	0.00	0.00
	complex	-36.03	-20.37	-12.74	-38.00
7c	ligand	-56.12	-39.92	0.00	0.00
	complex	-43.34	-21.25	-6.58	-37.05
14b	ligand	-87.00	-38.10	0.00	0.00
	complex	-64.00	-21.57	-20.99	-34.68
14c	ligand	-46.18	-41.06	0.00	0.00
	complex	-31.43	-23.30	-4.68	-35.00
14d	ligand	-51.79	-46.34	0.00	0.00
	complex	-30.46	-22.52	-17.55	-44.20
14e	ligand	-33.24	-41.74	0.00	0.00
	complex	-21.53	-26.21	-6.03	-31.92

Appendix G

LIE raw data for novel ligands.

Table G.1: LIE raw data for the novel ligands.

Compound		EL_w	vdW_w	EL_p	vdW_p
1_a	ligand	-42.33	-32.74	0.00	0.00
	complex	-23.70	-12.27	-12.74	-37.07
1_b	ligand	-42.34	-32.71	0.00	0.00
	complex	-23.64	-13.50	-14.22	-36.54
1_c	ligand	-42.37	-33.03	0.00	0.00
	complex	-22.60	-12.59	-13.03	-36.98
2_a	ligand	-43.60	-32.90	0.00	0.00
	complex	-20.44	-11.76	-16.15	-38.36
2_b	ligand	-43.89	-32.96	0.00	0.00
	complex	-21.58	-12.93	-14.77	-37.38
2_c	ligand	-48.85	-32.96	0.00	0.00
	complex	-23.50	-12.81	-15.36	-37 .55