UPTEC X 02 022
MAY 2002
ISSN 1401-2138

SINISA BJELIC

Computational study of Plasmodium falciparum plasmepsin II inhibitor binding

Master's degree project

Molecular Biotechnology Programme
Uppsala University School of Engineering

UPTEC X 02 022	**Date of issue 2000-05**

Author

Sinisa Bjelic

Title (English)

Computational study of P. falciparum plasmepsin II inhibitor binding

Title (Swedish)

Abstract

Prediction of free energy of binding for two inhibitors with four isomers each was performed against malaria parasite *Plasmodium falciparum* protease plasmepsin II by molecular dynamics simulations and linear interaction energy method. The binding energies were compared further for isomer dependent activity.

Keywords

Molecular dynamics, linear interaction energy method, malaria

Supervisors

Johan Åqvist
Uppsala Universitet

Examiner

Torsten Unge
Uppsala Universitet

Project name	Sponsors
Language **English**	Security
ISSN 1401-2138	Classification
Supplementary bibliographical information	Pages **30**

Biology Education Centre Biomedical Center Husargatan 3 Uppsala
Box 592 S-75124 Uppsala Tel +46 (0)18 4710000 Fax +46 (0)18 555217

Computational Study of P. falciparum Plasmepsin II inhibitor binding

Sinisa Bjelic

Sammanfattning

Malaria dödar 2-3 miljoner människor varje år och man beräknar att antalet smittade är cirka 300-500 miljoner. I kampen mot malaria används framförallt läkemedel eftersom någon fungerande vaccin inte finns. Dessa läkemedel inverkar på malariaparasitens nödvändiga livsfunktioner, till exempel i det här fallet skulle hemoglobinnedbrytningen blockeras genom att inhibitorerna binder till enzymet plasmepsin II.

Att kunna förutsäga inhibitorernas inverkan på ett enzym, skulle underlätta i vår kamp mot sjukdomar. Det finns olika metoder att göra detta, här används molekyldynamik för att simulera molekylernas rörelse och linjära interaktionsenergimetoden för att bestämma bindningsenergier.

Vid syntes av olika inhibitorer erhålls olika isomerer, av vilka oftast en uppvisar aktivitet. Att kunna simulera detta skulle spara både resurser och framförallt tid, vilket leder till att framställningen av nya läkemedel går snabbare.

Examensarbete 20 p i Molekylär bioteknikprogrammet

Uppsala universitet Maj 2002

CONTENTS

Abstract - in Swedish

1. Introduction 3

2. Theoretical background

2.1 Molecular dynamics 5

2.2 Free energy calculations 8

2.3 Molecular mechanics and the force fields 10

3. Methods 12

3.1 Linear interaction energy 12

3.2 Computer simulations 15

3.2.1 Initial configuration: starting structure and ligands 15

4. Results 20

5. Discussion 24

6. References 27

Appendix 1. 29

Appendix 2. 28

Appendix 3. 30

1. INTRODUCTION

Plasmodium falciparum, the most dangerous of human malaria parasites, is afflicting 300-400 million people world wide and killing 2-3 million annually. The indiscriminate use of anti-malarial drugs has rendered them less useful and some malaria strains have become resistant to all known drugs. The development of new drugs, targeting essential biological processes, has been a priority in the struggle against the disease. To be able to do this we have to understand the life cycle of the malaria parasite and the molecular mechanisms, still poorly known, behind it. The malaria parasite erythrocytic life stage [1] begins when merozoites invade erythrocytes. In erythrocytes the merozoites develop from ring stage organism to more metabolically active trophozoites and then to multinucleated schizonts. The erythrocytic life cycle is completed when mature schizonts rapture the erythrocyte (Figure 1.)

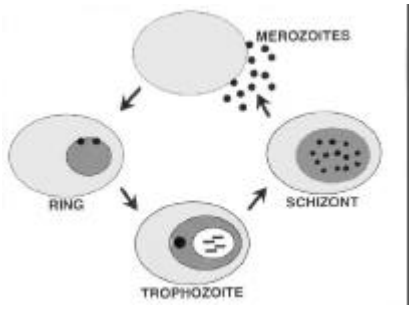


Figure 1. The malaria parasite life cycle.

The rapture and reinvasion of the host erythrocites by the parasite and the degradation of haemoglobin during the intraerythrocytic life stage is achieved by the use of proteases. The focus of this work has been on one of the haemoglobin degrading proteases, plasmepsin II (ppII) - PDB entry 1SME. The ppII is situated together with plasmepsin I in the food vacuole, where they are responsible for degradation of denatured respectively native haemoglobin. The plasmepsin II is an aspartic protease, and the whole family of aspartic proteases contain two aspartyl residues in the active site, that are responsible for the cleavage of the peptide bond. The residues are situated in the different domains, that together form a binding cleft between them (Figure 2.)

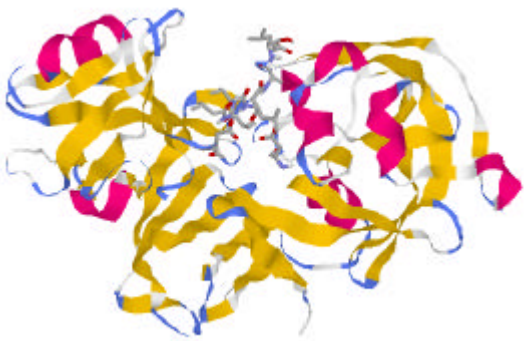


Figure 2. Structure of plasmepsin II, with bound pepstatin. The catalytic aspartates are situated in the respective domains.

In this work two inhibitors, with four different isomers each have been modelled into plasmepsin II, to investigate if the plasmepsin II stereospecificity could be determined theoretically and to predict the binding conformation and energy by using molecular dynamics (MD) together with the linear interaction energy method.

2. THEORETICAL BACKGROUND

2.1 Molecular Dynamics

In chemical sciences molecules are often presented as ball and sticks with rotation around bonds as the only degree of freedom. While this model is useful for predicting basic chemical interactions it is not sufficient for computational prediction and reproduction of experimental observables. Instead, a more dynamic system is needed that closely resembles a reality in which atoms in a molecule are in the thermal motion. Molecular dynamics (MD) can be defined as a method that simulates the motions of particles through time with respect to the forces that act on them. If we consider *N* atoms with spatial positions $\mathbf{r_1}(t)$, ..., $\mathbf{r_N}(t)$ acting under some kind of a potential energy $V(\mathbf{r_1}, ..., \mathbf{r_N})$, it is possible to calculate forces $\mathbf{F}_1$, ..., $\mathbf{F}_N$ that are acting on *N* particles specified by

$$\mathbf{F}_i = -\frac{\partial V(\mathbf{r}_1, \ldots, \mathbf{r}_N)}{\partial \mathbf{r}_i} \tag{1}$$

In molecular dynamics, Newton's laws of motion are integrated thus generating successive configurations of the system. This gives a trajectory that specifies how positions and velocities of the particles in the system vary with time. The trajectory is obtained by solving the differential equation, i.e. Newton's second law

$$\mathbf{F}_i = m_i \cdot \ddot{\mathbf{r}}_i \tag{2}$$

where m_1, ..., m_N are the masses of the *N* particles. The dynamic behaviour of the system can be calculated by solving the differential equation of Newton's second law. At time t equal to zero the initial velocity is the constant, and is given by random selection from the Maxwell-Boltzmann probability distribution p at the temperature of interest

$$p(\dot{\mathbf{r}}_i) = \left(\frac{m_i}{2\pi k_B T}\right)^{1/2} \cdot \exp\left[-\frac{m_i \dot{\mathbf{r}}_i^2}{2k_B T}\right]. \tag{3}$$

The Maxwell-Boltzmann distribution provides a probability that a particle i has a certain velocity at temperature T. The expression for velocity can be integrated once more giving the next configuration of the system

$$\mathbf{r}_i = (\dot{\mathbf{r}}_i)_{t=0} t + \frac{\ddot{\mathbf{r}}_i t^2}{2} + (\mathbf{r}_i)_{t=0} \tag{4}$$

where t equal to zero gives the current position. The expressions derived so far can be used in the finite difference method for generating molecular dynamics trajectories. The idea is that the integration is

numerically discretized and separated by a fixed time step Δt, and that positions (velocities and accelerations) can be approximated as Taylor series expansion

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \dot{\mathbf{r}}_i \cdot \Delta t + \ddot{\mathbf{r}}_i \cdot \frac{\Delta t^2}{2} + \ldots \tag{5}$$

This expression is truncated for higher order terms thus introducing an error at every time step. A number of algorithms exists that deals with finite time steps and truncates the series expansion at the different terms. The one used in here is leap-frog version of Verlet's algorithm [2]. Here it is assumed that the velocity, constant over a finite time step, is calculated at the midpoint thus giving the new positions

$$\dot{\mathbf{r}}_i(t+\frac{\Delta t}{2}) = \dot{\mathbf{r}}_i(t-\frac{\Delta t}{2}) + \ddot{\mathbf{r}}_i(t) \cdot \Delta t \tag{6}$$

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \dot{\mathbf{r}}_i(t+\frac{\Delta t}{2}) \cdot \Delta t \tag{7}$$

where acceleration is obtained from Newton's second law. The position equation can be integrated with the velocity equation into

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \dot{\mathbf{r}}_i(t-\frac{\Delta t}{2}) \cdot \Delta t + \frac{\mathbf{F}_i(t)}{m_i} \cdot \Delta t^2 . \tag{8}$$

The leap-frog version of Verlet's algorithm has the advantage of allowing easy temperature control since the velocity explicitly enters into the algorithm.

Now we have both the position and velocities of particles and can calculate an instantaneous value of a property $A(\mathbf{p}_1(t), \mathbf{r}_i(t))$,where $\mathbf{p}_1(t)$, ..., $\mathbf{p}_N(t)$ denotes momenta of the particles N at time t. The property A is calculated for the microscopic system comprising for instance, of a solvated protein with the bound ligand. On the other hand, during experiments, the properties are measured in a system containing a large number of molecules and the average value during a certain time t is estimated. The connection between the microscopic system and the macroscopic properties is made by statistical mechanics, which is based on the ensemble concept and that many individual microscopic configurations lead to the same macroscopic properties. Averaging over a large number of these systems that are in different microscopic configurations gives the macroscopic property formulated in terms of ensemble average.

$$< A >= \iint A(\mathbf{p},\mathbf{r}) r(\mathbf{p},\mathbf{r}) d\mathbf{p}^N d\mathbf{r}^N \tag{9}$$

the angle brackets <> indicate an ensemble average (expectation value) and ρ is the probability density, i.e. the probability of finding a configuration with the momenta **p** and positions **r**. There are different statistical ensembles depending on the thermodynamic properties that are held constant. In the canonical ensemble (N,V,T constant) the probability density takes the form of theBoltzmann distribution

$$r(\mathbf{p},\mathbf{r}) = \frac{e^{-E(\mathbf{p}_i\mathbf{r}_i)/(k_BT)}}{Q} \tag{10}$$

where E is the energy, and Q is the partition function. For the canonical ensemble the partition function is

$$Q = h^{-3N} \iint e^{-E/(k_BT)} d\mathbf{p}^N d\mathbf{r}^N \tag{11}$$

the energy E being the sum of kinetic and potential energies. The interesting part is the potential energy and it can be solved from the partition function since it is independent from the kinetic energy. Thus, the configurational, or non-kinetic part, can be expressed as

$$Z = \int e^{-V(\mathbf{r})/(k_BT)} d\mathbf{r}^N \tag{12}$$

where V is the potential energy and Z is the configurational integral. After integrating a property A over all possible conformations the ensemble average is determined. The fundamental principle in statistical mechanics, the ergodic hypothesis [3], states that the ensemble average is equal to the time average for a single system when $\tau \to 0$.

$$\lim_{t\to\infty} \frac{1}{t} \int_{t=0}^{t} A(\mathbf{p}(t),\mathbf{r}(t))dt = \iint A(\mathbf{p},\mathbf{r}) r(\mathbf{p},\mathbf{r}) d\mathbf{p}^N d\mathbf{r}^N . \tag{13}$$

This means that given an infinite amount of time the system will cover all the possible conformations and averaging over a trajectory of the system is equivalent to the averaging over the ensemble.
In molecular dynamics the time average over an observable A is obtained by calculating the statistical

average of a property sampled over a trajectory

$$\bar{A} = \frac{1}{M}\sum_{n=1}^{M} A(\mathbf{p}(t_n), \mathbf{r}(t_n)) \tag{14}$$

where M is the number of time steps and $A(\mathbf{p}(t_n),\mathbf{r}(t_n))$ is the instantaneous value of A at the time step t_n. If M goes to infinity the ergodic hypothesis is equivalent to the statement

$$\lim_{M\to\infty} \bar{A} = < A > \tag{15}$$

i.e. a single numerical trajectory can be used for calculating the ensemble average of an observable. A simplified picture of molecular simulation is that molecular dynamics simulates a time-dependent system on a microscopic level giving a time trajectory. This trajectory is equivalent to the true trajectory, as long the systems do not diverge too much. Under the ergodic hypothesis, a single trajectory can be used for calculating an ensemble average which , on the other hand, is connected to physical observables through the statistical mechanics .

2.2 Free energy calculation

The property A that is calculated depends on the system that is studied. During experiments we are interested in what is going to happen with different reactants, for example, as studied here, the binding of a ligand to a receptor. A way to quantify this is by determining the free energy difference between the two states, i.e. here between free ligand and enzyme and the ligand/enzyme complex. In thermodynamics, the free energy is calculated in two ways depending on what restraints we have imposed on the system. If the pressure and volume are held constant we get Gibbs free energy and Helmholtz free energy, respectively. A way of obtaining free energy from a simple chemical reaction of ligand L binding to a receptor R, goes by the dissociation constant K_d

$$R + L \xleftarrow[K_d]{} R - L \;\; <=> \;\; K_d = \frac{[R][L]}{[RL]} = e^{\left(\Delta G_{bind}/(RT)\right)} \tag{16}$$

where ΔG_{bind} is the free energy of binding. So the property A that is calculated from molecular

dynamics is in this case a free energy. The free energy, derived from statistical mechanics, is calculated by

$$A = -b^{-1} \ln Q \text{ or } \mathrm{G} = -b^{-1} \ln \Delta \tag{17}$$

$$b = \frac{1}{\mathrm{k_B T}} \tag{18}$$

where Δ is the partition function for the isothermal-isobaric (NPT) ensemble, A is Helmholtz free energy, and G is Gibbs free energy. In practice there is no difference between Gibbs free energy and Helmholtz free energy and no further distinction is made between them. The difference in free energy ΔA between the two different states X and Y is defined by

$$\Delta \mathrm{A} = \mathrm{A_Y} - \mathrm{A_X} = -b^{-1} \ln(\mathrm{Z_Y} / \mathrm{Z_X}) \tag{19}$$

where Q has been exchanged for the configuration integral Z. Substituting expression for Z in the previous equation gives the free energy in terms of potential energy $V(\mathbf{r}_1, ..., \mathbf{r}_N)$,

$$\Delta \mathrm{A} = -b^{-1} \ln\left(\frac{\int \mathrm{e}^{-b\mathrm{V_Y}} \mathrm{d}\mathbf{r}^{\mathrm{N}}}{\int \mathrm{e}^{-b\mathrm{V_X}} \mathrm{d}\mathbf{r}^{\mathrm{N}}}\right). \tag{20}$$

To get the expression in the form that can be used in computer simulations we have to be able to calculate the free energy in the form of ensembles, done by substituting $\exp(-\beta \mathrm{V_X}) \cdot \exp(+\beta \mathrm{V_X}) = 1$ in the numerator

$$\Delta \mathrm{A} = -b^{-1} \ln\left(\frac{\int \mathrm{e}^{-b\mathrm{V_Y}} \mathrm{e}^{-b\mathrm{V_X}} \mathrm{e}^{+b\mathrm{V_X}} \mathrm{d}\mathbf{r}^{\mathrm{N}}}{\int \mathrm{e}^{-b\mathrm{V_X}} \mathrm{d}\mathbf{r}^{\mathrm{N}}}\right) =$$

$$-b^{-1} \ln\left(\frac{\int \mathrm{e}^{-b(\mathrm{V_Y}-\mathrm{V_X})} \mathrm{e}^{-b\mathrm{V_X}} \mathrm{d}\mathbf{r}^{\mathrm{N}}}{\int \mathrm{e}^{-b\mathrm{V_X}} \mathrm{d}\mathbf{r}^{\mathrm{N}}}\right) = -b^{-1} \ln < e^{-b\Delta \mathrm{V}} >_{\mathrm{X}} \tag{21}$$

where ΔV is the difference of the potential energies for the two different states X and Y and the subscript X denotes averaging over the configurations representative of the initial state X. The previous equation is called the free energy perturbation formula generally attributed to Zwanzig [3].

2.3 Molecular mechanics and the force field

Computer simulations, to be performed at all with some accuracy, have to be done on the model that closely resembles reality. In molecular mechanics the smallest unit that is considered is the atom. Two non-bonded atoms can interact with each other in two different ways, polar or non-polar. Of course the interactions between atoms are of electromagnetic nature but the term is used here for the interactions that occur between charged or dipolar groups. The electrostatic interaction energies are described by Coulomb's law

$$V_{el} = \frac{1}{4\pi\varepsilon_0}\frac{q_i q_j}{r_{ij}} \tag{22}$$

where q is the partial charge on atom i and j respectively, r_{ij} is the distance between atoms i and j and ε_0 is the electric permittivity of vacuum. The non-polar or van der Waals interactions are modelled by the Lennard-Jones potential

$$V_{LJ} = A_{ij}r_{ij}^{-12} - B_{ij}r_{ij}^{-6} \tag{23}$$

A and B are Lennard-Jones interaction parameters for the interaction between atoms i and j. The first term describes a steeply rising repulsive potential energy when two atoms are close to each other, i.e. the repulsion is due to overlapping of the electron clouds. The second term is an attractive potential that is the result of favourable interaction of London dispersion forces.
The interactions discussed so far are also called the non-bonded ones, in contrast to the bonded interactions that arise from the forces that are acting on the covalently connected atoms or between atoms separated by two or three bonds. Two bonded atoms vibrate approximately in a harmonic way when they are close to their equilibrium value. The vibration can be modelled by a simple harmonic function

$$V_{bond} = \frac{k_b}{2}(r - r_0)^2 \tag{24}$$

where k_b is bond-stretching force constant, r is the distance between two bonded atoms and r_0 is the equilibrium bond length. Bond angles are treated in the same way as bond lengths

$$V_{angles} = \frac{k_q}{2}(q - q_0)^2 \tag{25}$$

k_θ is angle-bending force constant, θ is angle between two bonds and θ_0 is the equilibrium angle. The bond and angle forces are not enough to model all the possible intramolecular forces in the molecule. One of the most important forces for understanding structural properties of molecules is force affecting the rotation about chemical bonds. This type of motion is modelled by potential energy for torsion

$$V_{\text{torsion}} = K_j \left[1 + \cos(n j - d)\right] \tag{26}$$

where K_φ is the force constant for rotation around a dihedral angle, n is the number of minima per full turn of torsion angle φ, and the δ is the location of the first barrier. The torsion φ is defined as an angle between planes 1-2-3 and 2-3-4 (Figure 3.)

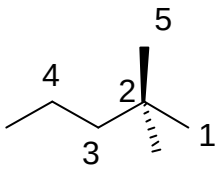


Figure 3. Definition of the torsion angles between planes 1-2-3 and 2-3-4 and the improper angles 1-2-3 and 1-2-5

The potential energy model for torsion angles is sufficient for most cases but sometimes atoms that have to be in the plane are bent out if it. To solve the problem the improper torsion angles between planes 1-2-3 and 1-2-5 (Figure 3.) or impropers are defined by the potential energy

$$V_{\text{improper}} = \frac{k_x}{2}\left(x - x_0\right)^2 \tag{27}$$

k_ξ is an out of plane bending force constant for the improper angle ξ and ξ_0 is the equilibrium angle. All interactions, both bonded and non-bonded, constitute together molecular mechanics potential energy function V [4], or a force field that is needed to model molecules in a way that resembles reality

$$\begin{aligned} V = & \sum_{\text{bonds}} \frac{k_b}{2}(r - r_0)^2 + \\ & \sum_{\text{angles}} \frac{k_q}{2}(q - q_0)^2 + \\ & \sum_{\text{dihedrals}} K_j\left[1 + \cos(n j - d)\right] + \\ & \sum_{\text{improper}} \frac{k_x}{2}(x - x_0)^2 + \\ & \sum_{\text{atompairs i, j}} \left(\frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{ij}} + A_{ij} r_{ij}^{-12} - B_{ij} r_{ij}^{-6} \right) \end{aligned} \tag{28}$$

3. METHODS

3.1 Linear interaction energy (LIE) method

The free energy of binding calculated by the LIE method [5], is obtained by two simulations: a solvated ligand and solvated protein/ligand complex (Figure 4.) The ligand interacts with its surrounding in a both polar and non-polar way.

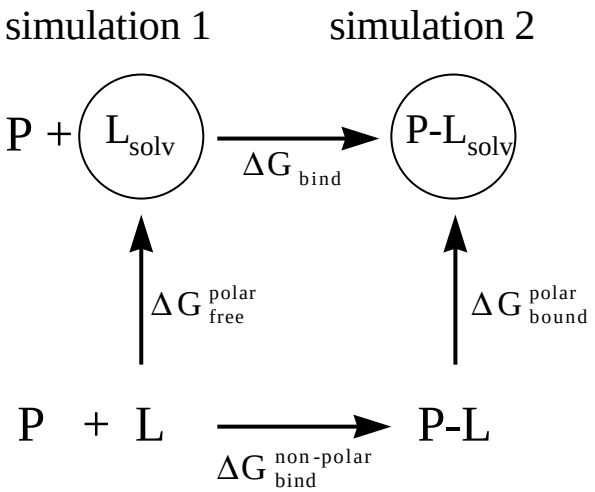


Figure 4. The schematic representation of Linear Interaction Energy Method

In LIE method the polar interaction energies are calculated by subtracting the ligand's electrostatic interactions in solvated protein with its electrostatic energies in water and multiplied with a constant. The process can be described as charging a ligand in the water or in a solvated protein. The constant depends on the linear response assumption, which means that the solvent polarisation (i.e. the mean force applied to the solvent) depends linearly on the changes in the solute field. A system that exhibits the linear response behaviour has a constant of 0.5 (ionic groups). This can be obtained by the expansion in the exponent and logarithm of the free energy differences between the two states, given by Zwanzig,

$$\Delta G_{el} = -b^{-1} \ln < e^{-b\Delta V} >_i \qquad (29)$$

$$\Updownarrow$$

$$\Delta G_{el} = < \Delta V >_i - \frac{b}{2} < (\Delta V - < \Delta V >_i)^2 >_i + \frac{b^2}{6} < (\Delta V - < \Delta V >_i)^3 >_i + \ldots \qquad (30).$$

The two different states A and B, i is either the A or B potential energy surface, are an uncharged solute Lennard - Jones cavity interacting with the solvent (l-s) with all intramolecular interactions (l-l), both electrostatic (el)and van der Waals (LJ), and solvent - solvent (s-s) interactions and a charged solute with the same interactions respectively. The energy difference between the two states is equal, thus with the electrostatic potential energy of the state B

$$\Delta V = V_B - V_A = (V_{l-l}^{gas} + V_{s-s} + V_{l-s}^{LJ} + V_{l-s}^{el})_B - (V_{l-l}^{gas} + V_{s-s} + V_{l-s}^{LJ})_A \qquad (31)$$

<·> denotes ensemble average sampled on either A or B. The calculation of free energy from the equation (30) will not work since higher order terms (energy fluctuations) converge very slowly in both molecular dynamics and Monte Carlo simulations. Adding the free energies for the two states A and B gives a more convenient expression

$$\Delta G_{el} = \frac{1}{2}(<\Delta V>_A + <\Delta V>_B) - \frac{b}{4}(<(\Delta V - <\Delta V>_A)^2>_A + <(\Delta V - <\Delta V>_B)^2>_B) + \ldots \qquad (32)$$

If the expression can be truncated after second - order terms and if we assume that the mean square energy fluctuations of the energy gap are the same for both A and B, the free energy can be calculated from average values of ΔV

$$\Delta G_{el} = \frac{1}{2}(<\Delta V_A> + <\Delta V_B>) \qquad (33)$$

As mentioned above state A is uncharged solute and the solvent does not experience the charge distribution and it is therefore reasonable to assume

$$<\Delta V>_A = <V_{l-s}^{el}>_A = 0. \qquad (34)$$

The electrostatic free energy can be now calculated [6] as

$$\Delta G_{el} = \frac{1}{2} <V_{l-s}^{el}>_B . \qquad (35)$$

The simulations of the charged ions confirmed the linear response assumption. However, investigations of its validity in the polar solvents showed that the electrostatic scaling factor β varies with the chemical composition of the solute. For the dipolar compounds the free energy of charging was lower then predicted by the linear response assumption. The β factor also varied with the number of hydroxyl groups, so that the increase in the number of hydroxyl groups was associated with the lower values of

β. The deviations from the linear response assumption are taken into the account by dividing the components into the four classes: charged, dipolar with no hydroxyl groups, dipolar with one hydroxyl group and dipolar with two or more hydroxyl groups and assigning them values of β=0.5, β=0.43, β=0.37 and β=0.33 respectively. These values of β were derived from the simulations of the typical compounds of the different classes for example sodium ion, acetone, ethanol and ethylene glycol respectively [6]. Now when the ligand has been solvated in the water or in protein/ligand complex the difference in the electrostatic binding energy can be obtained by subtracting the free energies needed for the charging process

$$\Delta G_{bind}^{el} = b \cdot (< V_{l-s}^{el} >_{bound} - < V_{l-s}^{el} >_{free}) \quad (36)$$

where β denotes a scaling factor that depends on the chemical composition of the ligand.
The solvation free energies of the non-polar compounds scale linearly with the molecular size measures as for example the chain length, volume and surface area. For instance for the hydrocarbons the free energy of solvation ΔG_{sol}=kn+l, depends approximately linearly on the length n of the carbon chain. This has been known from the experimental data, but has also been verified by the molecular dynamics computer simulations. On the other hand the van der Waals solute - solvent interactions scale also linearly with the same molecular size measures. It seems possible to estimate non - polar binding contribution by scaling the Lennart - Jones i.e. van der Waals interaction energies. The L-J interaction gives approximately the packing of the molecules around the ligand which obviously is not the origin of hydrophobic interactions, but it can been used to predict them. The scaling factor α is calibrated against experimental data. The non - polar contribution to the binding can be written in the analogy with the electrostatic binding energy

$$\Delta G_{bind}^{vdW} = a \cdot (< V_{l-s}^{vdW} >_{bound} - < V_{l-s}^{vdW} >_{free}). \quad (37)$$

The total binding energy given by LIE method is thus

$$\Delta G_{bind}^{total} = a \cdot (< V_{l-s}^{vdW} >_{bound} - < V_{l-s}^{vdW} >_{free}) + b \cdot (< V_{l-s}^{el} >_{bound} - < V_{l-s}^{el} >_{free}) + g \quad (38)$$

where an additional constant term γ is also included. The γ term arises due to the nonzero difference in the constant terms for solvation free energies in the protein and water. It is well established that the constant term may exist, for example extrapolation of solvation free energy of hydrocarbons towards zero chain length results in the non zero free energy ΔG_{sol}=l for n=0.

The model used here was optimised on a set of eighteen compounds, and incorporates specific deviations from the electrostatic linear response in the β term depending on the chemical composition of the inhibitor (Figure 5.). [7] The optimal α value was determined to 0.181 with the root mean square deviations of 0.84 kcal/mol. Here the LIE equation was optimised with the γ term equal to zero since the addition of the constant term γ did not improve the model in this case. Besides the calculated value of the γ term was 0.02 kcal/mol which is virtually zero.

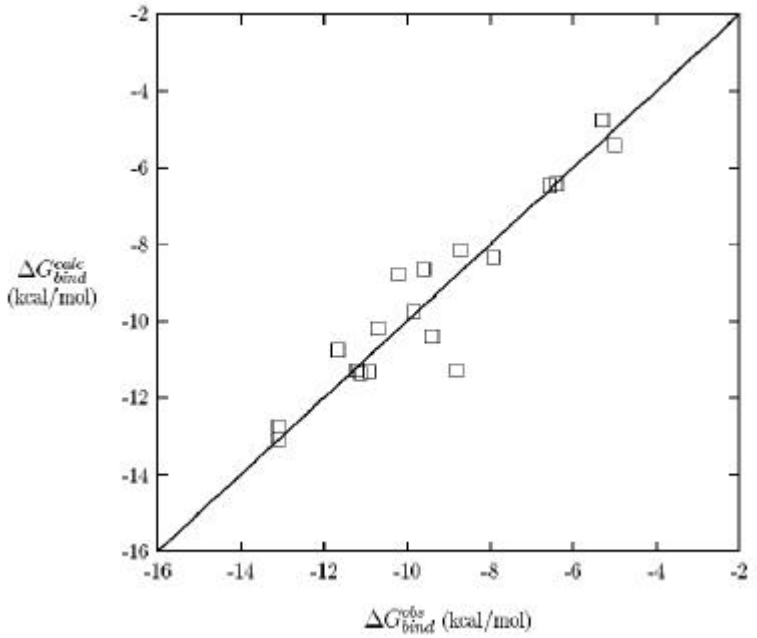


Figure 5. The calculated (calc) versus observed (obs) absolute free energies of binding for the eighteen compounds used for the calibration of the LIE equation with the deviations from the electrostatic linear response incorporated in the β term. [7]

The current model was applied to the pepstatin A/plasmepsin II complex to verify it. The calculated free energy of binding of pepstatin A to plasmepsin II was determined to ΔG_{bind}=-9.67 kcal/mol while the observed value was ΔG_{bind}=-15.4 kcal/mol (K_i=0.006 nM [8]). This implies the presence of the γ term here equal to -5.7 kcal/mol. Unfortunately, due to the insufficient binding information present on the inhibitor - plasmepsin II complexes the reparemetrisation of the LIE equation could not be carried out. The calculated values here are thus the relative free energies of binding.

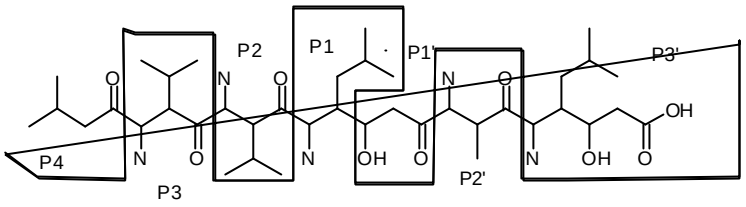
3.2 Computer simulations

3.2.1 Initial configuration: starting structure and ligands

To perform a computer simulation we need a structural model that closely resembles a system that we are interested in. The model can be derived in different ways for example from X-ray crystallography, NMR or by theoretical modelling. In my case the structure of plasmepsin II - obtained from Protein Data Bank (PDB) with ID number 1SME - was derived by X-ray diffraction [8] to a resolution of 2.70 Å and an

R-value of 0.195. The unit cell contains two copies of the protease ‘a’ and ‘b’, in the complex with pepstatin A, a general inhibitor of aspartic proteases. Each of the positions P in the inhibitor, counted from the point of cleavage as in the substrate with the prime side at the carboxyl group, occupies a corresponding subsite S located in the active site of the protease. The standard labelling of binding elements for proteases as proposed by Schechter and Berger [9] is given in Figure 6.

a)



b)

s4	s3	s2	s1	s1'	s2'	s3'
79	15	78	32	34	36	75
218	32	79	34	36	37	76
219	79	214	77	77	75	77
243	111	217	79	78	76	78
286	114	221	111	192	77	130
288	216	290	123	214	78	131
290	217	292	216	216	131	133
	218	300	217		192	191
	219					294

Figure 6. a) Pepstatin A subsite classification with each respective position P occupying a specific subsite S. b) The mapping out of the amino acids (numbers in the table correspond to the 1SME nomenclature) in the active site of pepstatin A/plasmepsin II complex is carried out by determining the amino acids in contact with the P moieties of pepstatin A and assigning them to corresponding subsites. One amino acid can belong to more than one subsite S and more amino acids can belong to a subsite but are not listed here since they do not interact with Pepstatin A.

The two copies do not display the same conformation and the root mean square deviation for the superposition of all 329 Cα atoms of both molecules is 0.93 Å.[10] The copy ‘a’ was chosen for ligand binding due to its lower B-factor determined by an inspection in RASMOL. The inhibition constant K_i for pepstatin A has been measured to 0.006 nM, which gives us a very potent inhibitor on which new ligands can be designed. The cell also contained 122 crystallographic waters that were all removed but the five of them HOH 804, HOH 820, HOH 826, HOH 851 and HOH 392 (1SME nomenclature). The water molecules removed were on the surface of the protein and were considered to be unimportant for the structural stability and ligand binding.

The catalytic site of plasmepsin II formed by the two aspartic residues, asp 34 and asp 214, was modelled in such way that a negative charge was assigned to asp 34 while asp 214 was protonated. This is not in accordance with accepted mechanism of pepsin-like enzyme function where the situation is opposite. [11] But in this case the asp 34 has two opportunities to H-bonding, ser 37:O-asp 34 and gly

36:N-asp 34, while asp 214 has only one, gly 216:N-asp214. Besides these two amino acids also asp 303 and arg 307 were modelled, by protonating aspartate and by assigning a positive charge to the arginine. The arginine was chosen since it is the closest amino acid to the negative aspartate that can be positively charged, and it neutralised the effects of the negative charge from aspartate on the protein structure. Asp 303 was protonated so it can participate in the H-bonding with the Ser 215.

The ligands, were built using InsightII software [12] with standard bond lengths and angles, using the implemented BUILDER module. They consist of four different isomers, with six chiral centres (the valines, the hydroxyl groups and fatty groups R) and a two-fold symmetry about the vertical axis passing between the hydroxyl groups. The R group is either a phenyl or a vinyl group. Thus altogether there are eight different inhibitors. The structural information about inhibitors is shown in Figure 7.

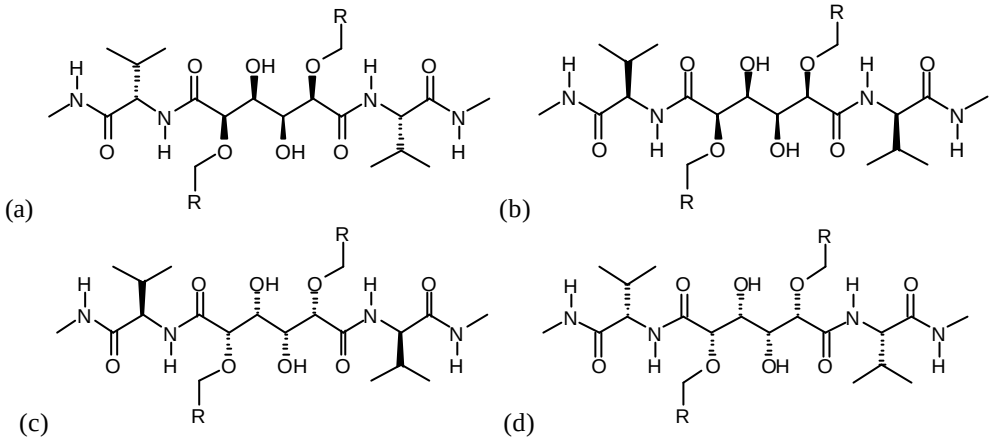


Figure 7. Depending if the L-mannitol or D- mannitol and L-valine or D- valine are used for the synthesis of the ligands, the isomers are allyloxy L,L or bensyloxy L,L (a), allyloxy L,D or bensyloxy L,D (b), allyloxy D,D and bensyloxy D,D (c), and allyloxy D,L or bensyloxy D,L (d) if the R-group is a vinyl or phenyl group respectively.

Allyloxy L,L (Figure 7a) was docked manually using InsightII software in three different conformations. The part of pepstatin A backbone extending through the subsite S'_2 and the end and the beginning of the subsites S'_2 and S'_3 respectively is the same as in the inhibitor so it was possible to overlay them closely in the all three conformations. In the first docking (**I**) the rest of the backbone was superimposed on the pepstatin A backbone, the fatty groups were fitted into the subsites S'_1 and the S_1 that is occupied by a central statin moiety in the pepstatin A. Furthermore this positions the hydroxyl group in the subsite S'_1 similar to the central statin hydroxyl group in the pepstatin A, which binds to the catalytic aspartates and is thought to be principal for the inhibition. The valines were modelled so that one of them is in the subsite S_2, occupying the same place as the valine in the pepstatin A, while the other one is in the subsite S'_2 instead of the pepstatin A methyl group. In the second docking (**II**) the other half of the backbone was modelled in the subsite S'_1, one fatty group pointing in the same direction as the papstatin A backbone and the other one is in the subsite S_1. In the third docking (**III**) the backbone occupies approximately the same position as in the first docking but the with fatty groups

pointing into the subsites S'_1 and S_3. The hydroxyl groups were also positioned towards each of the catalytic aspartates in the last two dockings. The dockings are shown in the Appendix 1. The docking that gave the best interaction energies as well as the binding conformation in the active site served as the starting point for the docking of the rest of the inhibitors.

During the docking the torsional angles were rotated which can lead to the introduction of the strain in the inhibitor. To relax the ligand an energy minimisation was performed in vacuum for 13.2 ps at the temperature of 10.0 K but the first 0.2 ps were done at 1.0 K. During the first 0.2 ps the system also had a very strong coupling of τ=0.2 fs to the temperature bath but was relaxed later to 10 fs (Table I.) While the protein and the water molecules were restrained to their initial positions by a harmonic potential force constant of 100 kcal/(mol·Å^2) the restraints on the inhibitor were relaxed with time from 100 kcal/(mol·Å^2) to 0 so that the inhibitor could be minimized. Moreover, atom distance restraints between the inhibitor and protein were introduced to keep the hydrogen bonds during minimisation. However, these restraints were removed during the last 4 ps of the minimisation.

The simulation was performed on the atoms inside the sphere of 18Å radius from the P_1 hydroxyl oxygen group. All atoms outside the sphere were restrained by a force constant of 200 kcal/(mol·Å^2).

Table I. The specifications for ligand minimisation.

	Steps	stepsize (fs)	temperature (K)	bath-coupling (fs)	inh-restraint (kcal/(mol·Å^2))
1	1000	0.2	1.0	0.2	100
2	1000	1.0	10.0	10.0	10
3	4000	1.0	10.0	10.0	5
4	4000	1.0	10.0	10.0	2
5	4000	1.0	10.0	10.0	0

Both the minimisation and the following molecular dynamics simulations were carried out using the program Q [4] its version of the GROMOS87 [13]force field. All non-bonded interactions are truncated at 10 Å cutoff distance, and the long-range electrostatics were treated by the local reaction field (LRF) method. In the MD simulations of the ligands, a sphere of 18 Å radius was filled with approximately 770 SPC water molecules, ca 2300 atoms. The ligands were held in the centre by a harmonic positional restraint on one of the hydroxyl oxygens. The protein/ligand complex simulations were carried in a sphere of the same size, with its centre at the hydroxyl oxygen in the subsite S_1. The sphere contained about 260 water molecules, which is ca 800 atoms, and ca 1100 protein atoms. Some 2000 protein atoms that extended out of the simulation sphere were held fixed by a harmonic potential force constant of 200 kcal/(mol·Å^2) while about 250 atoms in the innermost 1.5 Å shell from the simulation sphere are held by a force constant of 50 kcal/(mol·Å^2).

To relax the structure of the solvated protein/ligand complex an equilibration was performed by slow heating of the system from 1 K to 300 K during 59 ps. At the same time the harmonic potential restraints

on the complex were reduced from 100 kcal/(mol·Å^2) to 0. The equilibration of the ligand in water was performed in the same manner, except the first 9 ps which were not included due to a small system of about 50 atoms (Table II.) The ligand was held centred by a 100 kcal/(mol·Å^2) harmonic restraint on one of the central hydroxyl oxygens.

Table II. The equilibration of the ligand performed prior to the production phase.

	Steps	stepsize (fs)	temperature (K)	bath-coupling (fs)	complex restraint (kcal/(mol·Å^2))
1	1000	0.2	1	0.2	100
2	4000	0.2	50	10	10
3	4000	1	150	10	5
4	4000	1	300	10	2
5	50000	1	300	10	0

The equilibration was followed by the production phase during which the energy data was collected. During this phase, the molecular dynamics simulation was carried out for 200 ps (the allyloxy L,L was run for 100 ps extra since last half of the run had a much better interaction energy than the first half - mean energies were calculated over last 200 ps) with the step size of 1 fs and the energies were collected every 50 steps (Table III.) The energy sampling was also tried every 10 steps (performed on pp II/pepstatin A complex, not shown here) but the difference between the energy averages from the two runs was less then 0.03%.

Table III. The production phase of the molecular dynamics simulation

Steps	stepsize (fs)	temperature (K)	bath-coupling (fs)	complex restraint (kcal/(mol·Å^2))
50000	1	300	10	0
50000	1	300	10	0
50000	1	300	10	0
50000	1	300	10	0

.

4. RESULTS

The MD simulation was performed on the three initially docked structures, after the minimisation and equilibration. The electrostatic (el) and van der Waals (vdW) interactions were sampled during 200 ps of the MD simulation and the average values were calculated. The docking **I** of allyloxy L,L gave the best interactions energies (Table IV.) as well as H-bonding and root mean square (rms) deviation compared to the docking **II** and **III**.

Table IV. The el and vdW interaction energies V for the three different dockings of allyloxy L,L.

	I (kcal/mol)	**II** (kcal/mol)	**III** (kcal/mol)
$< V_{l-s}^{el} >_{bound}$	-65.39	-61.16	-61.47
$< V_{l-s}^{vdW} >_{bound}$	-61.84	-51.58	-58.52

The rms deviation was calculated between the initial structure and the end structure after the 200 ps of the MD simulation. The rms was determined for all heavy (non-hydrogen) atoms and Cα atoms only (Table V.) The large-scale protein motion outside the sphere was eliminated due to the restraints, and the motion inside the sphere gave the realistic interactions, that is no unnatural distorting forces were present due to the modelling and simulation of the complexes.

Table V. The root mean square deviations of the initial and end structures calculated for the Cα and heavy atoms for the docking **I**, **II** and **III** of allyloxy L,L.

	I (Å)	**II** (Å)	**III** (Å)
Cα	0.51	0.76	0.86
heavy atoms	0.80	1.02	1.30

The hydrogen bonding pattern between the ligand and enzyme in the simulated complex was determined and compared to the existing H-bond network from the pepstatinA in the crystal structure of plasmepsin II (Table VI.) The hydrogen bonds of docking **I** were similar to the pepstatin A, while dockings **II** and **III** had two respectively three H-bonds less than docking **I**. Moreover the docking **I** and **II,** contain an additional H-bond with Thr 217 which is not present in the crystal structure.

Table VI. The hydrogen bonding between pepstatin A and the allyloxy L,L docking **I**, **II** and **III** respectively and the plasmepsin II. The distances in Å are between the heavy atoms N-O or O-O.

	Pepstatin A	**I**	**II**	**III**
Asn76	3.0	3.2	3.3	3.0
Tyr192	2.4	2.9	-	2.8
Gly36	2.9	3.1	-	3.2
Val78	2.7	3.2	3.1	3.1
Val78	-	3.0	-	-
Asp34	2.8	2.7	2.6	2.7
Asp34	3.0	3.3	2.9	3.0
Ser218	2.9	3.4	-	-
Ser218	3.0	-	-	-
Thr217	-	2.8	3.1	-
Ser79	2.8	2.8	2.9	2.8
Ser79	2.6	-	2.8	-

The rest of the inhibitors (Figure 7b,c,d) were modelled into the active site in the same way as docking **I,** i.e. by trying to overlay the backbones of the ligand with the backbone of pepstatin A in the crystal structure and by modelling the fatty groups into the subsites S'_1 and S_1, while the valines were fitted into the subsites S_2 and S'_2. MD simulations were carried out in the same way as above. The interaction energies between the ligand and enzyme were measured and the relative binding energies were calculated by the LIE method

$$\Delta G_{\text{bind}}^{\text{total}} = \alpha \cdot (< V_{l-s}^{vdW} >_{\text{bound}} - < V_{l-s}^{vdW} >_{\text{free}}) + \beta \cdot (< V_{l-s}^{el} >_{\text{bound}} - < V_{l-s}^{el} >_{\text{free}}) + \gamma \quad (39)$$

where α=0181, β =0.33 and γ=0 (Table VII. a) and b)) To explore the configuration space of the free ligands four different simulations were carried out for allyloxy L,L, bensyloxy L,L and bensyloxy L,D in water simulation and the average values were calculated. Due to the symmetry between L,L - D,D and L,D - D,L isomers the mean value for these was also calculated (Table VII. b))

Table VII. The mean interaction energies for allyloxy L,L, allyloxy D,D, allyloxy L,D and allyloxy D,L and bensyloxy L,L, bensyloxy D,D, bensyloxy L,D and bensyloxy D,L in the complex with the solvated plasmepsin II (a) and in the solution (b). The free energy of binding ΔG_{bind}, is also determined for the respective complexes (c). The subscripts denote the different isomers of the allyloxy and bensyloxy and V^{el} and V^{vdW} stand for the electrostatic and van der Waals free energy of solvation.

a)

	$<V^{el}>_{L,L}$	$<V^{vdW}>_{L,L}$	$<V^{el}>_{D,D}$	$<V^{vdW}>_{D,D}$	$<V^{el}>_{L,D}$	$<V^{vdW}>_{L,D}$	$<V^{el}>_{D,L}$	$<V^{vdW}>_{D,L}$
ppII/allyloxy	-61.84	-65.39	-47.80	-62.27	-44.36	-64.61	-53.92	-62.87
ppII/bensyloxy	-67.71	-81.92	-53.47	-76.47	-64.9	-79.47	-52.4	-77.62

b)

	$<V^{el}>_{L,L-D,D}$	$<V^{vdW}>_{L,L-D,D}$	$<V^{el}>_{L,D-D,L}$	$<V^{vdW}>_{L,D-D,L}$
water/allyloxy	-63.41	-42.68	-69.11	-43.33
water/bensyloxy	-66.40	-54.56	-70.37	-51.39

c)

	$(\Delta G_{bind})_{L,L}$	$(\Delta G_{bind})_{D,D}$	$(\Delta G_{bind})_{L,D}$	$(\Delta G_{bind})_{D,L}$
ppII/allyloxy	-4.20	1.60	4.31	1.47
ppII/bensyloxy	-5.39	0.30	-3.27	1.18

* all interaction energies are in kcal/mol

From Table VII c) it can clearly be seen that allyloxy L,L and bensyloxy L,L have the best free energie of binding of -4.20 kcal/mol and -5.39 kcal/mol respectively, relative to the rest of the isomers in the respective groups. The experimentally active isomers i.e. the ones that inhibit plasmepsin II are allyloxy L,L and bensyloxy L,L which is reproduced here. These computer simulations predict that bensyloxy L,D also shows the activity with a free energy of binding of -3.27 kcal/mol. The structure of plasmepsin/bensyloxy L,D was investigated further, together with plasmepsin II/allyloxy L,L and plasmepsin II/bensyloxy L,L, in order to explain this and to determine the mode of binding in the active site of the enzyme. From these molecular dynamics simulations snapshot structures were sampled every

50 fs during last 50ps of the run thus giving 1000 snapshots of the complex from which average structures were calculated. The average structures were compared with the crystal structure of plasmepsin II in the complex with the pepstatin A (Figure 8.) The pepstatin-like part of the ligands binds similar to pepstatin A while the rest of the ligand seems to bind differently.

allyloxy L,L/pepstatinA bensyloxy L,L/pepstatinA bensyloxyL,D/pepstatinA

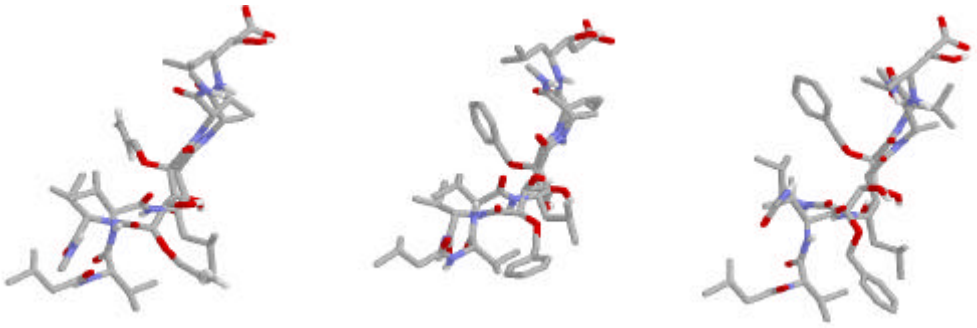


Figure 8. The comparison of the molecular dynamics average structures for allyloxy L,L, bensyloxy L,L, and bensyloxy L,D with the crystal structure conformation of the pepstatin A.

The hydrogen bonding pattern between plasmepsin II and the ligands allyloxy L,L, bensyloxy L,L and bensyloxy L,D was examined (Table VIII.) The difference in the binding energy between the active L,L isomers can partially be explained by the presence of the strong H-bond between Tyr192 and the bensyloxy ligand. This bond is present in the crystal structure and should be also expected in the pepstatin like part of the inhibitors. Bensyloxy L,D has, on the other hand, lost a H-bind to Asn76 and displays at the same time a free energy of binding that is less negative compared to the both other isomers due to the loss of the electrostatic interaction energy. It has also an internal H-bond between terminal N-H and second O in the pepstatin-like part of the backbone that can affect the stability of the ligand conformation.

Table VIII. The protein-ligand hydrogen bonding in complexes with between pepstatin A, allyloxy L,L, bensyloxy L,L and bensyloxy L,D determined from the average structures made from 50 ps MD simulations. The distances in Å are between the heavy atoms N-O or O-O..

	Pepstatin A	**allyloxy L,L**	**bensyloxy L,L**	**bensyloxy L,D**
Asn76	3.0	2.9	3.1	-
Tyr192	2.4	-	3.1	2.6
Gly36	2.9	3.0	3.2	2.9
Val78	2.7	2.8	2.9	3.2
Val78	-	-	3.2	3.2
Asp34	2.8	3.2	2.8	2.8
Asp34	3.0	2.9	-	-
Ser218	2.9	3.0	2.9	3.2
Ser218	3.0	-	-	3.1
Thr217	-	2.9	3.1	2.9
Ser79	2.8	2.8	3.3	3.1
Ser79	2.6	3.5	-	-

.

The rms deviation was calculated between the initial structure and the average structures for both all the heavy atoms and all Cα of the complex (Table IX.) Bensyloxy L,L in the complex with plasmepsin II has the smallest deviation for both the heavy atoms and Cα atoms compared to allyloxy L,L and bensyloxy L,D. This may imply that the best docking has been achieved by the bensyloxy L,L inhibitor since it has the best free energy of binding, the lowest rms deviation and very good H-bonding compared to other inhibitors.

Table IX. The root mean square deviation of the initial and the average structures of allyloxy L,L, bensyloxy L,L and bensyloxy L,D in complex with plsamepsin II calculated for the heavy and Cα atoms.

	ppII/allyloxy L,L (Å)	ppII/bensyloxy L,L (Å)	ppII/bensyloxy L,D (Å)
Cα	0.45	0.26	0.45
heavy atoms	0.75	0.51	0.86

The deviation in the active site was also analysed, and especially the amino acids in the S'_1 subsite, that is considered to be collapsed in the crystal structure, and the amino acids in the flap region - from Asn 76 to Thr 81 (Appendix 2.) The rest of the active site amino acids did not deviate considerably judging from the rms values. In the plasmepsin II/allyloxy L,L complex Val 78 moved 1.2 Å in the average structure compared to the crystal structure. Movement was also observed for the amino acids Phe 294 1.5 Å close to the subsite S'_3 and Tyr 192 0.91 Å that is in the S'_1 subsite, Asp 34 also displayed motion with 0.9 Å. The Plasmepsin II/bensyloxy L,L complex showed the smallest deviation from the crystal structure for example Val 78 moved by 0.42 Å while Phe 294 moved by 0.82 Å also Asp 34 moved by 0.25 Å. The plasmepsin II/bensyloxy L,D complex had the greatest deviations from the crystal structure, particularly the flap region with Val 78 moving 2.0 Å. It seems that allyloxy L,L expands the S'_1 subsite with the fatty group while bensyloxy L,L and L,D bind also partially in the S'_2. Bensyloxy L,L had again the smallest deviation, and confirmed that it may have the best docking relative to the rest of the inhibitors.

In Appendix 3. the protein-ligand electrostatic and van der Waals interaction energies are displayed for the allyloxy L,L, bensyloxy L,L and bensyloxy L,D complexes. Every point here was calculated by sampling the energies every 50 fs and than averaging them every picosecond. The interactions were more or less stable with the periodic variations around the average values, indicating that the ligand/plasmepsin II complex is equilibrated and varies around the minimum configuration.

.

5. DISCUSSION

The goal of this work was to examine if the LIE method could be used for predicting the stereo specificity of the enzyme plasmepsin II, which is an aspartic protease of the malaria parasite *Plasmodium. falciparum.* In this case the binding affinity was estimated for two inhibitors, each with four different stereo-isomers. If the two groups of inhibitors are compared, the molecular dynamics simulation and the LIE method predict that both allyloxy L,L and bensyloxy L,L are the active ligands with binding energies of -4.20 kcal/mol and -5.39 kcal/mol respectively. In the bensyloxy group also L,D isomer displayed activity with the free energy of binding of -3.27 kcal/mol. The structure of the bensyloxy L,D complex was investigated further, and it was found to bind similarly to bensyloxy L,L but with a larger rms deviation of the average MD structure compared to the initial structure, especially in the flap region and the S'_1 subsite. The predicted free energies of binding are the probably the relative ones since an additional term γ may be present.

The LIE model used here was optimised for a set of 18 compounds, using four different β values depending on the chemical composition of the inhibitor, and with the γ term equal to zero. The γ term can arise as the nonzero difference between the constant terms in the linear solvation energies but its introduction in the model used here did not offer any improvement during parametrisation on eighteen compounds described in methods section. This model has been successfully applied to predicting the absolute free energies of binding for a number of different proteins, for example arabinose binding protein, lysine binding protein and fatty acid binding protein (Figure 9.) [14]

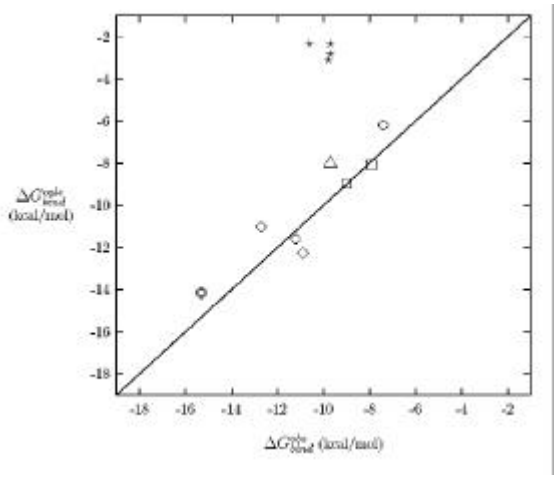


Figure 9. The calculated (calc) versus observed (obs) free energies of binding for the DHFR complexes (circles), arabinose binding protein complex (triangle), lysine binding protein complex (diamond), fatty acid binding protein complexes (squares) and RBP complexes (strars). [14]

These proteins have polar binding sites, while simulations of the proteins with hydrophobic binding sites gave completely wrong predictions with this LIE model. For the retinol binding protein (RBP),

the affinities of the set of hydrophobic retinoid ligands are offset with an error of +7 kcal/mol (Figure 9.) It can be also noted that human dihydrofolate reducatse (DHFR), circles in the Figure 9., shows a deviation from the observed binding energies possibly due to an additional γ term.

When the model was applied to eight thrombin inhibitors binding to human thrombin the energies were now offset by +3 kcal/mol. Least-square optimisation of the LIE model, by holding the α and β factors at their earlier determined values, gave a γ term equal with -2.916 kcal/mol and a mean unsigned error of 0.68 kcal/mol (Figure 10.) [15] The new model was able to predict accurately the experimental data.

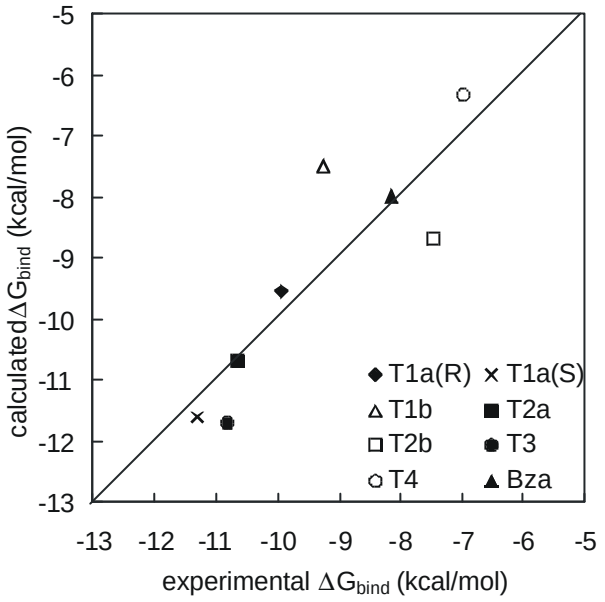


Figure 10. The calculated (calc) versus observed (obs) free energies of binding for the eight thrombin complexes. [15]

To investigate if the original model with γ=0 could be used to calculate the absolute binding energies in our study it was tested on the pepstatin A/plasmepsin II complex. The error turns out to be +5 kcal/mol which implies the presence of a nonzero γ term and the calculated values here are the relative free energies of binding. Unfortunately the reparametrisation of the LIE equation could not be carried out due to the insufficient binding data available for the inhibitor/plasmepsin II complexes.

The results can also be influenced by the starting structures. The pepstatinA/plasmepsin II had rather poor resolution of 2.7 Å, which can influence accuracy of the computer simulation. The other problem is the size of the configuration space that is accessible to the enzyme/ligand complex. The docking of the ligand is for that reason crucial for the result. To avoid this, the three different dockings were performed manually on the plasmepsin II/allyloxy L,L complex and the best one was chosen for the docking of the rest of the inhibitors. The manual docking should suffice since part of the ligand backbone is the identical to the pepstatin A. The docking, that gave the strongest interactions turned out to be for bensyloxy L,L which besides having the best interaction energies of all complexes also had the least rms deviation from the initial structure. The other problem is the flexibility of the ligand during the

simulation. The inhibitors may have no well defined structure in the aqueous solution, indicating that their free energy actually reflects a wide range of conformations, and it is possible that the present simulations do not sample the most probable conformation of the inhibitor.

In this work the mutual ranking between the different isomers has been predicted by the molecular dynamics simulation and the LIE method with both LL isomers being the active ones. The absolute free energies of binding have not been reproduced since the LIE equation could not be optimised for the current system. The inhibitor binding observed from the mean structures has a configuration that closely resembles that of pepstatin A in the crystal structure. The inhibitors modelled here are not very strong but they still can be used as the starting point for synthesis of the more potent inhibitors of plasmepsin II. The work presented here can be useful for the further study of the new compounds in the drug design against the malaria.

6. REFERENCES

1. Rosenthal,P.J., Proteases of Malaria Parasites: New Targets for Chemotherapy. (1998) Emerg. Infect. Dis., 4, 49-57.

2. Grant H.G. and Richards W.G., Computational Chemistry, Oxford University Press, Oxford, 1995.

3. Leach R.A., Molecular Modelling: Principles and Applications. Pearson Education Limited, Harlow, 2001.

4. Marelius J., Kolmodin K., Feierberg I. and Åqvist J., Q: a molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems. (1998) J Mol Graph Model.,16 ,213-225.

5. Åqvist, J., Medina, C and Samuelsson, J-E., A New Method for Predicting Binding Affinity in Computer-Aided Drug Design. (1994) Protein Eng., 7, 385-391.

6. Åqvist, J. and Hansson T. On the Validity of Electrostatic Linear Response in Polar Solvents. (1996) J. Phys. Chem., 100, 9512-9521.

7. Marelius, J., Hansson, T. And Åqvist, J., Calculation of Ligand Binding Free Energies from Molecular Dynamics Simulations. (1998) Int. J. Quantum Chem., 69, 77-78.

8. Silva A.M., Lee A.Y., Gulnik S.V., Maier P., Collins J., Bhat T.N., Collins P.J., Cachau R.E., Luker K.E., Gluzman I.Y., Francis S.E., Oksman A., Goldberg D.E. and Erickson J.W., Structure and Inhibition of Plasmepsin II, a Hemoglobin-Degrading Enzyme from *Plasmodium falciparum.* (1996) Proc Natl Acad Sci U S A., 19, 10034-10039.

9. Schechter I. and Berger A., On the Size of the Active Site in Proteases. I. Papain., (1967), Biochem. Biophys. Res. Commun., 27, 157-162.

10. Haque T.S., Skillman A.G., Lee C.E., Habashita H., Gluzman I.Y., Ewing T.J., Goldberg D.E., Kuntz I.D., Ellman J.A., Potent, Low-Molecular-Weight Non-Peptide Inhibitors of Malarial Aspartyl Protease Plasmepsin II. (999) J. Med. Chem. 8, 1428-1440.

11. Andreeva N.S. and Rumsh L.D., Analysis of Crystal Structures of Aspartatic Proteinases: On the Role of Amino Acid Residues Adjecent to the Catalytic Site of Pepsin-Like Enzymes. (2001) Protein. Sci., 10, 2439-2450.

12. InsightII, Biosym/MSI, San Diego, CA, USA, 1995.

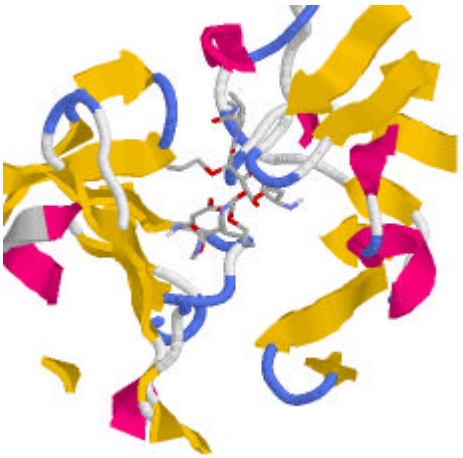
13. van Gunsteren, W.F. and Berendsen, H.J.C. (1987) Groningen Molecular Simulation (GROMOS) Library Manual, Biomos B.V., Groningen.

14. Aqvist J. and Marelius J., The linear interaction energy method for predicting ligand binding free energies. (2001) Comb. Chem. High Throughput Screen., 8, 613-26.

15. Ljungberg K.B., Marelius J., Musil D., Svensson P., Norden B. and Aqvist J., Computational Modelling of Inhibitor Binding to Human Thrombin. (2001) Eur. J. Pharm. Sci., 12, 441-446.

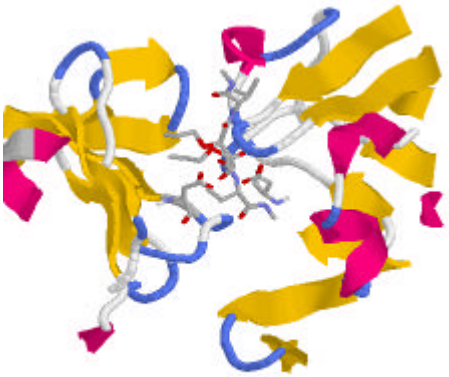
Appendix 1. The three different dockings of the allyloxy L,L.

I.



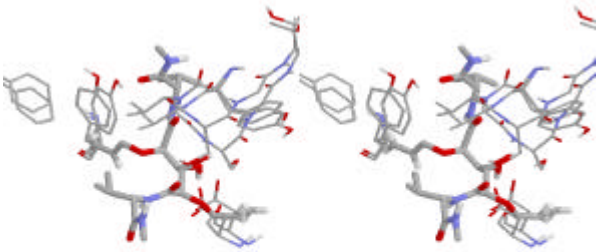
II.

III.

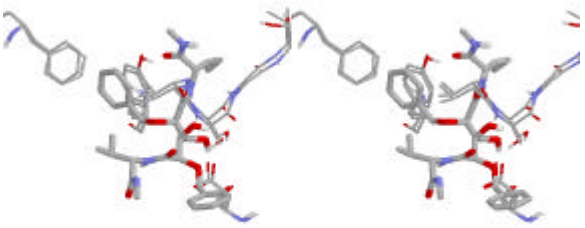


Appendix 2. The deviation of the active site of plasmepsin II with bound allyloxy L,L, bensyloxy L,L and bensyloxy L,D respectively.

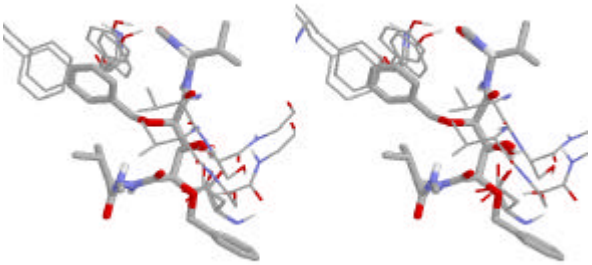
ppII/allyloxy L,L



ppII/bensyloxy L,L

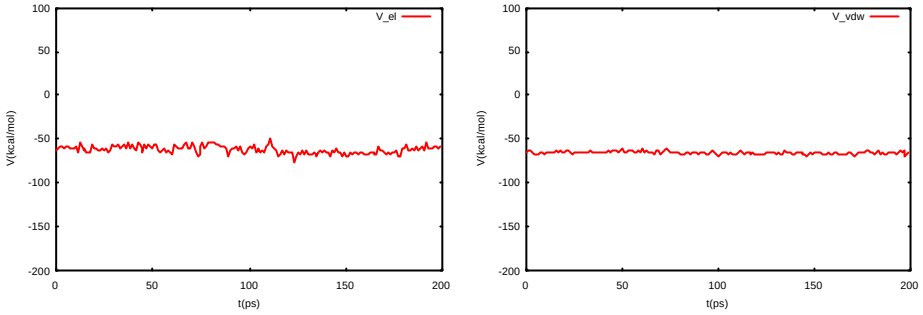


ppII/bensyloxy L,D

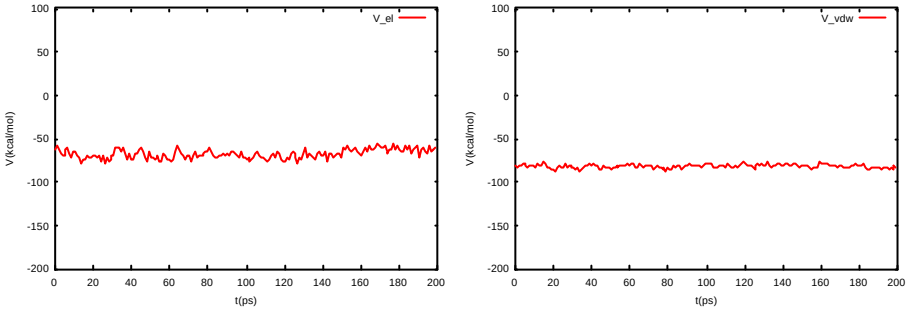


Appendix 3. The deviation of interaction energies between allyloxy L,L, bensyloxy L,L and bensyloxy L,D in complex with plasmepsin II respectively.

allyloxy L,L/ppII



bensyloxy L,L/ppII



bensyloxy L,D/ppII