Gerrit

## Introduction

The increasing resistance of malarial parasites to the existing antimalarial drugs, and in particular of *Plasmodium falciparum* (Pf), has focused efforts towards the discovery of more selective and potent drugs. [1] Hemoglobin(Hb)-degrading enzymes of Pf, emerge as very promising chemotherapeutics targets, because Hb degradation is an unique and critical process for Pf. [1, 2] Up to date, Hb degradation is catalyzed by the concerted action of aspartyl-,[3] cysteine-,[4] metallo-,[5] and a dipeptidyl amino peptidases [6] within the digestive vacuole (DV) of the parasite.[7] Earlier studies indicated that the aspartic proteases plasmepsins (Plms) play essential roles in Pf life cycle, due to the effectiveness of Plms inhibitors abolishing Hb degradation, erythrocyte rupture and parasite development. [1, 8] However, the redundant functional role of these enzymes in Hb digestion has been demonstrated by Plm deletion.[9, 10] This feature indicated that more effective drugs may be only obtained by blocking all Plms.[11]

*Plasmodium falciparum* PlmII has been the most extensively characterized of these enzymes with several crystal structures determined, [12, 13, 14, 15] and potent inhibitors developed. [16, 17] Approaches used in PlmII structure-based ligand design usually include some popular docking algorithms, [18, 19, 20, 21, 22, 23, 22] to predict conformations of inhibitor complexes, followed by a method to estimate the binding affinity.[17] However, most of the PlmII inhibitors obtained using these methods have generally shown a limited selectivity towards the human related protease Cathepsin D (hCatD).[24] On the other hand, the high degree of structural flexibility of the PlmII active site cavity [25, 26] allows the accommodation of different inhibitors scaffolds, [12, 13, 14, 15] which is a notable drawback for drug design using the traditional rigid docking approaches.

Synthetic peptide combinatorial libraries have been used to determine the substrate preference of malarial enzymes in the aspartic protease family.[27] An early study of Westling and coworkers suggested the major role of the S3 subsite pocket in the PlmII specificity, based on the higher differences in the kinetic parameters obtained by the amino acid variation in the substrate P3 position. [28] These authors found that PlmII prefers large hydrophobic residues within the S3 subsite pocket, while K or D substitutions in P3 were not well tolerated. [28] Conversely, the S3 subsite of the hCatD is a hydrophilic and a hydrophobic pocket and indeed, it has been shown that M, I, S and T are the preferred P3 residues for the human enzyme, with the hydrophilic residues binding to the Q14 side chain and the hydrophobic residues binding to hydrophobic amino acids of this pocket. [29] It should be noted that the replacement of the M15

(according to the PlmII numbering scheme) by the Q14 (according to the hCatD numbering scheme) is one of the most important differences between the S3 pocket residues of the parasite and the human protein, respectively. [26] However, the thermodynamic effect of the changes in the P3 position of well known PlmII substrates have not been studied by free energy calculations from atomic level simulations.

expand

Among the simulation protocols that aim at calculating free energy differences, perturbation approaches, [30, 31, 32] have grown in popularity over the last years. In these protocols the Hamiltonian $H$ is coupled to a parameter $\lambda$ which is used to drive a system from a state A ($\lambda = 0$) to a state B ($\lambda = 1$). Free energy differences between both states can be computed by equilibrium methods such as free energy perturbation [33] or thermodynamic integration,[34] or non-equilibrium methods such as those based on the work of Jarzynski,[35, 36] and Crooks. [37] Recently, Goette and Grübmuller showed that non-equilibrium methods outperformed the traditional equilibrium methods for test systems that include: *i*) the interconversion of ethane into methanol, *ii*) the switching from ($W \rightarrow G$) in a tripeptide, and *iii*) the binding of two different ligands to the globular protein snuportin 1. [38] Based on their results, they proposed a new non-equilibrium free energy method, termed the Crooks Gaussian Intersection (CGI), which combines the advantages of the available methods (see reference 38 for more details). This method has been used successfully to calculate the thermodynamic stability differences for 109 mutations in the microbial Ribonuclease Barnase. [39] Thus, the questions here are: *i*) can be the CGI method suitable applied to study the binding free energy differences of substitutions in the substrate P3 position over the PlmII substrate affinities?, and *ii*) What can we learn for these calculations for the PlmII inhibitor design process? This is not a strong motivation for a paper. Please rephrase

In this paper, we presented a detailed analysis of the free energy calculations with the CGI method of PlmII in complex with synthetic chromogenic substrates, to gain a better understanding of the preference of PlmII by large hydrophobic residues within the S3 subsite pocket. [28] Our results agreed with the experimental trend of Km values of PlmII with the assessed substrates. We also predicted the protonation state of the substrate P3 ionizable residues in the experimental assays. Finally, we showed the differential contributions of the enthalpic and entropy energy terms to the relative binding free energy differences of the changes from $F \rightarrow K^+$, and $F \rightarrow D^0$.

A good theoretical paper has at least a few predictions that can be tested experimentally! This is missing!!

## Results

### Analysis of PlmII:Substrate S3 subsite interactions

To better understand the known preference by hydrophobic residues of the PlmII S3 subsite pocket, [28] we calculated 3D models of the PlmII:Substrate complexes. Interestingly, we observed that the P3 position of the substrate was always solvent exposed (Figure 1). In addition, the analysis of the fraction of non-polar/polar contacts between each substrate P3 residue and the PlmII S3 subsite residues (Table I) showed that this rough parameter was not useful to explain the differences in the Michaelis constant values (Km) taken from the Westling and coworkers study.[28]

### Alchemy free energy calculations

Experimental binding free energy differences $\Delta\Delta G_{\exp}$ for P3 changes were calculated from the Michaelis constant values (Km).[28] The calculated binding free energy differences $\Delta\Delta G_{calc}$ were determined according to $\Delta G_4 - \Delta G_1$ (Figure 2). Therefore, changes in the P3 position of the substrate $(A \rightarrow B)$ that increases the affinity for the protein have a negative $\Delta\Delta G$, while changes that decrease the affinity have a positive $\Delta\Delta G$.

We reproduced the experimental trend of binding free differences for the changes in the P3 position of PlmII:Substrates complexes using the CGI method (Table II, Figure 3). As result, we obtained a Pearson product-moment correlation $r = 0.5$, with a mean absolute error $\langle |error| \rangle = 4.00 kJ/mol$, and a 60% of the calculated free energy differences within $\pm 4.18 kJ/mol$ of the experimental values (Figure 3A). In addition, we predicted that the lysine and aspartic acid residues must be in the protonated form in the experimental assays (TableII). It should be noted that taking in to account only the $\Delta\Delta G_{calc}$ from the protonation state of the ionizable residues that fit better with experimental $\Delta\Delta G_{\exp}$ values, we obtained a correlation $r = 0.75$, with a mean absolute error $\langle |error| \rangle = 2.58 kJ/mol$, and a 78% of the calculated free energy differences within $\pm 4.18 kJ/mol$ of the experimental values (Figure 3B).

To better understand the effect from changes in the P3 position over the PlmII:Substrates complexes affinity differences $\Delta\Delta G_{calc}$, we calculated the free energy differences $\Delta G$ in the bound and unbound simulation states (Table II). As can be seen, the change from $F \rightarrow I$ increase the affinity of PlmII for the substrate due to a higher unfavorable effect in the unbound system than in the complex system. Conversely, the

P3 changes from F to small hydrophobic (A), neutral polar (S, and $D^0$) or charged residues ($K^+$, and R) decrease the affinity of Plasmepsin II for the substrate. In this respect, our calculations show that this net effect is observed due to different consequences. The decrease in the affinity for the changes from $F \rightarrow A$, and $F \rightarrow K^+$ is caused by a higher unfavorable effect on the bound state than in the unbound state. On the contrary, the changes from $F \rightarrow S$, and $F \rightarrow D^0$ decrease the PlmII affinity for the substrate due to the lower favorable effect of these mutations on the complex system than in the solvent system. It should be noted that the decrease in the Plasmepsin II affinity for the change from $F \rightarrow R$ is the combination of an unfavorable effect in the complex with a favorable effect in the solvent system. Interesting, our calculations predict that the change for $F \rightarrow Y$ has a null effect over the PlmII:Substrate affinity thanks to a similar compensatory favorable effect in the complex and solvent systems. We have to remark that the P3 changes from other hydrophobic residues (I or A) to neutral polar (S, and $D^0$) or charged residues ($K^+$) decrease the affinity of Plasmepsin II for the substrate in a comparable fashion to the F mutation series. On the other hand, our calculations shown that the PlmII has a better affinity for neutral polar residues (S, and $D^0$) than for positive charge residue ($K^+$) because of a bigger favorable effect of the mutation on the complex system than on the solvent system.

**Entropy and enthalpy calculations**

To gain a better insight of the effect from the changes in the substrate P3 position over the PlmII:Substrates complexes affinities, we calculated the entropic and enthalpy contributions (Table III) to the binding free energy differences ($\Delta\Delta G_{calc}$), of the F mutations series, using the finite difference relationship of changes in the $\Delta S(T)$ with changes in the $\Delta G(T + \Delta T)$ and $\Delta G(T - \Delta T)$ (see equation 4, 5, and Table I of Supplementary Information). According to our calculations the change from $F \rightarrow I$ increase the affinity of PlmII for the substrate due to the raise in the entropy ($T\Delta\Delta S$) prevails over the enthalpy penalty ($\Delta\Delta H$). It should be noted that this mutation ($F \rightarrow I$) is an entropy driven process due to the combination of the enlarge in the entropy of the protein environment with the decrease of the entropy in the solvent environment (Table II). The P3 changes from F to small hydrophobic (A), polar neutral (S) and charge residues ($K^+$) reduces the affinity of the PlmII for the substrate because the enthalpy penalty ($\Delta\Delta H$) is higher than the raise in the entropy ($T\Delta\Delta S$) of the system (Table III).

origin

However, the raise in the entropy (T$\Delta\Delta S$) of these mutations have different ~~basis~~. In the case of the change from ($F \rightarrow K^+$) the raise in the entropy of the complex environment ~~prevails over~~ is higher than the raise in the entropy of the solvent environment. In the case of the change from ($F \rightarrow A$) the entropy penalty (T$\Delta S$) in the solvent environment is higher than in the protein environment. Conversely, the raise in the entropy of the change from ($F \rightarrow S$) is the combination of the raise of the entropy in the protein environment with the decrease in the solvent environment. On the other hand, the P3 change from ($F \rightarrow D^0$) reduce the affinity of PlmII for the substrate due to the entropy penalty (T$\Delta\Delta S$) overcomes the decrease in the enthalpy ($\Delta\Delta H$) of the system. Additionally, we have to remark that the entropy penalty (T$\Delta\Delta S$) in this mutation is the consequence of the drop off in the entropy of the protein environment.

This paragraph

To get a better understanding of the changes in the entropy of these systems due to the environment reorganization effects caused by the mutations in the P3 position of the substrate, we calculated the configurational entropy contributions ($T\Delta\Delta S^{conf}$, Table IV) to the binding entropy differences. It should be noted the relationship between the raise in the entropy of the environment reorganization ($T\Delta\Delta S - T\Delta\Delta S^{conf}$) with the increase of the bulky of the side chain mutated ($F \rightarrow K^+ \approx F \rightarrow I > F \rightarrow S > F \rightarrow A$). However, we observed that the entropy penalty of the P3 change from ($F \rightarrow D^0$) is determined by the entropy penalty of the environment reorganization effects.

→ This is selling in the right direction, but is too limited.

~~A compelling~~

~~descriptive → learning explanation~~

So far the paper is descriptive: we describe why the effect of mutations on P3. We should go beyond, and also try to describe what an ideal peptide inhibitor should look like.

In other words: what have we learned that can be used to design a better inhibitor?

## Discussion

### Understanding the PlmII S3 subsite preference

Understanding the specificity differences between PlmII and the human aspartic proteases is a useful requirement for designing specific inhibitors that will interact with the parasite protein and not with the human counterparts. Here, we studied the PlmII S3 subsite pocket preference by non-equilibrium free energy calculations using the CGI method. Earlier kinetic experiments indicated the major role of this subsite pocket in the PlmII specificity. [28] Our calculations showed that the P3 changes from hydrophobic residues (F, I or A) to neutral polar (S), acid ($D^0$) or charged residues ($K^+$, and R) decrease the affinity of Plasmepsin II for the substrate. This fit well with the experimental trend of relative binding free energy differences between PlmII:Substrate complexes obtained from Westling *et al* (Table II, Figure 3). [28] These authors found that PlmII prefers large hydrophobic residues within the S3 subsite pocket, while lysine or aspartic acid substitutions in P3 were not well tolerated. [28] Interestingly, we observed that the substrate P3 position was always solvent exposed in the PlmII:Substrate complexes. A previous study demonstrated the significant better accuracy of the CGI method for calculating the free energy change due to amino acid substitutions for surface residues (78.4% within ± 4.18 kJ/mol of experimental values) than for core residues (65.5%). [39]

We predicted that the lysine and aspartic acid residues in the substrate P3 position must be in the protonated form in the experimental assays carried out at pH = 4.4. [28] This prediction allowed explaining: *i*) the 2.5-fold decrease in Km of the PlmII M15E mutant for a substrate with D in the P3 position with respect to the wild-type protein, and *ii*) the 7-fold decrease in Km of this mutant for a substrate with K in P3 with respect to the wild-type protein. [28] These results demonstrated that the M15 residue contributes to the S3 subsite specificity of PlmII. [28] In addition, the M15E mutant data support the theory that the presence of a Glu at this position in aspartic proteinases enhances the binding of substrates/inhibitors with Lys in P3. [40, 41] Furthermore, because PlmII differs at position 15 from the human enzymes: cathepsin D, cathepsin E, and pepsin, [26] specific inhibitors may be rationally designed to take advantage of this specificity distinction. [28]

Despite previous studies indicated that PlmII has a stringent requirement for hydrophobic amino acid in the P3 position of peptide substrates. [28; 42, 43] It was demonstrated that the PlmII M15E mutant processes a hemoglobin-based substrate that

contains an Arg in the P3 position, with an equal efficiency as the wild-type enzyme. [28] Gulnik *et al* observed a similar behavior for the wild-type enzyme with another substrate containing Arg in the P3 position. [44] These authors suggested that the preference of PlmII for certain amino acids in particular subsites may depend strongly on the substrate sequence context. [44] On the other hand, our predictions indicated a higher decrease in the PlmII affinity for the change in the P3 position from $F \rightarrow R$ than from $F \rightarrow K^+$. This agreed well with the experimental data obtained by Beyer *et al.* [43]

Despite the limitations, the synthetic peptide combinatorial libraries have been widely used to determine the substrate preference of the Pf DV Plms.[27, 43] However, these approaches can not elucidate the thermodynamic origin of the low tolerance for the acid or basic residues within the PlmII S3 subsite. Here, we showed the different basis of the relative binding free energy differences due to the changes in the substrate P3 position. It should be remarked that the unfavorable effect of the change from $F \rightarrow D^0$ over the PlmII:Substrate complex affinity is the consequence of the lower favorable effect of this mutation on the bound state than in the unbound state. In the case of the change from $F \rightarrow R$ the decrease in the Plasmepsin II affinity is the combination of an unfavorable effect in the complex with a favorable effect in the solvent system. Conversely, the change from $F \rightarrow K^+$ reduces the affinity of PlmII for the substrate due to the higher unfavorable effect of this mutation on the complex system than in the solvent system (Table II).This results indicated that to improve the affinity of PlmII inhibitors is whished: *i*) mutations more favorable on the complex state than in the solvent state, *ii*) mutations less unfavorable on the complex state than in the solvent state, or ideally *iii*) mutations more favorable on the complex state and more unfavorable on the solvent state.

**What can we learn for PlmII inhibitor design?**

Designing more potent ligands to a known receptor needs to consider how to introduce new functional groups in the ligand structure that improve the transfer process of the ligand from the solution (free state) into the binding site of the solvated receptor (bound state). The binding affinity of a compound can be improved by generating a favorable binding enthalpy, favorable solvation entropy, or by minimizing the unfavorable conformational entropy.[45] It should be remarked that extremely high affinity is achieved when the three factors are optimized simultaneously.

Earlier thermodynamic studies of PlmII with several inhibitors, derived from the allophenylnorstatine scaffold, and ranging in the nM Ki affinities, revealed that the entropic contribution is the dominant driving force for binding. [11] According to our calculations the change from $F \rightarrow I$ is the only one that increases the affinity ($\Delta\Delta G$) of PlmII for the substrate due to the higher unfavorable effect of this mutation in the unbound state than in the bound state. At the thermodynamic level the binding affinity is determined by the magnitude of the Gibbs energy (ΔG), an additive function (ΔG = ΔH - TΔS) of the enthalpy (ΔH) and entropy changes (ΔS).[46] The analysis of the entropy and enthalpic contributions of the change from $F \rightarrow I$ showed that this mutation is an entropy driven process because the raise in the entropy ($T\Delta\Delta S$) prevails over the enthalpy penalty ($\Delta\Delta H$). It should be remarked that the binding entropy is defined by two major terms, the first one is the solvation entropy associated with the burial from the solvent of hydrophobic groups, and the second one is the conformational entropy, which usually reflects the loss of conformational degrees upon binding. [46] Based on our results the raise in the entropy of the environment reorganization ($T\Delta\Delta S - T\Delta\Delta S^{conf}$) due to the change from $F \rightarrow I$ overcomes the loss of the conformational degrees upon binding of this side chain mutation.

From the engineering point of view, a favorable enthalpy change is obtained from good geometric complementarities between drug and target, and the proper location of hydrogen bond donors and acceptors is an important contributor to the drug affinity and selectivity.[45] The structural analysis of the PlmII:Substrate complexes indicated that the hydroxyl group of the Tyr side chain in the P3 position interacts with the solvent and not with the PlmII S3 subsite pocket residues. This allowed explaining our prediction of the null effect over the PlmII:Substrate affinity from the change $F \rightarrow Y$. This suggested the high degree of difficulty associated to optimize the enthalpy contributions of PlmII peptidomimetic inhibitors by introducing polar groups in the P3 position.

**Concluding remarks**

In this paper, we found that the CGI method can be suitable applied to study the binding free energy differences of substitutions in the substrate P3 position over the PlmII:Substrate complexes affinities. We focused on the study of the PlmII S3 substrate preference considering previous kinetic experiments which suggested the major role of this subsite pocket in the protease specificity. Our results agreed well with the relative experimental trend of relative binding free energy differences from these systems, and

is this a new finding?

allowed explaining the energetic basis of low tolerance of PlmII S3 subsite pocket by the lysine and aspartic acid residues in the substrate P3 position. This fit well with previous works that describe the highly hydrophobic character of this subsite pocket. The structural analysis of the PlmII:Substrate complexes showed that the substrate P3 position is solvent exposed. This suggested the high degree of difficulty to optimize the enthalpy contribution of PlmII peptidomimetic inhibitors by the introduction of polar groups in P3.

Additionally, we predicted that the lysine and aspartic acid residues in the P3 position of the substrate must be in the protonated form in the experimental assays.

## Methods

### Comparative 3D Modeling

Three dimensional models for PlmII:Substrate complexes were generated with the MODELLER software, [47] using as template the crystallography structure of PlmII complexed with a peptide-based inhibitor (PDB code: 2r9b, R = 2.8 Å). We calculated 20 models for each complex with the spatial restraints extracted from the target-template alignment. These models were assessed using the UCLA web server tools (http://nihserver.mbi.ucla.edu/SAVS/); and the DOPE energy function[48] provided with the modeling package. [47]

### Alchemy free energy calculations

We calculated relative binding free energy differences for 22 point mutations in the P3 positions of PlmII:Substrates complexes (K-P-X-E-F*Nph-R-L, where Nph = para-nitrophenylalanine, * indicates the cleavage point, and X the amino acid variation in P3). [28] Each mutation was automatically setup with the PYMACS package (http://www.user.gwdg.de/~dseelig/pymacs.html), a collection of scripts which, 1) replace residues in the structure file by the appropriate mutation, and 2) modify the GROMACS topology files to ensure the correct force field parameters for each state.[39] All Molecular Dynamics (MD) simulations were carried out using the GROMACS software package (version 4.0.7)[49] with the AMBER99sb force field[50], and the TIP3P water model[51]. As simulation protocol, we chose non-equilibrium fast-growth thermodynamic integration (FGTI) runs. For determining the $\Delta\Delta G$ binding value of each PlmII:Substrate P3 mutation, equilibrium ensembles at $\lambda$=0 and at $\lambda$=1 are required for the bound and unbound states. The bound simulation systems consist of the complex solvated in a dodecahedron box with ≈ 14192 water molecules and NaCl was added to achieve a 100 mM solution. The unbound simulation systems consist of the substrate solvated in a similar box with ≈ 3193 water molecules and 100 mM of NaCl. Both the A- and B-states were sampled for 100 ns using the stochastic dynamic integrator at 310 K and constant pressure of 1 atm using the Parrinello-Rahman barostat. [52] The electrostatic interactions were calculated at every step with the particle-mesh Ewald method [53, 54], and short-range repulsive and attractive dispersion interactions simultaneously described by a Lennard-Jones potential with a cutoff of 1.1 nm and a switching function that was used between 1.0 and 1.1 nm. Dispersion correction for energy and pressure was applied. The SETTLE [55] algorithm was used to constrain

bonds and angles of water molecules., and LINCS [56] was used for all other bonds, allowing a time step of 2 fs.

From these ensembles, the first 2 ns were discarded, 196 snapshots were taken, and short simulations were performed in which λ was changed from zero to one, respectively. For the FGTI simulations, we used the leap-frog integrator[57] with the velocity rescaling thermostat. [58] Energy calculations, timestep, and pressure coupling were setup analogous to the equilibrium runs. To account the occurrence of atomic overlaps close to λ = 0 and λ = 1, soft-core potentials were used for both electrostatics and Lennard-Jones interactions as implemented in GROMACS-4.0 with α = 0.3, σ = 0.25, and a soft-core power of 1. The complete switching from λ = 0 to λ = 1 was done within 100 ps and the derivates of the Hamiltonian with respect to λ were recorded at every step. Free energies were calculated from the work distribution obtained from integration according to:

$$W = \int_{\lambda=0}^{\lambda=1} \frac{\partial H_{\lambda}}{\delta\lambda} \partial\lambda \quad [1]$$

and calculating the intersection of the forward and backward work distributions according to the Crooks-Gaussian-intersection method previously described. [38] Binding free energy differences $\Delta\Delta G = \Delta G_3 - \Delta G_2 = \Delta G_4 - \Delta G_1$, were calculated from the difference of the free energies between the bound (complex) and unbound (substrate) mutation simulations according to the thermodynamic cycle shown in the Figure 2.

Experimental binding free energies differences($\Delta\Delta G_{exp}$) of each P3 mutation in PlmII:Substrates complexes were calculated from the experimental Michaelis constant values (Km) [28] at 310 K considering that:

$$\Delta\Delta G_{\exp} = RT \times \ln\left(\frac{Km_B}{Km_A}\right). \quad [2]$$

To measure the overall fit between $\Delta\Delta G_{\exp}$ and $\Delta\Delta G_{calc}$, we calculated the Pearson product-moment correlation coefficient (r) using equation 3:

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}} \quad [3]$$

where n is the sample size, and $x_i$ and $y_i$ are the $\Delta\Delta G_{\exp}$, and the $\Delta\Delta G_{calc}$ values for each P3 mutation, respectively.

**Entropy and enthalpy calculations**

The constant-pressure binding entropy differences $\Delta S$ (310K, 1 atm) were obtained from the finite difference relationship:

$$\Delta S(T) = -\frac{\Delta G(T + \Delta T) - \Delta G(T - \Delta T)}{2\Delta T} \quad [4]$$

where we used $\Delta T$ = 25 K. The binding free energies $\Delta G(T + \Delta T)$ and $\Delta G(T - \Delta T)$ were calculated according the thermodynamic cycle shown in Figure 2. The constant-pressure binding enthalpy was calculated according to the equation 5:

$$\Delta G = \Delta H - T\Delta S \quad [5]$$

The configurational binding entropy differences $\Delta S^{conf}$ (310K, 1 atm) from the equilibration MD runs of the bound and unbound states were estimated based on the use of the quasiharmonic analysis with the covariance related to the quantum-mechanical entropy formula. [59]

explain letter

**Binding Interaction analyses**

We calculated the heavy atom-atom contacts between the P3 residue of the substrate and the binding site residues using 0.4 nm as distance cutoff criteria along the MD trajectories. We considered as stable contacts pair each one, we a frequency (f) $f \geq \frac{1}{m}$, where m is the mean contact per atom. Then, we calculated the fraction of non-polar/polar contacts for each P3 residue of the substrate with the S3 subsite residues of the protein.

**Acknowledgements**

PAV acknowledges the DAAD and EMBO financial support for performing a research stay in the Max Planck Institute of Biophysical Chemistry, Göttingen, Germany.

## References


1. Coombs, G. H., Goldberg, D. E., Klemba, M., Berry, C., Kay, J.,and Mottram, J. C. (2001). Aspartic proteases of Plasmodium falciparum and other parasitic protozoa as drug targets. *Trends Parasitol* **1**, 532-7.
2. Kappe, S. H., Vaughan, A. M., Boddey, J. A. & Cowman, A. F. (2010). That Was Then But This Is Now: Malaria Research in the Time of an Eradication Agenda. . *Science* **328**, 862-866.
3. Klemba, M., and Goldberg, D. E. (2002). Biological roles of proteases in parasitic protozoa. *Annu Rev Biochem* **7**, 275-305.
4. Shenai, B. R., Sijwali, P. S., Singh, A. & Rosenthal, P. J. (2000 Sep). Characterization of native and recombinant falcipain-2, a principal trophozoite cysteine protease and essential hemoglobinase of Plasmodium falciparum. *J Biol Chem* **275**, 29000-10.
5. Murata, C. E., and Goldberg, D. E. (2003). Plasmodium falciparum falcilysin: A metalloprotease with dual specificity. *J Biol Chem* **278**, 38022-38028.
6. Klemba, M., Gluzman, I. & Goldberg, D. E. (2004). A Plasmodium falciparum dipeptidyl aminopeptidase I participates in vacuolar hemoglobin degradation. *J Biol Chem* **279**, 43000-43007.
7. Francis, S. E., Sullivan, D. J. J. & Goldberg, D. E. (1997 ). Hemoglobin metabolism in the malaria parasite Plasmodium falciparum. *Annu Rev Microbiol* **51**, 97-123.
8. Francis, S. E., Gluzman, I. Y., Oksman, A., Knickerbocker, A., Mueller, R., Bryant, M. L., Sherman, D. R., Russell, D. G. & Goldberg, D. E. (1994 Jan). Molecular characterization and inhibition of a Plasmodium falciparum aspartic hemoglobinase. *EMBO J* **13**, 306-17.
9. Liu, J., Gluzman, I. Y., Drew, M. E. & Goldberg, D. E. (2005 Jan). The role of Plasmodium falciparum food vacuole plasmepsins. *J Biol Chem* **280**, 1432-7.
10. Liu, J., Istvan, E. S., Gluzman, I. Y., Gross, J. & Goldberg, D. E. (2006 Jun). Plasmodium falciparum ensures its amino acid supply with multiple acquisition pathways and redundant proteolytic enzyme systems. *Proc Natl Acad Sci U S A* **103**, 8840-5.
11. Nezami, A., Kimura, T., Hidaka, K., Kiso, A., Liu, J., Kiso, Y., Goldberg, D. E. & Freire, E. (2003 ). High-affinity inhibition of a family of Plasmodium falciparum proteases by a designed adaptive inhibitor. *Biochemistry* **42**, 8459-8464.
12. 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. & Erickson, J. W. (1996 ). Structure and inhibition of plasmepsin II, a hemoglobin-degrading enzyme from Plasmodium falciparum. *Proc. Natl. Acad. Sci. USA* **93**, 10034-10039.
13. Asojo, O. A., Afonina, E., Gulnik, S. V., Yu, B., Erickson, J. W., Randad, R., Medjahed, D. & Silva, A. M. (2002 ). Structures of Ser205 mutant plasmepsin II from Plasmodium falciparum at 1.8 A in complex with the inhibitors rs367 and rs370. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 2001-2008.
14. Asojo, O. A., Gulnik, S. V., Afonina, E., Yu, B., Ellman, J. A., Haque, T. S. & Silva, A. M. (2003). Novel uncomplexed and complexed structures of plasmepsin II, an aspartic protease from Plasmodium falciparum. *J Mol Biol* **327**, 173-181.
15. Prade, L., Jones, A. F., Boss, C., Richard-Bildstein, S., Meyer, S., Binkert, C. & Bur, D. (2005 ). X-ray structure of plasmepsin II complexed with a potent achiral inhibitor. *J Biol Chem* **280**, 23837-23843.
16. Ersmark, K., Samuelsson, B. & Hallberg, A. (2006 ). Plasmepsins as potential targets for new antimalarial therapy. *Med Res Rev* **26**, 626-66.
17. Bjelic, S., Nervall, M., Gutierrez-de-Teran, H., Ersmark, K., Hallberg, A. & Aqvist, J. (2007 Sep). Computational inhibitor design against malaria plasmepsins. *Cell Mol Life Sci* **64**, 2285-305.
18. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K. & Olson, A. J. (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem* **19**, 1639-1642.
19. Ewing, T. J. A. & Kuntz, I. D. (1997). Critical evaluation of search algorithms for automated molecular docking and data-base screening. . *J. Comput. Chem* **18**, 1175-1189.
20. Rarey, M., Kramer, B., Lengauer, T. & Klebe, G. (1996). A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **261**, 470-489.
21. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P. & Shenkin, P. S. (2004). Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. . *J. Med. Chem.* **47**, 1739-1749.

22. Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T. & Banks, J. L. (2004). Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **47**, 1750-1759.
23. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol* **267**, 727-748.
24. Ersmark, K., Feierberg, I., Bjelic, S., Hamelink, E., Hackett, F., Blackman, M. J., Hulten, J., Samuelsson, B., Aqvist, J. & Hallberg, A. (2004 ). Potent inhibitors of the Plasmodium falciparum enzymes plasmepsin I and II devoid of cathepsin D inhibitory activity. *J Med Chem* **47**, 110-22.
25. Bhargavi, R., Sastry, G. M., Murty, U. S. & Sastry, G. N. (2005 ). Structural and active site analysis of plasmepsins of Plasmodium falciparum: potential anti-malarial targets. *Int J Biol Macromol* **37**, 73-84.
26. Valiente, P. A., Batista , P. R., Pupo, A., Pons, T., Valencia, A. & Pascutti, P. G. (2008). Predicting functional residues in Plasmodium falciparum plasmepsins by combining sequence and structural analysis with molecular dynamics simulations. *Proteins* **73**, 440-457.
27. Dunn, B. M. & Hung, S. (2000 Mar). The two sides of enzyme-substrate specificity: lessons from the aspartic proteinases. *Biochim Biophys Acta* **1477**, 231-40.
28. Westling, J., Cipullo, P., Hung, S. H., Saft, H., Dame, J. B. & Dunn, B. M. (1999 Oct). Active site specificity of plasmepsin II. *Protein Sci* **8**, 2001-9.
29. Brinkworth, R. I., Prociv, P., Loukas, A. & Brindley, P. J. (2001). Hemoglobin-degrading, aspartic proteases of blood-feeding parasites: substrate specificity revealed by homology models. *J Biol Chem* **276** 38844-51.
30. Jorgensen, W. & Ravimohan, C. (1985). Monte Carlo simulation of differences in free energies of hydration. *J.Chem. Phys* **83**, 3050-3054.
31. Beveridge, D. L. & Dicapua, F. M. (1989). Free energy via molecular simulations: Application to chemical and biochemical system. *Annu Rev Biophys Biophys Chem* **18**, 431-492.
32. Jorgensen, W. (2004). The many roles of computation in drug discovery. *Sci. STKE.* **303**, 1813-1818.
33. Zwanzig, R. W. (1954). High-temperature equations of state by perturbation method. I. Nonpolar gases. *Journal of Chemical Physics* **22**, 1420-1426.
34. Straatsma, T. & Berendsen., H. (1988). Free energy of ionic hydration: analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations. *J. Chem. Phys.* **89**, 5876-5886.
35. Jarzynski, C. (1997). Nonequilibrium equality for free energy difference. *Phys. Rev. Lett* **78**, 2690-2693.
36. Jarzynski, C. (1997). Equilibrium free-energy differences from nonequilibrium measurements: a master-equation approach. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics.* **56**, 5018-5035.
37. Crooks, G. (1998). Nonequilibrium measurements of free energy differences for microscopically reversible Markovian systems. *J. Stat. Phys.* **90**, 1481-1487.
38. Goette, M. & Grubmüller, H. (2009). Accuracy and convergence of free energy differences calculated from nonequilibrium switching processes. *Journal of Computational Chemistry* **30**, 447-456.
39. Seeliger, D. & de Groot, B. L. Protein Thermostability Calculations Using Alchemical Free Energy Simulations. *Biophysical journal* **98**, 2309-2316.
40. Lowther, W. T., Majer, P. & Dunn, B. M. (1995). Engineering the substrate specificity of rhizopuspepsin: The role of Asp 77 of fungal aspartic proteinases in facilitating the cleavage of oligopeptide substrates with lysine in P1. *Protein Sci* **4**, 689-702.
41. Scarborough, P. E., Guruprasad, K., Topham, C., Richo, G. R., Conner, G. E., Blundell, T. L. & Dunn, B. M. (1993). Exploration of subsite binding specificity of human cathepsin D through kinetics and rule-based molecular modeling. *Protein Sci* **2**, 264-276.
42. Westling, J., Yowell, C. A., Majer, P., Erickson, J. W., Dame, J. B. & Dunn, B. M. (1997). Plasmodium falciparum, P. vivax, and P. malariae: a comparison of the active site properties of plasmepsins cloned and expressed from three different species of the malaria parasite. *Exp Parasitol* **87**, 185-93.
43. Beyer, B. B., Johnson, J. V., Chung, A. Y., Li, T., Madabushi, A., Agbandje-McKenna, M., McKenna, R., Dame, J. B. & Dunn, B. M. (2005). Active-site specificity of digestive aspartic peptidases from the four species of Plasmodium that infect humans using chromogenic combinatorial peptide libraries. *Biochemistry* **44**, 1768-79.

44. Gulnik, S. V., Suvorok, L. I., Majer, P. & Erickson, J. W. (2000). Sensitive fluorogenic substrates for plasmepsin 2. *Protein and Peptide Letters* **7**, 219-223.
45. Ruben, A. J., Kiso, Y. & Freire, E. (2006). Overcoming Roadblocks in Lead Optimization: A Thermodynamic Perspective. *Chemical Biology & Drug Design* **67**, 2-4.
46. Chaires, J. B. (2008). Calorimetry and Thermodynamics in Drug Design. *Annu. Rev. Biophys.* **37**, 135-151.
47. Sali, A. & Blundell, T. L. (1993). Comparative protein modeling by satisfaction of spatial restraints. *J Mol Biol* **234**, 779-815.
48. Shen, M. Y. & Sali, A. (2006). Statistical potential for assessment and prediction of protein structures. *Protein Sci* **15**, 2507-24.
49. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **4**, 435-447.
50. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. & Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* **65**, 712-725.
51. William, L. J., Jayaraman, C., Jeffry, D. M., Roger, W. I. & Michael, L. K. (1983). Comparison of simple potential functions for simulating liquid water, Vol. 79, pp. 926-935. AIP.
52. Parrinello, M. & Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method, Vol. 52, pp. 7182-7190. AIP.
53. Darden, T., York, D., Pedersen, L. . (1993). Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J Chem Phys* **98**, 10089-10092.
54. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., Pedersen, L. G. . (1995). A smooth particle mesh ewald potential. . *J Chem Phys* **103**, 8577-8592.
55. Miyamoto, S. & Kollman, P. A. (1992). Settle an analytical version of the shake and rattle algorithm for rigid water models. . *J. Comput. Chem.* **13**, 952-962.
56. Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997). LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **18**, 1463-1472.
57. Hockney, R. W. (1970). The potential calculation and some applications. *Methods in Computational Physics* **9**, 136-211.
58. Giovanni, B., Davide, D. & Michele, P. (2007). Canonical sampling through velocity rescaling, Vol. 126, pp. 014101. AIP.
59. Ioan, A. & Martin, K. (2001). On the calculation of entropy from covariance matrices of the atomic fluctuations, Vol. 115, pp. 6289-6292.

## Figure and Table Legends

**Fig1** Top view representation of the PlmII:Substrate complex (A). For clarity PlmII is represented as a blue cartoon diagram, and the substrate as a red licorice diagram. The black box highlighted the substrate P3 position. Surface side view representation of the interaction between the PlmII S3 subsite residues and different amino acids in the substrate P3 position: B) Phe, C) Ile, D) Ala, E) Ser, F) Asp, and J) Lys. For clarity the substrate P3 position is represented as a licorice diagram, and the non polar residues in the PlmII S3 subsite pocket were displayed as gray, while the polar residues were displayed as green.

**Fig 2** Thermodynamic cycle for the free energy calculations. S and Sm represented the state A and state B from of the substrate in solution, respectively. P:S and P:Sm represented the state A and state B of the PlmII:Substrate complex, respectively. The binding free energy differences $\Delta\Delta G = \Delta G_3 - \Delta G_2$ for the amino acid variations in the substrate P3 position of the PlmII:Substrate complexes can be calculated via $\Delta\Delta G = \Delta G_4 - \Delta G_1$.

**Fig 3** Scatter diagram of calculated ($\Delta\Delta G_{calc}$) versus experimental ($\Delta\Delta G_{exp}$) binding free energy differences of PlmII:Substrate complexes using the CGI method. The two red lines parallel to the diagonal line represent deviations of ± 4.18 kJ/mol. (A). Calculated

binding free energy differences in the P3 positions of PlmII:Substrates complexes considering both protonation of the ionizable residues. (B) Calculated binding free energy differences in the P3 positions of PlmII:Substrates complexes considering the protonation of the ionizable residues that fit better with experimental results.

**Table I** Summary of the atomic interactions differences between the substrate P3 position and the PlmII S3 subsite contact residues

**Table II** Calculated and experimental binding free energy differences of mutations in the substrate P3 position of PlmII-Substrates complexes

**Table III** Entropy and enthalpy binding contributions of mutations in the P3 position of the PlmII-Substrates complexes

**Table IV** Calculated configurational entropy differences of mutations in the P3 position of the substrate in the free and complex MD simulations

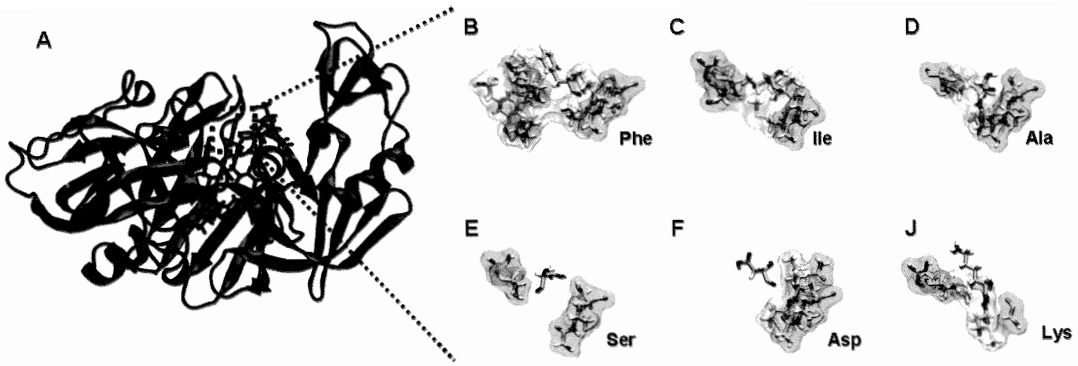


**Fig 1** Top view representation of the PlmII:Substrate complex (A). For clarity PlmII is represented as a blue cartoon diagram, and the substrate as a red licorice diagram. The black box highlighted the substrate P3 position. Surface side view representation of the interaction between the PlmII S3 subsite residues and different amino acids in the substrate P3 position: B) Phe, C) Ile, D) Ala, E) Ser, F) Asp, and J) Lys. For clarity the substrate P3 position is represented as a licorice diagram, and the non polar residues in the PlmII S3 subsite pocket were displayed as gray, while the polar residues were displayed as green.

Lucy!

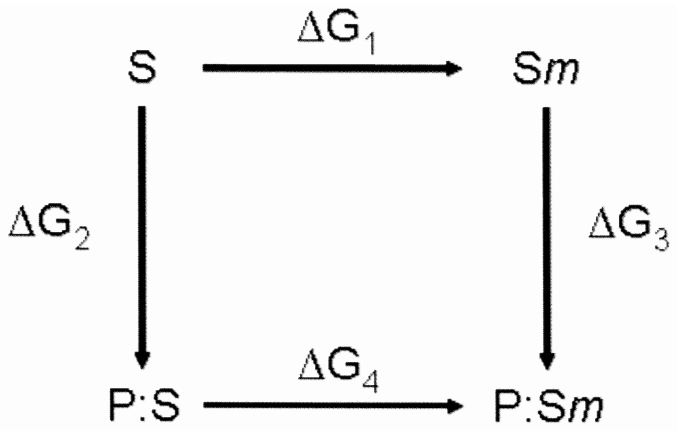


**Fig 2** Thermodynamic cycle for the free energy calculations. S and Sm represented the state A and state B from of the substrate in solution, respectively. P:S and P:Sm represented the state A and state B of the PlmII:Substrate complex, respectively. The binding free energy differences $\Delta\Delta G = \Delta G_3 - \Delta G_2$ for the amino acid variations in the substrate P3 position of the PlmII:Substrate complexes can be calculated via $\Delta\Delta G = \Delta G_4 - \Delta G_1$.

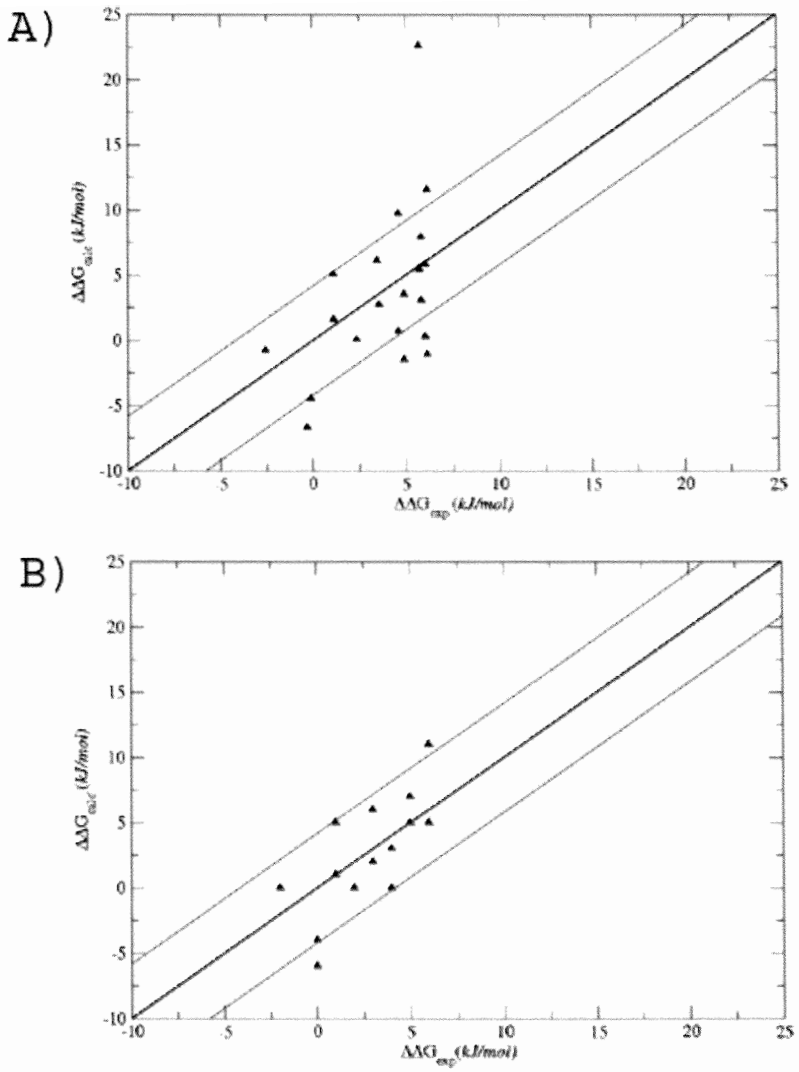


**Fig 3** Scatter diagram of calculated ($\Delta\Delta G_{calc}$) versus experimental ($\Delta\Delta G_{exp}$) binding free energy differences of PlmII:Substrate complexes using the CGI method. The two red lines parallel to the diagonal line represent deviations of ± 4.18 kJ/mol. (A). Calculated binding free energy differences in the P3 positions of PlmII:Substrates complexes considering both protonation of the ionizable residues. (B) Calculated binding free energy differences in the P3 positions of PlmII:Substrates complexes considering the protonation of the ionizable residues that fit better with experimental results.

**Table I** Summary of the atomic interactions differences between the substrate P3 position and the PlmII S3 subsite contact residues

| P3 residue | Fraction Non Polar | Fraction Polar | Km (μM)[a] |
|---|---|---|---|
| *F* | 0.65 | 0.35 | 24 ± 3 |
| *I* | 0.47 | 0.53 | 23 ± 3 |
| A | 0.27 | 0.73 | 37 ± 8 |
| *S* | 0.38 | 0.62 | 92 ± 9 |
| *K* | 0.56 | 0.44 | 250 ± 60 |
| *D* | 0.60 | 0.40 | 221 ± 57 |

a) The experimental Km values were taken from reference 28.

Why is the $\Delta G^{b}_{substrate}$ so low for the F → R mutation?

**Table II** Calculated and experimental binding free energy differences of mutations in the substrate P3 position of PlmII-Substrates complexes.

| **P3 Mutations** $(A \rightarrow B)$ | | $\Delta G^{a}_{complex}$ $(kJ/mol)$ | $\Delta G^{b}_{substrate}$ $(kJ/mol)$ | $\Delta\Delta G^{c}_{calc}$ $(kJ/mol)$ | $\Delta\Delta G^{d}_{\exp}$ $(kJ/mol)$ | $\lvert error \rvert^{e}$ $(kJ/mol)$ |
|---|---|---|---|---|---|---|
| $F \rightarrow A$ | | 30.01 ± 0.42 | 28.40 ± 0.25 | 1.61 ± 0.49 | 1.12 | 0.49 |
| $F \rightarrow S$ | | -21.29 ± 0.48 | -27.43 ± 0.35 | 6.14 ± 0.59 | 3.47 | 2.67 |
| $F \rightarrow I$ | | 8.94 ± 0.50 | 13.43 ± 0.30 | -4.49 ± 0.59 | -0.11 | 4.38 |
| $F \rightarrow K$ | $K^{+}$ | 688.56 ± 1.08 | 682.72 ± 0.62 | 5.84 ± 1.25 | 6.04 | 0.20 |
| | $K^{0}$ | -59.24 ± 1.54 | -59.55 ± 0.95 | 0.31 ± 1.81 | | 5.73 |
| $F \rightarrow D$ | $D^{-}$ | -1239.38 ± 1.29 | -1261.97 ± 0.76 | 22.59 ± 1.49 | 5.74 | 16.85 |
| | $D^{0}$ | -213.29 ± 0.60 | -218.75 ± 0.30 | 5.48 ± 0.67 | | 0.26 |
| $F \rightarrow R$ | | 3.37± 1.57 | -4.12 ± 0.73 | 7.49 ± 1.73 | - | - |
| $F \rightarrow Y$ | | -101.27 ± 0.29 | -101.26 ± 0.18 | -0.01± 0.34 | - | - |
| $I \rightarrow A$ | | 25.26 ± 0.43 | 20.21 ± 0.20 | 5.05 ± 0.47 | 1.11 | 3.94 |
| $I \rightarrow S$ | | -28.87 ± 0.52 | -31.60 ± 0.24 | 2.73 ± 0.57 | 3.57 | 0.84 |
| $I \rightarrow K$ | $K^{+}$ | 681.70 ± 1.02 | 670.14 ± 0.46 | 11.56 ± 1.57 | 6.15 | 5.41 |
| | $K^{0}$ | -74.29 ±1.09 | -73.20 ± 0.64 | -1.09 ± 1.26 | | 7.24 |
| $I \rightarrow D$ | $D^{-}$ | -1268.14 ± 1.10 | -1271.23 ± 0.96 | 3.09 ± 1.46 | 5.83 | 2.74 |
| | $D^{0}$ | -219.01 ± 0.63 | -226.93 ± 0.40 | 7.92 ± 0.75 | | 2.09 |
| $A \rightarrow S$ | | -53.82 ± 0.22 | -53.86 ± 0.14 | 0.04 ± 0.27 | 2.35 | 2.31 |
| $A \rightarrow K$ | $K^{+}$ | 657.05 ± 0.71 | 653.51 ± 0.47 | 3.54 ± 0.85 | 4.92 | 1.38 |
| | $K^{0}$ | -90.83 ± 0.86 | -89.37 ± 0.63 | -1.46 ± 1.06 | | 6.38 |
| $A \rightarrow D$ | $D^{-}$ | -1280.59 ± 0.79 | -1290.33 ± 0.88 | 9.74 ± 1.18 | 4.61 | 5.13 |
| | $D^{0}$ | -246.20 ± 0.40 | -246.89 ± 0.33 | 0.69 ± 0.52 | | 3.92 |
| $K \rightarrow S$ | | -882.43 ± 0.78 | -881.55 ± 0.58 | -0.78 ± 0.97 | -2.57 | 1.79 |
| $K \rightarrow D$ | $D^{-}$ | - | - | - | -0.32 | - |
| | $D^{0}$ | -1080.42 ± 1.03 | -1073.73 ± 0.74 | -6.69± 1.27 | | 6.37 |

a) $\Delta G^{a}_{complex}$ of each P3 mutations in the PlmII-substrate complexes were obtained with the CGI method.

b) $\Delta G^{b}_{substrate}$ of each P3 mutations in the substrate-solvent systems were obtained with the CGI method

c) Calculated binding free energy differences ($\Delta\Delta G^{c}_{calc}$) of each P3 mutation in the systems were calculated using the relationship $\Delta\Delta G_{calc} = \Delta G_{complex} - \Delta G_{substrate}$

d) Experimental binding free energy differences ($\Delta\Delta G_{exp}$) of each P3 mutation in PlmII-substrates complexes at 310 K was calculated using the relationship $\Delta\Delta G_{\exp} = RT \times \ln\left(\frac{Km_B}{Km_A}\right)$. The experimental Michaelis constant values (*Km*) of each complex were taken from ref. 28.

e) Absolute error $\lvert error \rvert$ refer to deviations, in kJ/mol, between the calculated and experimental values of each P3 mutation in the PlmII-substrate complexes.

**Table III** Entropy and enthalpy binding contributions of mutations in the P3 position of the PlmII-Substrates complexes

| **P3 mutations** $(A \rightarrow B)$ | $T\Delta S_{310K}^{complex}$ $(kJ/mol)$ | $T\Delta S_{310K}^{substrate}$ $(kJ/mol)$ | $T\Delta\Delta S_{310K}$ $(kJ/mol)$ | $\Delta H_{310K}^{complex}$ $(kJ/mol)$ | $\Delta H_{310K}^{substrate}$ $(kJ/mol)$ | $\Delta\Delta H_{310K}$ $(kJ/mol)$ |
|---|---|---|---|---|---|---|
| $F \rightarrow A$ | -6.51 | -12.40 | 5.89 | 23.50 | 16.00 | 7.50 |
| $F \rightarrow S$ | 10.54 | -4.65 | 15.31 | -10.69 | -32.14 | 21.45 |
| $F \rightarrow I$ | 7.44 | -7.13 | 14.76 | 16.50 | 6.23 | 10.27 |
| $F \rightarrow K^+$ | 49.91 | 42.16 | 7.76 | 738.53 | 724.93 | 13.60 |
| $F \rightarrow D^0$ | -31.00 | 0.62 | -31.62 | -244.35 | -218.25 | -26.10 |

a) The constant-pressure binding entropy differences $\Delta S$ (310K, 1 atm) was obtained from the finite difference relationship

$\Delta S(T) = -\frac{\Delta G(T+\Delta T) - \Delta G(T-\Delta T)}{2\Delta T}$ where we used $\Delta T$ = 25 K. The binding free energies $\Delta G(T + \Delta T)$ and $\Delta G(T - \Delta T)$ were calculated using the CGI method. (see Table SI)