Inhibitor Binding to the Plasmepsin IV Aspartic Protease from *Plasmodium falciparum*[†]

Hugo Gutiérrez-de-Terán,[‡] Martin Nervall,[‡] Karolina Ersmark,[§] Peng Liu,[ǁ] Linda K. Janka,[ǁ] Ben Dunn,[ǁ] Anders Hallberg,[§] and Johan Åqvist*[,‡]

Department of Cell and Molecular Biology, Uppsala University, BMC, Box 596, 751 24 Uppsala, Sweden, Department of Medicinal Chemistry, Uppsala University, BMC, Box 574, 751 23 Uppsala, Sweden, and Department of Biochemistry and Molecular Biology, University of Florida College of Medicine, P.O. Box 100245, Gainesville, Florida 32610-0245

ABSTRACT: Plasmepsin IV (Plm IV) is one of the aspartic proteases present in the food vacuole of the malaria parasite *Plasmodium falciparum* involved in host hemoglobin degradation by the parasite. Using a series of previously synthesized plasmepsin inhibitors [Ersmark, K., et al. (2005) *J. Med. Chem. 48*, 6090−106], we report here experimental data and theoretical analysis of their inhibitory activity toward Plm IV. All compounds share a 1,2-dihydroxyethylene unit as the transition state mimic. They possess symmetric P1 and P1′ side chains and either a diacylhydrazine, a five-membered oxadiazole ring, or a retroamide at the P2 and P2′ positions. Experimental binding affinities are compared to those predicted by the linear interaction energy (LIE) method and an empirical scoring function, using both a crystal structure and a homology model for the enzyme. Molecular dynamics (MD) simulations of the modeled complexes allow a rational interpretation of the structural determinants for inhibitor binding. A ligand bearing a P2 and P2′ symmetric oxadiazole which is devoid of amide bonds is identified both experimentally and theoretically as the most potent inhibitor of Plm IV. For the P2 and P2′ asymmetric compounds, the results are consistent with earlier predictions regarding the mode of binding of this class of inhibitors to Plm II. Theoretical estimation of selectivity for some compounds is also reported. Significant features of the Plm IV binding pocket are discussed in comparison to related enzymes, and the results obtained here should be helpful for further optimization of inhibitors.

Malaria, one of the major infectious diseases in the world, is caused by protozoan parasites of the genus *Plasmodium*. Four species can infect humans, namely, *Plasmodium falciparum*, *Plasmodium malariae*, *Plasmodium ovale*, and *Plasmodium vivax*. Of these, *P. falciparum* is the most lethal and is therefore the main target for drug intervention (*1*). There is a growing need for effective drugs with new mechanisms of action, due to the high rate of mutation of the parasite, which leads to the development of resistance. The complete genome of *P. falciparum* has recently been published (*2*), and structural and functional characterization of the proteins involved in parasite development is an important task.

One of the critical stages of the life cycle of the parasite during human infection is the degradation of hemoglobin which provides nutrients for its own growth and maturation (*3*). The digestion of hemoglobin occurs within the acidic food vacuole of the parasite through the use of several proteases that act in a consecutive manner. A family of aspartic proteases, known as the plasmepsins, appears to be involved in the initial steps of the degradation pathway (*4*) and therefore makes for an attractive antimalarial drug target. Up to 10 genes encode plasmepsins in *P. falciparum*, numbered Plm I−Plm X[1] (*3*). Among them, only Plm I, Plm II, HAP (histoaspartic protease, corresponding to Plm III), and Plm IV are located in the food vacuole. Plm II has been the most extensively characterized, with several crystal structures determined (*5−7*) and potent inhibitors developed (*8−12*). However, recent studies point to the necessity of blocking more than one plasmepsin to attain effective drugs. This is due not only to the demonstrated viability of parasites that lack any one of the four food vacuole plasmepsins (*13*) but also to the fact that drug resistance is then less likely to arise (*3*). Although the level of sequence identity among the aspartic proteases of the *Plasmodium* species is rather high (∼60%), response to inhibitors can be different (*14−17*), indicating variations in the specific binding interactions among the different plasmepsins (*18*). Therefore, considerable efforts have been devoted lately to gathering structural and biochemical knowledge about the other plasmepsins present in the food vacuole. For example, theoretical models for the substrate binding mode and mechanism of the histoaspartic protease (HAP) (*19*) have recently been proposed (*20, 21*), and a crystal structure for the Plm IV enzyme has been deposited in the Protein Data Bank (entry 1LS5),

* To whom correspondence should be addressed. Phone: +46 18 471 4109. Fax: +46 18 53 69 71. E-mail: aqvist@xray.bmc.uu.se.
[‡] Department of Cell and Molecular Biology, Uppsala University.
[§] Department of Medicinal Chemistry, Uppsala University.
[ǁ] University of Florida College of Medicine.

[1] Abbreviations: Plm, plasmepsin; Cat D, cathepsin D; MD, molecular dynamics; LIE, linear interaction energy; rms, root-mean-square.

although unpublished, and considerable biochemical and pharmacological data for Plm IV have emerged in the past years (*14*, *15*, *17*, *22*, *23*). The interest in Plm IV is also motivated by the fact that it is the only plasmepsin located in the food vacuole of *P. falciparum* which has orthologs in the other three *Plasmodium* species infecting humans (*22*). In this sense, the crystal structure of one of these orthologs, the plasmepsin from *P. malariae* in complex with an inhibitor, has been reported very recently (*24*). Consequently, a characterization of the binding site of Plm IV should be helpful in the design of new compounds effective in the treatment of malaria.

Some attempts to design inhibitors with a broad spectrum toward plasmepsins have been made (*16*), although these inhibitors had limited selectivity against the human related protease cathepsin D (Cat D) (*12*). However, high-affinity inhibitors for Plm I and II have been synthesized, which are devoid of activity against Cat D (*9*, *10*). We have recently reported a series of inhibitors designed against Plm II, some of which exhibited good binding properties also for Plm IV (*25*). In particular, compound **6** in Table 1 constitutes the most potent nonpeptide inhibitor for Plm IV which is totally devoid of amide bonds, while maintaining high selectivity over Cat D. In this work, we explore the binding properties of Plm IV for a series of previously synthesized inhibitors, both in vitro and in silico. The aim of this approach is to give a rationale for the binding of these compounds and to elucidate features of the binding site of Plm IV that may be further explored in inhibitor design. For this purpose, binding to Plm IV is evaluated computationally using the linear interaction energy (LIE) (*26*, *27*) approach combined with molecular dynamics (MD) sampling. Additionally, an empirical scoring function is used to analyze the MD trajectories. We discuss the correlation obtained with the experimental data reported here and point out the characteristics of the binding site of Plm IV responsible for its particular profile.

MATERIALS AND METHODS

Assays and K_i Determinations. The expression and purification of Plm IV were carried out on the basis of the protocol for Plm II preparation (*28*). The production of Plm I will be published elsewhere (P. Liu et al., manuscript in preparation). The concentrations of pro-Plm IV and mature Plm I in the 250 μL reaction system were approximately 10 and 800 nM, respectively. Chromogenic octapeptide substrate 1 (Lys-Pro-Ile-Glu-Phe*Nph-Arg-Leu, where an asterisk denotes a cleavage site and Nph is *p*-nitrophenylalanine) and substrate 2 (Lys-Pro-Ile-Leu-Phe*Nph-Arg-Leu) were hydrolyzed by Plm IV and Plm I, respectively. Substrate 1 was dissolved in 10% formic acid and 90% distilled deionized water, and substrate 2 was dissolved in 20% DMSO, 10% formic acid, and 70% distilled deionized water. Substrate stock solutions (approximately 5 mg/mL) were quantified by amino acid analysis. Substrate stock solutions were quantified by amino acid analysis. For kinetic assays, the Plm IV zymogen needed a preincubation in 0.1 M sodium acetate buffer (pH 4.5) at 37 °C for 5 min to complete autoactivation, while mature Plm I was equilibrated in 0.1 M sodium citrate buffer (pH 5.5) at 37 °C for 3 min. The resulting enzymes were then mixed with the corresponding

substrate, and the catalyzed hydrolyses were monitored using a Cary 50 Bio UV−visible spectrophotometer. The initial rates (absorbance units per second) of enzymatic cleavage of at least six different concentrations of a chromogenic peptide substrate in the presence of at least two different concentrations of inhibitors were measured, and the K_i value was determined by fitting the initial hydrolysis rates and related substrate and inhibitor concentrations to the equation $v = ([S]V_{max})/[[S] + K_m(1 + [I]/K_i)]$ of the single-substrate−single-inhibitor (competitive) program in Enzyme Kinetic Module 1.0 of SigmaPlot 2000 (version 6.10).

Homology Model of Plm IV. A homology model of Plm IV was built using the automated SWISS-MODEL server (*29*). This method allows the consideration of up to four templates in the modeling process. The structures of three different complexes of Plm II were selected together with the structure of Plm *vivax* complexed to pepstatin A (see Table 2). The quality of the generated model was checked with the WHAT-IF suite of programs (*30*).

Docking of Inhibitors. Each compound was built up in an extended conformation with ViewerPro (*31*). Automated docking was preformed with GOLD version 2.2 (*32*), allowing full flexibility for the ligand while keeping the protein fixed. The docking exploration consisted of 20 independent runs of the docking algorithm with each compound, using the default genetic algorithm (GA) search parameters and the GoldScore scoring function. The automated assignment of protein and ligand atom types was checked to ensure a correct treatment of the potential hydrogen bond donors and acceptors.
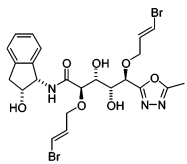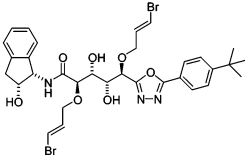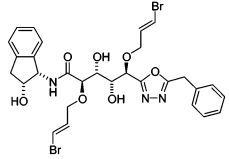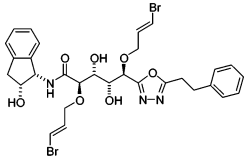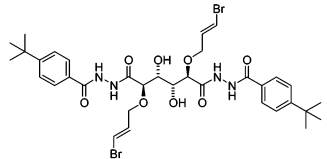
Binding Free Energy Calculations. Binding affinities were calculated using the LIE method, described in detail elsewhere (*26*, *27*). Basically, this approach estimates the ligand free energy of binding from the difference in the ligand and surrounding interaction energies in both its bound and free state. The relationship between the ligand intermolecular interaction energies and the free energy of binding is given by the equation
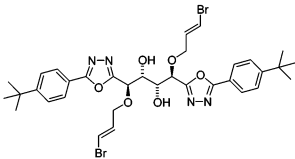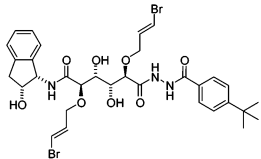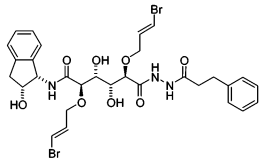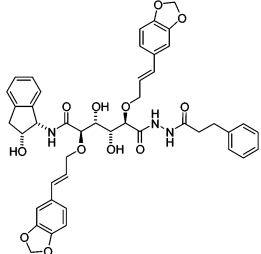
$$\Delta G_{bind} = \alpha\Delta\langle V_{1-s}^{vdW}\rangle + \beta\Delta\langle V_{1-s}^{el}\rangle + \gamma \qquad (1)$$

where V_{1-s}^{vdW} and V_{1-s}^{el} denote the Lenard-Jones and electrostatic interactions between the ligand and its surroundings (l−s), respectively. These interactions are evaluated as energy averages (denoted by the broken brackets) from separate MD simulations of the free (solvated in water) and bound states. The difference (Δ) between such averages for each type of potential is scaled by different coefficients (see ref *27*), giving the polar and nonpolar contributions to the binding free energy. For the nonpolar contribution, this coefficient (α) has been empirically set to 0.181, while for the polar contribution, the scaling factor is dependent on the chemical nature of the ligand. For ligands with at least two hydroxyl groups, like the ones handled in this study, the β value is 0.33 (*27*). Finally, γ is a constant term which can be different from zero to account for absolute binding free energies.

The starting position of the ligand in the protein binding pocket should be sufficiently accurate that the values averaged during the MD sampling can converge within a reasonable calculation time. In earlier work, the binding mode for some of the compounds studied here (**1**, **4**, **8**, and **9**) was predicted for Plm II (*25*). Thus, the initial docking position

Table 1: Experimental Inhibitory Activities of the Compounds Studied

Compound	Structure	Enzyme K_i (nM)				%inh @4 μM (cells)[a]
		Plm I	Plm II[a]	Plm IV	Cat D[a]	
1		3000	n.d.	10000	>6000	n.d.
2		259	n.d.	1200	>6000	10
3		871	2100	2000	>6000	0
4		684	2900	12000	>6000	n.d.
5		104	1500	478	3000	0
6		430	n.d.	35	>5000	25
7		111	142	14000	>6000	11
8		19	7	2000	>6000	0
9		57	11	397	800	n.d.

[a] Data from ref 25.

for **4** and **9** was determined by superposition of the Plm II–ligand complex on the Plm IV structure so that the interactions of the ligands with conserved residues were maintained. The remaining compounds were built up in the Plm IV binding site by mutation of the corresponding ligand structures using ViewerPro (*31*).

Molecular dynamics simulations were carried out using Q (*33*) and the GROMOS87 force field implemented therein (*34*). The parameters needed for the ligands that were not present in the original version of the force field were retrieved from earlier work (*25*). The system was solvated with a 20 Å radius simulation sphere of SPC water (*35*), centered on the carbon of the transition state mimic hydroxyl interacting with the catalytic aspartates. The sphere of simulation was increased to a 22 Å radius in the case of pepstatin A, due to its larger size. The water surface of this sphere was subjected to radial and polarization restraints (*36*) to mimic bulk water at the sphere boundary. Nonbonded interaction energies were calculated up to a 10 Å cutoff, except for the ligand atoms for which no cutoff was used. Beyond the cutoff, long-range electrostatics were treated with the local reaction field (LRF) multipole expansion method (*37*). Protein atoms outside the simulation sphere were restrained to their initial positions and only interacted with the system through bonds, angles, and torsions. For the ligand–protein simulations, a heating and equilibration procedure was applied before the data collection phase. The equilibration protocol started with 1000 steps of MD using a very short time step (0.2 fs) at 1 K, coupled to a strong bath (0.2 fs bath coupling) with positional restraints on heavy atoms. Then the system was gradually heated to 300 K, relaxing the bath coupling to 100 fs and increasing the time step to 1 fs. During this equilibration process, hydrogen bonding constraints were imposed between the central hydroxyls and the catalytic dyad, Asp34 and Asp214, as well as for the hydrogen bonds involving the amide bond isosteres. All these restraints were then gradually released along the equilibration phase, until no restraint was applied. Similarly, positional restraints which were initially imposed on the protein heavy atoms were gradually released during this phase. Unrestrained molecular dynamics then followed for more than 2 ns, with energies collected at regular intervals of 25 fs. Energy averaging was performed on the energetically stable phase of this collection period (never shorter than 1.5 ns). Stability was addressed by comparing the average binding free energy values of the first and second halves of the data collection period.

During MD sampling, a snapshot of the system was recorded at regular intervals of 1 ps, and the X-score scoring function was calculated for each snapshot belonging to the collection period. An average scoring value was then obtained for each ligand. The estimated affinity using this scoring function has the general form (*38*)

$$\Delta G_{bind} = \Delta G_{vdW} + \Delta G_{H-bond} + \Delta G_{rot} + \Delta G_{hydrophobic} + \Delta G_0 \ (2)$$

where ΔG_{vdW} accounts for the van der Waals energy between ligand and protein, using an 8-4 Lennard-Jones potential. ΔG_{H-bond} is a directional hydrogen bond term. The hydrophobic effect ($\Delta G_{hydrophobic}$) is calculated as a combination of three different algorithms (named hydrophobic contact, hydrophobic surface, and hydrophobic matching). The term

Table 2: Crystal Structures of Plasmepsins Used in This Work

PDB entry	protein, inhibitor	resolution (Å)	ref
1QS8	Plm *vivax*, pepstatin A	2.5	*50*
1LF2	Plm II, RS370	1.8	*7*
1M43	Plm II, pepstatin A	2.4	*6*
1LEE	Plm II, RS367	1.9	*7*
1LS5	Plm IV, pepstatin A	2.8	–

ΔG_{rot} accounts for the rotational entropy of the ligand. Finally, ΔG_0 is a regression constant.

The MD sampling of the free ligand was again carried out with a 20 Å SPC water sphere (22 Å for pepstatin A). The solvent was initially equilibrated with a 10 ps MD trajectory at 300 K, in which the heavy atoms of the ligand were restrained through a 10 kcal mol^{-1} Å^{-2} force constant to their original position. MD followed for more than 2.5 ns under the same conditions that were used for the bound state, but keeping the initial position of the central atom of the ligand fixed, to ensure a homogeneous solvation.

RESULTS

Enzyme Inhibition. The inhibitory activities of compounds **1**–**9** (Table 1) were determined in Plm I and Plm IV enzyme assays. The activities against Plm II and Cat D have been reported (*25*). Three compounds (**5**, **6**, and **9**) exhibit activity toward Plm IV in the nanomolar range. Among them, **6** is the most potent non-peptidomimetic inhibitor reported so far for this enzyme and also exhibits an interesting selectivity. On the other hand, compounds **8** and **9**, which demonstrated Plm II activities in the low nanomolar range, appear to be 300 and 30 times less active in the Plm IV assay, respectively. Among the compounds with systematic variations of the substitution on the 1,3,4-oxadiazole group (compounds **1**–**4**), only **2** and **3** exhibited moderate activity toward Plm IV. Interestingly, the 4-*tert*-butylphenyl substitutent of the most active compounds, **5** and **6**, is less effective when present only in one of the P2 and P2′ positions as in compound **2**.

Structural Analysis and Homology Modeling. The experimental affinity data presented in Table 1 clearly indicate that there must be differences in the binding site of Plm IV compared to those of Plm I and II. According to earlier structural and sequence analysis (*6, 16, 39*), there are some structural differences between the binding sites of these plasmepsins, which we confirmed through a direct superposition of different experimental structures of Plm II and Plm IV summarized in Table 2 (see Figure 1). First, the distance between the flap region (residues 71–83) and the proline-rich loop (residues 290–297) varies between the two enzymes. In all the complexes examined in Table 2, Plm II has a longer distance between Cα atoms in position 78 and Cα atoms in position 292 (12.6–9.9 Å) than the Plm IV–pepstatin A complex (8.9 Å). This originates from the following differences in sequence: Val78Gly, Leu292Val, and Phe294Ile (with Plm II residues first and Plm IV residues last), with the smaller side chains consistently belonging to Plm IV. The flexible loop (residues 236–245) is folded toward the S3 binding site in Plm IV, resulting in more closed P2–P3 pockets. The question of the possible biological relevance of the 2-fold dimeric association of plasmepsins, which occurs through the aforementioned loop (*6*), has been
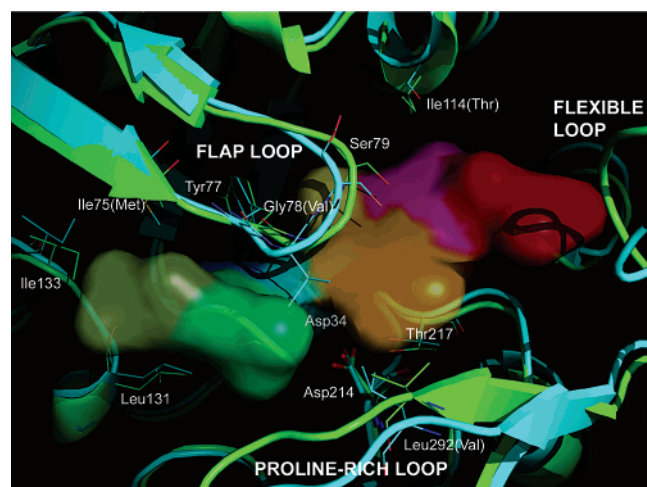
FIGURE 1: Binding site of plasmepsins. Plm IV (PDB entry 1LS5) is colored green and Plm II (PDB entry 1M43) cyan. The surface representation of pepstatin A illustrates the different inhibitor sites: P4 (red), P3 (magenta), P2 (yellow), P1 (orange), P1′ (cyan), P2′ (blue), P3′ (green), and P4′ (dark green). Residues are labeled according to Plm IV sequence, with the corresponding mutation in Plm II in parentheses.

Table 3: Relative Binding Free Energy Estimates (kilocalories per mole) of Pepstatin A

protonation model	1LS5	Plm IV-m
Asp⁻34/AspH214	-13.0 ± 1.6	-11.4 ± 0.9
AspH34/Asp⁻214	-13.0 ± 0.9	-11.3 ± 0.5

addressed. Recent studies also point to the importance of this loop for the catalytic efficiency in plasmepsin II, suggesting its relevance for enzyme inhibition (40). Another difference is due to the Met75Ile substitution in the S2′ pocket of Plm IV. Previous computational and experimental work on Plm II has shown the importance of the flexibility of the Met side chain at this position for the accommodation of large P2′ substitutions (7, 25). Finally, the geometry of the hydrogen bond network between the catalytic aspartates (Asp34 and Asp214) and the central hydroxyl group of pepstatin A is not conserved. The double hydrogen bond (Asp34···ROH···Asp214) present in all inhibitor complexes in Table 2 is not observed in the Plm IV crystal structure. In this case, the distances and angles between the hydroxyl group of the inhibitor and the OD atoms of Asp214 are not optimal (pairwise distances of 3.1 and 3.5 Å, with the hydroxyl of the inhibitor displaced 1.9 Å out of the plane formed by the carboxylates).

To check the importance of these particularities of the 1LS5 structure, we decided to build a homology model for Plm IV, hereafter called Plm IV-m. Three different complexed structures of Plm II and the crystal structure of Plm *vivax* in complex with pepstatin A (Table 2) were chosen as templates, with an aim of combining all the structural information available for close homologues of Plm IV. In this respect, Plm II offers information about the adaptation of a protein to different inhibitors, while Plm *vivax* is the closest ortholog of Plm IV with a known three-dimensional structure (78% level of sequence identity). The homology model performs somewhat better than the 1LS5 crystal structure in terms of several quality indicators, as reported by WHAT-IF analysis (30). This is perhaps not so surprising since the crystal structure has only "intermediate" resolution of 2.8 Å. In particular, the Ramachandran plot for the homology model showed no residues in the forbidden region and less than 5% of the residues in unfavorable regions, while this value is 12% for the 1LS5 structure. The overall rmsd of the Cα trace between this homology model and 1LS5 is 1.19 Å, with the largest differences associated with the

conformation of the flexible loop, which in Plm IV-m resembles the conformation adopted by Plm II. The conformation of the side chains from conserved residues in the plasmepsin family is also in agreement with the conformations observed in the available experimental structures, including the orientation of the catalytic aspartates.

Pepstatin Binding. Pepstatin A is a canonical peptide-mimetic inhibitor of all aspartic proteases. Its binding constant for Plm IV has been reported in the literature (K_i = 20 pM) (15), and the crystal structure of the complex (1LS5) is available. We therefore also examined the binding of this ligand mainly as a check of docking protocols, the homology model, and the protonation state of the catalytic aspartates. GOLD (32) was capable of redocking pepstatin A into the 1LS5 structure, illustrating its potential for docking inhibitors into Plm IV. It should be noted that, due to the large size of pepstatin A, one H-bond constraint between the central hydroxyl group of the inhibitor and the carboxylate of Asp214 was needed to obtain the correct solution, with a rmsd with respect to the native conformation of 1.3 Å. With regard to binding free energy calculations with the LIE approach, the case of pepstatin A was examined under different conditions. First, both the crystal structure (1LS5) and the homology-generated model (Plm IV-m) were used as starting protein conformations. This was done in an attempt to assess the validity of the generated homology model for detailed binding simulations. Second, the protonation state of the catalytic dyad was explored. This is a nontrivial aspect of aspartic proteases, which has been addressed by a number of computational studies (see refs *41−43* and ref *44* for a review). However, it is hard to draw an unambiguous conclusion from these studies, and it seems that both the protonation state and the location of the proton-(s) may be dependent on the inhibitor that is bound. In earlier works by us and others (9, 45), a monoprotonated form was assumed, with the proton located on the oxygen of Asp34 (OD1) which is closest to Asp214 (i.e., AspH34/Asp⁻214). However, since the 1LS5 structure shows deviations from ideal hydrogen bonding distances, the other possible location for the proton on Asp214 (OD1, the oxygen closest to Asp34) was also examined (i.e., Asp⁻34/AspH214). The terminal P3′ carboxylate group of pepstatin, which is solvent-exposed, was neutralized in the calculations since an adequate solvation of this net charge would require a larger simulation sphere and since we are not concerned with its absolute binding affinity.

The results (Table 3) show that very similar binding affinities are predicted from the 1LS5 structure for the two monoprotonated models of the catalytic aspartates. When the Plm IV-m structure was used as starting geometry for the enzyme, the calculated relative free energy of binding was higher than in the former case but again comparable between both protonation models. Two conclusions can be derived from these results. First, our data do not favor either location of the proton in the catalytic dyad, since both alternatives

yield the same result. Second, the homology model seems to perform reasonably well in binding calculations. It is perhaps not surprising that the relative binding energy is lower in the simulation starting from the crystal structure, since the protein—ligand interatomic distances presumably are optimal in the crystal structure of the pepstatin A complex. As noted above, the carboxylate group of the P3 side chain was kept neutral in these calculations, wherefore its effect on the absolute binding constant is not included. The relative affinities discussed here should, however, not be affected since the carboxylate is solvent-exposed in all cases.

Docking of Inhibitors. All the compounds in our series of inhibitors share a 1,2-dihydroxyethylene moiety acting as a transition state analogue, which is located close to the catalytic dyad. However, several possible docking orientations may fulfill this condition; i.e., the positions of the P1, P1′, P2, and P2′ substituents in the available S1, S1′, S2, and S2′ binding pockets, respectively, are not straightforwardly assigned. Another common feature of these inhibitors is the symmetry of the P1 and P1′ substitutions (bisvinylbromide in all compounds except **9**), which are assumed to bind in the S1 and S1′ pockets of the enzyme (*9, 10, 25*). With regard to the P2 and P2′ side chains, only compounds **5** and **6** present symmetry for these positions. For the seven remaining compounds, it is not trivial to decide whether to place the P2 or P2′ side chain on the enzyme prime side. Computational studies of Plm II predicted the indanol substituent on the nonprime side (*25*), but this need not necessarily be true for Plm IV. Therefore, we performed automated docking on the whole series of compounds, searching for a consensus solution for the unsymmetrical compounds. It should be noted that several docking orientations did not attain all the expected interactions (i.e., hydrogen bonding pattern), and the fit of a given ligand substituent (P1, P1′, P2, or P2′) in a given protein cavity (S1, S1′, S2, or S2′) was used as a criterion for the identification of the binding modes. We will refer to the two major alternate binding modes as the *prime* mode, if the amino indanol substitution is on the prime side, or *nonprime* if it is positioned in the S2 site. Unexpectedly, when the 1LS5 structure was used as the rigid protein conformation, GOLD did not find any solution that placed the P1 and P1′ side chains docked into the S1 and S1′ pockets, respectively, but it mainly placed the amino indanol in the S1 pocket and the vinylbromide in the S2 pocket. Since the binding pocket is less collapsed in the Plm IV-m structure, we decided to use this conformation also for flexible ligand docking. This time GOLD found both *prime* and *nonprime* binding modes along the series, although the score and the distribution of the solutions were different for each inhibitor. The *nonprime* orientation was the solution found for six of seven compounds (all except **1**). Among these, the *prime* orientation was also identified in three cases (compounds **3, 8**, and **9**), but only in the case of **8** did this binding mode have a higher score. Hence, the docking experiments reveal the *nonprime* mode as the consensus binding mode within the series, and this is therefore in agreement with the previously proposed binding mode in Plm II (*25*).

Binding Free Energy Calculations. The binding affinity of all compounds in the series was estimated with the LIE method (*26, 27*). On the basis of the docking results, the

nonprime binding mode was selected as the hypothesized binding mode. Reasonable starting structures were generated by superposition of the previously modeled Plm II—**9** and Plm II—**4** complexes (*25*) into the Plm IV (1LS5 or Plm IV-m) structure, and the remaining compounds were built up in the Plm IV binding site by mutation of the corresponding ligand structures. Manual docking was followed by an equilibration phase in which the conserved hydrogen bonds were constrained, allowing the ligands to fit into the target position. In addition to LIE, the MD trajectories were analyzed with the scoring function X-score (*38*) to obtain an average binding estimate for each ligand.

When we used the 1LS5 structure as a model for the protein, the correlation with experiments was excellent for seven compounds, as can be seen from Table 4 and Figure 2A, but two well-defined outliers were identified (compounds **3** and **8**). It should be noted here that the optimal value of the constant term γ in the LIE equation (eq 1) was found to be 1.2 kcal/mol, while it is 0.0 in Plm II (*10, 25*). This constant, which has been shown to be protein specific (*46*), does not affect the relative affinity values for a given protein but can be relevant for selectivity.

The inability to find a consensus docking orientation for this series using the 1LS5 protein structure suggested that the novel ligand structures do not necessarily fit well in the preorganized binding pocket of the Plm IV—pepstatin A complex. We felt therefore encouraged to repeat the LIE simulations using Plm IV-m as a starting structure for the protein. The initial docking position for each ligand was the same that was used in the 1LS5 calculations. The results obtained with this model system are also summarized in Table 4 and Figure 2B. In this case, we find that the optimal value of the constant γ is 1.6 kcal/mol, which encouragingly is very close the value obtained with the 1LS5 model. As can be observed in Figure 2B, the ability to distinguish between good and bad binders is clear, and no real outlier is identified in this case.

The two structural models 1LS5 and Plm IV-m thus give very reasonable predictions of the experimental data with rms errors of 0.98 and 0.77 kcal/mol, respectively. However, a significant improvement is obtained when the energetic data from the two structures are combined, and we will refer to this as the "combined model" (Table 4). This gives an impressively low rms error of only 0.55 kcal/mol, and the corresponding plot of calculated versus experimental binding free energies is shown in Figure 2C (the constant γ in eq 2 is 1.4 kcal/mol in this case). The fact that combining the energetic data from MD trajectories with different initial models leads to such an improvement clearly shows that increased conformational sampling is translated into a gain in statistical accuracy. It therefore suggests that the use of several experimental and/or homology models may be a generally viable strategy for enhancing the predictive power of binding free energy calculations.

Some structure—activity relationships can be understood with these models as follows. Compounds with the oxadiazole moiety (the symmetrical **6** and the series **1**—**4**) are generally well predicted in terms of relative free energies of binding. Compounds **1**—**4** offer a systematic exploration of the P2′ side chain, which yields substantial variations in the experimental affinities. The trend of this small series of inhibitors is reasonably well predicted using all protein

Table 4: Experimental and Calculated Energetics (kilocalories per mole) of Plm IV Inhibition Using the Homology Model (Plm IV-m), the Crystal Structure (1LS5), and the Combined Results from the Model and X-ray Structure

compound	combined		1LS5		ligand-surrounding interactions		Plm IV-m		ligand-surrounding interactions		water ligand-surrounding interactions	
	expt[a] ΔG_{bind}	LIE[b] ΔG_{bind}	LIE ΔG_{bind}[c]	X-score ΔG_{bind}	$\langle V^{vdW}_{l-s}\rangle_p$	$\langle V^{el}_{l-s}\rangle_p$	LIE ΔG_{bind}[d]	X-score ΔG_{bind}	$\langle V^{vdW}_{l-s}\rangle_p$	$\langle V^{el}_{l-s}\rangle_p$	$\langle V^{vdW}_{l-s}\rangle_p$	$\langle V^{el}_{l-s}\rangle_p$
1	−6.9	−6.3 ± 0.5	−6.5 ± 0.6	−9.8	−83.0 ± 0.8	−79.9 ± 0.7	−6.1 ± 0.4	−7.8	−79.5 ± 0.5	−81.9 ± 0.3	−47.2 ± 0.3	−76.3 ± 0.4
2	−8.1	−7.4 ± 0.5	−7.3 ± 0.6	−11.2	−106.1 ± 0.9	−76.8 ± 0.6	−7.5 ± 0.4	−11.1	−101.4 ± 0.3	−81.1 ± 0.3	−61.0 ± 0.3	−75.8 ± 0.6
3	−7.8	−8.7 ± 0.5	−9.1 ± 0.6	−11.1	−101.0 ± 0.0	−81.2 ± 1.0	−8.2 ± 0.4	−10.3	−96.4 ± 0.4	−82.1 ± 0.0	−54.8 ± 0.6	−75.1 ± 0.6
4	−6.8	−7.0 ± 0.3	−6.4 ± 0.1	−10.1	−90.9 ± 0.0	−80.7 ± 0.2	−7.6 ± 0.4	−9.9	−94.6 ± 0.6	−83.5 ± 0.6	−56.3 ± 0.1	−76.7 ± 0.1
5	−8.7	−9.2 ± 0.4	−8.7 ± 0.3	−11.8	−108.4 ± 0.1	−75.2 ± 0.5	−9.6 ± 0.4	−11.9	−109.9 ± 0.0	−78.4 ± 0.8	−69.6 ± 0.5	−66.4 ± 0.1
6	−10.2	−9.9 ± 0.3	−9.3 ± 0.3	−12.7	−110.5 ± 0.8	−76.4 ± 0.4	−10.6 ± 0.3	−11.4	−103.4 ± 0.0	−85.5 ± 0.7	−67.5 ± 0.2	−68.3 ± 0.1
6—Plm II[e]	−8.7[f]	−9.5 ± 0.2		−12.3	−112.9 ± 0.1	−72.2 ± 0.3					−67.5 ± 0.2	−68.3 ± 0.1
7	−6.7	−6.5 ± 0.4	−5.6 ± 0.3	−10.3	−100.9 ± 0.5	−72.2 ± 0.3	−7.5 ± 0.4	−10.8	−104.2 ± 0.2	−77.3 ± 0.7	−63.9 ± 0.4	−71.7 ± 0.2
8	−7.8	−8.4 ± 0.4	−9.7 ± 0.4	−10.2	−99.1 ± 0.5	−82.2 ± 0.6	−7.1 ± 0.3	−10.0	−99.0 ± 0.5	−75.3 ± 0.3	−55.3 ± 0.1	−73.1 ± 0.1
8—Plm II[e]	−11.2	−9.7 ± 0.4		−11.6	−108.2 ± 0.4	−73.4 ± 0.9					−55.3 ± 0.1	−73.1 ± 0.1
9	−8.8	−8.3 ± 0.3	−9.0 ± 0.3	−12.3	−119.9 ± 0.0	−75.6 ± 0.4	−7.6 ± 0.2	−12.0	−120.8 ± 0.1	−72.3 ± 0.2	−67.4 ± 0.3	−73.6 ± 0.2
9—Plm II[e]	−10.9	−11.8 ± 0.4		−13.6	−128.9 ± 0.3	−75.8 ± 0.7					−67.4 ± 0.3	−73.6 ± 0.2

[a] The binding free energy calculated from experimentally determined K_i values using the relation $\Delta G_{bind,exp}° = RT \ln K_i$. [b] Average binding free energy from the 1LS5 and Plm IV-m structures (the constant γ term in eq 2 is 1.4 kcal/mol). [c] Calculated binding free energy with the 1LS5 structure (the constant γ term in eq 2 is 1.2 kcal/mol). [d] Calculated binding free energy with the Plm IV-m structure (the constant γ term in eq 2 is 1.6 kcal/mol). [e] LIE values in Plm II for compound 6 are calculated as described in the text. Protein simulation values for compounds 8 and 9 have been taken from ref 25 (the constant γ term in eq 2 is 0.0 kcal/mol). [f] The experimental value from the Plm I assay is used since Plm I and Plm II affinities have usually been found to be very similar (10, 25).

models. In the MD simulations, we observed that the S2′ pocket (Ile75, Thr77, Leu131, and Ile133) is optimally filled with the 4-*tert*-butylphenyl substitution present in compound **2**. Smaller side chains (**1**) reduce the predicted binding affinity of the inhibitors, due to instability of the hydrogen bonds observed on the prime side, in agreement with the experimental data. Larger side chains (**4**) also reduce the activity, which is correctly predicted using the 1LS5 structure but not observed with the Plm IV-m model. On the other hand, compound **3**, with an activity very similar to that of **2**, is overpredicted by 1.3 kcal/mol with the 1LS5 structure, due to exaggerated electrostatic interactions with the protein.

The symmetrical compound **6**, in which the P2′ position of ligand **2** is replicated at the P2 position, is the most potent ligand of the series (K_i = 35 nM), and its high affinity is reproduced in our models. In addition to the good fit of the P2′ substituent in the S2′ pocket, we also observed a shape complementarity of the P2 side chain to the S2 and S3 hydrophobic pocket formed by residues Leu14, Met15, and Ile114. This pocket undergoes structural reorganization upon binding of this compound, detected after a few hundred picoseconds of the MD trajectories, and the 111−114 loop approaches the 4-*tert*-butylphenyl group of the ligand. The resulting conformation of the protein displays good shape complementarity with the ligand, resulting in favorable nonpolar interactions (in particular with Ile114) and also contributing to keeping the ligand in the initial position, thus making electrostatic interactions in the complex more favorable (Figure 3). These involve double hydrogen bonding between the central diol system and the catalytic aspartates and a single hydrogen bond between each of the oxadiazole rings to the flap loop, with the backbone of residues Gly78 and Ser79. This pattern of interactions does not differ too much between the MD simulations performed with the two protein models, even though the Plm IV-m structure yields a more pronounced electrostatic binding component.

Our data explain why the P2 substitution in **6** works better than the indanol substitution present in the **1**−**4** series. In that series, the ligand can make three additional hydrogen bonds to the protein, observed between the amide group and the backbones of Gly78 (flap) and Ser218, and a third hydrogen bond is occasionally observed between the side chain of Thr219 and the hydroxyl of the indanol. This might seem at variance with the single hydrogen bond between the oxadiazole ring and Ser79 on the flap predicted for ligand **6**. However, the indanol amide group is more polar than the 5-(4-*tert*-butylphenyl)-[1,3,4-oxadiazole] of **6**, which is translated into more negative electrostatic solvation energies in water for the former. The resulting electrostatic (polar) contribution to the free energy of binding is therefore less favorable. A comparison of the interaction energies in the bound and unbound states (Table 4) for ligand **6** and **2**, which only differ in this P2 substitution, illustrates this point.

It is interesting to note that two of the key residues that support the binding of compound **6** are not conserved in Plm II. Ile75 of Plm IV is a Met in Plm II, which was found to be flexible and to affect the shape of the S2′ pocket in that protein (7, 25). The Gly78 at the tip of the flap is a Val in Plm II, and the hydrophobic Ile114 is replaced by a Thr in Plm II, in the S3 pocket. All three of these positions have been shown to exhibit high variability among the different plasmepsins from *P. falciparum* (16), while there is a higher
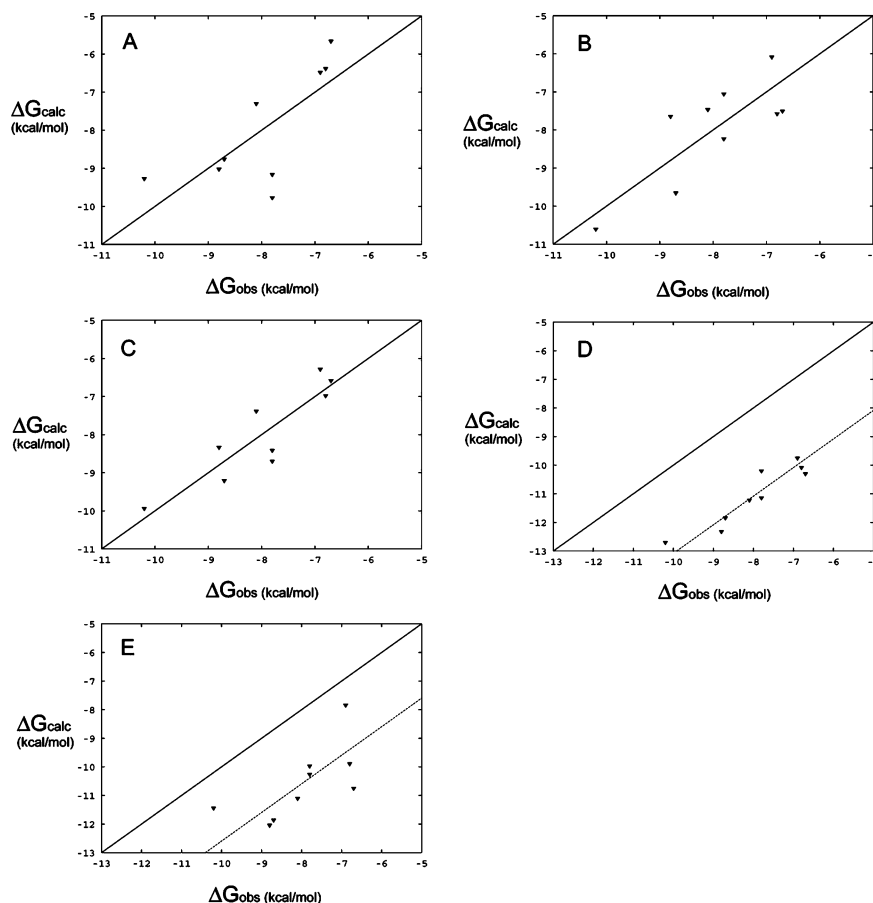
FIGURE 2: Scatter diagram of calculated vs experimental binding free energies using the different models and methods discussed in the text. LIE results are given for simulations utilizing (A) the 1LS5 structure, (B) the Plm IV-m structure, and (C) the combined model. X-Score results averaged over MD trajectories are shown for (D) the 1LS5 structure and (E) the Plm IV-m structure. The dotted lines show least-squares fitting of calculated X-score values with a constant offset (3.0 kcal/mol for panel D and 2.7 kcal/mol for panel E).
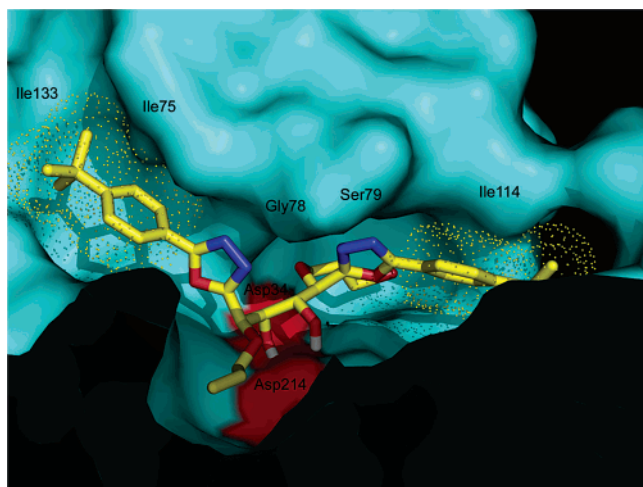

FIGURE 3: Surface representation of Plm IV in complex with compound **6**. The shape complementarity of the 4-*tert*-butylphenyl substitutions with both S2 and S2′ sites of the enzyme can be appreciated. The positions of the catalytic aspartates are colored red.

degree of conservation among the Plm IV orthologs from different *Plasmodium* species (*22*). To investigate the role of these differences, we have calculated the binding affinity for compound **6** also in Plm II. The crystal structure of Plm II in complex with pepstatin A (1SME) (*5*) was used in these calculations. As can be seen from the data reported in Table 4, the affinity of **6** for Plm II is also well reproduced in our

calculations. However, selectivity for Plm IV is only predicted if we consider the calculations using the Plm IV-m structure or the combined model. The hydrophobic interactions are slightly more pronounced in Plm II, due to favorable interactions with Val78 on the flap loop, while the electrostatic component of binding is more pronounced in Plm IV, especially when the homology model was used as the input structure. Again, the explanation comes from a more tightened position in the binding site of Plm IV, due to the relatively more closed S2−S3 and S2′ sites. However, the structural difference between the two proteins at the S2′ site is smaller than expected, since Met75 in Plm II maintained a closed conformation of this pocket along the MD simulation.

With regard to the compounds sharing a diazylhydrazine linker in P2′ (**5** and **7−9**), there is also a broad spectrum of experimental activities, and some interesting differences with the previously reported values for Plm II. Our calculations generally explain the experimental values, although some inconsistencies are found between the MD trajectories from the two protein models. These variations are smoothed when the combined model is considered, leading to more robust results. Two compounds (**5** and **9**) exhibit nanomolar activity in Plm IV, and of these, compound **9** is a very potent inhibitor of Plm II. These experimental results are correctly reproduced in our calculations using any of the three LIE models. The binding mode of compound **9** is depicted in Figure 4. It can be appreciated that the ligand folds the P2′
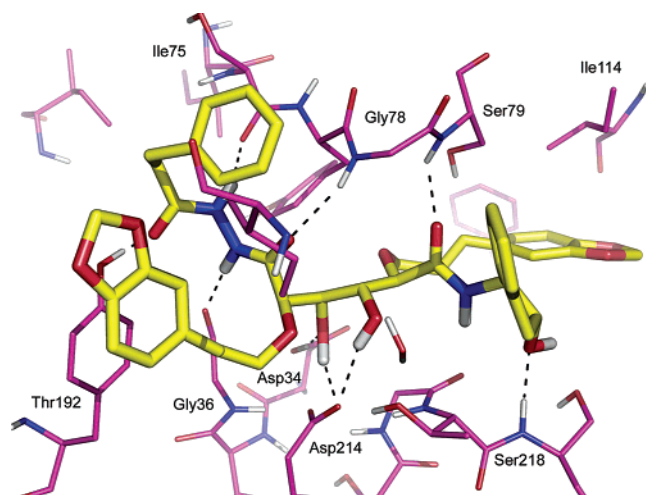
FIGURE 4: Complex of Plm IV with compound **9**. Hydrogen bonds are represented by dashed lines. Note the folding of the P2′ substitution toward the solvent since there is no space in the rigid S2′ pocket to accommodate it.

side chain back onto the rest of the ligand, since the rigid S2′ pocket of Plm IV does not allow expansion of the large phenylethyl group, as previously observed in Plm II (*25*). This is translated into a decrease in just the nonpolar binding component, since the hydrogen bond network between the ligand and protein is conserved with respect to previous results for Plm II (see Table 4). The same applies to compound **8**, which differs from **9** in the P1 and P1′ side chains and has a marked selectivity for Plm II. However, the calculation using the 1LS5 structure overpredicts the activity of this particular compound due to an overestimation of the polar binding component. In fact, the electrostatic interactions for **8** stand out from the rest of the series and also from the Plm II−**8** simulations previously reported (*25*) (see Table 4).

It is interesting to compare **5** and **6**, since the only difference between the two symmetric ligands is the linker (oxadiazole in **6** and diazylhydrazine in **5**). The replacement of the oxadiazole with the diazylhydrazine leads to a 10-fold decrease in the observed binding affinity (35 nM vs 478 nM). This effect is reproduced in our models, which confirms the good shape complementarity of the 4-*tert*-butylphenyl substituent with either oxadiazole or diazylhy-drazine linkers, being slightly better in the first case. Surprisingly, the combination of this side chain in P2′ with an indanol amine at P2 leads to a huge decrease in the predicted affinity for compound **7**. The simulations using 1LS5 show a conformational change in the P2′ side chain, which leads to a considerable loss of interactions with the protein and also affects the geometry of the double hydrogen bonding with the catalytic dyad. However, the compound remains more stable around the initial docking position if the Plm IV-m structure is used, and the compound is not properly ranked as the less active binder in this case.

The binding estimates using the empirical scoring function were also very good in the relative ranking of the compounds (see Figure 2D,E). The absolute values for the predicted affinities were, however, shifted by an average overprediction of 3.0 and 2.7 kcal/mol for 1LS5 and Plm IV-m, respectively. The overprediction in the estimation of binding affinity using scoring functions for this class of compounds has been

observed previously (*9*), and our interest is rather focused on the relative values. It should also be noted here that the scoring with X-score was calculated as an average value of snapshots from the MD trajectories discussed above. With this, we not only increased the precision of the results compared to single-point scoring on the initial positions (Figure 5) but also ensured that the ligand binding mode we were evaluating was the same as in the LIE calculations (i.e., relaxation of the initial complex was allowed).

It can be appreciated that the ranking is slightly better when the trajectories generated with the 1LS5 structure are used, with only small deviations from the experimental ranking in the less active compounds (Figure 2D, rms error = 0.40 kcal/mol). The predictions with the scoring function using the trajectories generated with the homology model also reproduce the trend of the series rather well (Figure 2E, rms error = 1.29 kcal/mol). The three compounds with nanomolar activity (**5**, **6**, and **9**) are the top-ranked ones. However, the ranking among them is not correct, and compound **9** is estimated to be the most active in the series. There is also a difficulty in the prediction of inactive ligand **7**, which has a very stable conformation along the MD trajectories.

Selectivity. It is a major challenge in the drug design field to understand the reasons for specificity and selectivity of binding. This issue can be addressed by molecular modeling techniques, like the ones used herein. Thus, the LIE equation can easily be rewritten to obtain relative binding affinities of a given compound in different proteins. If LIE equations for two different proteins (p1 and p2) are subtracted, the contributions from the free state will cancel, yielding

$$\Delta\Delta G_{bind} = \alpha(\langle V_{1-s}^{vdW}\rangle_{p1} - \langle V_{1-s}^{vdW}\rangle_{p1}) + \beta(\langle V_{1-s}^{el}\rangle_{p1} - \langle V_{1-s}^{el}\rangle_{p1}) + \gamma_{p1-p2} \quad (3)$$

where $\gamma_{p1-p2} = \gamma_{p1} - \gamma_{p2}$. In this case, since $\gamma_{p1} = 0$ for Plm II, the constant parameter in eq 1 equals $-\gamma_{p2}$, that is, the calculated γ for Plm IV in the particular protein model used with an inverted sign. With this formulation, we can obtain selectivity estimates avoiding the water simulations, which are commonly the main source of variance in the computed average energies (*35*).

Selectivity is not necessarily desirable between different plasmepsins, since recent investigations point toward the concurrent function of these proteases in the hemoglobin degradation pathway (*13*). However, our compounds present some difference in the affinities between Plm I/Plm II and Plm IV. This property might be used as a pharmacological tool for the characterization of differences between the two enzymes and the specific role of each in the hemoglobin degradation. One of the inhibitors (compound **6**) exhibited the highest affinity for Plm IV, being 10 times less potent toward Plm I and Plm II. As shown in Table 5, this property is qualitatively reproduced in both Plm IV-m and the combined model, but not observed using the 1LS5 structure. With regard to the other two compounds with different affinities between the two proteins (**8** and **9**), the results again show a better correlation if the Plm IV-m structure is used for calculations on Plm IV. The nonoptimal fitting of the P2′ substituent in the case of compound **8** accounts for the difference with respect to Plm II, while in the case of **9**, this
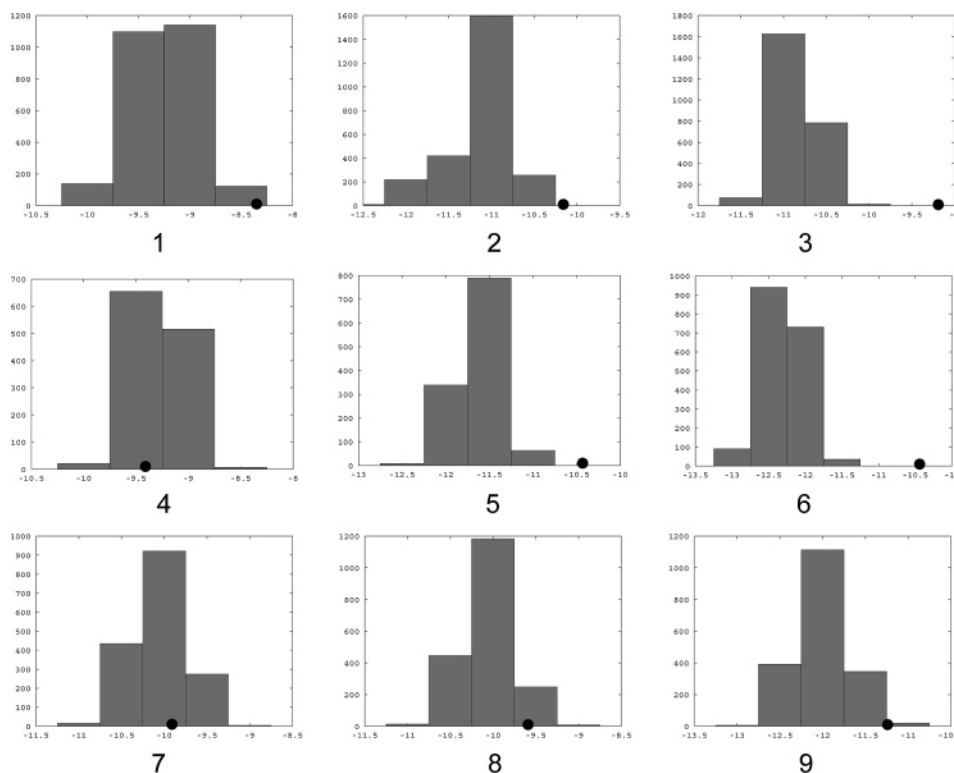
FIGURE 5: Distribution of calculated X-score affinities along MD trajectories using the 1LS5 structure. Frequencies are calculated in bins of 0.5 kcal/mol. The value for the single-point calculation using the initial structure is represented with a dot.

Table 5: Experimental and Calculated Selectivity ($\Delta\Delta G_{bind}$) between Plm IV and Plm II[a]

| compound | $\Delta\Delta G_{obs}$ | $\Delta\Delta G_{calc,LIE}$[b] | | | $\Delta\Delta G_{calc,X-score}$ | |
		1LS5	Plm IV-m	combined	1LS5	Plm IV-m
6	+1.5	−0.2	+1.4	+0.6	+0.4	−0.9
8	−3.4	+0.3	−2.6	−1.2	−1.4	−1.6
9	−2.1	−2.8	−4.2	−3.6	−1.3	−1.6

[a] As a convention, selectivity for Plm IV is denoted by a positive sign for $\Delta\Delta G_{bind}$, while a negative sign indicates selectivity for Plm II. [b] $\Delta\Delta G_{bind}$ as calculated with eq 3.

effect is partially counterbalanced by the good fit of the P1 and P1′ substitutions in the S1 and S1′ pockets of Plm IV. The combined model, which was shown to be the most predictive model for the relative ranking of compounds, is also affective in the characterization of selectivities between the two plasmepsins considered in this work. The selectivities calculated from the X-score values are also given in Table 5 for comparison. In this case, the strong selectivity of **6** for Plm IV is not particularly well reproduced.

DISCUSSION

We have investigated the binding properties of Plm IV with a series of previously reported inhibitors initially designed against Plm II. The determination of the affinities for Plm IV revealed a broad spectrum of activities. Three compounds exhibited K_i activities in the nanomolar range, and the most potent inhibitor of the series (**6**) was also selective for Plm IV. On the other hand, some compounds with high activity toward Plm II (**8** and **9**) had lower affinities for Plm IV. The selectivity that these compounds show makes them interesting in our attempt to understand the structural differences between the active sites of plasmepsins.

Our approach to analyzing features governing binding included homology modeling, automated docking, and calculations of binding affinities for Plm IV, comparing the obtained results with previous studies of Plm II. As a preliminary step, direct comparison of X-ray structures of both enzymes revealed important differences in the P2 and P2′ binding sites, as well as some discrepancies in the hydrogen bond pattern involving the catalytic aspartates. A major problem with highly flexible proteins, like the malarial aspartic proteases considered here, is the difficulty in modeling a given class of inhibitors using three-dimensional information from structurally unrelated inhibitors (6, 25). Our docking studies supported this idea, since it was not possible to find a consistent solution for our series of inhibitors using the crystal structure of the Plm IV−pepstatin A complex. We thus created a homology model based on four different templates, aiming to incorporate information about the flexibility. The use of different templates was motivated by the observation that there are indeed big differences between the pepstatin complexes and the complexes with larger inhibitors (5−7). Three different conformations of Plm II were then used together with the crystal structure of Plm *vivax* (the closest ortholog of Plm IV). The crystal structure of Plm *malariae* (24) has been released during the preparation of this work and, therefore, could not be considered as a template in this work. The resulting model was found to be a useful tool for identifying the binding mode and for estimating the binding of a chemically diverse set of compounds.

Each compound was simulated by MD using the *nonprime* binding mode as a starting point in parallel calculations, utilizing both the crystal structure and the homology model. This binding mode was determined by automated docking

using the homology model and was in agreement with previous studies on Plm II. The trajectories generated by MD simulations were used both for the estimation of binding free energies with the LIE approach and for calculating average scores using the empirical X-score function (*38*). Analysis of the binding profile with the LIE method allowed a reasonable interpretation of the experimental data. All the compounds exhibited a high component of nonpolar binding, while the electrostatic interactions with the protein mainly counterbalanced the desolvation process, adding only a small contribution to the total free energy of binding. This is in agreement with previous observations that electrostatic (dipolar) interactions are mainly important for specificity, while a large part of the overall affinity is attained through nonpolar contacts (*47*).

When the homology model was used for the calculations, the results showed more robustness within the series, and no outlier was identified. These results support the idea that homology models can be used for detailed ligand binding calculations with the LIE method, as was shown in recent studies of these and other systems (*20, 48, 49*). However, to our knowledge, this is the first direct evaluation of the effect that different protein models (e.g., experimental vs modeled protein structures) can have on the calculated estimates of ligand binding constants. It is encouraging that if no suitable crystal structure is available, homology modeling can give a reliable answer to the problem. However, a major question in molecular modeling is the extent to which the accuracy of the protein conformation can affect the calculations. Here we examine the influence of the initial starting point on the average MD values obtained with both the LIE method and a scoring function. Our results indicate that some variation can be found in the predicted values from different protein models, but an overall examination of the results points to the same conclusions for the structure−activity relationship of the series of inhibitors. Furthermore, it is encouraging that the combined model performs significantly better than either of the underlying X-ray and homology models.

When the binding affinities were evaluated with the scoring function, the ranking of the series was reasonable in the case of the homology model and excellent with the crystal structure. It should be noted that the results improved substantially when the score was calculated as an average over MD frames along the trajectory, in comparison to the static initial docking orientations. Figure 5 shows Gaussian-like distributions of the calculated values along the MD sampling for each ligand, together with the single-point calculation made on the initial structure. It is clear that for many compounds, the scoring value obtained by the initial single-point calculation is far from the representative mean value obtained by MD sampling. Therefore, more accurate results can be achieved by conformational sampling of each complex through MD simulation. Empirical scoring functions are obviously faster to evaluate than force field energies, and they do not require simulations of the unbound ligand state. However, it is more difficult to reliably elucidate properties that govern the binding process with a purely statistical method, like the scoring function used in this work. In that sense, we find the more physical, force field-based LIE approach to be more informative.

The observation of a unique trend in affinities through the series using two very different methods for binding affinity predictions (LIE and X-score), each of them applied to two different protein models, strongly argues in favor of the binding mode proposed for this series of compounds. The comparison of the results with recent data for the same type of compounds in Plm II also supports this binding mode. Furthermore, with the more accurate LIE approach (using Plm IV-m or the combined model), the experimental selectivity is well reproduced for the three cases that are considered (see Table 5).

CONCLUDING REMARKS

In this work, we have reported three potent inhibitors for Plm IV with K_i values in the nanomolar range. One of these inhibitors (compound **9**) also exhibits a high potency against Plm I and Plm II, thus constituting a hit in the identification of multitarget inhibitors against plasmepsins, which may be desirable for therapeutic success (*13*). On the other hand, the most potent compound of the series exhibited some selectivity for Plm IV. The structural characterization of the binding of this compound to Plm IV from *P. falciparum* can be valuable for lead optimization of novel inhibitors of this enzyme and the relative orthologs in other *Plasmodium* species. The quantitative prediction of inhibitor selectivity between closely related proteins by computational methods remains a major challenge, but as shown here, several variants of current scoring procedures can provide useful results. We have shown that an exhaustive sampling improves the accuracy of the binding estimates for both the scoring function and LIE methodologies. Our results point to the importance of taking conformational flexibility of both ligand and receptor into account for accurate computational predictions of ligand binding affinities.

REFERENCES

1. Breman, J. G. (2001) The ears of the hippopotamus: Manifestations, determinants, and estimates of the malaria burden, *Am. J. Trop. Med. Hyg. 64*, 1−11.

2. Gardner, M. J., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R. W., Carlton, J. M., Pain, A., Nelson, K. E., Bowman, S., Paulsen, I. T., James, K., Eisen, J. A., Rutherford, K., Salzberg, S. L., Craig, A., Kyes, S., Chan, M.-S., Nene, V., Shallom, S. J., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M. W., Vaidya, A. B., Martin, D. M. A., Fairlamb, A. H., Fraunholz, M. J., Roos, D. S., Ralph, S. A., McFadden, G. I., Cummings, L. M., Subramanian, G. M., Mungall, C., Venter, J. C., Carucci, D. J., Hoffman, S. L., Newbold, C., Davis, R. W., Fraser, C. M., and Barrell, B. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*, *Nature 419*, 498−511.

3. 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. 17*, 532−7.

4. Goldberg, D. E., Slater, A. F., Beavis, R., Chait, B., Cerami, A., and Henderson, G. B. (1991) Hemoglobin degradation in the human malaria pathogen *Plasmodium falciparum*: A catabolic pathway initiated by a specific aspartic protease, *J. Exp. Med. 173*, 961−9.

5. Silva, A. M., Lee, A. Y., Gulnik, S. V., Majer, 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. (1996) Structure and inhibition of plasmepsin II, a

hemoglobin-degrading enzyme from *Plasmodium falciparum*, *Proc. Natl. Acad. Sci. U.S.A. 93*, 10034−9.

6. Asojo, O. A., Gulnik, S. V., Afonina, E., Yu, B., Ellman, J. A., Haque, T. S., and Silva, A. M. (2003) Novel uncomplexed and complexed structures of plasmepsin II, an aspartic protease from *Plasmodium falciparum*, *J. Mol. Biol. 327*, 173−81.

7. Asojo, O. A., Afonina, E., Gulnik, S. V., Yu, B., Erickson, J. W., Randad, R., Medjahed, D., and Silva, A. M. (2002) Structures of ser205 mutant plasmepsin II from *Plasmodium falciparum* at 1.8 angstrom in complex with the inhibitors rs367 and rs370, *Acta Crystallogr. D58*, 2001−8.

8. Boss, C., Richard-Bildstein, S., Weller, T., Fischli, W., Meyer, S., and Binkert, C. (2003) Inhibitors of the *Plasmodium falciparum* parasite aspartic protease plasmepsin II as potential antimalarial agents, *Curr. Med. Chem. 10*, 883−907.

9. Ersmark, K., Feierberg, I., Bjelic, S., Hulten, J., Samuelsson, B., Aqvist, J., and Hallberg, A. (2003) C-2-symmetric inhibitors of *Plasmodium falciparum* plasmepsin II: Synthesis and theoretical predictions, *Bioorg. Med. Chem. 11*, 3723−33.

10. Ersmark, K., Feierberg, I., Bjelic, S., Hamelink, E., Hackett, F., Blackman, M. J., Hulten, J., Samuelsson, B., Aqvist, J., and 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.

11. Haque, T. S., Skillman, A. G., Lee, C. E., Habashita, H., Gluzman, I. Y., Ewing, T. J. A., Goldberg, D. E., Kuntz, I. D., and Ellman, J. A. (1999) Potent, low-molecular-weight non-peptide inhibitors of malarial aspartyl protease plasmepsin II, *J. Med. Chem. 42*, 1428−40.

12. Kiso, A., Hidaka, K., Kimura, T., Hayashi, Y., Nezami, A., Freire, E., and Kiso, Y. (2004) Search for substrate-based inhibitors fitting the S-2′ space of malarial aspartic protease plasmepsin II, *J. Pept. Sci. 10*, 641−7.

13. Liu, J., Gluzman, I. Y., Drew, M. E., and Goldberg, D. E. (2005) The role of *Plasmodium falciparum* food vacuole plasmepsins, *J. Biol. Chem. 280*, 1432−7.

14. Wyatt, D. M., and Berry, C. (2002) Activity and inhibition of plasmepsin IV, a new aspartic proteinase from the malaria parasite, *Plasmodium falciparum*, *FEBS Lett. 513*, 159−62.

15. Li, T., Yowell, C. A., Beyer, B. B., Hung, S. H., Westling, J., Lam, M. T., Dunn, B. M., and Dame, J. B. (2004) Recombinant expression and enzymatic subsite characterization of plasmepsin 4 from the four *Plasmodium* species infecting man, *Mol. Biochem. Parasitol. 135*, 101−9.

16. Nezami, A., Kimura, T., Hidaka, K., Kiso, A., Liu, J., Kiso, Y., Goldberg, D. E., and Freire, E. (2003) High-affinity inhibition of a family of *Plasmodium falciparum* proteases by a designed adaptive inhibitor, *Biochemistry 42*, 8459−64.

17. Beyer, B. B., Johnson, J. V., Chung, A. Y., Li, T., Madabushi, A., Agbandje-McKenna, M., McKenna, R., Dame, J. B., and 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.

18. Westling, J., Yowell, C. A., Majer, P., Erickson, J. W., Dame, J. B., and 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.

19. Berry, C., Humphreys, M. J., Matharu, P., Granger, R., Horrocks, P., Moon, R. P., Certa, U., Ridley, R. G., Bur, D., and Kay, J. (1999) A distinct member of the aspartic proteinase gene family from the human malaria parasite *Plasmodium falciparum*, *FEBS Lett. 447*, 149−54.

20. Bjelic, S., and Aqvist, J. (2004) Computational prediction of structure, substrate binding mode, mechanism, and rate for a malaria protease with a novel type of active site, *Biochemistry 43*, 14521−8.

21. Andreeva, N., Bogdanovich, P., Kashparov, I., Popov, M., and Stengach, M. (2004) Is histoaspartic protease a serine protease with a pepsin-like fold? *Proteins 55*, 705−10.

22. Dame, J. B., Yowell, C. A., Omara-Opyene, L., Carlton, J. M., Cooper, R. A., and Li, T. (2003) Plasmepsin 4, the food vacuole aspartic proteinase found in all *Plasmodium* spp. infecting man, *Mol. Biochem. Parasitol. 130*, 1−12.

23. Ersmark, K., Nervall, M., Gutiérrez-de-Terán, H., Hamelink, E., Janka, L. K., Clemente, J. C., Dunn, B. M., Gogoll, A., Samuelsson, B., Aqvist, J., and Hallberg, A. (2006) Macrocyclic

24. Clemente, J. C., Govindasamy, L., Madabushi, A., Fisher, S. Z., Moose, R. E., Yowell, C. A., Hidaka, K., Kimura, T., Hayashi, Y., Kiso, Y., Agbandje-McKenna, M., Dame, J. B., Dunn, B. M., and McKenna, R. (2006) Structure of the aspartic protease plasmepsin 4 from the malarial parasite *Plasmodium malariae* bound to an allophenylnorstatine-based inhibitor, *Acta Crystallogr. D62*, 246−52.

25. Ersmark, K., Nervall, M., Hamelink, E., Janka, L. K., Clemente, J. C., Dunn, B. M., Blackman, M. J., Samuelsson, B., Aqvist, J., and Hallberg, A. (2005) Synthesis of malarial plasmepsin inhibitors and prediction of binding modes by molecular dynamics simulations, *J. Med. Chem. 48*, 6090−106.

26. Aqvist, J., Medina, C., and Samuelsson, J. E. (1994) A new method for predicting binding affinity in computer-aided drug design, *Protein Eng. 7*, 385−91.

27. Hansson, T., Marelius, J., and Aqvist, J. (1998) Ligand binding affinity prediction by linear interaction energy methods, *J. Comput.-Aided Mol. Des. 12*, 27−35.

28. Westling, J., Cipullo, P., Hung, S. H., Saft, H., Dame, J. B., and Dunn, B. M. (1999) Active site specificity of plasmepsin II, *Protein Sci. 8*, 2001−9.

29. Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003) SWISS-MODEL: An automated protein homology-modeling server, *Nucleic Acids Res. 31*, 3381.

30. Hooft, R. W., Vriend, G., Sander, C., and Abola, E. E. (1996) Errors in protein structures, *Nature 381*, 272.

31. Accelerys, Inc. (2002) *ViewerPro*, San Diego.

32. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., and Taylor, R. D. (2003) Improved protein−ligand docking using GOLD, *Proteins 52*, 609−23.

33. Marelius, J., Kolmodin, K., Feierberg, I., and Aqvist, J. (1999) Q: An MD program for free energy calculations and empirical valence bond simulations in biomolecular systems, *J. Mol. Graphics Modell. 16*, 213−25.

34. van Gunsteren, W. F. B., and Berendsen, H. J. C. (1987) *Groningen molecular simulation (gromos) library manual*, Nijenborgh, Groningen, The Netherlands.

35. Marelius, J., Hansson, T., and Aqvist, J. (1998) Calculation of ligand binding free energies from molecular dynamics simulations, *Int. J. Quantum Chem. 69*, 77−88.

36. King, G., and Warshel, A. (1989) A surface constrained all-atom solvent model for effective simulations of polar solutions, *J. Chem. Phys. 91*, 3647−61.

37. Lee, F. S., and Warshel, A. (1992) A local reaction field method for fast evaluation of long-range electrostatic interactions in molecular simulations, *J. Chem. Phys. 97*, 3100−7.

38. Wang, R., Lai, L., and Wang, S. (2002) Further development and validation of empirical scoring functions for structure-based binding affinity prediction, *J. Comput.-Aided Mol. Des. 16*, 11.

39. Nezami, A., and Freire, E. (2002) The integration of genomic and structural information in the development of high affinity plasmepsin inhibitors, *Int. J. Parasitol. 32*, 1669−76.

40. Istvan, E. S., and Goldberg, D. E. (2005) Distal substrate interactions enhance plasmepsin activity, *J. Biol. Chem. 280*, 6890−6.

41. Northrop, D. B. (2001) Follow the protons: A low-barrier hydrogen bond unifies the mechanisms of the aspartic proteases, *Acc. Chem. Res. 34*, 790−7.

42. Piana, S., Sebastiani, D., Carloni, P., and Parrinello, M. (2001) Ab initio molecular dynamics-based assignment of the protonation state of pepstatin A/HIV-1 protease cleavage site, *J. Am. Chem. Soc. 123*, 8730−7.

43. Trylska, J., Grochowski, P., and McCammon, J. A. (2004) The role of hydrogen bonding in the enzymatic reaction catalyzed by HIV-1 protease, *Protein Sci. 13*, 513−28.

44. Dunn, B. M. (2002) Structure and mechanism of the pepsin-like family of aspartic peptidases, *Chem. Rev. 102*, 4431−58.

45. Huo, S., Wang, J., Cieplak, P., Kollman, P. A., and Kuntz, I. D. (2002) Molecular dynamics and free energy analyses of cathepsin D-inhibitor interactions: Insight into structure-based ligand design, *J. Med. Chem. 45*, 1412−9.

46. Almlöf, M., Brandsdal, B. O., and Aqvist, J. (2004) Binding affinity prediction with different force fields: Examination of the linear interaction energy method, *J. Comput. Chem. 25*, 1242−54.

47. Hulten, J., Bonham, N. M., Nillroth, U., Hansson, T., Zuccarello, G., Bouzide, A., Aqvist, J., Classon, B., Danielson, U. H., Karlen,

A., Kvarnstrom, I., Samuelsson, B., and Hallberg, A. (1997) Cyclic HIV-1 protease inhibitors derived from mannitol: Synthesis, inhibitory potencies, and computational predictions of binding affinities, *J. Med. Chem. 40*, 885−97.

48. Gutiérrez-de-Terán, H., Pastor, M., Centeno, N. B., Aqvist, J., and Sanz, F. (2004) Comparative analysis of putative agonist binding modes in the human A_1 adenosine receptor, *ChemBioChem 5*, 841−9.

49. Österberg, F., and Åqvist, J. (2005) Exploring blocker binding to a homology model of the open hERG K^+ channel using docking and molecular dynamics methods, *FEBS Lett. 579*, 2939.

50. Bernstein, N. K., Cherney, M. M., Yowell, C. A., Dame, J. B., and James, M. N. (2003) Structural insights into the activation of *P. vivax* plasmepsin, *J. Mol. Biol. 329*, 505−24.